
Javad Mohammadpour
Karolos M. Grigoriadis
Editors

Efficient Modeling and Control of Large-Scale Systems

Efficient Modeling and Control of Large-Scale Systems

Javad Mohammadpour • Karolos M. Grigoriadis
Editors

Efficient Modeling and Control of Large-Scale Systems

Editors

Javad Mohammadpour
University of Houston
Dept. Mechanical Engineering
Calhoun Road 4800
77204-4006 Houston Texas
N204, Engineering Bldg. 1
USA
jmohammadpour@uh.edu

Karolos M. Grigoriadis
University of Houston
Dept. Mechanical Engineering
Calhoun Road 4800
77204-4006 Houston Texas
N204, Engineering Bldg. 1
USA
karolos@uh.edu

ISBN 978-1-4419-5756-6 e-ISBN 978-1-4419-5757-3
DOI 10.1007/978-1-4419-5757-3
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010929392

© Springer Science+Business Media, LLC 2010

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

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

Foreword

Almost five decades have passed from the organized formulation of Large-Scale Systems (LSS) engineering as a branch of control and systems engineering. In the 1960s modeling, analysis and synthesis of complex systems were being pursued by a number of research teams in the US and overseas. It is impossible to identify all of them here. Among those there were the MIT team headed by Professor Michael Athans and many of his colleagues and his former doctoral students; the other was at the University of Illinois, headed by Professor Joe Cruz, Jr. with able peers like Petar Kokotovic and Bill Perkins and their many doctoral students; another was the work done by Ted Davison and his peers at the University of Toronto in Canada and Andy Sage at the University of Virginia and later on at George Mason University. At the time I was a doctoral student of Petar Kokotovic at the University of Illinois. These were not the only ones of course, but they were very visible in the annual IEEE CDC and ACC meetings. The approaches of LSS theory at Illinois were based on sensitivity of system with respect to inherent model's parameters and by determining the sensitivity of inputs and outputs and thereby deal with many attributes of an LSS – be it nonlinearity, dimension, time delays, etc. Others were concerned with robust control or multi-variable control approaches. Then it was in 1978 that a special issue of IEEE Transaction on Automatic Control (Volume AC-23, 1978) guest edited by Mike Athans, Nil Sandell, Pravin Varaiya and Mike Safonov, the control of complex systems were designated *Large-Scale Control Systems*.

Three definitions of LSS can be documented, as this author enumerated in his 1983 textbook on Large-Scale Systems – Modeling and Control (North-Holland Publishers, New York, 1983). First a system is considered an LSS with a hierarchical structure so that behavior of subsystems can be coordinated from above and feasibility and optimality of the overall system can be guaranteed by iterative processes based on optimization principles (Minimum Principle, Dynamic Programming or Lagrangian optimization). This definition led to “hierarchical control”, which was pursued extensively by M.D. Mesraovic et al. at the Case Western Reserve University. The second definition is that a system whose components or subsystems are not centralized in one location is an LSS. This approach was researched by many, including the Davison group in Toronto. The third one is that a

system whose attributes (nonlinear, stochastic, time-varying, delays, dimensionality, etc.) are such that standard theories and approaches of classical or modern control theories need to be re-invented. This definition led to new paradigms for modeling and control of LSS which was based on sensitivity functional techniques and tools that were pursued at Illinois.

Now, nearly 50 years later, at the beginning of the twenty-first century, systems and control engineering fields are being challenged once again to deal with integration and interoperability of a group of legacy or other systems to make the full use of our global wireless *flat* world to quote Thomas Friedman of *New York Times*. This new observation of complex systems is called either *System of Systems* (SoS) or *Cyber-Physical Systems* (CPS). Interoperability or system integration adds another complexity dimension to an already complex system and for the first time control engineers need to design a feedback control paradigm where a wireless Ad-hoc network lies in the feedback loop. Here one has a marriage between the “Cyberspace” and “Physical Space”. This new class of complex systems is really not new and does in fact follow the 2nd definition of LSS of decentralize nature of many complex systems. The added complexity is the communication protocol among SoS constituents and Internet (network) – induced time delays and their potential adverse effect on the system performance. These time-delays are not necessarily even Gaussian and are often best estimated by using “alpha-stable” distribution like a non-Gaussian noise.

This volume, *Efficient Modeling and Control of Large-Scale Systems*, edited by Javad Mohammadpour and Karolos Grigoriadis, Springer, 2010 is a welcome new addition to a collection of published books on LSS. Scanning the chapters of the book, it serves, in a way, as a bridge from LSS to SoS class of complex systems. Topics covered are modeling, model reduction, and distributed (decentralized) control, network, among others. In the rest of this century, system integration will appear in all aspects of our modern lives. Our *iPhoneTM*, laptop, *BlackberryTM*, a “smart-grid” distributed power generation systems, etc. are all examples of twenty-first century cyber-physical (or SoS) systems. One may say that LSS is moving into a new phase of “universal integration” systems thanks to the advances in wireless communication. The authors and editors of this volume deserve our thanks for adding a volume to this re-invigorated branch of control and systems engineering.

San Antonio, TX
November 2009

Prof. Mo Jamshidi

Preface

The complexity and dynamic order of controlled engineering systems is constantly increasing. Complex large scale dynamic systems – where “large” reflects the system’s order and not necessarily its physical size – appear in many engineering fields, such as, aerospace, micro-electro-mechanics, manufacturing, civil engineering and power systems. Modeling of these systems often results in high-order models imposing great challenges to their analysis, design, and feedback control. This edited volume is aimed at providing comprehensive contributions on recent analytical and computational methods for addressing the model reduction, performance analysis, and feedback control design of such systems. The book presents new theoretical developments, computational approaches, and illustrative applications to various fields. The scope of the book is intended to be interdisciplinary emphasizing the commonality and applicability of the methods to various fields.

The book comprises 12 chapters.

The first chapter, by A.C. Antoulas, C.A. Beattie, and S. Gugercin, presents a detailed and thorough exposition of interpolatory model order reduction methods for large-scale systems. Roughly speaking, these methods seek to obtain reduced-order models whose transfer function interpolates that of the original system at selected interpolation points in the frequency domain along selected input or output tangent directions. The question of the selection of interpolation points and tangential directions is examined in the context of meeting desired approximation goals. Finally, the methods are applied to systems with a generalized coprime factorization (such as second-order systems and systems with delays) and systems with a structured dependence on parameters that need to be retained in the reduced-order model. Examples from various disciplines conclude the chapter.

The chapter by R. Colgren examines the efficient model reduction of large-scale systems for control design. The motivation is based on meeting computational requirements for control and meeting practical implementation constraints of full-order dynamic controllers. A summary of the evolution of model reduction methods for control is presented with an emphasis on approaches that are easily applicable to large generalized models obtained by Computer Aided Design (CAD) and Finite Element Analysis (FEA) tools. A large-scale model of an air vehicle with significant aero-servo-elastic coupling is used to illustrate the methods.

The chapter by M.C. de Oliveira and A.S. Wroldsen presents a thorough study of the dynamic modeling and simulation of large tensegrity structural systems. Tensegrity structures are composed of just rods and strings, so that the rods are in compression and the strings in tension. Such large tensegrity systems can find applications as lightweight, deployable, and shape controllable structures. The authors derive the dynamic models for a general Class k tensegrity system and examine the implementation of these models in a computer simulation environment.

The chapter by X. Dai, T. Breikin, H. Wang, G. Kulikov, and V. Arkov investigates the data-driven reduced-order modeling of gas turbine engines for the purpose of on-board condition monitoring and fault detection. Model-based condition monitoring relies on the long-term predictive capabilities of reduced-order models that are compatible with limited on-board computational resources. A data-driven reduced-order model development is presented for gas turbine systems based on dynamic nonlinear least-squares identification. Methods to accelerate the speed of the identification algorithm are proposed using iterative calculation of the gradient and Hessian approximation. Data gathered in an aero-engine test bed are used to illustrate the reduced-order modeling approach.

The chapter by J. Lavaei, S. Sojoudi, and A.G. Aghdam examines the robust controllability and observability of a large-scale uncertain system with the objective of selecting dominant inputs and outputs to obtain a simplified control structure. The problem is formulated as the minimization of the smallest singular value of the corresponding Gramian in a polynomial uncertain system. It is shown that for such a system the Gramian is a rational function that can be approximated by a matrix polynomial. Subsequently, sum-of-squares (SoS) optimization techniques are employed for efficient computational solution, and simulation studies are used to validate the effectiveness of the proposed results.

The chapter by P. Krishnamurthy and F. Khorrami addresses the decentralized output feedback control of a class of interconnected nonlinear uncertain large-scale systems using a dynamic high-gain scaling approach. The method provides a unified framework for observer/controller design based on the solution of coupled state-dependent Lyapunov inequalities. Stability and disturbance attenuation of the proposed decentralized control solution are discussed in the input-to-output practical stability and integral-input-to-output practical stability frameworks.

The chapter by J. Xiong, V.A. Ugrinovskii, and I.R. Petersen examines the guaranteed-cost output feedback control problem for a class of large-scale uncertain stochastic systems with random parameters. It is assumed that the system uncertainties satisfy integral quadratic constraints and the random parameters follow a Markov process. Sufficient conditions are derived for decentralized controller synthesis that guarantees stability and a suboptimal level of quadratic performance. The control law uses local, in general non-Markovian, subsystem outputs and local subsystem operation modes to produce local subsystem control actions. Design conditions are provided in terms of rank constrained linear matrix inequalities (LMIs).

The chapter by M.S. Stankovic, D.M. Stipanovic, and S.S. Stankovic presents novel algorithms for decentralized overlapping control of large-scale complex systems. The approach is based on multiagent networks in which the agents utilize

dynamic consensus strategies to reach the agreement upon their actions. Different decentralized control structures are proposed and different algorithms are derived depending on the local control laws implemented by the agents. Properties and performance of the algorithms are discussed and an application is presented to the decentralized control of formations of unmanned aerial vehicles (UAVs).

The chapter by D. Zelazo and M. Mesbahi develops a network-centric analysis and synthesis framework for certain classes of large-scale interconnected systems. The systems under consideration involve linear dynamic subsystems that interact with other subsystems via an interconnection topology. The chapter examines the controllability and observability of such networked systems and investigates the network performance with respect to its \mathcal{H}_2 system norm. The effect of the structural properties of the network interconnection on its controllability, observability, and \mathcal{H}_2 norm performance are delineated. An algorithm for synthesizing optimal networks in the \mathcal{H}_2 system norm setting is presented.

The chapter by R.R. Negenborn, G. Hug-Glazmann, B. De Schutter, and G. Andersson, illustrates a novel coordination strategy for coordinating multiple control agents that control overlapping subnetworks in a network. The motivation stems from the distributed control of large-scale power networks. A simulation study on an adjusted IEEE 57-bus power network with Flexible Alternating Current Transmission Systems (FACTS) devices as controlled entities is used to validate the co-ordination strategy.

The chapter by J. Rice, P. Massioni, T. Keviczky, and M. Verhaegen, examines recently developed distributed control techniques for designing controllers for large-scale systems with sparse structure. The methods rely on the structural properties of the system and the associated control problem by assuming a sequentially semiseparable, decomposable, or identical subsystem architecture. A benchmark problem of the control of an infinite-dimensional car platoon in an \mathcal{H}_2 norm setting is used for a comparative study of the proposed methods.

The chapter by M. Meisami-Azad, J. Mohammadpour-Velni, K. Hiramoto, and K.M. Grigoriadis, investigates the integrated plant parameter and control parameter design in controlled large-scale structural systems. The objective is to integrate the structural parameter and the controller gain design steps to achieve an improved final design. To this end, an explicit upper bound of the $\mathcal{H}_2/\mathcal{H}_{\infty}$ norm of a collocated structural system is developed along with a parameterization of output feedback gains that guarantee such bounds. The proposed bounds are subsequently used in a convex optimization framework to design structural damping parameters and feedback control gains that meet closed-loop performance specifications.

We thank Steven Elliot, Senior Editor – Engineering in Springer for his assistance and most importantly the authors of the individual chapters for their contribution to this effort.

Houston, TX
May 2010

*Javad Mohammadpour
Karolos M. Grigoriadis*

Contents

Part I Model Reduction, Large-Scale System Modeling and Applications

Interpolatory Model Reduction of Large-Scale Dynamical Systems	3
Athanasios C. Antoulas, Christopher A. Beattie, and Serkan Gugercin	
1 Introduction	3
2 Problem Setting	4
2.1 The General Interpolation Framework	6
2.2 Model Reduction via Projection	8
2.3 Interpolatory Projections	10
2.4 Error Measures	14
3 Interpolatory Optimal \mathcal{H}_2 Approximation	18
3.1 An Algorithm for Interpolatory Optimal \mathcal{H}_2 Model Reduction	21
3.2 Numerical Results for IRKA	22
4 Interpolatory Passivity Preserving Model Reduction	25
4.1 An Example of Passivity Preserving Model Reduction	27
5 Structure-Preserving Model Reduction Using Generalized Coprime Factorizations	30
5.1 A Numerical Example: Driven Cavity Flow	33
5.2 Second-Order Dynamical Systems	35
6 Model Reduction of Parametric Systems	37
6.1 Numerical Example	40
7 Model Reduction from Measurements	41
7.1 Motivation: S-Parameters	41
7.2 The Loewner Matrix Pair and Construction of Interpolants	42
7.3 Loewner and Pick Matrices	46
7.4 Examples	47
8 Conclusions	53
References	53

Efficient Model Reduction for the Control of Large-Scale Systems	59
Richard Colgren	
1 Introduction	59
2 Spectral Decomposition	60
3 Simultaneous Gradient Error Reduction	62
4 Balancing	64
4.1 Techniques Not Requiring Balancing	66
4.2 Balancing Over A Disk	68
5 Example: Large-Scale System Application	69
References	71
Dynamics of Tensegrity Systems	73
Maurício C. de Oliveira and Anders S. Woldsen	
1 Introduction and Motivation	73
2 Dynamics of a Single Rigid Rod	74
2.1 Nodes as Functions of the Configuration	76
2.2 String Forces	77
2.3 Generalized Forces and Torques	78
2.4 Equations of Motion	78
3 Class 1 Tensegrity Systems	79
4 Class k Tensegrity Systems	82
4.1 A Class 2 Tensegrity Cable Model	82
5 Concluding Remarks	87
References	87
Modeling a Complex Aero-Engine Using Reduced Order Models	89
Xuewu Dai, Timofei Breikin, Hong Wang, Gennady Kulikov, and Valentin Arkov	
1 Introduction	89
2 Gas Turbine System	90
2.1 Nonlinear Static Model of Gas Turbine	91
2.2 Nonlinear Dynamic Model of Gas Turbine	92
3 Problem Formulation for Reduced Order Data Driven Modeling	93
3.1 Criterion Selection	94
3.2 Model Selection: EE vs. OE	96
4 NLS for OE Parameter Identification	99
4.1 Calculation of $\partial V(\theta_k)/\partial \theta$ and the Jacobian	100
4.2 Approximation of $R(k)$ and the Hessian	101
5 Application and Results	104
5.1 First-Order Model	105
5.2 Second-Order Model	109
6 Summary	110
References	110

Part II Large-Scale Systems Control and Applications

Robust Control of Large-Scale Systems: Efficient Selection of Inputs and Outputs	115
Javad Lavaei, Somayeh Sojoudi, and Amir G. Aghdam	
1 Introduction	115
2 Preliminaries and Problem Formulation	117
2.1 Background on Sum-of-Squares	118
3 Robust Controllability Degree	119
3.1 Special Case: A Polytopic Region	124
3.2 Comparison with Existing Results	128
4 Numerical Example	129
5 Summary	132
References	132
Decentralized Output-Feedback Control of Large-Scale Interconnected Systems via Dynamic High-Gain Scaling	135
P. Krishnamurthy and F. Khorrami	
1 Introduction	135
2 Decentralized Control Based on The Adaptive Dual Dynamic High-Gain Scaling Paradigm	137
2.1 Assumptions	137
2.2 Observer and Controller Designs	140
2.3 Stability Analysis	141
3 Generalized Scaling: Application to Decentralized Control	151
3.1 Assumptions	152
3.2 Observer Design	154
3.3 Controller Design	155
3.4 Stability Analysis	158
References	164
Decentralized Output Feedback Guaranteed Cost Control of Uncertain Markovian Jump Large-Scale Systems: Local Mode Dependent Control Approach	167
Junlin Xiong, Valery A. Ugrinovskii, and Ian R. Petersen	
1 Introduction	167
2 Problem Formulation	170
3 Guaranteed Cost Controller Design	173
3.1 Design Methodology	173
3.2 Design of Global Mode Dependent Controllers	176
3.3 The Main Result: Design of Local Mode Dependent Controllers	180
3.4 Design Procedure	182
4 An Illustrative Example	183
5 Conclusions	186

Appendix 1	187
Appendix 2	194
References	195
Consensus Based Multi-Agent Control Algorithms	197
Miloš S. Stanković, Dušan M. Stipanović, and Srdjan S. Stanković	
1 Introduction	197
2 Problem Formulation	199
3 Consensus at the Control Input Level	200
3.1 Algorithms Derived from the Local Dynamic Output Feedback Control Laws	200
3.2 Algorithms Derived from the Local Static Feedback Control Laws	205
4 Consensus at the State Estimation Level	207
5 Consensus Based Decentralized Control of UAV Formations	210
5.1 Formation Model	210
5.2 Global LQ Optimal State Feedback	212
5.3 Decentralized State Estimation	213
5.4 Experiments	215
References	217
Graph-Theoretic Methods for Networked Dynamic Systems:	
Heterogeneity and \mathcal{H}_2 Performance	219
Daniel Zelazo and Mehran Mesbahi	
1 Introduction	219
1.1 Preliminaries and Notations	221
2 Canonical Models of Networked Dynamic Systems	224
3 Analysis and Graph-Theoretic Performance Bounds	229
3.1 Observability and Controllability of NDS	229
3.2 Graph-Theoretic Bounds on NDS Performance	232
4 Topology Design for NDS	241
4.1 \mathcal{H}_2 Topology Design for NDS Coupled at the Output	242
4.2 Sensor Placement with \mathcal{H}_2 Performance for NDS Coupled at the State	244
5 Concluding Remarks	246
References	247
A Novel Coordination Strategy for Multi-Agent Control Using Overlapping Subnetworks with Application to Power Systems	251
R.R. Negenborn, G. Hug-Glazmann, B. De Schutter, and G. Andersson	
1 Introduction	251
1.1 Multi-Agent Control of Power Networks	252
1.2 Control of Subnetworks	253
1.3 Optimal Power Flow Control	255
1.4 Goal and Outline of This Chapter	256

2	Modeling of Network Characteristics and Control Objectives	256
2.1	Network Characteristics	256
2.2	Control Objectives	257
2.3	Definition of Subnetworks	257
3	Multi-Agent Control of Touching Subnetworks	258
3.1	Internal and External Nodes	258
3.2	Control Problem Formulation for One Agent	259
3.3	Control Scheme for Multiple Agents	262
4	Multi-Agent Control for Overlapping Subnetworks	263
4.1	Common Nodes	263
4.2	Control Problem Formulation for One Agent	264
4.3	Control Scheme for Multiple Agents	267
5	Application: Optimal Flow Control in Power Networks	268
5.1	Parameters of the Power Network	268
5.2	Steady-state Characteristics of Power Networks	268
5.3	Control Objectives	273
5.4	Setting Up the Control Problems	273
5.5	Simulations	274
6	Conclusions and Future Research	277
	References	277
Distributed Control Methods for Structured Large-Scale Systems		279
Justin Rice, Paolo Massioni, Tamás Keviczky, and Michel Verhaegen		
1	Introduction	279
2	Problem Statement	280
2.1	\mathcal{H}_2 Problem and Exact Solution	281
3	A Rational Laurent Operator Structure Preserving Iterative Approach to Distributed Control	282
3.1	L-Operator Sign Function	284
3.2	Definition	285
3.3	Convergence	285
3.4	Applications	286
3.5	Numerical Difficulties	287
3.6	Application to the Example Problem	288
4	Distributed Control Design for Decomposable Systems	288
4.1	General Description	289
4.2	Application to the Example Problem	292
5	Distributed LQR of Identical Systems	294
5.1	Special Properties of LQR for Dynamically Decoupled Systems	295
5.2	Application to the Example Problem	297
6	Numerical Results of the Car Platoons Benchmark Problem	299
7	Conclusions and Open Problems	301
	References	302

Integrated Design of Large-Scale Collocated Structural System and Control Parameters Using a Norm Upper Bound Approach	305
Mona Meisami-Azad, Javad Mohammadpour Velti, Kazuhiko Hiramoto, and Karolos M. Grigoriadis	
1 Introduction	305
2 Symmetric Output Feedback Control of Collocated Systems	307
3 Upper Bounds on Collocated Structural System Norms	308
4 Integrated Damping and Control Design Using the Analytical Bound Approach	311
4.1 Integrated Design Based on an \mathcal{H}^∞ Specification	311
4.2 Integrated Design Based on an \mathcal{H}^2 Specification	312
4.3 Integrated Design Based on a Mixed $\mathcal{H}^2/\mathcal{H}^\infty$ Specification	314
4.4 Decentralized Control Using the Norm Upper Bound Formulation	315
4.5 Additional Remarks	317
5 Simulation Results	317
6 Concluding Remarks	326
References	326
Index	329

List of Contributors

Amir G. Aghdam

Department of Electrical and Computer Engineering, Concordia University,
Montreal, Canada, H3G 2W1, e-mail: aghdam@ece.concordia.ca

Göran Andersson

Power Systems Laboratory, ETH Zürich, Physikstrasse 3, 8092 Zürich, Switzerland,
e-mail: andersson@eeh.ee.ethz.ch

Athanasis C. Antoulas

Department of Electrical and Computer Engineering, Rice University, Houston,
TX 77005-1892, USA, e-mail: aca@rice.edu

Valentin Arkov

Department of Automated Control Systems, Ufa State Aviation Technical
University, K. Marx Street 12, Ufa, 450000, Russia,
e-mail: t.breikin@manchester.ac.uk

Christopher A. Beattie

Department of Mathematics, Virginia Tech, Blacksburg, VA 24061-0123, USA,
e-mail: beattie@vt.edu

Timofei Breikin

Control Systems Center, School of Electrical and Electronic Engineering,
University of Manchester, Manchester, UK, M60 1QD,
e-mail: t.breikin@manchester.ac.uk

Richard D. Colgren

Viking Aerospace LLC, 100 Riverfront Rd., Suite B, Lawrence, KS 66044, USA,
e-mail: rcolgren@gmail.com

Xuewu Dai

School of Electronic and Information Engineering, Southwest University,
Chongqing, China, 400715, e-mail: dxw.dai@gmail.com

Maurício C. de Oliveira

Department of Mechanical and Aerospace Engineering, University of California San Diego, 9500 Gilman Dr., La Jolla CA, 92093-0411, USA,
e-mail: mauricio@ucsd.edu

Bart De Schutter

Delft Center for Systems and Control & Department of Marine and Transport Technology, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands, e-mail: b@deschutter.info

Karolos M. Grigoriadis

Department of Mechanical Engineering, University of Houston, Houston, TX 77204, USA, e-mail: karolos@uh.edu

Serkan Gugercin

Department of Mathematics, Virginia Tech, Blacksburg, VA 24061-0123, USA,
e-mail: gugercin@math.vt.edu

Kazuhiko Hiramoto

Department of Mechanical and Production Engineering, Niigata University, 8050 Ikarashi-2-no-cho, Nishi-ku, Niigata 950-2181, Japan,
e-mail: hiramoto@eng.niigata-u.ac.jp

Gabriela Hug-Glantzmann

Department of Electrical and Computer Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, USA, e-mail: ghug@andrew.cmu.edu

Tamás Keviczky

Delft Center for Systems and Control, Delft University of Technology, 2628 CD, Delft, The Netherlands, e-mail: t.keviczky@tudelft.nl

Farshad Khorrami

Control/Robotics Research Laboratory (CRRL), Department of Electrical and Computer Engineering, Polytechnic Institute of NYU, Brooklyn, NY 11201, USA, e-mail: khorrami@smart.poly.edu

Prashanth Krishnamurthy

Control/Robotics Research Laboratory (CRRL), Department of Electrical and Computer Engineering, Polytechnic Institute of NYU, Brooklyn, NY 11201, USA, e-mail: pk@crrl.poly.edu

Gennady Kulikov

Department of Automated Control Systems, Ufa State Aviation Technical University, K. Marx Street 12, Ufa, 450000, Russia,
e-mail: t.breikin@manchester.ac.uk

Javad Lavaei

Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125, USA, e-mail: lavaei@cds.caltech.edu

Paolo Massioni

Delft Center for Systems and Control, Delft University of Technology, 2628 CD,
Delft, The Netherlands, e-mail: p.massioni@tudelft.nl

Mona Meisami-Azad

Department of Mechanical Engineering, University of Houston, Houston,
TX 77204, USA, e-mail: mmeisami@mail.uh.edu

Mehran Mesbahi

Department of Aeronautics and Astronautics, University of Washington,
Box 352400, Seattle, WA 98195, USA, e-mail: mesbahi@aa.washington.edu

Javad Mohammadpour

Department of Mechanical Engineering, University of Houston, Houston,
TX 77204, USA, e-mail: jmohammadpour@uh.edu

Rudy R. Negenborn

Delft Center for Systems and Control, Delft University of Technology, Mekelweg
2, 2628 CD Delft, The Netherlands, e-mail: r.r.negenborn@tudelft.nl

Ian R. Petersen

School of Engineering and Information Technology, University of New South
Wales at the Australian Defence Force Academy, Northcott Drive, Canberra,
ACT 2600, Australia, e-mail: i.r.petersen@gmail.com

Justin Rice

Delft Center for Systems and Control, Delft University of Technology, 2628 CD,
Delft, The Netherlands, e-mail: j.k.rice@tudelft.nl

Somayeh Sojoudi

Control and Dynamical Systems, California Institute of Technology, Pasadena,
CA 91125, USA, e-mail: sojoudi@cds.caltech.edu

Miloš S. Stanković

ACCESS Linnaeus Center, School of Electrical Engineering, Royal Institute
of Technology, 100-44 Stockholm, Sweden, e-mail: milsta@kth.se

Srdjan S. Stanković

School of Electrical Engineering, University of Belgrade, Serbia,
e-mail: stankovic@etf.rs

Dušan M. Stipanović

Department of Industrial and Enterprise Systems Engineering and the Coordinated
Science Laboratory, University of Illinois at Urbana-Champaign, IL, USA,
e-mail: dusan@illinois.edu

Valery A. Ugrinovskii

School of Engineering and Information Technology, University of New South
Wales at the Australian Defence Force Academy, Northcott Drive, Canberra,
ACT 2600, Australia, e-mail: v.ugrinovskii@gmail.com

Michel Verhaegen

Delft Center for Systems and Control, Delft University of Technology, 2628 CD,
Delft, The Netherlands, e-mail: m.verhaegen@moesp.org

Hong Wang

Control Systems Center, School of Electrical and Electronic Engineering,
University of Manchester, Manchester M60 1QD, UK,
e-mail: hong.wang@manchester.ac.uk

Anders S. Wroldsen

Center for Ships and Ocean Structures, Norwegian University of Science
and Technology, Trondheim, Norway, e-mail: wroldsen@ntnu.no

Junlin Xiong

Department of Automation, University of Science and Technology of China,
Hefei 230026, China, e-mail: junlin.xiong@gmail.com

Daniel Zelazo

Institute for Systems Theory and Automatic Control, University of Stuttgart,
Pfaffenwaldring 9, 70550 Stuttgart, Germany,
e-mail: Daniel.Zelazo@ist.uni-stuttgart.de

Part I

**Model Reduction, Large-Scale System
Modeling and Applications**

Interpolatory Model Reduction of Large-Scale Dynamical Systems

Athanasis C. Antoulas, Christopher A. Beattie, and Serkan Gugercin

1 Introduction

Large scale dynamical systems are a common framework for the modeling and control of many complex phenomena of scientific interest and industrial value, with examples of diverse origin that include signal propagation and interference in electric circuits, storm surge prediction before an advancing hurricane, vibration suppression in large structures, temperature control in various media, neuro-transmission in the nervous system, and behavior of micro-electro-mechanical systems. Direct numerical simulation of underlying mathematical models is one of few available means for accurate prediction and control of these complex phenomena. The need for ever greater accuracy compels inclusion of greater detail in the model and potential coupling to other complex systems leading inevitably to very large-scale and complex dynamical models. Simulations in such large-scale settings can make untenable demands on computational resources and efficient model utilization becomes necessary. *Model reduction* is one response to this challenge, wherein one seeks a simpler (typically lower order) model that nearly replicates the behavior of the original model. When high fidelity is achieved with a reduced-order model, it can then be used reliably as an efficient surrogate to the original, perhaps replacing it as a component in larger simulations or in allied contexts such as development of simpler, faster controllers suitable for real time applications.

A.C. Antoulas

Department of Electrical and Computer Engineering, Rice University, Houston,
TX 77005-1892, USA

e-mail: aca@rice.edu

C.A. Beattie

Department of Mathematics, Virginia Tech, Blacksburg, VA 24061-0123, USA
e-mail: beattie@vt.edu

S. Gugercin

Department of Mathematics, Virginia Tech, Blacksburg, VA 24061-0123, USA
e-mail: gugercin@math.vt.edu

Interpolatory model reduction methods have emerged as effective strategies for approximation of large-scale linear dynamical systems. These methods produce reduced models whose transfer function interpolates the original system transfer function at selected interpolation points. They are closely related to rational Krylov-based methods and indeed, are precisely that for single input/single output systems. One main reason why interpolatory model reduction has become the method of choice in true large-scale settings is that this class of methods is numerically stable and well-suited to large scale computations; the model reduction process does not require any dense matrix transformations such as singular value decompositions. Moreover, the approximation theory behind this approach is similar to that of rational interpolants used to approximate meromorphic functions. The goal of this contribution is to present a survey of model reduction by interpolation and provide a selection of the most recent developments in this direction.

2 Problem Setting

Linear dynamical systems are principally characterized through their input–output map $\mathcal{S} : \mathbf{u} \mapsto \mathbf{y}$, mapping inputs \mathbf{u} to outputs \mathbf{y} via a state-space realization given as:

$$\mathcal{S} : \begin{cases} \mathbf{E}\dot{\mathbf{x}}(t) = \mathbf{Ax}(t) + \mathbf{Bu}(t) \\ \mathbf{y}(t) = \mathbf{Cx}(t) + \mathbf{Du}(t) \end{cases} \quad \text{with } \mathbf{x}(0) = \mathbf{0}, \quad (1)$$

where $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times m}$, $\mathbf{C} \in \mathbb{R}^{p \times n}$, and $\mathbf{D} \in \mathbb{R}^{p \times m}$ are constant matrices. In (1), $\mathbf{x}(t) \in \mathbb{R}^n$, $\mathbf{u}(t) \in \mathbb{R}^m$ and $\mathbf{y}(t) \in \mathbb{R}^p$ are, respectively, an internal variable (the state if \mathbf{E} is non-singular), the input and the output of the system \mathcal{S} . We refer to \mathcal{S} as a *single-input/single-output* (SISO) system when $m = p = 1$ (scalar-valued input, scalar-valued output) and as a *multi-input/multi-output* (MIMO) system otherwise.

Systems of the form (1) with extremely large state-space dimension n arise in many disciplines; see [2] and [57] for a collection of such examples. Despite large state-space dimension, in most cases, the state space trajectories, $\mathbf{x}(t)$, hew closely to subspaces with substantially lower dimensions and evolve in ways that do not fully occupy the state space. The original model \mathcal{S} behaves nearly as if it had many fewer internal degrees-of-freedom, in effect, much lower state-space dimension. *Model reduction* seeks to produce a surrogate dynamical system that evolves in a much lower dimensional state space (say, dimension $r \ll n$), yet is able to mimic the original dynamical system, recovering very nearly the original input–output map. We want the reduced input–output map, $\mathcal{S}_r : \mathbf{u} \mapsto \mathbf{y}_r$, to be close to \mathcal{S} in an appropriate sense.

Being a smaller version of the original dynamical model, the input–output map \mathcal{S}_r is described by the reduced system in state-space form as:

$$\mathcal{S}_r : \begin{cases} \mathbf{E}_r \dot{\mathbf{x}}_r(t) = \mathbf{A}_r \mathbf{x}_r(t) + \mathbf{B}_r \mathbf{u}(t) \\ \mathbf{y}_r(t) = \mathbf{C}_r \mathbf{x}_r(t) + \mathbf{D}_r \mathbf{u}(t) \end{cases} \quad \text{with } \mathbf{x}_r(0) = \mathbf{0}, \quad (2)$$

where $\mathbf{A}_r, \mathbf{E}_r \in \mathbb{R}^{r \times r}$, $\mathbf{B}_r \in \mathbb{R}^{r \times m}$, $\mathbf{C}_r \in \mathbb{R}^{p \times r}$ and $\mathbf{D}_r \in \mathbb{R}^{p \times m}$. Note that the number of inputs, m , and the number of outputs, p , are the same for both the original and reduced models; only the internal state-space dimensions differ: $r \ll n$. A successful reduced-order model should meet the following criteria:

Goals for Reduced Order Models

1. The reduced input–output map S_r should be uniformly close to S in an appropriate sense. That is, when presented with the same inputs, $\mathbf{u}(t)$, the difference between full and reduced system outputs, $\mathbf{y} - \mathbf{y}_r$, should be *small* with respect to a physically relevant norm over a wide range of system inputs, such as over all \mathbf{u} with bounded energy (e.g., in the unit ball of $L^2([0, \infty), \mathbb{R}^m)$).
2. Critical system features and structure should be preserved in the reduced order system. This can include passivity, Hamiltonian structure, subsystem interconnectivity, or second-order structure.
3. Strategies for obtaining $\mathbf{E}_r, \mathbf{A}_r, \mathbf{B}_r, \mathbf{C}_r$, and \mathbf{D}_r should lead to robust, numerically stable algorithms and furthermore require minimal application-specific tuning with little to no expert intervention. It is important that model reduction methods be computationally efficient and reliable so that very large problems remain tractable; they should be robust and largely automatic to allow the broadest level of flexibility and applicability in complex multiphysics settings.

Whenever the input $\mathbf{u}(t)$ is exponentially bounded – that is, when there is a fixed $\gamma \in \mathbb{R}$ such that $\|\mathbf{u}(t)\| \sim \mathcal{O}(e^{\gamma t})$, then $\mathbf{x}(t)$ and $\mathbf{y}(t)$ from (1) and $\mathbf{x}_r(t)$ and $\mathbf{y}_r(t)$ from (2) will also be exponentially bounded and the Laplace transform can be applied to (1) and (2) to obtain

$$\hat{\mathbf{y}}(s) = (\mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1} \mathbf{B} + \mathbf{D}) \hat{\mathbf{u}}(s), \quad (3)$$

$$\hat{\mathbf{y}}_r(s) = (\mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{B}_r + \mathbf{D}_r) \hat{\mathbf{u}}(s), \quad (4)$$

where we have denoted Laplace transformed quantities with “ $\hat{\cdot}$ ”. We define transfer functions accordingly – from (3) and (4)

$$\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1} \mathbf{B} + \mathbf{D} \quad (5)$$

$$\mathbf{H}_r(s) = \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{B}_r + \mathbf{D}_r \quad (6)$$

so that

$$\hat{\mathbf{y}}(s) - \hat{\mathbf{y}}_r(s) = [\mathbf{H}(s) - \mathbf{H}_r(s)] \hat{\mathbf{u}}(s)$$

For a given input, the loss in fidelity in passing from the full model to a reduced model can be associated with the difference between respective system transfer functions over a range of inputs. This is evaluated more precisely in Sect. 2.4.

It is a fact of life that many dynamical systems of critical interest are *nonlinear*. By contrast, we will consider here only linear dynamical systems and the methods we discuss fundamentally exploit this linearity. We do this unapologetically but note that by narrowing the scope of the problems that one considers in this way, one is able to be far more ambitious and demanding in the quality of outcomes that are achieved, and indeed, the techniques we discuss here are often dramatically successful. To the extent that many methods for approaching nonlinear systems build upon the analysis of carefully chosen linearizations of these systems, the methods we describe here can be expected to play a role in evolving strategies for the reduction of large scale nonlinear dynamical systems as well. See, for instance, [12].

2.1 The General Interpolation Framework

Interpolation is a very simple and yet effective approach that is used ubiquitously for the general approximation of complex functions using simpler ones. The construction of such approximations is easy: indeed, school children learn how to construct piecewise linear approximations and Taylor polynomial approximations to functions through interpolation. The accuracy of the resulting approximations and the connections with strategic placement of interpolating points has been studied in many broad contexts – indeed, in the case of interpolation of meromorphic functions by polynomials or rational functions, the associated error analysis constitutes a large body of work tied in closely with potential theory and classical complex analysis. Many of these recognized advantages of approximation via interpolation translate immediately into boons in our setting as well as we seek to approximate a “difficult” transfer function, $\mathbf{H}(s)$, with a simpler one, $\mathbf{H}_r(s)$.

Simply stated, our overarching goal is to produce a low order transfer function, $\mathbf{H}_r(s)$, that approximates a large order transfer function, $\mathbf{H}(s)$ with high fidelity: for order $r \ll n$, we want $\mathbf{H}_r(s) \approx \mathbf{H}(s)$ in a sense that will be made precise later. Interpolation is the primary vehicle with which we approach this problem. Naively, we could select a set of points $\{\sigma_i\}_{i=1}^r \subset \mathbb{C}$ and then seek a reduced order transfer function, $\mathbf{H}_r(s)$, such that $\mathbf{H}_r(\sigma_i) = \mathbf{H}(\sigma_i)$ for $i = 1, \dots, r$. This is a good starting place for SISO systems but turns out to be overly restrictive for MIMO systems, since the condition $\mathbf{H}_r(\sigma_i) = \mathbf{H}(\sigma_i)$ in effect imposes $m \cdot p$ scalar conditions at each interpolation point. This may not be realizable.

A far more advantageous formulation is to consider interpolation conditions that are imposed at specified interpolation points as above but that only act in specified directions, that is, *tangential interpolation*. We state this formally as part of our first problem of focus:

Problem 1: Model Reduction Given State Space System Data

Given a full-order model (1) specified through knowledge of the system matrices, \mathbf{E} , \mathbf{A} , \mathbf{B} , \mathbf{C} , and \mathbf{D} and given

left interpolation points:

$$\{\mu_i\}_{i=1}^q \subset \mathbb{C},$$

with corresponding

left tangent directions:

$$\{\tilde{\mathbf{c}}_i\}_{i=1}^q \subset \mathbb{C}^p,$$

right interpolation points:

$$\{\sigma_j\}_{j=1}^r \subset \mathbb{C}$$

with corresponding

right tangent directions:

$$\{\tilde{\mathbf{b}}_j\}_{j=1}^r \subset \mathbb{C}^m.$$

Find a reduced-order model (2) through specification of reduced system matrices \mathbf{E}_r , \mathbf{A}_r , \mathbf{B}_r , \mathbf{C}_r , and \mathbf{D}_r such that the associated transfer function, $\mathbf{H}_r(s)$, in (6) is a *tangential interpolant* to $\mathbf{H}(s)$ in (5):

$$\tilde{\mathbf{c}}_i^T \mathbf{H}_r(\mu_i) = \tilde{\mathbf{c}}_i^T \mathbf{H}(\mu_i) \quad \text{and} \quad \mathbf{H}_r(\sigma_j) \tilde{\mathbf{b}}_j = \mathbf{H}(\sigma_j) \tilde{\mathbf{b}}_j, \quad (7)$$

for $i = 1, \dots, q$, for $j = 1, \dots, r$,

Interpolation points and tangent directions are selected to realize the model reduction goals described on page 5.

It is of central concern to determine effective choices for interpolation points and tangent directions. In Sect. 3, we discuss the selection of interpolation points and tangent directions that can yield *optimal* reduced order models with respect to \mathcal{H}_2 quality measures of system approximation. In Sect. 4, we describe selection of interpolation points and tangent directions that guarantee passivity is preserved in the reduced-order model. Preservation of other types of system structure such as second-order systems, systems that involve delays or memory terms, and systems having a structured dependence on parameters is discussed in Sects. 5 and 6.

Tangential interpolation is a remarkably flexible framework within which to consider model reduction when one considers the significance of the interpolation data. Note that if $\tilde{\mathbf{y}} = \mathbf{H}(\sigma) \tilde{\mathbf{b}}$ then $e^{\sigma t} \tilde{\mathbf{y}}$ is precisely the response of the full order system to a pure input given by $\mathbf{u}(t) = e^{\sigma t} \tilde{\mathbf{b}}$, so the tangential interpolation conditions that characterize $\mathbf{H}_r(s)$ could (at least in principle) be obtained from *measured input-output data* drawn directly from observations on the original system. For example, if $\sigma = i\omega_0$, one is observing in $\tilde{\mathbf{y}}$, the sinusoidal response of the system to a pure tone input of frequency ω_0 . Similarly, if the *dual* dynamical system were driven by an input given by $e^{\mu t} \tilde{\mathbf{c}}$ producing an output $e^{\mu t} \tilde{\mathbf{z}}$ then $\tilde{\mathbf{c}}^T \mathbf{H}(\mu) = \tilde{\mathbf{z}}^T$. This creates the following alternative problem setting that we consider based entirely on observed input-output response data and requiring no other a priori information about system-related quantities.

Problem 2: Model Reduction Given Input–Output Data

Given a set of input–output response measurements on the full-order system specified by

left driving frequencies:

$$\{\mu_i\}_{i=1}^q \subset \mathbb{C},$$

using *left input directions:*

$$\{\tilde{\mathbf{c}}_i\}_{i=1}^q \subset \mathbb{C}^p,$$

producing *left responses:*

$$\{\tilde{\mathbf{z}}_i\}_{i=1}^q \subset \mathbb{C}^m,$$

right driving frequencies:

$$\{\sigma_i\}_{i=1}^r \subset \mathbb{C}$$

using *right input directions:*

$$\{\tilde{\mathbf{b}}_i\}_{i=1}^r \subset \mathbb{C}^m$$

producing *right responses:*

$$\{\tilde{\mathbf{y}}_i\}_{i=1}^r \subset \mathbb{C}^p$$

Find a dynamical system model (2) by specifying (reduced) system matrices \mathbf{E}_r , \mathbf{A}_r , \mathbf{B}_r , \mathbf{C}_r , and \mathbf{D}_r such that the associated transfer function, $\mathbf{H}_r(s)$, in (6) is a *tangential interpolant* to the given data:

$$\begin{aligned} \tilde{\mathbf{c}}_i^T \mathbf{H}_r(\mu_i) &= \tilde{\mathbf{z}}_i^T & \text{and} & \mathbf{H}_r(\sigma_j) \tilde{\mathbf{b}}_j &= \tilde{\mathbf{y}}_j, \\ \text{for } i = 1, \dots, q, & & & \text{for } j = 1, \dots, r, & \end{aligned} \quad (8)$$

Interpolation points and tangent directions are determined (typically) by the availability of experimental data.

The structure of Problem 2 introduces difficulties involved in the requirement that noisy and inconsistent data must be accommodated appropriately – in this case, we describe how to construct a reduced order system described by the system matrices \mathbf{E}_r , \mathbf{A}_r , \mathbf{B}_r , \mathbf{C}_r , and \mathbf{D}_r so that experimentally observed response data (reformulated as approximate interpolation conditions) are matched at least approximately: $\mathbf{H}_r(\sigma_i) \tilde{\mathbf{b}}_i \approx \tilde{\mathbf{y}}_i$ and $\tilde{\mathbf{c}}_j^T \mathbf{H}_r(\mu_j) \approx \tilde{\mathbf{z}}_j^T$.

Numerous issues arise spontaneously for both problem types and we discuss here what we feel is a representative subset. Note for both problems, it is necessary to have a computationally stable method for constructing the (reduced) system matrices \mathbf{E}_r , \mathbf{A}_r , \mathbf{B}_r , \mathbf{C}_r , and \mathbf{D}_r that produces an associated transfer function, $\mathbf{H}_r(s)$, satisfying the interpolation conditions. We discuss this in Sects. 2.2 and 2.3 for Problem 1 and Sect. 7.2 for Problem 2.

2.2 Model Reduction via Projection

When the full-order dynamical system \mathcal{S} described in (1) has been specified via the state-space matrices \mathbf{E} , \mathbf{A} , \mathbf{B} , \mathbf{C} , and \mathbf{D} , most model reduction methods proceed with some variation of a Petrov–Galerkin projective approximation to construct a reduced-order model \mathcal{S}_r . We motivate this approach by describing the evolution of the full order model (1) in an indirect way – although the systems described in

(1) and (2) typically are real-valued, we find it convenient to allow all quantities involved to be complex-valued:

Find $\mathbf{x}(t)$ contained in \mathbb{C}^n such that

$$\mathbf{E}\dot{\mathbf{x}}(t) - \mathbf{A}\mathbf{x}(t) - \mathbf{B}\mathbf{u}(t) \perp \mathbb{C}^n \text{ (i.e., } = 0).$$

Then the associated output is $\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t)$.

The Petrov-Galerkin projective approximation proceeds by replacing \mathbb{C}^n in the above description with (nominally small) subspaces of \mathbb{C}^n : we choose an r -dimensional trial subspace, the *right modeling subspace*, $\mathcal{V}_r \subset \mathbb{C}^n$, and an r -dimensional test subspace, the *left modeling subspace*, $\mathcal{W}_r \subset \mathbb{C}^n$ – and then we describe the evolution of a reduced order model in this way:

Find $\mathbf{v}(t)$ contained in \mathcal{V}_r such that

$$\mathbf{E}\dot{\mathbf{v}}(t) - \mathbf{A}\mathbf{v}(t) - \mathbf{B}\mathbf{u}(t) \perp \mathcal{W}_r. \quad (9)$$

Then the associated output is $\mathbf{y}_r(t) = \mathbf{C}\mathbf{v}(t) + \mathbf{D}\mathbf{u}(t)$.

The dynamics described by (9) can be represented as a dynamical system evolving in a reduced order state-space as in (2) once bases are chosen for the two subspaces \mathcal{V}_r and \mathcal{W}_r . Let $\text{Ran}(\mathbf{R})$ denote the range of a matrix \mathbf{R} . Let $\mathbf{V}_r \in \mathbb{C}^{n \times r}$ and $\mathbf{W}_r \in \mathbb{C}^{n \times r}$ be matrices defined so that $\mathcal{V}_r = \text{Ran}(\mathbf{V}_r)$ and $\mathcal{W}_r = \text{Ran}(\overline{\mathbf{W}}_r)$. We can represent the reduced system trajectories as $\mathbf{v}(t) = \mathbf{V}_r \mathbf{x}_r(t)$ with $\mathbf{x}_r(t) \in \mathbb{C}^r$ for each t and the Petrov-Galerkin approximation (9) can be rewritten as

$$\mathbf{W}_r^T (\mathbf{E}\mathbf{V}_r \dot{\mathbf{x}}_r(t) - \mathbf{A}\mathbf{V}_r \mathbf{x}_r(t) - \mathbf{B}\mathbf{u}(t)) = \mathbf{0} \quad \text{and} \quad \mathbf{y}_r(t) = \mathbf{C}\mathbf{V}_r \mathbf{x}_r(t) + \mathbf{D}\mathbf{u}(t),$$

leading to the reduced order state-space representation (2) with

$$\begin{aligned} \mathbf{E}_r &= \mathbf{W}_r^T \mathbf{E} \mathbf{V}_r, & \mathbf{B}_r &= \mathbf{W}_r^T \mathbf{B}, \\ \mathbf{A}_r &= \mathbf{W}_r^T \mathbf{A} \mathbf{V}_r, & \mathbf{C}_r &= \mathbf{C} \mathbf{V}_r, \end{aligned} \quad \text{and} \quad \mathbf{D}_r = \mathbf{D}. \quad (10)$$

We note that the definitions of reduced-order quantities in (10) are invariant under change of basis for the original state space, so the quality of reduced approximations evidently will depend only on effective choices for the right modeling space $\mathcal{V}_r = \text{Ran}(\mathbf{V}_r)$ and the left modeling space $\mathcal{W}_r = \text{Ran}(\overline{\mathbf{W}}_r)$. We choose the modeling subspaces to enforce interpolation which allows us to shift our focus to how best to choose effective interpolation points and tangent directions.

Since $\mathbf{D} \in \mathbb{R}^{p \times m}$ and both p and m are typically of only modest size, usually \mathbf{D} does not play a significant role in the cost of simulation and $\mathbf{D}_r = \mathbf{D}$ is both a common choice and a natural one arising in the context of a Petrov-Galerkin approximation as described in (10). Note that if \mathbf{E} and \mathbf{E}_r are both nonsingular then the choice $\mathbf{D}_r = \mathbf{D}$ also enforces interpolation at infinity: $\lim_{s \rightarrow \infty} \mathbf{H}(s) = \lim_{s \rightarrow \infty} \mathbf{H}_r(s) = \mathbf{D}$, facilitating, in effect, a good match between true and reduced system outputs for sufficiently high frequency inputs. The case that \mathbf{E} is singular and other performance goals can motivate choices for $\mathbf{D}_r \neq \mathbf{D}$. This is discussed in some detail on p. 13.

2.3 Interpolatory Projections

Given the full-order dynamical system $\mathbf{H}(s)$, a set of r right interpolation points $\{\sigma_i\}_{i=1}^r \in \mathbb{C}$, with right directions $\{b_i\}_{i=1}^r \in \mathbb{C}^m$, and a set of q left interpolation points $\{\mu_j\}_{j=1}^q \in \mathbb{C}$, with left-tangential directions $\{c_i\}_{i=1}^q \in \mathbb{C}^p$, interpolatory model reduction involves finding a reduced-order model $\mathbf{H}_r(s)$ that tangentially interpolates $\mathbf{H}(s)$ using this given data: find $\mathbf{H}_r(s)$ so that

$$\begin{aligned}\mathbf{H}(\sigma_i)b_i &= \mathbf{H}_r(\sigma_i)b_i, & \text{for } i = 1, \dots, r, \\ c_j^T \mathbf{H}(\mu_j) &= c_j^T \mathbf{H}_r(\mu_j), & \text{for } j = 1, \dots, r,\end{aligned}\tag{11}$$

We wish to interpolate $\mathbf{H}(s)$ without ever computing the quantities to be matched since these numbers are numerically ill-conditioned, as Feldman and Freund [33] illustrated for the special case of single-input/single-output dynamical systems. Remarkably, this can be achieved by employing Petrov-Galerkin projections with carefully chosen test and trial subspaces.

Interpolatory projections for model reduction were initially introduced by Skelton et. al. in [30, 90, 91]. Grimme [42], later, modified this approach into a numerically efficient framework by utilizing the rational Krylov subspace method of Ruhe [79]. The projection framework for the problem setting we are interested in, that is, for rational tangential interpolation of MIMO dynamical systems, has been recently developed by Gallivan et al. [39] for the case of Problem 1 and by Mayo and Antoulas [66] for the case of Problems 2 as laid out in §2.1. We will be using these frameworks throughout the manuscript.

We first illustrate and prove how to solve the rational tangential interpolation problem in (11) by projection. The general case for matching derivatives of the transfer function (in effect, generalized Hermite interpolation) will be presented in Theorem 1.2. As we point out above, the Petrov-Galerkin approximation described in (10) leads to the choice of $\mathbf{D}_r = \mathbf{D}$ in a natural way, but other choices are possible. We will first assume $\mathbf{D}_r = \mathbf{D}$, so that $\mathbf{H}_r(s)$ in Theorems 1.1 and 1.2 below is assumed to be obtained as in (10).

Theorem 1.1. *Let $\sigma, \mu \in \mathbb{C}$ be such that $s\mathbf{E} - \mathbf{A}$ and $s\mathbf{E}_r - \mathbf{A}_r$ are invertible for $s = \sigma, \mu$. Also let $\mathbf{V}_r, \mathbf{W}_r \in \mathbb{C}^{n \times r}$ in (10) have full-rank. If $\mathbf{b} \in \mathbb{C}^m$ and $\mathbf{c} \in \mathbb{C}^\ell$ are fixed nontrivial vectors then*

- (a) if $(\sigma\mathbf{E} - \mathbf{A})^{-1} \mathbf{B}\mathbf{b} \in \text{Ran}(\mathbf{V}_r)$, then $\mathbf{H}(\sigma)\mathbf{b} = \mathbf{H}_r(\sigma)\mathbf{b}$;
- (b) if $\left(c^T \mathbf{C}(\mu\mathbf{E} - \mathbf{A})^{-1}\right)^T \in \text{Ran}(\mathbf{W}_r)$, then $c^T \mathbf{H}(\mu) = c^T \mathbf{H}_r(\mu)$; and
- (c) if both (a) and (b) hold, and $\sigma = \mu$, then $c^T \mathbf{H}'(\sigma)\mathbf{b} = c^T \mathbf{H}'_r(\sigma)\mathbf{b}$ as well.

Proof. We follow the recent proofs provided in [49] and [16]. Define

$$\mathcal{P}_r(z) = \mathbf{V}_r(z\mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{W}_r^T (z\mathbf{E}_r - \mathbf{A}) \quad \text{and}$$

$$\mathcal{Q}_r(z) = (z\mathbf{E} - \mathbf{A}) \mathcal{P}_r(z) (z\mathbf{E} - \mathbf{A})^{-1} = (z\mathbf{E} - \mathbf{A}) \mathbf{V}_r (z\mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{W}_r^T.$$

Both $\mathcal{P}_r(z)$ and $\mathcal{Q}_r(z)$ are analytic matrix-valued functions in neighborhoods of $z = \sigma$ and $z = \mu$. It is easy to verify that both $\mathcal{P}_r(z)$ and $\mathcal{Q}_r(z)$ are projectors, i.e. $\mathcal{P}_r^2(z) = \mathcal{P}_r(z)$ and $\mathcal{Q}_r^2(z) = \mathcal{Q}_r(z)$. Moreover, for all z in a neighborhood of σ , $\mathcal{V}_r = \text{Ran}(\mathcal{P}_r(z)) = \text{Ker}(\mathbf{I} - \mathcal{P}_r(z))$ and $\mathcal{W}_r^\perp = \text{Ker}(\mathcal{Q}_r(z)) = \text{Ran}(\mathbf{I} - \mathcal{Q}_r(z))$ where $\text{Ran}(\mathbf{R})$ denotes the kernel of a matrix \mathbf{R} . First observe that

$$\mathbf{H}(z) - \mathbf{H}_r(z) = \mathbf{C}(z\mathbf{E} - \mathbf{A})^{-1} (\mathbf{I} - \mathcal{Q}_r(z)) (z\mathbf{E} - \mathbf{A}) (\mathbf{I} - \mathcal{P}_r(z)) (z\mathbf{I} - \mathbf{A})^{-1} \mathbf{B} \quad (12)$$

Evaluating this expression at $z = \sigma$ and post-multiplying by \mathbf{b} yields the first assertion; evaluating (12) at $z = \mu$ and pre-multiplying by \mathbf{c}^T yields the second. Note

$$\begin{aligned} ((\sigma + \varepsilon)\mathbf{E} - \mathbf{A})^{-1} &= (\sigma\mathbf{E} - \mathbf{A})^{-1} - \varepsilon(\sigma\mathbf{E} - \mathbf{A})^{-1}\mathbf{E}(\sigma\mathbf{E} - \mathbf{A})^{-1} + \mathcal{O}(\varepsilon^2) \\ ((\sigma + \varepsilon)\mathbf{E}_r - \mathbf{A}_r)^{-1} &= (\sigma\mathbf{E}_r - \mathbf{A}_r)^{-1} - \varepsilon(\sigma\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{E}_r(\sigma\mathbf{E}_r - \mathbf{A}_r)^{-1} + \mathcal{O}(\varepsilon^2) \end{aligned}$$

so evaluating (12) at $z = \sigma + \varepsilon$, premultiplying by \mathbf{c}^T , and postmultiplying by \mathbf{b} under the hypotheses of the third assertion yields

$$\mathbf{c}^T \mathbf{H}(\sigma + \varepsilon) \mathbf{b} - \mathbf{c}^T \mathbf{H}_r(\sigma + \varepsilon) \mathbf{b} = \mathcal{O}(\varepsilon^2).$$

Now observe that since $\mathbf{c}^T \mathbf{H}(\sigma) \mathbf{b} = \mathbf{c}^T \mathbf{H}_r(\sigma) \mathbf{b}$,

$$\frac{1}{\varepsilon} (\mathbf{c}^T \mathbf{H}(\sigma + \varepsilon) \mathbf{b} - \mathbf{c}^T \mathbf{H}(\sigma) \mathbf{b}) - \frac{1}{\varepsilon} (\mathbf{c}^T \mathbf{H}_r(\sigma + \varepsilon) \mathbf{b} - \mathbf{c}^T \mathbf{H}_r(\sigma) \mathbf{b}) \rightarrow 0,$$

as $\varepsilon \rightarrow 0$, which proves the third assertion. \square

Theorem 1.1 shows that given a set of distinct right interpolation points (“right shifts”) $\{\sigma_i\}_{i=1}^r$, a set of distinct interpolation left points (“left shifts”) $\{\mu_j\}_{j=1}^r$, left tangential directions $\{\mathbf{c}_k\}_{k=1}^r \in \mathbb{C}^p$, and right tangential directions $\{\mathbf{b}_k\}_{k=1}^r \in \mathbb{C}^m$, the solution of the tangential rational Hermite interpolation problem is straightforward and constructive. One simply computes matrices \mathbf{V}_r and \mathbf{W}_r such that

$$\mathbf{V}_r = [(\sigma_1\mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \mathbf{b}_1, \dots, (\sigma_r\mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \mathbf{b}_r] \text{ and} \quad (13)$$

$$\mathbf{W}_r^T = \begin{bmatrix} \mathbf{c}_1^T \mathbf{C}(\mu_1\mathbf{E} - \mathbf{A})^{-1} \\ \vdots \\ \mathbf{c}_r^T \mathbf{C}(\mu_r\mathbf{E} - \mathbf{A})^{-1} \end{bmatrix}. \quad (14)$$

The reduced order system $\mathbf{H}_r(s) = \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{B}_r$ defined by (10) then solves the tangential Hermite interpolation problem provided that $\sigma_i\mathbf{E}_r - \mathbf{A}_r$ and $\mu_i\mathbf{E}_r - \mathbf{A}_r$ are nonsingular for each $i = 1, \dots, r$.

For completeness, we describe how to solve the analogous generalized Hermite interpolation problem, which involves matching transfer function values and higher derivative values. The m th derivative of $\mathbf{H}(s)$ with respect to s evaluated at $s = \sigma$, will be denoted by $\mathbf{H}^{(m)}(\sigma)$.

Theorem 1.2. Let $\sigma, \mu \in \mathbb{C}$ be such that $s\mathbf{E} - \mathbf{A}$ and $s\mathbf{E}_r - \mathbf{A}_r$ are invertible for $s = \sigma, \mu$. If $\mathbf{b} \in \mathbb{C}^m$ and $\mathbf{c} \in \mathbb{C}^\ell$ are fixed nontrivial vectors then

- (a) if $\left((\sigma\mathbf{E} - \mathbf{A})^{-1}\mathbf{E}\right)^{j-1}(\sigma\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\mathbf{b} \in \text{Ran}(\mathbf{V}_r)$ for $j = 1, \dots, N$
then $\mathbf{H}^{(\ell)}(\sigma)\mathbf{b} = \mathbf{H}_r^{(\ell)}(\sigma)\mathbf{b}$ for $\ell = 0, 1, \dots, N-1$
- (b) if $\left((\mu\mathbf{E} - \mathbf{A})^{-T}\mathbf{E}^T\right)^{j-1}(\mu\mathbf{E} - \mathbf{A})^{-T}\mathbf{C}^T\mathbf{c} \in \text{Ran}(\mathbf{W}_r)$ for $j = 1, \dots, M$,
then $\mathbf{c}^T\mathbf{H}^{(\ell)}(\mu) = \mathbf{c}^T\mathbf{H}_r^{(\ell)}(\mu)\mathbf{b}$ for $\ell = 0, 1, \dots, M-1$;
- (c) if both (a) and (b) hold, and if $\sigma = \mu$, then $\mathbf{c}^T\mathbf{H}^{(\ell)}(\sigma)\mathbf{b} = \mathbf{c}^T\mathbf{H}_r^{(\ell)}(\sigma)\mathbf{b}$,
for $\ell = 1, \dots, M+N+1$

Theorem 1.2 solves the rational tangential interpolation problem via projection. All one has to do is to construct the matrices \mathbf{V}_r and \mathbf{W}_r according to the theorem, and the reduced order model is guaranteed to satisfy the interpolation conditions. Note that Theorems 1.1 and 1.2 solve the interpolation problem without ever explicitly computing the values that are interpolated, since that computation is known to be prone to poor numerical conditioning.

Model reduction by interpolation (also called Krylov-based model reduction or moment matching in the single-input/single-output case) has been recognized previously, especially in the circuit simulation community; see, for example, [9–11, 33, 37, 73] and the references therein. [74, 76] also investigated the interpolation problem for efficient circuit analysis, however required explicit computation of the transfer function values and derivatives to be matched.

Despite their computational efficiency compared to Gramian-based model reduction techniques such as balanced-truncation [70, 72] and optimal Hankel norm approximation [40], interpolatory model reduction used to have the main drawback of depending on an ad hoc selection of interpolation points and tangential directions, i.e. how to choose the interpolation points and tangential directions optimally or at least effectively were not known until very recently. This has raised some criticism of interpolatory methods for not guaranteeing stability and not providing error bounds or error estimates for the resulting reduced-order model. To solve the stability issue, an implicit restart technique was proposed by Grimme et al. [43] which removed the unstable poles via an implicitly restarted Arnoldi process; however the reduced-model no longer interpolated the original one, but rather a nearby one. Gugercin and Antoulas [47] proposed combining interpolatory projection methods with Gramian-based approaches that guaranteed stability of the reduced-order model. Recently, reformulation of the Fourier model reduction approach of Willcox and Megretzki [87] in the interpolation framework by Gugercin and Willcox [50] has shown this method is indeed a stability preserving interpolatory model reduction method with an \mathcal{H}_∞ upper bound, not necessarily easily computable. Bai [93], and Gugercin and Antoulas [46] have also focused on the error estimates and provided *error expressions* (not error bounds) for the interpolatory model reduction. However, despite these efforts, the optimal interpolation points and tangential direction selection remained unresolved until very recently.

This optimal point selection strategy together with some other very recent developments in interpolation-based model reduction will be presented in the remaining of the manuscript: Sect. 3 will illustrate how to choose the interpolation points to minimize the \mathcal{H}_2 norm of the error systems. Passivity of the reduced-model is vital when dealing with model coming from, especially, circuit theory. Section 4 will show how to construct high-quality passivity-preserving reduced-order models using interpolation. Obtaining reduced-order models solely from input-output measurements without the knowledge of the original state-space quantities has been the focus of recent research. In Sect. 7, we will illustrate how to achieve this in the interpolation framework. Even though the state-space formulation of the original system in (1) is quite general, and indeed, will include, with suitable reformulation, the majority of linear dynamical systems, some dynamical systems might have a natural description that takes a form quite different from this standard realization; such as systems with internal delays, memory terms, etc. Section 5 will extend the interpolatory model reduction theory described here to settings where the transfer functions $\mathbf{H}(s)$ take a much more general form than that of (5), allowing interpolatory model reduction for a much richer category of dynamical systems.

Constructing Interpolants with $\mathbf{D}_r \neq \mathbf{D}$. Certain circumstances in model reduction require that a reduced-order model have a \mathbf{D}_r term different than \mathbf{D} . One such case is that of a singular \mathbf{E} with a non-defective eigenvalue at 0. This occurs when the internal system dynamics has an auxiliary algebraic constraint that must always be satisfied (perhaps representing rigid body translations or rotations or fluid incompressibility, for example). The dynamical system description given in (1) is then a *differential algebraic equation* (DAE) and our condition that \mathbf{E} have a non-defective eigenvalue at 0 amounts to the requirement that the DAE be of index 1 (see [60]). In this case, $\lim_{s \rightarrow \infty} \mathbf{H}(s) \neq \mathbf{D}$. If we want $\mathbf{H}_r(s)$ to match $\mathbf{H}(s)$ asymptotically at high frequencies then we should choose

$$\mathbf{D}_r = \lim_{s \rightarrow \infty} (\mathbf{H}(s) - \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r)$$

If $r < \text{rank}(\mathbf{E})$ then it will often happen that \mathbf{E}_r will be nonsingular, in which case then $\lim_{s \rightarrow \infty} \mathbf{H}_r(s) = \mathbf{D}_r$ and we may assign $\mathbf{D}_r = \lim_{s \rightarrow \infty} \mathbf{H}(s)$ if we wish $\mathbf{H}_r(s)$ to match $\mathbf{H}(s)$ asymptotically at high frequencies. See, for example, [21] where a reduced-order model with $\mathbf{D}_r \neq \mathbf{D}$ constructed in case of a singular \mathbf{E} to match $\mathbf{H}(s)$ and its higher derivatives around $s = \infty$.

One may be less concerned with $\mathbf{H}_r(s)$ matching $\mathbf{H}(s)$ well at high frequencies but instead may wish to minimize the maximum mismatch between $\mathbf{H}(s)$ and $\mathbf{H}_r(s)$ over the imaginary axis (best \mathcal{H}_{∞} approximation). Flexibility in choosing the \mathbf{D}_r term is necessary in this case as well. How to incorporate arbitrary choices of \mathbf{D}_r without losing interpolation properties, is described in the next theorem, first presented in [66] and later generalized in [16]. For simplicity, we assume $\mathbf{D} = \mathbf{0}$, i.e.

$$\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}. \quad (15)$$

The general case with $\mathbf{D} \neq \mathbf{0}$ is recovered by replacing \mathbf{D}_r with $\mathbf{D}_r - \mathbf{D}$.

Theorem 1.3. Given $\mathbf{H}(s)$ as in (15), $2r$ distinct points, $\{\mu_i\}_{i=1}^r \cup \{\sigma_j\}_{j=1}^r$, in the right halfplane, together with $2r$ nontrivial vectors, $\{\mathbf{c}_i\}_{i=1}^r \subset \mathbb{C}^p$ and $\{\mathbf{b}_j\}_{j=1}^r \subset \mathbb{C}^m$, let $\mathbf{V}_r \in \mathbb{C}^{n \times r}$ and $\mathbf{W}_r \in \mathbb{C}^{n \times r}$ be as in (13) and (14), respectively. Define $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{C}}$ as

$$\tilde{\mathbf{B}} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_r] \quad \text{and} \quad \tilde{\mathbf{C}}^T = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_r]^T$$

For any $\mathbf{D}_r \in \mathbb{C}^{p \times m}$, define

$$\begin{aligned} \mathbf{E}_r &= \mathbf{W}_r^T \mathbf{E} \mathbf{V}_r, & \mathbf{A}_r &= \mathbf{W}_r^T \mathbf{A} \mathbf{V}_r + \tilde{\mathbf{C}}^T \mathbf{D}_r \tilde{\mathbf{B}}, \\ \mathbf{B}_r &= \mathbf{W}_r^T \mathbf{B} - \tilde{\mathbf{C}}^T \mathbf{D}_r, & \text{and } \mathbf{C}_r &= \mathbf{C} \mathbf{V}_r - \mathbf{D}_r \tilde{\mathbf{B}} \end{aligned} \quad (16)$$

Then the reduced-order model $\mathbf{H}_r(s) = \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r + \mathbf{D}_r$ satisfies

$$\mathbf{H}(\sigma_i)\mathbf{b}_i = \mathbf{H}_r(\sigma_i)\mathbf{b}_i \quad \text{and} \quad \mathbf{c}_i^T \mathbf{H}(\mu_i) = \mathbf{c}_i^T \mathbf{H}_r(\mu_i) \quad \text{for } i = 1, \dots, r.$$

Theorem 1.3 describes how to construct an interpolant with an *arbitrary* \mathbf{D}_r term; hence the reduced-order model \mathbf{H}_r solves the rational interpolation problem for any \mathbf{D}_r that can be chosen to satisfy specific design goals.

It may happen that the state space representation of the reduced system is not minimal (for an example see [31]), in which case interpolation of some of the data is lost. Theorem 1.3 provides a remedy: there always exists \mathbf{D}_r , such that the realization given by (16) is minimal, i.e., both controllable and observable, and hence interpolation of all the data is guaranteed.

2.4 Error Measures

The loss in accuracy in replacing a full-order model with a reduced model is measured by the difference between respective system outputs over a range of inputs, which is best understood in the frequency domain. For the full-order system described by the transfer function $\mathbf{H}(s)$ as in (5) and the reduced-order system by $\mathbf{H}_r(s)$ as in (6), the *error system*, in the time domain, is defined by $S_{\text{error}} : \mathbf{u}(t) \mapsto [\mathbf{y} - \mathbf{y}_r](t)$. Equivalently, we can represent the error in the frequency domain by $\hat{\mathbf{y}}(s) - \hat{\mathbf{y}}_r(s) = [\mathbf{H}(s) - \mathbf{H}_r(s)] \hat{\mathbf{u}}(s)$; hence the error system can be considered having the transfer function $\mathbf{H}(s) - \mathbf{H}_r(s)$.

We consider the reduced-order system to be a good approximation of the full-order system if the error system S_{error} is small in a sense we wish to make precise. Ultimately, we want to produce a reduced order system that will yield outputs, $\mathbf{y}_r(t)$, such that $\mathbf{y}_r(t) \approx \mathbf{y}(t)$ uniformly well over a large class of inputs $\mathbf{u}(t)$. Different measures of approximation and different choices of input classes lead to different model reduction goals. Indeed, the “size” of linear systems can be measured in various ways – the two most common metrics are the so-called \mathcal{H}_∞ and \mathcal{H}_2 norms.

The \mathcal{H}_∞ Norm The \mathcal{H}_∞ norm of a stable linear system associated with a transfer function, $\mathbf{H}(s)$, is defined as

$$\|\mathbf{H}\|_{\mathcal{H}_\infty} = \max_{\omega \in \mathbb{R}} \|\mathbf{H}(\imath\omega)\|_2,$$

where $\|\mathbf{R}\|_2$ here denotes the usual induced 2-norm of a matrix \mathbf{R} . If the matrix \mathbf{E} is singular, we must require in addition that 0 is a nondefective eigenvalue of \mathbf{E} so that $\mathbf{H}(s)$ is bounded as $s \rightarrow \infty$. Suppose one wants to ensure that the output error $\mathbf{y}(t) - \mathbf{y}_r(t)$ is small in a root mean square sense for $t > 0$ (that is, we want $(\int_0^\infty \|\mathbf{y}(t) - \mathbf{y}_r(t)\|_2^2 dt)^{1/2}$ to be small) uniformly over all inputs, $\mathbf{u}(t)$, having bounded “energy,” $\int_0^\infty \|\mathbf{u}(t)\|_2^2 dt \leq 1$. Observe first that $\widehat{\mathbf{y}}(s) - \widehat{\mathbf{y}}_r(s) = [\mathbf{H}(s) - \mathbf{H}_r(s)] \widehat{\mathbf{u}}(s)$ so then by the Parseval relation:

$$\begin{aligned} \int_0^\infty \|\mathbf{y}(t) - \mathbf{y}_r(t)\|_2^2 dt &= \frac{1}{2\pi} \int_{-\infty}^\infty \|\widehat{\mathbf{y}}(\imath\omega) - \widehat{\mathbf{y}}_r(\imath\omega)\|_2^2 d\omega \\ &\leq \frac{1}{2\pi} \int_{-\infty}^\infty \|\mathbf{H}(\imath\omega) - \mathbf{H}_r(\imath\omega)\|_2^2 \|\widehat{\mathbf{u}}(\imath\omega)\|_2^2 d\omega \\ &\leq \max_{\omega} \|\mathbf{H}(\imath\omega) - \mathbf{H}_r(\imath\omega)\|_2^2 \left(\frac{1}{2\pi} \int_{-\infty}^\infty \|\widehat{\mathbf{u}}(\imath\omega)\|_2^2 d\omega \right)^{1/2} \\ &\leq \max_{\omega} \|\mathbf{H}(\imath\omega) - \mathbf{H}_r(\imath\omega)\|_2^2 \stackrel{\text{def}}{=} \|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_\infty}^2 \end{aligned}$$

Hence, proving that the $\|\mathbf{H}\|_{\mathcal{H}_\infty}$ norm is the L^2 induced operator norm of the associated system mapping, $\mathcal{S} : \mathbf{u} \mapsto \mathbf{y}$.

The \mathcal{H}_2 Norm. The \mathcal{H}_2 norm of a linear system associated with a transfer function, $\mathbf{H}(s)$, is defined as

$$\|\mathbf{H}\|_{\mathcal{H}_2} := \left(\frac{1}{2\pi} \int_{-\infty}^\infty \|\mathbf{H}(\imath\omega)\|_F^2 d\omega \right)^{1/2} \quad (17)$$

where now $\|\mathbf{R}\|_F^2 = \text{trace}(\overline{\mathbf{R}}\mathbf{R}^T)$ denotes the Frobenius norm of a complex matrix \mathbf{R} . Note that we must require in addition that if \mathbf{E} is singular then 0 is a nondefective eigenvalue of \mathbf{E} and that $\lim_{s \rightarrow \infty} \mathbf{H}(s) = 0$ in order that the \mathcal{H}_2 norm of the system to be finite. For more details, see [2] and [94].

Alternatively, suppose one wants to ensure that each component of the output error $\mathbf{y}(t) - \mathbf{y}_r(t)$ remains small for all $t > 0$ (that is, we want $\max_{t>0} \|\mathbf{y}(t) - \mathbf{y}_r(t)\|_\infty$ to be small) again uniformly over all inputs, $\mathbf{u}(t)$ with $\int_0^\infty \|\mathbf{u}(t)\|_2^2 dt \leq 1$.

$$\begin{aligned} \max_{t>0} \|\mathbf{y}(t) - \mathbf{y}_r(t)\|_\infty &= \max_{t>0} \left\| \frac{1}{2\pi} \int_{-\infty}^\infty (\widehat{\mathbf{y}}(\imath\omega) - \widehat{\mathbf{y}}_r(\imath\omega)) e^{i\omega t} d\omega \right\|_\infty \\ &\leq \frac{1}{2\pi} \int_{-\infty}^\infty \|\widehat{\mathbf{y}}(\imath\omega) - \widehat{\mathbf{y}}_r(\imath\omega)\|_\infty d\omega \\ &\leq \frac{1}{2\pi} \int_{-\infty}^\infty \|\mathbf{H}(\imath\omega) - \mathbf{H}_r(\imath\omega)\|_F \|\widehat{\mathbf{u}}(\imath\omega)\|_2 d\omega \end{aligned}$$

$$\begin{aligned} &\leq \left(\int_{-\infty}^{\infty} \|\mathbf{H}(\imath\omega) - \mathbf{H}_r(\imath\omega)\|_F^2 d\omega \right)^{1/2} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \|\widehat{\mathbf{u}}(\imath\omega)\|_2^2 d\omega \right)^{1/2} \\ &\leq \left(\int_{-\infty}^{+\infty} \|\mathbf{H}(\imath\omega) - \mathbf{H}_r(\imath\omega)\|_F^2 d\omega \right)^{1/2} \stackrel{\text{def}}{=} \|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2} \end{aligned}$$

It should be noted that in the single-input single-output case, the above relation holds with equality sign, because the \mathcal{H}_2 norm is equal to the $(2, \infty)$ induced norm of the convolution operator (see [2] for details).

\mathcal{H}_2 denotes the set of matrix-valued functions, $\mathbf{G}(z)$, with components that are analytic for z in the open right half plane, $\operatorname{Re}(z) > 0$, such that for $\operatorname{Re}(z) = x > 0$, $\mathbf{G}(x + \imath y)$ is square integrable as a function of $y \in (-\infty, \infty)$:

$$\sup_{x>0} \int_{-\infty}^{\infty} \|\mathbf{G}(x + \imath y)\|_F^2 dy < \infty.$$

\mathcal{H}_2 is a Hilbert space and holds our interest because transfer functions associated with stable finite dimensional dynamical systems are elements of \mathcal{H}_2 . Indeed, if $\mathbf{G}(s)$ and $\mathbf{H}(s)$ are transfer functions associated with stable dynamical systems having the same input and output dimensions, the \mathcal{H}_2 -inner product can be defined as

$$\langle \mathbf{G}, \mathbf{H} \rangle_{\mathcal{H}_2} \stackrel{\text{def}}{=} \frac{1}{2\pi} \int_{-\infty}^{\infty} \operatorname{Tr}(\overline{\mathbf{G}(\imath\omega)} \mathbf{H}(\imath\omega)^T) d\omega = \int_{-\infty}^{\infty} \operatorname{Tr}(\overline{\mathbf{G}(-\imath\omega)} \mathbf{H}(\imath\omega)^T) d\omega, \quad (18)$$

with a norm defined as

$$\|\mathbf{G}\|_{\mathcal{H}_2} \stackrel{\text{def}}{=} \left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} \|\mathbf{G}(\imath\omega)\|_F^2 d\omega \right)^{1/2}. \quad (19)$$

where $\operatorname{Tr}(\mathbf{M})$ and $\|\mathbf{M}\|_F$ denote the trace and Frobenius norm of \mathbf{M} , respectively. Notice in particular that if $\mathbf{G}(s)$ and $\mathbf{H}(s)$ represent real dynamical systems then $\langle \mathbf{G}, \mathbf{H} \rangle_{\mathcal{H}_2} = \langle \mathbf{H}, \mathbf{G} \rangle_{\mathcal{H}_2}$ so that $\langle \mathbf{G}, \mathbf{H} \rangle_{\mathcal{H}_2}$ must be real.

Suppose $f(s)$ is a meromorphic function (analytic everywhere except at isolated poles of finite order). λ is a simple pole of $f(s)$ if $\lim_{s \rightarrow \lambda} (s - \lambda)^\ell f(s) = 0$ for $\ell \geq 2$ and the residue is nontrivial: $\operatorname{res}[f(s), \lambda] = \lim_{s \rightarrow \lambda} (s - \lambda)f(s) \neq 0$. For matrix-valued meromorphic functions, $\mathbf{F}(s)$, we say that λ is a simple pole of $\mathbf{F}(s)$ if $\lim_{s \rightarrow \lambda} (s - \lambda)^\ell \mathbf{F}(s) = 0$ for $\ell \geq 2$ and $\operatorname{res}[\mathbf{F}(s), \lambda] = \lim_{s \rightarrow \lambda} (s - \lambda)\mathbf{F}(s)$ has rank 1. λ is a semi-simple pole of $\mathbf{F}(s)$ if $\lim_{s \rightarrow \lambda} (s - \lambda)^\ell \mathbf{F}(s) = 0$ for $\ell \geq 2$ and $\operatorname{res}[\mathbf{F}(s), \lambda] = \lim_{s \rightarrow \lambda} (s - \lambda)\mathbf{F}(s)$ has rank larger than 1.

If $\|\mathbf{F}(s)\|_2$ remains bounded as $s \rightarrow \infty$ then $\mathbf{F}(s)$ only has a finite number of poles. If all these poles are either simple or semi-simple then we define the *order* or *dimension* of $\mathbf{F}(s)$ by $\dim \mathbf{F} = \sum_{\lambda} \operatorname{rank}(\operatorname{res}[\mathbf{F}(s), \lambda])$ where the sum is taken over all poles λ . In this case, we can represent $\mathbf{F}(s)$ as

$$\mathbf{F}(s) = \sum_{i=1}^{\dim \mathbf{F}} \frac{1}{s - \lambda_i} \mathbf{c}_i \mathbf{b}_i^T,$$

where λ_i are indexed according to multiplicity as indicated by $\operatorname{rank}(\operatorname{res}[\mathbf{F}(s), \lambda_i])$.

Lemma 1.1. Suppose that $\mathbf{G}(s)$ and $\mathbf{H}(s)$ are stable (poles contained in the open left halfplane) and suppose that $\mathbf{H}(s)$ has poles at $\mu_1, \mu_2, \dots, \mu_m$. Then

$$\langle \mathbf{G}, \mathbf{H} \rangle_{\mathcal{H}_2} = \sum_{k=1}^m \text{res}[\text{Tr}(\overline{\mathbf{G}}(-s)\mathbf{H}(s)^T), \mu_k]. \quad (20)$$

In particular, if $\mathbf{H}(s)$ has only simple or semi-simple poles at $\mu_1, \mu_2, \dots, \mu_m$ and $m = \dim \mathbf{H}$ then $\mathbf{H}(s) = \sum_{i=1}^m \frac{1}{s-\mu_i} \mathbf{c}_i \mathbf{b}_i^T$ and

$$\langle \mathbf{G}, \mathbf{H} \rangle_{\mathcal{H}_2} = \sum_{k=1}^m \mathbf{c}_k^T \overline{\mathbf{G}}(-\mu_k) \mathbf{b}_k$$

and

$$\|\mathbf{H}\|_{\mathcal{H}_2} = \left(\sum_{k=1}^m \mathbf{c}_k^T \overline{\mathbf{H}}(-\mu_k) \mathbf{b}_k \right)^{1/2}.$$

Proof. Notice that the function $\text{Tr}(\overline{\mathbf{G}}(-s)\mathbf{H}(s)^T)$ has singularities in the left half plane only at $\mu_1, \mu_2, \dots, \mu_m$. For any $R > 0$, define the semicircular contour in the left halfplane:

$$\Gamma_R = \{z \mid z = \iota\omega \text{ with } \omega \in [-R, R]\} \cup \left\{ z \mid z = Re^{i\theta} \text{ with } \theta \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right] \right\}.$$

Γ_R bounds a region that for sufficiently large R contains all the system poles of $\mathbf{H}(s)$ and so, by the residue theorem

$$\begin{aligned} \langle \mathbf{G}, \mathbf{H} \rangle_{\mathcal{H}_2} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \text{Tr}(\overline{\mathbf{G}}(-\iota\omega)\mathbf{H}(\iota\omega)^T) d\omega \\ &= \lim_{R \rightarrow \infty} \frac{1}{2\pi\iota} \int_{\Gamma_R} \text{Tr}(\overline{\mathbf{G}}(-s)\mathbf{H}(s)^T) ds \\ &= \sum_{k=1}^m \text{res}[\text{Tr}(\overline{\mathbf{G}}(-s)\mathbf{H}(s)^T), \mu_k]. \end{aligned}$$

The remaining assertions follow from the definition. \square

Lemma 1.1 immediately yields a new expression, recently introduced by Beattie and Gugercin in [17], for the \mathcal{H}_2 error norm for MIMO dynamical systems based on the poles and residues of the transfer function. A similar expression for SISO systems was first introduced by Krajewski et al. [58], and rediscovered later by Gugercin and Antoulas [2, 44, 46]:

Theorem 1.4. Given a full-order real system, $\mathbf{H}(s)$ and a reduced model, $\mathbf{H}_r(s)$, having the form $\mathbf{H}_r(s) = \sum_{i=1}^r \frac{1}{s-\hat{\lambda}_i} \mathbf{c}_i \mathbf{b}_i^T$ (\mathbf{H}_r has simple poles at $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_r$ and rank-1 residues $\mathbf{c}_1 \mathbf{b}_1^T, \dots, \mathbf{c}_r \mathbf{b}_r^T$), the \mathcal{H}_2 norm of the error system is given by

$$\|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2}^2 = \|\mathbf{H}\|_{\mathcal{H}_2}^2 - 2 \sum_{k=1}^r \mathbf{c}_k^T \mathbf{H}(-\hat{\lambda}_k) \mathbf{b}_k + \sum_{k,\ell=1}^r \frac{\mathbf{c}_k^T \mathbf{c}_\ell \mathbf{b}_\ell^T \mathbf{b}_k}{-\hat{\lambda}_k - \hat{\lambda}_\ell} \quad (21)$$

3 Interpolatory Optimal \mathcal{H}_2 Approximation

Given a full-order system $\mathbf{H}(s)$, the optimal \mathcal{H}_2 model reduction problem seeks to find a reduced-order model $\mathbf{H}_r(s)$ that minimizes the \mathcal{H}_2 error; i.e.

$$\|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2} = \min_{\substack{\dim(\tilde{\mathbf{H}}_r) = r \\ \tilde{\mathbf{H}}_r : \text{stable}}} \left\| \mathbf{H} - \tilde{\mathbf{H}}_r \right\|_{\mathcal{H}_2}. \quad (22)$$

This optimization problem is nonconvex – obtaining a global minimizer can be intractable and is at best a hard task. The common approach is to find reduced order models that satisfy first-order necessary optimality conditions. Many researchers have worked on this problem. These efforts can be grouped into two categories: Lyapunov-based optimal \mathcal{H}_2 methods such as [53, 54, 81, 88, 89, 95]; and interpolation-based optimal \mathcal{H}_2 methods such as [15, 17, 25, 45, 48, 49, 59, 68, 83]. The main drawback of most of the Lyapunov-based methods is that they require solving a series of Lyapunov equations and so rapidly become infeasible as dimension increases. We note that even though recent developments have made approximate solution of large order Lyapunov equations possible, solving a sequence of them through an iteration remains daunting and the effect of only having approximate solutions throughout the iteration has not yet been assessed. On the other hand, interpolatory approaches can take advantage of the sparsity structure of \mathbf{E} and \mathbf{A} and have proved numerically very effective. Moreover, both problem frameworks are equivalent, as shown by Gugercin et al. [49], and this has further motivated expansion of interpolatory approaches to optimal \mathcal{H}_2 approximation.

Interpolation-based \mathcal{H}_2 -optimality conditions were originally developed by Meier and Luenberger [68] for SISO systems. Based on these conditions, an effective numerical algorithm for interpolatory optimal \mathcal{H}_2 approximation, called the **Iterative Rational Krylov Algorithm** (IRKA), was introduced in [45, 48]. Analogous \mathcal{H}_2 -optimality conditions for MIMO systems within a tangential interpolation framework have recently been developed by [25, 49, 83] leading to an analogous algorithm for the MIMO case; descriptions can be found in [25, 49]. Below, we present a new proof of the interpolatory first-order conditions for the optimal \mathcal{H}_2 model reduction problem. These conditions are the starting point for the numerical algorithm presented in the next section.

Theorem 1.5. Suppose $\mathbf{H}(s)$ is a real stable dynamical system and that $\mathbf{H}_r(s) = \sum_{i=1}^r \frac{1}{s - \hat{\lambda}_i} c_i b_i^T$ is a real dynamical system that is the best stable r th order approximation of \mathbf{H} with respect to the \mathcal{H}_2 norm. (\mathbf{H}_r has simple poles at $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_r$ and rank-1 residues $c_1 b_1^T, \dots, c_r b_r^T$.) Then

$$(a) \mathbf{H}(-\hat{\lambda}_k) \mathbf{b}_k = \mathbf{H}_r(-\hat{\lambda}_k) \mathbf{b}_k, \quad (b) c_k^T \mathbf{H}(-\hat{\lambda}_k) = c_k^T \mathbf{H}_r(-\hat{\lambda}_k), \quad (23)$$

$$\text{and} \quad (c) c_k^T \mathbf{H}'(-\hat{\lambda}_k) \mathbf{b}_k = c_k^T \mathbf{H}'_r(-\hat{\lambda}_k) \mathbf{b}_k \quad \text{for } k = 1, 2, \dots, r.$$

Proof. Suppose $\tilde{\mathbf{H}}_r(s)$ is a transfer function associated with a stable r th order dynamical system. Then

$$\begin{aligned} \|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2}^2 &\leq \|\mathbf{H} - \tilde{\mathbf{H}}_r\|_{\mathcal{H}_2}^2 = \|\mathbf{H} - \mathbf{H}_r + \mathbf{H}_r - \tilde{\mathbf{H}}_r\|_{\mathcal{H}_2}^2 \\ &= \|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2}^2 + 2\Re e \langle \mathbf{H} - \mathbf{H}_r, \mathbf{H}_r - \tilde{\mathbf{H}}_r \rangle_{\mathcal{H}_2} + \|\mathbf{H}_r - \tilde{\mathbf{H}}_r\|_{\mathcal{H}_2}^2 \end{aligned}$$

so that $0 \leq 2\Re e \langle \mathbf{H} - \mathbf{H}_r, \mathbf{H}_r - \tilde{\mathbf{H}}_r \rangle_{\mathcal{H}_2} + \|\mathbf{H}_r - \tilde{\mathbf{H}}_r\|_{\mathcal{H}_2}^2$ (24)

By making judicious choices in how $\tilde{\mathbf{H}}_r$ is made to differ from \mathbf{H}_r , we arrive at tangential interpolation conditions via Lemma 1.1.

Toward this end, pick an arbitrary unit vector $\xi \in \mathbb{C}^m$, $\varepsilon > 0$, and for some ℓ , define $\theta = \pi - \arg \xi^T (\mathbf{H}(-\hat{\lambda}_\ell) - \mathbf{H}_r(-\hat{\lambda}_\ell)) \mathbf{b}_\ell$, and so that

$$\mathbf{H}_r(s) - \tilde{\mathbf{H}}_r(s) = \frac{\varepsilon e^{i\theta}}{s - \hat{\lambda}_\ell} \xi \mathbf{b}_\ell^T$$

and using Lemma 1.1, $\langle \mathbf{H} - \mathbf{H}_r, \mathbf{H}_r - \tilde{\mathbf{H}}_r \rangle_{\mathcal{H}_2} = -\varepsilon |\xi^T (\mathbf{H}(-\hat{\lambda}_\ell) - \mathbf{H}_r(-\hat{\lambda}_\ell)) \mathbf{b}_\ell|$.

Now (24) leads to

$$0 \leq |\xi^T (\mathbf{H}(-\hat{\lambda}_\ell) - \mathbf{H}_r(-\hat{\lambda}_\ell)) \mathbf{b}_\ell| \leq \varepsilon \frac{\|\mathbf{b}_\ell\|_2^2}{-2\Re e(\hat{\lambda}_\ell)}$$

which by taking ε small implies first that

$$\xi^T (\mathbf{H}(-\hat{\lambda}_\ell) - \mathbf{H}_r(-\hat{\lambda}_\ell)) \mathbf{b}_\ell = 0$$

but then since ξ was chosen arbitrarily, we must have that

$$(\mathbf{H}(-\hat{\lambda}_\ell) - \mathbf{H}_r(-\hat{\lambda}_\ell)) \mathbf{b}_\ell = 0.$$

A similar argument yields (23b).

If (23c) did not hold then $c_\ell^T \mathbf{H}'(-\hat{\lambda}_\ell) \mathbf{b}_\ell \neq c_\ell^T \mathbf{H}'_r(-\hat{\lambda}_\ell) \mathbf{b}_\ell$ and we may pick $0 < \varepsilon < |\Re e(\hat{\lambda}_\ell)|$ and $\theta = -\arg c_\ell^T (\mathbf{H}'(-\hat{\lambda}_\ell) - \mathbf{H}'_r(-\hat{\lambda}_\ell)) \mathbf{b}_\ell$ in such a way that $\mu = \hat{\lambda}_\ell + \varepsilon e^{i\theta}$ does not coincide with any reduced order poles $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_r$. Note that $\Re e(\mu) < 0$. Define $\tilde{\mathbf{H}}_r(s)$ so that

$$\mathbf{H}_r(s) - \tilde{\mathbf{H}}_r(s) = \left(\frac{1}{s - \hat{\lambda}_\ell} - \frac{1}{s - \mu} \right) c_\ell \mathbf{b}_\ell^T.$$

$\tilde{\mathbf{H}}_r(s)$ has the same poles and residues as $\mathbf{H}_r(s)$ aside from μ which replaces $\hat{\lambda}_\ell$ as a pole in $\tilde{\mathbf{H}}_r(s)$ with no change in the associated residue. Because of (23a-b), we calculate

$$\begin{aligned}\langle \mathbf{H} - \mathbf{H}_r, \mathbf{H}_r - \tilde{\mathbf{H}}_r \rangle_{\mathcal{H}_2} &= \mathbf{c}_\ell^T \left(\mathbf{H}(-\hat{\lambda}_\ell) - \mathbf{H}_r(-\hat{\lambda}_\ell) \right) \mathbf{b}_\ell - \mathbf{c}_\ell^T (\mathbf{H}(-\mu) - \mathbf{H}_r(-\mu)) \mathbf{b}_\ell \\ &= -\mathbf{c}_\ell^T (\mathbf{H}(-\mu) - \mathbf{H}_r(-\mu)) \mathbf{b}_\ell\end{aligned}$$

Then (24) leads to

$$0 \leq -2\Re(\mathbf{c}_\ell^T (\mathbf{H}(-\mu) - \mathbf{H}_r(-\mu)) \mathbf{b}_\ell) - \frac{|\mu - \hat{\lambda}_\ell|^2}{|\mu + \hat{\lambda}_\ell|^2} \frac{\Re(\mu + \hat{\lambda}_\ell)}{2\Re(\hat{\lambda}_\ell)\Re(\mu)} \|\mathbf{c}_\ell\|_2^2 \|\mathbf{b}_\ell\|_2^2 \quad (25)$$

Now, easy manipulations yield first a resolvent identity

$$(-\mu \mathbf{E} - \mathbf{A})^{-1} = (-\hat{\lambda}_\ell \mathbf{E} - \mathbf{A})^{-1} + (\mu - \hat{\lambda}_\ell)(-\hat{\lambda}_\ell \mathbf{E} - \mathbf{A})^{-1} \mathbf{E}(-\mu \mathbf{E} - \mathbf{A})^{-1}$$

and then resubstituting,

$$\begin{aligned}(-\mu \mathbf{E} - \mathbf{A})^{-1} &= (-\hat{\lambda}_\ell \mathbf{E} - \mathbf{A})^{-1} + (\mu - \hat{\lambda}_\ell)(-\hat{\lambda}_\ell \mathbf{E} - \mathbf{A})^{-1} \mathbf{E}(-\hat{\lambda}_\ell \mathbf{E} - \mathbf{A})^{-1} \\ &\quad + (\mu - \hat{\lambda}_\ell)^2 (-\hat{\lambda}_\ell \mathbf{E} - \mathbf{A})^{-1} \mathbf{E}(-\mu \mathbf{E} - \mathbf{A})^{-1} \mathbf{E}(-\hat{\lambda}_\ell \mathbf{E} - \mathbf{A})^{-1} \quad (26)\end{aligned}$$

Premultiplying by \mathbf{C} and postmultiplying by \mathbf{B} , (26) implies

$$\begin{aligned}\mathbf{H}(-\mu) &= \mathbf{H}(-\hat{\lambda}_\ell) + (\mu - \hat{\lambda}_\ell) \mathbf{H}'(-\hat{\lambda}_\ell) \\ &\quad + (\mu - \hat{\lambda}_\ell)^2 \mathbf{C}(-\hat{\lambda}_\ell \mathbf{E} - \mathbf{A})^{-1} \mathbf{E}(-\mu \mathbf{E} - \mathbf{A})^{-1} \mathbf{E}(-\hat{\lambda}_\ell \mathbf{E} - \mathbf{A})^{-1} \mathbf{B}\end{aligned}$$

and analogous arguments yield

$$\begin{aligned}\mathbf{H}_r(-\mu) &= \mathbf{H}_r(-\hat{\lambda}_\ell) + (\mu - \hat{\lambda}_\ell) \mathbf{H}'_r(-\hat{\lambda}_\ell) \\ &\quad + (\mu - \hat{\lambda}_\ell)^2 \mathbf{C}_r(-\hat{\lambda}_\ell \mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{E}_r(-\mu \mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{E}_r(-\hat{\lambda}_\ell \mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{B}_r\end{aligned}$$

Using these expressions in (25) yields

$$0 \leq -2\epsilon |\mathbf{c}_\ell^T (\mathbf{H}'(-\hat{\lambda}_\ell) - \mathbf{H}'_r(-\hat{\lambda}_\ell)) \mathbf{b}_\ell| + \mathcal{O}(\epsilon^2).$$

As $\epsilon \rightarrow 0$ a contradiction occurs unless $|\mathbf{c}_\ell^T (\mathbf{H}'(-\hat{\lambda}_\ell) - \mathbf{H}'_r(-\hat{\lambda}_\ell)) \mathbf{b}_\ell| = 0$. \square

One byproduct of the proof of Theorem 1.5 is that if $\mathbf{H}_r(s)$ satisfies the necessary conditions (23), then $\mathbf{H}_r(s)$ is guaranteed to be an *optimal* approximation of $\mathbf{H}(s)$ relative to the \mathcal{H}_2 norm among all reduced order systems having the same reduced system poles $\{\hat{\lambda}_i\}_{i=1}^r$. This follows from the observation that the set of all systems having the all same poles $\{\hat{\lambda}_i\}_{i=1}^r$ comprise a subspace of \mathcal{H}_2 and the projection theorem is both necessary and sufficient for $\mathbf{H}_r(s)$ to be a minimizer out of a subspace of candidate minimizers.

Conversely, the set of all stable r th order dynamical systems is not only not a subspace but it is not even convex. Indeed, the original problem (22) may have

multiple minimizers and there may be “local minimizers” that do not solve (22) yet do satisfy the first order conditions (23). To clarify, a reduced order system \mathbf{H}_r , is said to be a *local minimizer* for (22), if for all $\varepsilon > 0$ sufficiently small,

$$\|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2} \leq \|\mathbf{H} - \tilde{\mathbf{H}}_r^{(\varepsilon)}\|_{\mathcal{H}_2}, \quad (27)$$

for all stable dynamical systems, $\tilde{\mathbf{H}}_r^{(\varepsilon)}$ such that both $\dim(\tilde{\mathbf{H}}_r^{(\varepsilon)}) = r$ and $\|\mathbf{H}_r - \tilde{\mathbf{H}}_r^{(\varepsilon)}\|_{\mathcal{H}_2} \leq C\varepsilon$ for some constant C .

As a practical matter, global minimizers that solve (22) are difficult to obtain reliably and with certainty; current approaches favor seeking reduced order models that satisfy local (first-order) necessary conditions for optimality and so may produce models that are only certain to be local minimizers. Even though such strategies do not guarantee global minimization, they often produce very effective reduced order models, nonetheless.

3.1 An Algorithm for Interpolatory Optimal \mathcal{H}_2 Model Reduction

Theorem 1.5 states that first-order conditions for \mathcal{H}_2 optimality are tangential interpolation conditions at mirror images of reduced-order poles. However, one cannot simply construct the corresponding matrices \mathbf{V}_r and \mathbf{W}_r since the interpolation points and the tangential directions depend on the reduced-model; hence are not known a priori. The *Iterative Rational Krylov Algorithm* (**IRKA**) introduced in [49] resolves this problem by iteratively correcting the interpolation points and the directions as outlined below.

Let \mathbf{Y}^* and \mathbf{X} denote, respectively, the left and right eigenvectors for $\lambda\mathbf{E}_r - \mathbf{A}_r$ so that $\mathbf{Y}^*\mathbf{A}_r\mathbf{X} = \text{diag}(\hat{\lambda}_i)$ and $\mathbf{Y}^*\mathbf{E}_r\mathbf{X} = \mathbf{I}_r$. Denote the columns of $\mathbf{C}_r\mathbf{X}$ as $\hat{\mathbf{c}}_i$ and the rows of $\mathbf{Y}^*\mathbf{B}_r$ as $\hat{\mathbf{b}}_i^T$. Then in **IRKA** interpolation points used in the next step are chosen to be the mirror images of the current reduced order poles, i.e., the eigenvalues, $\lambda(\mathbf{A}_r, \mathbf{E}_r)$, of the pencil $\lambda\mathbf{E}_r - \mathbf{A}_r$ in the current step. The tangent directions are corrected in a similar way, using the residues of the previous reduced-ordered model until (23) is satisfied. A brief sketch of **IRKA** is described in Algorithm 1:

Upon convergence, $\{\sigma_i\}$ will be the mirror images of the eigenvalues of the reduced pencil $\lambda\mathbf{E}_r - \mathbf{A}_r$; and $\hat{\mathbf{b}}_i^T = \mathbf{e}_i^T \mathbf{Y}^* \mathbf{B}_r$ and $\hat{\mathbf{c}}_i = \mathbf{C}_r \mathbf{X} \mathbf{e}_i$ for $i = 1, \dots, r$. Consequently from Steps 4.(d) and 4.(e), the reduced-order transfer function satisfies (23), first-order conditions for \mathcal{H}_2 optimality. The main computational cost of this algorithm involves solving $2r$ linear systems to generate \mathbf{V}_r and \mathbf{W}_r . Computing the eigenvectors \mathbf{Y} and \mathbf{X} , and the eigenvalues of the reduced pencil $\lambda\mathbf{E}_r - \mathbf{A}_r$ are cheap since the dimension r is small.

IRKA has been remarkably successful in producing high fidelity reduced-order approximations; it is numerically effective and has been successfully applied to finding \mathcal{H}_2 -optimal reduced models for systems of high order, $n > 160,000$, see [56].

MIMO \mathcal{H}_2 Optimal Tangential Interpolation Method

1. Make an initial r -fold shift selection: $\{\sigma_1, \dots, \sigma_r\}$ that is closed under conjugation (i.e., $\{\sigma_1, \dots, \sigma_r\} \equiv \{\overline{\sigma_1}, \dots, \overline{\sigma_r}\}$ viewed as sets) and initial tangent directions $\hat{b}_1, \dots, \hat{b}_r$ and $\hat{c}_1, \dots, \hat{c}_r$, also closed under conjugation.
2. $\mathbf{V}_r = [(\sigma_1 \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \hat{b}_1 \dots (\sigma_r \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \hat{b}_r]$
 $\mathbf{W}_r = [(\sigma_1 \mathbf{E} - \mathbf{A}^T)^{-1} \mathbf{C}^T \hat{c}_1 \dots (\sigma_r \mathbf{E} - \mathbf{A}^T)^{-1} \mathbf{C}^T \hat{c}_1].$
3. while (not converged)
 - a. $\mathbf{A}_r = \mathbf{W}_r^T \mathbf{A} \mathbf{V}_r$, $\mathbf{E}_r = \mathbf{W}_r^T \mathbf{E} \mathbf{V}_r$, $\mathbf{B}_r = \mathbf{W}_r^T \mathbf{B}$, and $\mathbf{C}_r = \mathbf{C} \mathbf{V}_r$
 - b. Compute $\mathbf{Y}^* \mathbf{A}_r \mathbf{X} = \text{diag}(\tilde{\lambda}_i)$ and $\mathbf{Y}^* \mathbf{E}_r \mathbf{X} = \mathbf{I}_r$ where \mathbf{Y}^* and \mathbf{X} are the left and right eigenvectors of $\lambda \mathbf{E}_r - \mathbf{A}_r$.
 - c. $\sigma_i \leftarrow -\lambda_i(\mathbf{A}_r, \mathbf{E}_r)$ for $i = 1, \dots, r$, $\hat{b}_i^* \leftarrow \mathbf{e}_i^T \mathbf{Y}^* \mathbf{B}_r$ and $\hat{c}_i \leftarrow \mathbf{C}_r \mathbf{X} \mathbf{e}_i$.
 - d. $\mathbf{V}_r = [(\sigma_1 \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \hat{b}_1 \dots (\sigma_r \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \hat{b}_r]$
 - e. $\mathbf{W}_r = [(\sigma_1 \mathbf{E} - \mathbf{A}^T)^{-1} \mathbf{C}^T \hat{c}_1 \dots (\sigma_r \mathbf{E} - \mathbf{A}^T)^{-1} \mathbf{C}^T \hat{c}_1].$
4. $\mathbf{A}_r = \mathbf{W}_r^T \mathbf{A} \mathbf{V}_r$, $\mathbf{E}_r = \mathbf{W}_r^T \mathbf{E} \mathbf{V}_r$, $\mathbf{B}_r = \mathbf{W}_r^T \mathbf{B}$, $\mathbf{C}_r = \mathbf{C} \mathbf{V}_r$

Remark 1.1. Note that the \mathcal{H}_2 error formulae in Theorem 1.4 expresses the error as a function of reduced-order system poles and residues. Based on this expression, Beattie and Gugercin [17] recently developed a trust-region based descent algorithm where the poles and residues of $\mathbf{H}_r(s)$ are the optimization parameters as opposed to the interpolation points and tangential directions in **IRKA**. The algorithm reduces the \mathcal{H}_2 error at each step of the iteration and globally converges to a local minimum. Even though the method does not use the interpolation point as the parameters, it is worth noting that the resulting reduced-order model indeed is an interpolatory approximation satisfying the interpolatory \mathcal{H}_2 conditions of Theorem 1.5. Hence, the method of [17] achieves the interpolation conditions while using the reduced-order poles and residues as the variables.

3.2 Numerical Results for IRKA

This problem arises during a cooling process in a rolling mill and is modeled as boundary control of a two dimensional heat equation. A finite element discretization results in a descriptor system state-dimension $n = 79,841$, i.e., $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{79841 \times 79841}$, $\mathbf{B} \in \mathbb{R}^{79841 \times 7}$, $\mathbf{C} \in \mathbb{R}^{6 \times 79841}$. For details regarding this model, see [18, 20].

Using **IRKA**, we reduce the order of the full-order system, $\mathbf{H}(s)$, to $r = 20$ to obtain the \mathcal{H}_2 optimal reduced model, $\mathbf{H}_{\text{IRKA}}(s)$. Figure 1 illustrates the convergence behavior of **IRKA** through the history of the relative \mathcal{H}_∞ error, $\frac{\|\mathbf{H} - \mathbf{H}_{\text{IRKA}}\|_{\mathcal{H}_\infty}}{\|\mathbf{H}\|_{\mathcal{H}_\infty}}$, as it evolves through the course of the iteration. Even though the figure shows the method converging after approximately 11-12 iterations, one may also see that it nearly achieves the final error value much earlier in the iteration; already around the sixth or seventh step.

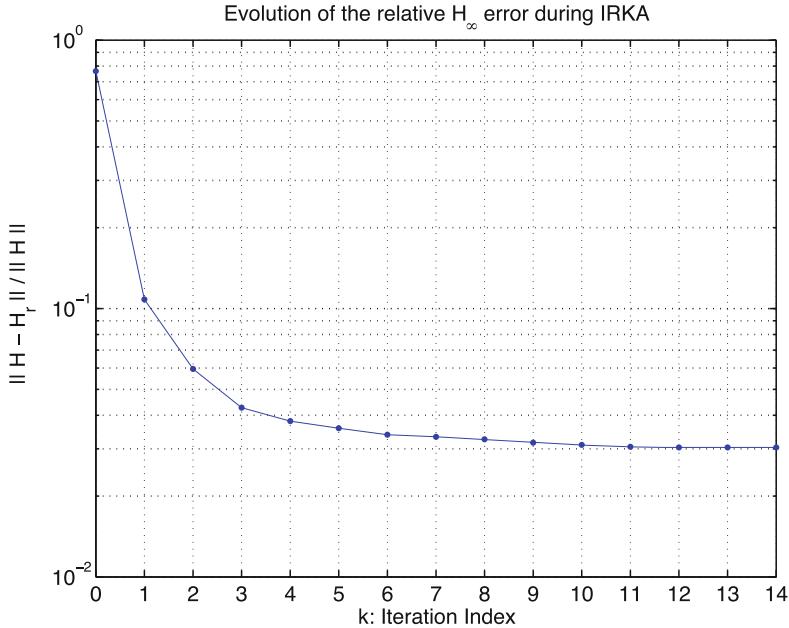


Fig. 1 Evolution of IRKA

Starting from a large initial relative error (nearly 100%), the method automatically and without any user intervention corrects both interpolation points and tangent directions, reaching within a few iterations a near-optimal solution with a relative error of around 3×10^{-2} .

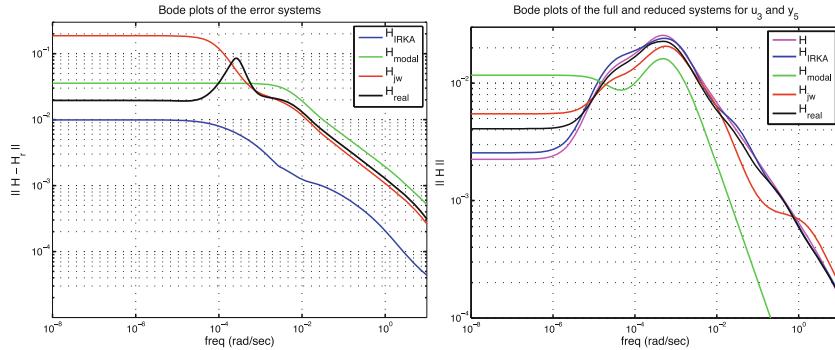
We compare our approach with other commonly used model reduction techniques.

1. Modal Approximation: We reduce the order $r = 20$ using 20 dominant modes of $\mathbf{H}(s)$. The reduced model is denoted by $\mathbf{H}_{\text{modal}}$.
2. Interpolation using points on the imaginary axis: Based on the bode plot of $\mathbf{H}(s)$, we have placed interpolation points on the imaginary axis where $\|\mathbf{H}(j\omega)\|$ is dominant. This was a common approach for choosing interpolation points before the \mathcal{H}_2 -optimal shift selection strategy was developed. The reduced model is denoted by $\mathbf{H}_{i\omega}$.
3. 20 real interpolation points are chosen as the mirror images of the poles of $\mathbf{H}(s)$. This selection is a good initialization for IRKA in the SISO case, see [49]. The reduced model is denoted by \mathbf{H}_{real} .

In our experiments, we found that in order to be able to find reasonable interpolation points and directions for $\mathbf{H}_{i\omega}$ and \mathbf{H}_{real} , we needed first to run several experiments. Many of them either resulted in unstable systems or very poor performance

Table 1 Error norms

	\mathbf{H}_{IRKA}	$\mathbf{H}_{\text{modal}}$	$\mathbf{H}_{t\omega}$	\mathbf{H}_{real}
Relative \mathcal{H}_∞ error	3.03×10^{-2}	1.03×10^{-1}	5.42×10^{-1}	2.47×10^{-1}

**Fig. 2** Comparison of reduced-order models. (a) Error system bode plots. (b) Reduced model bode plots for u_3 and y_5

results. Here we are presenting the best selection we were able to find. This is the precise reason why **IRKA** is superior. We initiate it once, randomly in this case, and the algorithm automatically finds the optimal points and directions. There is no need for an ad hoc search. Table 1 shows the relative \mathcal{H}_∞ error norms for each reduced model. Clearly, **IRKA** is the best one, yielding an error one order of magnitude smaller than those the other four approaches yield. Note from Fig. 1 that the initial guess for **IRKA** has a higher error than all the other methods. However, even after only two steps of the iteration long before convergence, the **IRKA** iterate has already a smaller error norm than all other three approaches. Note that $\mathbf{H}(s)$ has 7 inputs and 6 outputs; hence there are 42 input-output channels. \mathbf{H}_{IRKA} with order $r = 20$, less than the total number of input-output channels, is able to replicate these behaviors with a relative accuracy of order 10^{-2} . Even though **IRKA** is an \mathcal{H}_2 -based approach, superior \mathcal{H}_∞ performance is also observed and is not surprising. It is an efficient general purpose \mathcal{H}_2 and \mathcal{H}_∞ model reduction method, not only \mathcal{H}_2 optimal. Figure 2a depicts the bode plots of the error systems. It is clear from the error plots that \mathbf{H}_{IRKA} outperforms the rest of the methods. To give an example how the reduced models match the full-order model for a specific input/output channel, we show the bode plots for the transfer function between the third input u_3 and the fifth output y_5 . Clearly, \mathbf{H}_{IRKA} yields the best match. We note that while Table 1 lists the relative error norms, the error norms in Fig. 2 shown are the absolute error values.

Since **IRKA** yielded a much lower error value, we have checked what lowest order model from **IRKA** would yield similar error norms as the other approaches. We have found that **IRKA** for order $r = 2$ yields a relative \mathcal{H}_∞ error of 2.26×10^{-1} ; already better than 20th order $\mathbf{H}_{t\omega}$ and \mathbf{H}_{real} . For $r = 6$, for example, **IRKA** yielded a relative \mathcal{H}_∞ error of 1.64×10^{-1} , a number close to the one obtained by 20th order $\mathbf{H}_{\text{modal}}$. These numbers vividly illustrate the strength of the optimal \mathcal{H}_2 shift selection.

4 Interpolatory Passivity Preserving Model Reduction

Passive systems are an important class of dynamical systems that can only absorb power – for which the work done by the inputs in producing outputs must always be positive. A system is called *passive* if

$$\Re e \int_{-\infty}^t \mathbf{u}(\tau)^* \mathbf{y}(\tau) d\tau \geq 0,$$

for all $t \in \mathbb{R}$ and for all $\mathbf{u} \in \mathcal{L}_2(\mathbb{R})$. In this section, we will show how to create interpolatory projections so that passivity of reduced models is preserved. For further details we refer to [3].

A rational (square) matrix function $\mathbf{H}(s)$ is called *positive real* if:

1. $\mathbf{H}(s)$, is analytic for $\Re e(s) > 0$,
2. $\overline{\mathbf{H}(s)} = \mathbf{H}(\bar{s})$ for all $s \in \mathbb{C}$, and
3. the Hermitian part of $\mathbf{H}(s)$, i.e., $\mathbf{H}(s) + \mathbf{H}^T(\bar{s})$, is positive semi-definite for all $\Re e(s) \geq 0$.

Dynamical systems as given in (1) are passive if and only if the associated transfer function $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$, is positive real. An important consequence of positive realness is the existence of a *spectral factorization*: a matrix function $\Phi(s)$ for which $\mathbf{H}(s) + \mathbf{H}^T(-s) = \Phi(s)\Phi^T(-s)$, and the poles as well as the (finite) zeros of $\Phi(s)$ are all stable. The *spectral zeros* of the system represented by $\mathbf{H}(s)$ are defined to be those values, λ , for which $\Phi(\lambda)$ (and hence $\mathbf{H}(\lambda) + \mathbf{H}^T(-\lambda)$) loses rank.

Assume for simplicity that λ is a spectral zero with multiplicity one (so that $\text{nullity}(\Phi(\lambda)) = 1$). Then there is a right spectral zero direction, \mathbf{z} , such that $(\mathbf{H}(\lambda) + \mathbf{H}^T(-\lambda))\mathbf{z} = \mathbf{0}$. Evidently, if (λ, \mathbf{z}) is a right spectral zero pair for the system represented by $\mathbf{H}(s)$, then $(-\bar{\lambda}, \mathbf{z}^*)$ is a left spectral zero pair: $\mathbf{z}^*(\mathbf{H}(-\bar{\lambda}) + \mathbf{H}^T(\bar{\lambda})) = \mathbf{0}$.

The key result that is significant here was shown in [3] that if interpolation points are chosen as spectral zeros, passivity is preserved.

Theorem 1.6. Suppose the dynamical system given in (1) and represented by the transfer function $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$ is stable and passive. Suppose that for some index $r \geq 1$, $\lambda_1, \dots, \lambda_r$ are stable spectral zeros of \mathbf{H} with corresponding right spectral zero directions $\mathbf{z}_1, \dots, \mathbf{z}_r$.

If a reduced order system $\mathbf{H}_r(s)$ tangentially interpolates $\mathbf{H}(s)$ as in (11) with $\sigma_i = \lambda_i$, $\mathbf{b}_i = \mathbf{z}_i$, $\mu_i = -\bar{\lambda}_i$, and $\mathbf{c}_i^T = \mathbf{z}^*$ for $i = 1, \dots, r$, then $\mathbf{H}_r(s)$ is stable and passive.

As a practical matter, the first task that we face is computation of the spectral zeros of the system represented by $\mathbf{H}(s)$. This can be formulated as a structured

eigenvalue problem. Following the development of [80], we define the following pair of matrices which has *Hamiltonian* structure:

$$\mathcal{H} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{B} \\ \mathbf{0} & -\mathbf{A}^T & -\mathbf{C}^T \\ \mathbf{C} & \mathbf{B}^T & \mathbf{D} + \mathbf{D}^T \end{bmatrix} \text{ and } \mathcal{E} = \begin{bmatrix} \mathbf{E} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

The *spectral zeros* of the system are the generalized eigenvalues of the pencil: $\mathcal{H}\mathbf{x}_i = \lambda_i \mathcal{E}\mathbf{x}_i$. To see this, partition the eigenvector \mathbf{x}_i into components \mathbf{v}_i , $\overline{\mathbf{w}_i}$, and \mathbf{z}_i such that

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{B} \\ \mathbf{0} & -\mathbf{A}^T & -\mathbf{C}^T \\ \mathbf{C} & \mathbf{B}^T & \mathbf{D} + \mathbf{D}^T \end{bmatrix} \begin{bmatrix} \mathbf{v}_i \\ \overline{\mathbf{w}_i} \\ \mathbf{z}_i \end{bmatrix} = \lambda_i \begin{bmatrix} \mathbf{E} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v}_i \\ \overline{\mathbf{w}_i} \\ \mathbf{z}_i \end{bmatrix},$$

Then note that as a consequence,

$$\mathbf{v}_i = (\lambda_i \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \mathbf{z}_i, \quad \mathbf{w}_i^T = \mathbf{z}_i^* \mathbf{C} (-\overline{\lambda}_i \mathbf{E} - \mathbf{A})^{-1},$$

and $(\mathbf{H}(\lambda_i) + \mathbf{H}^T(-\lambda_i)) \mathbf{z}_i = \mathbf{0}$.

Thus, λ_i are spectral zeros of $\mathbf{H}(s)$ associated with the right spectral zero directions, \mathbf{z}_i , for $i = 1, \dots, r$. Furthermore, basis vectors for the right and left modeling subspaces used to impose the tangential interpolation conditions required by Theorem 1.6, are determined immediately by the remaining two components of \mathbf{x}_i : $\mathbf{V}_r = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r]$ and $\mathbf{W}_r = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r]$.

Since \mathcal{H} and \mathcal{E} are real, the eigenvalues of $\mathcal{H}\mathbf{x} = \lambda \mathcal{E}\mathbf{x}$ occur in complex conjugate pairs. Furthermore, if we define $\mathbf{J} = \begin{bmatrix} \mathbf{0} & -\mathbf{I} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}$ Then $(\mathbf{J}\mathcal{H})^* = \mathbf{J}\mathcal{H}$ and $(\mathbf{J}\mathcal{E})^* = -\mathbf{J}\mathcal{E}$ and it is easy to see that if $\mathbf{x}^T = [\mathbf{v}^T \ \overline{\mathbf{w}}^T \ \mathbf{z}^T]$ is a (right) eigenvector associated with λ : $\mathcal{H}\mathbf{x} = \lambda \mathcal{E}\mathbf{x}$, then $\mathbf{y}^* = [\mathbf{w}^T \ -\mathbf{v}^* \ \mathbf{z}^*]$ is a left eigenvector of the pencil associated with $-\overline{\lambda}$: $\mathbf{y}^* \mathcal{H} = -\overline{\lambda} \mathbf{y}^* \mathcal{E}$. Thus, if λ is an eigenvalue of $\mathcal{H}\mathbf{x} = \lambda \mathcal{E}\mathbf{x}$ then so is each of $\overline{\lambda}$, $-\overline{\lambda}$, and $-\lambda$.

Thus, the spectral zeros, associated spectral zero directions, and bases for the left and right modeling subspaces can be obtained by means of the above Hamiltonian eigenvalue problem.

The question remains of which r spectral zeros to choose. The concept of *dominance* arising in modal approximation is useful in distinguishing effective choices of spectral zero sets. For details we refer to [55].

Assume for simplicity that $\mathbf{D} + \mathbf{D}^T$ is invertible, take $\Delta = (\mathbf{D} + \mathbf{D}^T)^{-1}$ and define

$$\mathcal{B} = \begin{bmatrix} \mathbf{B} \\ -\mathbf{C}^T \\ \mathbf{0} \end{bmatrix} \Delta, \quad \mathcal{C} = -\Delta [\mathbf{C}, \ \mathbf{B}^T, \ \mathbf{0}],$$

It can be checked (with some effort) that

$$\mathbf{G}(s) \triangleq [\mathbf{H}(s) + \mathbf{H}^T(-s)]^{-1} = \Delta + \mathcal{C}(s\mathcal{E} - \mathcal{H})^{-1}\mathcal{B}.$$

Let the partial fraction expansion of $\mathbf{G}(s)$ be

$$\mathbf{G}(s) = \sum_{j=1}^{2n} \frac{\mathbf{R}_j}{s - \lambda_j}, \text{ with } \mathbf{R}_j = \frac{1}{\mathbf{y}_j^* \mathcal{E} \mathbf{x}_j} \mathcal{C} \mathbf{x}_j \mathbf{y}_j^* \mathcal{B},$$

where λ_i are the spectral zeros of the original system (poles of the associated Hamiltonian system) and \mathbf{R}_j are the *residues*. The left and right eigenvectors of \mathbf{y}_j , \mathbf{x}_j , are computed from $\mathcal{H}\mathbf{x}_j = \lambda_j \mathcal{E}\mathbf{x}_j$ and $\mathbf{y}_j^*\mathcal{H} = \lambda_j \mathbf{y}_j^*\mathcal{E}$.

A spectral zero λ_i is *dominant* over another spectral zero λ_j , if

$$\frac{\|\mathbf{R}_i\|_2}{|\operatorname{Re}(\lambda_i)|} > \frac{\|\mathbf{R}_j\|_2}{|\operatorname{Re}(\lambda_j)|}.$$

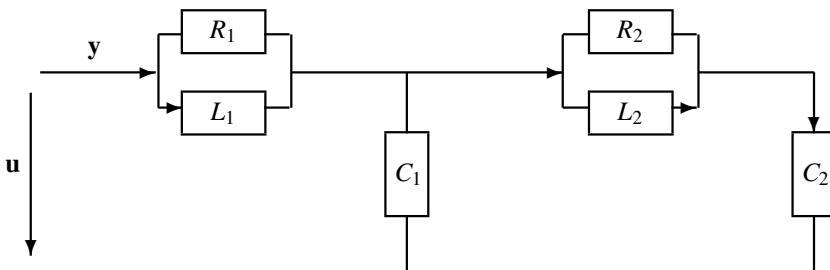
To efficiently compute the *r most dominant* spectral zeros of a dynamical system represented by $\mathbf{H}(s)$, we use an algorithm introduced in [77], and referred to as SADPA (Subspace Accelerated Dominant Pole Algorithm).

To summarize, the benefits of the spectral zero method are:

- (1) The dominant spectral zeros λ_j can be computed automatically
- (2) The resulting eigenvectors are readily available to construct projection matrices, \mathbf{V}_r and \mathbf{W}_r , for the right and left modeling subspaces
- (3) The dominance criterion is flexible and easily adjustable and
- (4) Iterative methods are suitable for large scale applications For a detailed description of the resulting algorithm and all the missing details see [55].

4.1 An Example of Passivity Preserving Model Reduction

We conclude this section with a simple example which illustrates the concepts discussed above. Consider the following RLC circuit:



Using the voltages across C_1 , C_2 , and the currents through L_1 , L_2 , as state variables, $\mathbf{x}_i = 1, 2, 3, 4$, respectively, we end up with equations of the form $\mathbf{E}\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$, $\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t)$, where

$$\mathbf{E} = \begin{pmatrix} C_1 & 0 & -G_1 L_1 & G_2 L_2 \\ 0 & C_2 & 0 & -G_2 L_2 \\ 0 & 0 & L_1 & 0 \\ 0 & 0 & 0 & L_2 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{C} = [-G_1, 0, 1, 0], \quad \mathbf{D} = G_1,$$

and $G_i = \frac{1}{R_i}$, $i = 1, 2$, are the corresponding conductances. By construction, the system is passive (for non-negative values of the parameters), and it is easy to see that its transfer function has a zero at $s = 0$. Hence the system has a double spectral zero at $s = 0$. According to the definition of dominance mentioned above, among all finite spectral zeros, those on the imaginary axis are dominant. Hence we will compute a second order reduced system by using the eigenpairs of $(\mathcal{H}, \mathcal{E})$, corresponding to the double zero eigenvalue. The Hamiltonian pair is:

$$\mathcal{H} = \begin{pmatrix} 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & G_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ -G_1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 2G_1 \end{pmatrix},$$

and

$$\mathcal{E} = \begin{pmatrix} C_1 & 0 & -G_1 L_1 & G_2 L_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & C_2 & 0 & -G_2 L_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & L_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & L_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & C_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & C_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -G_1 L_1 & 0 & L_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & G_2 L_2 & -G_2 L_2 & 0 & L_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

It follows that although the algebraic multiplicity of this eigenvalue is two, its geometric multiplicity is only one. Hence we need a Jordan chain of eigenvectors \mathbf{x}_1 , \mathbf{x}_2 , corresponding to this eigenvalue. In particular, \mathbf{x}_1 satisfies $\mathcal{H}\mathbf{x}_1 = 0$, while \mathbf{x}_2 , satisfies $\mathcal{H}\mathbf{x}_2 = \mathcal{E}\mathbf{x}_1$. These eigenvectors are:

$$[\mathbf{x}_1, \mathbf{x}_2] = \left(\begin{array}{c|c} \begin{matrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & \frac{C_2}{C_1+C_2} \end{matrix} & \\ \hline \begin{matrix} -1 & 0 \\ -1 & 0 \\ -G_1 & -1 \\ 0 & \frac{-C_2}{C_1+C_2} \end{matrix} & \\ \hline \begin{matrix} 1 & 0 \end{matrix} & \end{array} \right).$$

Thus the projection is defined by means of $\mathbf{V}_r = [\mathbf{x}_1(1 : 4, :), \mathbf{x}_2(1 : 4, :)]$ and $\mathbf{W}_r = -[\mathbf{x}_1(5 : 8, :), \mathbf{x}_2(5 : 8, :)]$, that is

$$\mathbf{V}_r = \left(\begin{array}{c|c} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & \frac{C_2}{C_1+C_2} \end{array} \right), \quad \mathbf{W}_r = \left(\begin{array}{c|c} 1 & 0 \\ 1 & 0 \\ G_1 & 1 \\ 0 & \frac{C_2}{C_1+C_2} \end{array} \right).$$

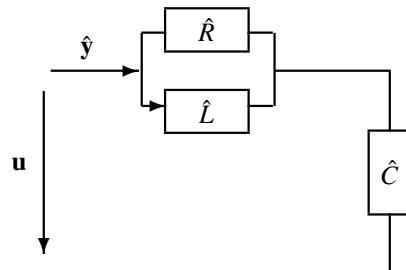
Therefore by (10) the reduced quantities are:

$$\begin{aligned} \mathbf{E}_r &= \mathbf{W}_r^* \mathbf{E} \mathbf{V}_r = \begin{pmatrix} C_1 + C_2 & 0 \\ 0 & L_1 + \frac{C_2^2}{(C_1 + C_2)^2} L_2 \end{pmatrix}, \quad \mathbf{A}_r = \mathbf{W}_r^* \mathbf{A} \mathbf{V}_r = \begin{pmatrix} -G_1 & 1 \\ -1 & 0 \end{pmatrix}, \\ \mathbf{B}_r &= \mathbf{W}_r^* \mathbf{B} = \begin{pmatrix} G_1 \\ 1 \end{pmatrix}, \quad \mathbf{C}_r = \mathbf{C} \mathbf{V}_r = (-G_1 \ 1), \quad \mathbf{D}_r = \mathbf{D}. \end{aligned}$$

The corresponding transfer function $\mathbf{H}_r(s) = \mathbf{D} + \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r$, can be expressed as follows:

$$\mathbf{H}_r^{-1}(s) = \frac{1}{s(C_1 + C_2)} + \frac{1}{G_1 + \frac{\kappa}{s}} \quad \text{where} \quad \kappa = \frac{L_1 L_2 (C_1 + C_2)^2}{L_1 C_2^2 + L_2 (C_1 + C_2)^2}.$$

From this expression or from the state space matrices, we can read-off an RLC realization. The reduced order circuit contains namely, a capacitor $\hat{C} = C_1 + C_2$, an inductor $\hat{L} = \frac{1}{\kappa} = L_1 + \frac{C_2^2}{(C_1 + C_2)^2} L_2$, and a resistor of value $\hat{R} = R_1$. Thereby the capacitor is in series with a parallel connection of the inductor and the resistor as shown below.



Hence in this particular case, after reduction besides passivity, the structure (topology) of the circuit using the spectral zero reduction method, is preserved.

5 Structure-Preserving Model Reduction Using Generalized Coprime Factorizations

The model reduction framework we have described for reducing an original system of the form (1) to a reduced-order model of similar form (2) is quite general. With suitable generalizations and reformulations, this framework will encompass many linear dynamical systems of practical importance. However, there exist many linear dynamical systems whose natural descriptions take a form rather different than the standard canonical form (1).

We motivate the by considering a model of the dynamic response of a viscoelastic body. One possible continuum model for forced vibration of an isotropic incompressible viscoelastic solid could be described as

$$\partial_{tt} \mathbf{w}(x, t) - \eta \Delta \mathbf{w}(x, t) - \int_0^t \rho(t-\tau) \Delta \mathbf{w}(x, \tau) d\tau + \nabla \varpi(x, t) = \mathbf{b}(x) \cdot \mathbf{u}(t), \\ \nabla \cdot \mathbf{w}(x, t) = 0 \text{ determining } \mathbf{y}(t) = [\mathbf{w}(\hat{x}_1, t), \dots, \mathbf{w}(\hat{x}_p, t)]^T, \quad (28)$$

where $\mathbf{w}(x, t)$ is the displacement field for the body, $\varpi(x, t)$ is the associated pressure field and $\nabla \cdot \mathbf{w} = 0$ represents the incompressibility constraint (see e.g., [63]). Here, $\rho(\tau)$ is a presumed known “relaxation function” satisfying $\rho(\tau) \geq 0$ and $\int_0^\infty \rho(\tau) d\tau < \infty$; $\eta > 0$ is a known constant associated with the “initial elastic response”. The term $\mathbf{b}(x) \cdot \mathbf{u}(t) = \sum_{i=1}^m b_i(x) u_i(t)$ is a superposition of the m inputs $\mathbf{u}(t) = \{u_1(t), \dots, u_m(t)\}^T \in \mathbb{R}^m$. Displacements at $\hat{x}_1, \dots, \hat{x}_p$ are the outputs. Semi-discretization of (28) with respect to space produces a large order linear dynamical system of the form:

$$\mathbf{M} \ddot{\mathbf{x}}(t) + \eta \mathbf{K} \dot{\mathbf{x}}(t) + \int_0^t \rho(t-\tau) \mathbf{K} \mathbf{x}(\tau) d\tau + \mathbf{D} \varpi(t) = \mathbf{B} \mathbf{u}(t), \\ \mathbf{D}^T \mathbf{x}(t) = \mathbf{0}, \quad \text{which determines } \mathbf{y}(t) = \mathbf{C} \mathbf{x}(t). \quad (29)$$

where $\mathbf{x} \in \mathbb{R}^{n_1}$ is the discretization of the continuous displacement field \mathbf{w} ; $\varpi \in \mathbb{R}^{n_2}$ is the discretization of the continuous pressure field ϖ . The matrices \mathbf{M} and \mathbf{K} are $n_1 \times n_1$ real, symmetric, positive-definite matrices, \mathbf{D} is an $n_1 \times n_2$ matrix, $\mathbf{B} \in \mathbb{R}^{n_1 \times m}$, and $\mathbf{C} \in \mathbb{R}^{p \times n_2}$. The state space dimension is $n = n_1 + n_2$, an aggregate of $\mathbf{x}(t)$ and $\varpi(t)$.

Applying the Laplace transform to (29) yields

$$\widehat{\mathbf{y}}(s) = [\mathbf{C} \quad \mathbf{0}] \begin{bmatrix} s^2 \mathbf{M} + (\widehat{\rho}(s) + \eta) \mathbf{K} & \mathbf{D} \\ \mathbf{D}^T & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{B} \\ \mathbf{0} \end{bmatrix} \widehat{\mathbf{u}}(s) = \mathcal{H}(s) \widehat{\mathbf{u}}(s), \quad (30)$$

defining the transfer function, $\mathcal{H}(s)$. The form of the transfer function is clearly different from the standard form (5). The system described in (29) is a *descriptor* system described by differential-algebraic equations with a hereditary damping; and a reformulation of (29) into the standard realization as in (1) is generally not possible (unless $\widehat{\rho}(s)$ has a simple form) and even when possible may not be desirable.

An effective reduced model for (29) should take into account the structure associated with distributed material properties. Therefore, we wish to consider reduced models similar to (29):

$$\begin{aligned} \mathbf{M}_r \ddot{\mathbf{x}}_r(t) + \eta \mathbf{K}_r \mathbf{x}_r(t) + \int_0^t \rho(t-\tau) \mathbf{K}_r \mathbf{x}_r(\tau) d\tau + \mathbf{D}_r \boldsymbol{\varpi}_r &= \mathbf{B}_r \mathbf{u}(t) \\ \mathbf{D}_r^T \mathbf{x}_r(t) &= \mathbf{0} \quad \text{which determines} \quad \mathbf{y}_r(t) = \mathbf{C}_r \mathbf{x}_r(t) \end{aligned} \quad (31)$$

where \mathbf{M}_r and \mathbf{K}_r are now $r_1 \times r_1$ real, symmetric, positive-definite matrices, \mathbf{D}_r is an $r_1 \times r_2$ matrix, $\mathbf{B}_r \in \mathbb{R}^{r_1 \times m}$ and $\mathbf{C}_r \in \mathbb{R}^{p \times r_2}$ with $r_1 \ll n_1$ and $r_2 \ll n_2$ and the reduced state space dimension is $r = r_1 + r_2$. Note that the reduced-model (31) has the same structure with the same hereditary damping as the original model. Systems with such structure may be expected to produce dynamic responses that are physically plausible and consistent with behaviors of viscoelastic bodies.

We proceed to construct reduced-order matrices in a similar way as we did in the standard case: We define matrices of trial vectors $\mathbf{U}_r \in \mathbb{R}^{n_1 \times r_1}$ and $\mathbf{Z}_r \in \mathbb{R}^{n_2 \times r_2}$; use the ansatz $\mathbf{x}(t) \approx \mathbf{U}_r \mathbf{x}_r(t)$ and $\boldsymbol{\varpi}(t) \approx \mathbf{Z}_r \boldsymbol{\varpi}_r(t)$; and force a Petrov-Galerkin orthogonality condition on the reduced state-space trajectory residuals to obtain finally reduced coefficient matrices for (31) as

$$\begin{aligned} \mathbf{M}_r &= \mathbf{U}_r^T \mathbf{M} \mathbf{U}_r, \quad \mathbf{K}_r = \mathbf{U}_r^T \mathbf{K} \mathbf{U}_r, \quad \mathbf{D}_r = \mathbf{U}_r^T \mathbf{D} \mathbf{Z}_r, \\ \mathbf{B}_r &= \mathbf{U}_r^T \mathbf{B}, \quad \text{and} \quad \mathbf{C}_r = \mathbf{C} \mathbf{U}_r. \end{aligned} \quad (32)$$

Note that this construction prevents mixing displacement state variables and pressure state variables. Also notice that both symmetry and positive-definiteness of \mathbf{M}_r and \mathbf{K}_r are preserved automatically as long as \mathbf{U}_r has full column rank. The transfer function $\mathbf{H}_r(s)$ for the reduced-model can be obtained similarly as in (30). Now, in addition to preserving the structure we want to choose \mathbf{U}_r and \mathbf{Z}_r so that the reduced model $\mathbf{H}_r(s)$ interpolates $\mathbf{H}(s)$ as in the generic case. The ideas of Theorem 1.2 of §2.3 must be extended to this more general setting.

To proceed, we follow the discussion of Beattie and Gugercin [16] and consider MIMO systems with the following state space description in the Laplace transform domain (which includes the form of (30)):

$$\begin{aligned} \text{Find } \widehat{\mathbf{v}}(s) \text{ such that} \quad \mathcal{K}(s) \widehat{\mathbf{v}}(s) &= \mathcal{B}(s) \widehat{\mathbf{u}}(s) \\ \text{then} \quad \widehat{\mathbf{y}}(s) &\stackrel{\text{def}}{=} \mathcal{C}(s) \widehat{\mathbf{v}}(s) + \mathbf{D} \widehat{\mathbf{u}}(s) \\ \text{yielding a transfer function:} \quad \mathbf{H}(s) &= \mathcal{C}(s) \mathcal{K}(s)^{-1} \mathcal{B}(s) + \mathbf{D}, \end{aligned} \quad (33)$$

where $\mathbf{D} \in \mathbb{R}^{p \times m}$, both $\mathcal{C}(s) \in \mathbb{C}^{p \times n}$ and $\mathcal{B}(s) \in \mathbb{C}^{n \times m}$ are analytic in the right half plane; and $\mathcal{K}(s) \in \mathbb{C}^{n \times n}$ is analytic and full rank throughout the right half plane. The goal is to find a reduced transfer function of the same form using a Petrov-Galerkin projection:

$$\mathbf{H}_r(s) = \mathcal{C}_r(s)\mathcal{K}_r(s)^{-1}\mathcal{B}_r(s) + \mathbf{D}_r \quad (34)$$

where $\mathbf{W}_r, \mathbf{V}_r \in \mathbb{C}^{n \times r}$, $\mathcal{C}_r(s) = \mathcal{C}(s)\mathbf{V}_r \in \mathbb{C}^{p \times r}$; $\mathcal{B}_r(s) = \mathbf{W}_r^T\mathcal{B}(s) \in \mathbb{C}^{r \times m}$, and $\mathcal{K}_r(s) = \mathbf{W}_r^T\mathcal{K}(s)\mathbf{V}_r \in \mathbb{C}^{n \times n}$.

For simplicity we only consider the case $\mathbf{D} = \mathbf{D}_r$, which is equivalent with respect to the interpolation conditions to $\mathbf{D} = \mathbf{D}_r = \mathbf{0}$. The general case with $\mathbf{D} \neq \mathbf{D}_r$ (similar to Theorem 1.3) can be found in the original source [16]. Following the notation in [16], we write $\mathcal{D}_\sigma^\ell f$ to denote the ℓ^{th} derivative of the univariate function $f(s)$ evaluated at $s = \sigma$ with the usual convention for $\ell = 0$, $\mathcal{D}_\sigma^0 f = f(\sigma)$.

Theorem 1.7. Suppose that $\mathcal{B}(s)$, $\mathcal{C}(s)$, and $\mathcal{K}(s)$ are analytic at $\sigma \in \mathbb{C}$ and $\mu \in \mathbb{C}$. Also let $\mathcal{K}(\sigma)$, $\mathcal{K}(\mu)$, $\mathcal{K}_r(\sigma) = \mathbf{W}_r^T\mathcal{K}(\sigma)\mathbf{V}_r$, and $\mathcal{K}_r(\mu) = \mathbf{W}_r^T\mathcal{K}(\mu)\mathbf{V}_r$ have full rank. Let nonnegative integers M and N be given as well as nontrivial vectors, $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{c} \in \mathbb{R}^p$.

- (a) If $\mathcal{D}_\sigma^i[\mathcal{K}(s)^{-1}\mathcal{B}(s)]\mathbf{b} \in \text{Ran}(\mathbf{V}_r)$ for $i = 0, \dots, N$
then $\mathbf{H}^{(\ell)}(\sigma)\mathbf{b} = \mathbf{H}_r^{(\ell)}(\sigma)\mathbf{b}$ for $\ell = 0, \dots, N$.
- (b) If $\left(\mathbf{c}^T \mathcal{D}_\mu^j [\mathcal{C}(s)\mathcal{K}(s)^{-1}] \right)^T \in \text{Ran}(\mathbf{W}_r)$ for $j = 0, \dots, M$
then $\mathbf{c}^T \mathbf{H}^{(\ell)}(\mu) = \mathbf{c}^T \mathbf{H}_r^{(\ell)}(\mu)$ for $\ell = 0, \dots, M$.
- (c) If both (a) and (b) hold and if $\sigma = \mu$,
then $\mathbf{c}^T \mathbf{H}^{(\ell)}(\sigma)\mathbf{b} = \mathbf{c}^T \mathbf{H}_r^{(\ell)}(\sigma)\mathbf{b}$ for $\ell = 0, \dots, M+N+1$.

Using this result, one can easily construct a reduced-model satisfying the desired interpolation conditions. For example, for $\mathbf{H}(s) = \mathcal{C}(s)\mathcal{K}(s)^{-1}\mathcal{B}(s)$, let r interpolation points $\{\sigma_i\}_{i=1}^r$, the left-directions $\{\mathbf{c}_i\}_{i=1}^r$ and the right-directions $\{\mathbf{b}_i\}_{i=1}^r$ be given. Construct

$$\mathbf{V}_r = [\mathcal{K}(\sigma_1)^{-1}\mathcal{B}(\sigma_1)\mathbf{b}_1, \dots, \mathcal{K}(\sigma_r)^{-1}\mathcal{B}(\sigma_r)\mathbf{b}_r] \quad (35)$$

$$\text{and } \mathbf{W}_r = [\mathcal{K}(\sigma_1)^{-T}\mathcal{C}(\sigma_1)^T\mathbf{c}_1, \dots, \mathcal{K}(\sigma_r)^{-T}\mathcal{C}(\sigma_r)^T\mathbf{c}_r]^T. \quad (36)$$

The reduced transfer function, $\mathbf{H}_r(s) = \mathcal{C}_r(s)\mathcal{K}_r(s)^{-1}\mathcal{B}_r(s)$ satisfies bi-tangential interpolation conditions, i.e., $\mathbf{H}(\sigma_i)\mathbf{b}_i = \mathbf{H}_r(\sigma_i)\mathbf{b}_i$, $\mathbf{c}_i^T \mathbf{H}(\sigma_i) = \mathbf{c}_i^T \mathbf{H}_r(\sigma_i)$ and $\mathbf{c}_i^T \mathbf{H}'(\sigma_i)\mathbf{b}_i = \mathbf{c}_i^T \mathbf{H}'_r(\sigma_i)\mathbf{b}_i$ for $i = 1, \dots, r$.

Theorem 1.7 achieves the desired result; it extends the interpolation theory of Theorem 1.2 from the canonical first-order setting to a general setting where we have a general coprime factorization, $\mathbf{H}(s) = \mathcal{C}(s)\mathcal{K}(s)^{-1}\mathcal{B}(s) + \mathbf{D}$ and guarantees that the reduced transfer function will have a similarly structured coprime factorization: $\mathbf{H}_r(s) = \mathcal{C}_r(s)\mathcal{K}_r(s)^{-1}\mathcal{B}_r(s) + \mathbf{D}_r$.

5.1 A Numerical Example: Driven Cavity Flow

We illustrate the concepts of Theorem 1.7 with a model of driven cavity flow in two dimensions, a slightly modified version of a model in (28). Consider a square domain $\Omega = [0, 1]^2$, as shown below, representing a volume filled with a viscoelastic material with boundary separated into a movable top edge (“lid”), $\partial\Omega_1$, and the complement, $\partial\Omega_0$ (stationary bottom, left, and right edges). The material in the cavity is excited through shearing forces on the material caused by transverse displacement of the lid, $u(t)$. We are interested in the displacement response of the material, $\mathbf{w}(\hat{x}, t)$, at the center of Ω , i.e. $\hat{x} = (0.5, 0.5)$.

$$\partial_{tt}\mathbf{w}(x, t) - \eta_0 \Delta \mathbf{w}(x, t) - \eta_1 \partial_t \int_0^t \frac{\Delta \mathbf{w}(x, \tau)}{(t - \tau)^\alpha} d\tau + \nabla \varpi(x, t) = 0 \text{ for } x \in \Omega$$

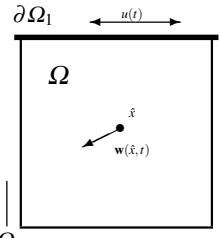
$$\begin{aligned} \nabla \cdot \mathbf{w}(x, t) &= 0 & \text{and} & \mathbf{w}(x, t) = 0 \text{ for } x \in \partial\Omega_0 \\ \text{for } x \in \Omega & & \mathbf{w}(x, t) &= u(t) \text{ for } x \in \partial\Omega_1 \\ \text{which determines the output } \mathbf{y}(t) &= \mathbf{w}(\hat{x}, t), \end{aligned}$$

We use a model of a viscoelastic material given by

Bagley and Torvik in [6] and take as material constants,

$$\eta_0 = 1.05 \times 10^6, \quad \eta_1 = 2.44 \times 10^5, \quad \text{and} \quad \alpha = 0.519, \quad \partial\Omega_0$$

corresponding to experimentally derived values for the polymer butyl B252.



We discretize the continuous model using Taylor-Hood finite elements defined over a uniform triangular mesh on Ω . The transfer function of the discretized system is given by $\mathbf{H}(s) = \mathcal{C}(s)\mathcal{K}(s)^{-1}\mathcal{B}(s)$ where

$$\mathcal{K}(s) = \begin{bmatrix} s^2 \mathbf{M} + \widehat{\rho}(s) \mathbf{K} & \mathbf{D} \\ \mathbf{D}^T & \mathbf{0} \end{bmatrix}, \quad \mathcal{C}(s) = [\mathbf{C} \ \mathbf{0}], \quad \text{and} \quad \mathcal{B}(s) = \begin{bmatrix} s^2 \mathbf{m} + \widehat{\rho}(s) \mathbf{k} \\ \mathbf{0} \end{bmatrix}.$$

$\mathbf{C} \in \mathbb{R}^{n \times 2}$ corresponds to measuring the horizontal and vertical displacement at $\hat{x} = (0.5, 0.5)$, \mathbf{m} and \mathbf{k} are sums of the columns of the free-free mass and stiffness matrix associated with x -displacement degrees of freedom on the top lid boundary and $\widehat{\rho}(s) = \eta_0 + \eta_1 s^\alpha$. This produces a frequency dependent input-to-state map, $\mathcal{B}(s)$, since the system input is a boundary *displacement* as opposed to a boundary *force*. Note the nonlinear frequency dependency in the state-to-input map; however Theorem 1.7 can be still applied to produce an interpolatory reduced-order model of the same form as shown next: Suppose an interpolation point $\sigma \in \mathbb{C}$ and a direction $c \in \mathbb{C}^p$ are given. Since the input is a scalar, there is no need for a right tangential direction. We consider bitangential Hermite interpolating conditions:

$$\mathcal{H}_r(\sigma) = \mathcal{H}(\sigma), \quad c^T \mathcal{H}_r(\sigma) = c^T \mathcal{H}(\sigma), \quad \text{and} \quad c^T \mathcal{H}'_r(\sigma) = c^T \mathcal{H}'(\sigma). \quad (37)$$

Following Theorem 1.7, we solve the following two linear systems of equations:

$$\begin{bmatrix} \mathbf{F}(\sigma) & \mathbf{D} \\ \mathbf{D}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{z}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{N}(\sigma) \\ \mathbf{0} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathbf{F}(\sigma) & \mathbf{D} \\ \mathbf{D}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u}_2 \\ \mathbf{z}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{C}^T \mathbf{c} \\ \mathbf{0} \end{bmatrix}.$$

where $\mathbf{F}(\sigma) = \sigma^2 \mathbf{M} + \hat{\rho}(\sigma) \mathbf{K}$ and $\mathbf{N}(\sigma) = s^2 \mathbf{m} + \hat{\rho}(\sigma) \mathbf{k}$. Define the matrices $\mathbf{U}_r = [\mathbf{u}_1, \mathbf{u}_2]$ and $\mathbf{Z}_r = [\mathbf{z}_1, \mathbf{z}_2]$. Then the reduced system matrices defined in (32) with the modification that $\mathcal{B}_r(s) = s^2 \mathbf{U}_r^T \mathbf{m} + \hat{\rho}(s) \mathbf{U}_r^T \mathbf{k}$ corresponds to the choice $\mathbf{V}_r = \mathbf{W}_r = \mathbf{U}_r \oplus \mathbf{Z}_r$ and satisfies the bitangential Hermite interpolation conditions of (37).

We compare three different models:

1. \mathbf{H}_{fine} , using a fine mesh Taylor-Hood FEM discretization with 51,842 displacement degrees of freedom and 6,651 pressure degrees of freedom (mesh size $h = \frac{1}{80}$). We treat this as the “full-order model”.
2. $\mathbf{H}_{\text{coarse}}$, for a coarse mesh discretization with 29,282 displacement degrees of freedom and 3721 pressure degrees of freedom (mesh size $h = \frac{1}{60}$);
3. \mathbf{H}_{30} , a generalized interpolatory reduced order model as defined in (31)–(32) with $r = 30$, corresponding to 30 reduced displacement degrees of freedom and 30 reduced pressure degrees of freedom satisfying the bitangential Hermite interpolation conditions at each interpolation point.

The resulting frequency response plots shown in Figure 3:

In this example, this new framework allows interpolatory model reduction of a descriptor system with a hereditary damping term. Moreover, the reduced model is also a descriptor system with the same damping structure. The reduced model

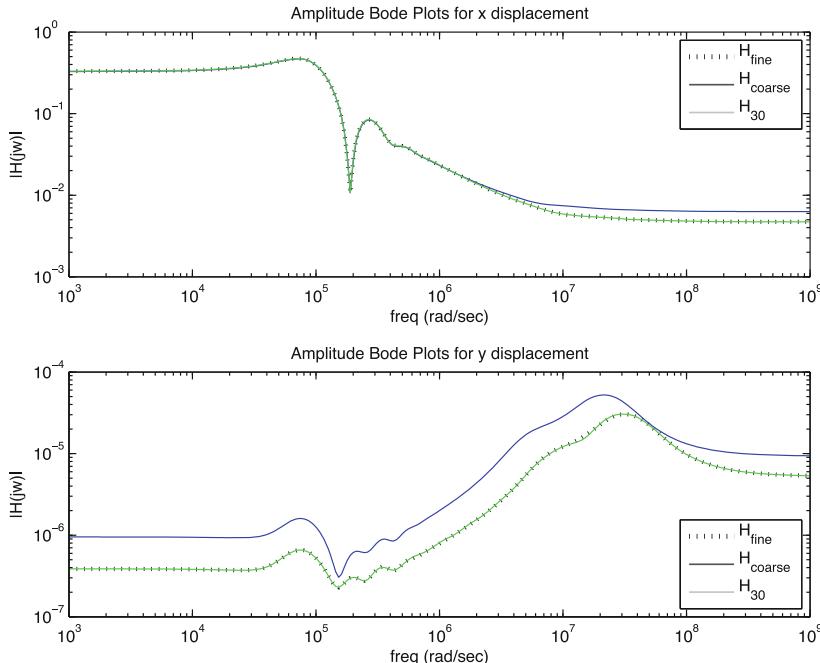


Fig. 3 Bode plots of \mathbf{H}_{fine} , $\mathbf{H}_{\text{coarse}}$ and reduced models \mathbf{H}_{20} and \mathbf{H}_{30}

\mathbf{H}_{30} having a total of 60 degrees of freedom has an input-output response that is virtually indistinguishable from the high fidelity model \mathbf{H}_{fine} having nearly 58,500 total degrees of freedom. Moreover, \mathbf{H}_{30} outperforms $\mathbf{H}_{\text{coarse}}$, which corresponds to 29,282 displacement and 3,721 pressure degrees of freedom (more than 500 times the order of \mathbf{H}_{30}).

5.2 Second-Order Dynamical Systems

One important variation on the dynamical system described in (33) arises in modeling the forced vibration of an n degree-of-freedom mechanical structure:

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{G}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) = \mathbf{B}\mathbf{u}(t), \quad \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t), \quad (38)$$

$$\text{with the transfer function } \mathbf{H}(s) = \mathbf{C}(s^2\mathbf{M} + s\mathbf{G} + \mathbf{K})^{-1}\mathbf{B} \quad (39)$$

where \mathbf{M} , \mathbf{G} , and $\mathbf{K} \in \mathbb{R}^{n \times n}$ are positive (semi)-definite symmetric matrices describing, respectively, inertial mass distribution, energy dissipation, and elastic strain energy (stiffness) distribution throughout the structure. The input $\mathbf{u}(t) \in \mathbb{R}^m$ is a time-dependent force applied along degrees-of-freedom specified in $\mathbf{B} \in \mathbb{R}^{n \times m}$ and $\mathbf{y}(t) \in \mathbb{R}^p$ is a vector of output measurements defined through observation matrix $\mathbf{C} \in \mathbb{R}^p$. Second order systems of the form (38) arise naturally in analyzing other phenomena aside from structural vibration such as electrical circuits and micro-electro-mechanical systems, as well; see [7, 8, 26–28, 57, 75, 84], and references therein.

We wish to generate, for some $r \ll n$, an r^{th} order reduced *second-order system* of the same form:

$$\mathbf{M}_r\ddot{\mathbf{x}}_r(t) + \mathbf{G}_r\dot{\mathbf{x}}_r(t) + \mathbf{K}_r\mathbf{x}_r(t) = \mathbf{B}_r\mathbf{u}(t), \quad \mathbf{y}_r(t) = \mathbf{C}_r\mathbf{x}_r(t),$$

where $\mathbf{M}_r, \mathbf{G}_r, \mathbf{K}_r \in \mathbb{R}^{r \times r}$ are positive (semi)-definite symmetric matrices, $\mathbf{B} \in \mathbb{R}^{r \times m}$, and $\mathbf{C} \in \mathbb{R}^{p \times r}$. In order to preserve the symmetry and positive definiteness of \mathbf{M} , \mathbf{G} and \mathbf{K} in the course of model reduction, one-sided reduction is applied, i.e. one takes $\mathbf{W}_r = \mathbf{V}_r$, resulting in

$$\mathbf{M}_r = \mathbf{V}_r^T \mathbf{M} \mathbf{V}_r, \quad \mathbf{G}_r = \mathbf{V}_r^T \mathbf{G} \mathbf{V}_r, \quad \mathbf{K}_r = \mathbf{V}_r^T \mathbf{K} \mathbf{V}_r, \quad \mathbf{B}_r = \mathbf{V}_r^T \mathbf{B}, \quad \mathbf{C}_r = \mathbf{C} \mathbf{V}_r.$$

One standard approach to model reduction of second-order systems involves the application of standard (first-order) model reduction techniques to a standard first-order realization of the system. However, this destroys the original second-order system structure and obscures the physical meaning of the reduced-order states. Once the reduction is performed in the first-order framework, it will not always be possible to convert this back to a corresponding second-order system of the form (38), see [69]. Even when this is possible, one typically cannot guarantee that structural properties such as positive definite symmetric reduced mass, damping, and stiffness

matrices, will be retained. Keeping the original structure is crucial both to preserve physical meaning of the states and to retain physically significant properties such as stability and passivity.

There have been significant efforts in this direction. Building on the earlier work of [82], Bai and Su [13] introduced “second-order” Krylov subspaces and showed how to obtain a reduced-order system directly in a second-order framework while still satisfying interpolation conditions at selected points for single-input/single-output systems. Other second-order structure preserving interpolatory reduction techniques were introduced by Chahlaoui *et al.* [26] and Freund [38], though these approaches require a first-order state-space realization.

The second-order transfer function $\mathbf{H}(s)$ in (39) fits perfectly in the framework of Theorem 1.7. Indeed, due to the simple structures of $\mathcal{C}(s) = \mathbf{C}$, $\mathcal{B}(s) = \mathbf{B}$ and $\mathcal{K}(s) = s^2\mathbf{M} + s\mathbf{G} + \mathbf{K}$, one obtains a simple three-term recurrence for the rational tangential interpolation of MIMO second-order systems directly in a second-order framework. This is presented in Algorithm 4.1.

From Theorem 1.7, the resulting reduced second-order model tangentially interpolates the original model up through derivatives of order $J_i - 1$ at the selected (complex) frequencies σ_i while preserving the original second-order structure. The second-order recursion of Bai and Su [13] is a special case of this algorithm for a single-input/single-output system using a single interpolation point σ .

Algorithm 4.1. Second-order Tangential MIMO Order Reduction

Given interpolation points $\{\sigma_1, \sigma_2, \dots, \sigma_N\}$, directions $\{b_1, b_2, \dots, b_N\}$, and interpolation orders $\{J_1, J_2, \dots, J_N\}$ (with $r = \sum_{i=1}^N J_i$).

1. For $i = 1, \dots, N$
 - a. For each shift σ_i and tangent direction b_i , define

$$\mathcal{K}_0^{(i)} = \sigma_i^2 \mathbf{M} + \sigma_i \mathbf{G} + \mathbf{K}, \quad \mathcal{K}_1^{(i)} = 2\sigma_i \mathbf{M} + \mathbf{G}, \text{ and } \mathcal{K}_2 = \mathbf{M}.$$
 - b. Solve $\mathcal{K}_0^{(i)} \mathbf{f}_{i1} = \mathbf{B} \mathbf{b}_i$ and set $\mathbf{f}_{i0} = 0$.
 - c. For $j = 2 : J_i$
 - Solve $\mathcal{K}_0^{(i)} \mathbf{f}_{ij} = -\mathcal{K}_1^{(i)} \mathbf{f}_{i,j-1} - \mathcal{K}_2 \mathbf{f}_{i,j-2}$,
2. Take $\mathbf{V}_r = [\mathbf{f}_{11}, \mathbf{f}_{12}, \dots, \mathbf{f}_{1J_1}, \mathbf{f}_{21}, \dots, \mathbf{f}_{2J_2}, \dots, \mathbf{f}_{N1}, \dots, \mathbf{f}_{NJ_N}]$ and then $\mathbf{M}_r = \mathbf{V}_r^T \mathbf{M} \mathbf{V}_r$, $\mathbf{G}_r = \mathbf{V}_r^T \mathbf{G} \mathbf{V}_r$, $\mathbf{K}_r = \mathbf{V}_r^T \mathbf{K} \mathbf{V}_r$, $\mathbf{B}_r = \mathbf{V}_r^T \mathbf{B}$, and $\mathbf{C}_r = \mathbf{C} \mathbf{V}_r$.

Remark 1.2. The general framework presented here can handle second-order systems of the form

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{G}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) = \mathbf{B}_1\dot{\mathbf{u}}(t) + \mathbf{B}_0\mathbf{u}(t), \quad \mathbf{y}(t) = \mathbf{C}_1\dot{\mathbf{x}}(t) + \mathbf{C}_0\mathbf{x}(t), \quad (40)$$

where the input $\mathbf{u}(t) \in \mathbb{R}^m$ is a time-dependent force or displacement, the output $\mathbf{y}(t) \in \mathbb{R}^p$ depends potentially not only on displacements but also the velocities. We note that second-order systems of the form (38) cannot be converted to generic first-order form (1) in a straightforward way when $\mathbf{B}_1 \neq 0$.

Remark 1.3. The discussion above and the algorithm can be easily generalized to higher order constant coefficient ordinary differential equations as well where the system dynamics follow

$$\mathbf{A}_0 \frac{d^\ell \mathbf{x}}{dt^\ell} + \mathbf{A}_1 \frac{d^{\ell-1} \mathbf{x}}{dt^{\ell-1}} + \cdots + \mathbf{A}_\ell \mathbf{x}(t) = \mathbf{B} \mathbf{u}(t) \quad \text{and} \quad \mathbf{y}(t) = \mathbf{C} \mathbf{x}(t). \quad (41)$$

6 Model Reduction of Parametric Systems

Frequently dynamical systems under study are varied by adjusting a comparatively small number of parameters repeatedly during the course of a set of simulations. Parameters can enter a model representing material properties, variations in shape of the structure or more generally of the solution domain, and strength of coupling in various types of boundary conditions. Micro-electromechanical systems, chip design and interconnect modeling are two of the most prominent examples where parameterized systems naturally arise; see, e.g., [29, 36, 51, 78]. It will be of value to have methods that will produce high fidelity reduced order models retaining the parametric structure of the original system. This leads to the concept of parameterized model reduction. The goal is to construct a high fidelity parametric reduced order models which recover the response of the original full order parametric system throughout the range of variation of interest of the design parameters.

Parametric model order reduction is at an early stage of the development. For interpolatory model reduction of parametric systems, see [14, 19, 22, 29, 32, 34–36, 51, 52, 64, 65, 71, 85, 86], and reference therein. These methods treat parametric dynamical systems in a standard first-order state-space form. Below, we extend these results to the generalized setting similar to that of Sect. 5.

We consider a multi-input/multi-output linear dynamical system that is parameterized with q parameters $\mathbf{p} = [\mathbf{p}_1, \dots, \mathbf{p}_q]$:

$$\mathbf{H}(s, \mathbf{p}) = \mathcal{C}(s, \mathbf{p}) \mathcal{K}(s, \mathbf{p})^{-1} \mathcal{B}(s, \mathbf{p}) \quad (42)$$

with $\mathcal{K}(s, \mathbf{p}) \in \mathbb{C}^{n \times n}$ and $\mathcal{B}(s, \mathbf{p}) \in \mathbb{C}^{n \times m}$ and $\mathcal{C}(s, \mathbf{p}) \in \mathbb{C}^{p \times n}$. We assume that

$$\begin{aligned} \mathcal{K}(s, \mathbf{p}) &= \mathcal{K}^{[0]}(s) + a_1(\mathbf{p}) \mathcal{K}^{[1]}(s) + \dots + a_v(\mathbf{p}) \mathcal{K}^{[v]}(s) \\ \mathcal{B}(s, \mathbf{p}) &= \mathcal{B}^{[0]}(s) + b_1(\mathbf{p}) \mathcal{B}^{[1]}(s) + \dots + b_v(\mathbf{p}) \mathcal{B}^{[v]}(s), \\ \mathcal{C}(s, \mathbf{p}) &= \mathcal{C}^{[0]}(s) + c_1(\mathbf{p}) \mathcal{C}^{[1]}(s) + \dots + c_v(\mathbf{p}) \mathcal{C}^{[v]}(s). \end{aligned} \quad (43)$$

where $a_1(\mathbf{p}), a_2(\mathbf{p}), \dots, b_1(\mathbf{p}), \dots, c_v(\mathbf{p})$ are scalar-valued parameter functions that could be linear or non-linear. Equation (42) represents a system structure that we may wish to retain – that is, our goal is to generate, for some $r \ll n$, a reduced-order system with dimension r having the same parametric structure. Suppose matrices $\mathbf{V}_r \in \mathbb{C}^{n \times r}$ and $\mathbf{W}_r \in \mathbb{C}^{n \times r}$ are specified and consider:

$$\mathbf{H}_r(s, \mathbf{p}) = \mathcal{C}_r(s, \mathbf{p}) \mathcal{K}_r(s, \mathbf{p})^{-1} \mathcal{B}_r(s, \mathbf{p}) \quad (44)$$

with $\mathcal{K}_r(s, \mathbf{p}) = \mathbf{W}_r^T \mathcal{K}(s, \mathbf{p}) \mathbf{V}_r$, $\mathcal{B}_r(s, \mathbf{p}) = \mathbf{W}_r^T \mathcal{B}(s, \mathbf{p})$, and $\mathcal{C}_r(s, \mathbf{p}) = \mathcal{C}(s, \mathbf{p}) \mathbf{V}_r$. We say that the reduced model has *the same parametric structure* in the sense that

$$\begin{aligned}\mathcal{K}_r(s, \mathbf{p}) &= \left(\mathbf{W}_r^T \mathcal{K}^{[0]}(s) \mathbf{V}_r \right) + a_1(\mathbf{p}) \left(\mathbf{W}_r^T \mathcal{K}^{[1]}(s) \mathbf{V}_r \right) + \dots + a_v(\mathbf{p}) \left(\mathbf{W}_r^T \mathcal{K}^{[v]}(s) \mathbf{V}_r \right) \\ \mathcal{B}_r(s, \mathbf{p}) &= \left(\mathbf{W}_r^T \mathcal{B}^{[0]}(s) \right) + b_1(\mathbf{p}) \left(\mathbf{W}_r^T \mathcal{B}^{[1]}(s) \right) + \dots + b_v(\mathbf{p}) \left(\mathbf{W}_r^T \mathcal{B}^{[v]}(s) \right), \\ \mathcal{C}_r(s, \mathbf{p}) &= \left(\mathcal{C}^{[0]}(s) \mathbf{V}_r \right) + c_1(\mathbf{p}) \left(\mathcal{C}^{[1]}(s) \mathbf{V}_r \right) + \dots + c_v(\mathbf{p}) \left(\mathcal{C}^{[v]}(s) \mathbf{V}_r \right).\end{aligned}\quad (45)$$

with exactly the *same* parameter functions $a_1(\mathbf{p}), \dots, c_v(\mathbf{p})$ as in (43), but with smaller coefficient matrices. Significantly, all reduced coefficient matrices can be *precomputed* before the reduced model is put into service. The next result extends Theorem 1.7 to the parameterized dynamical system setting:

Theorem 1.8. Suppose $\mathcal{K}(s, \mathbf{p})$, $\mathcal{B}(s, \mathbf{p})$, and $\mathcal{C}(s, \mathbf{p})$ are analytic with respect to s at $\sigma \in \mathbb{C}$ and $\mu \in \mathbb{C}$, and are continuously differentiable with respect to \mathbf{p} in a neighborhood of $\hat{\mathbf{p}} = [\hat{p}_1, \dots, \hat{p}_q]$. Suppose further that both $\mathcal{K}(\sigma, \hat{\mathbf{p}})$ and $\mathcal{K}(\mu, \hat{\mathbf{p}})$ are nonsingular and matrices $\mathbf{V}_r \in \mathbb{C}^{n \times r}$ and $\mathbf{W}_r \in \mathbb{C}^{n \times r}$ are given such that both $\mathcal{K}_r(\sigma, \hat{\mathbf{p}}) = \mathbf{W}_r^T \mathcal{K}(\sigma, \hat{\mathbf{p}}) \mathbf{V}_r$ and $\mathcal{K}_r(\mu, \hat{\mathbf{p}}) = \mathbf{W}_r^T \mathcal{K}(\mu, \hat{\mathbf{p}}) \mathbf{V}_r$ are also nonsingular. For nontrivial tangential directions $\mathbf{b} \in \mathbb{C}^m$ and $\mathbf{c} \in \mathbb{C}^p$:

- (a) If $\mathcal{K}(\sigma, \hat{\mathbf{p}})^{-1} \mathcal{B}(\sigma, \hat{\mathbf{p}}) \mathbf{b} \in \text{Ran}(\mathbf{V}_r)$ then $\mathbf{H}(\sigma, \hat{\mathbf{p}}) \mathbf{b} = \mathbf{H}_r(\sigma, \hat{\mathbf{p}}) \mathbf{b}$
- (b) If $(\mathcal{C}^T(\mu, \hat{\mathbf{p}}) \mathcal{K}(\mu, \hat{\mathbf{p}})^{-1})^T \in \text{Ran}(\mathbf{W}_r)$ then $\mathbf{c}^T \mathbf{H}(\mu, \hat{\mathbf{p}}) = \mathbf{c}^T \mathbf{H}_r(\mu, \hat{\mathbf{p}})$
- (c) If both (a) and (b) hold and if $\sigma = \mu$, then

$$\nabla_{\mathbf{p}} \mathbf{c}^T \mathbf{H}(\sigma, \hat{\mathbf{p}}) \mathbf{b} = \nabla_{\mathbf{p}} \mathbf{c}^T \mathbf{H}_r(\sigma, \hat{\mathbf{p}}) \mathbf{b} \quad \text{and} \quad \mathbf{c}^T \mathbf{H}'(\sigma, \hat{\mathbf{p}}) \mathbf{b} = \mathbf{c}^T \mathbf{H}'_r(\sigma, \hat{\mathbf{p}}) \mathbf{b}$$

Proof. We prove the first two assertions only. The proof of the third assertion is rather technical. We refer the reader to [14] for the proof of this fact for the special case of $\mathcal{K}(s, \mathbf{p}) = s\mathbf{E}(\mathbf{p}) - \mathbf{A}(\mathbf{p})$.

Define the projections

$$\mathbf{P}_r(s, \mathbf{p}) = \mathbf{V}_r \mathcal{K}_r(s, \mathbf{p})^{-1} \mathbf{W}_r^T \mathcal{K}(s, \mathbf{p}) \quad \text{and}$$

$$\mathbf{Q}_r(s, \mathbf{p}) = \mathcal{K}(s, \mathbf{p}) \mathbf{V}_r \mathcal{K}_r(s, \mathbf{p})^{-1} \mathbf{W}_r^T$$

Also, define the vector $\mathbf{f}_0 = \mathcal{K}(\sigma, \hat{\mathbf{p}})^{-1} \mathcal{B}(\sigma, \hat{\mathbf{p}}) \mathbf{b}$. Note that the assumption of the first assertion implies that $\mathbf{f}_0 \in \text{Ran}(\mathbf{P}_r(\sigma, \hat{\mathbf{p}}))$. Then, a direct computation yields

$$\mathbf{H}(\sigma, \hat{\mathbf{p}}) \mathbf{b} - \mathbf{H}_r(\sigma, \hat{\mathbf{p}}) \mathbf{b} = \mathcal{C}(\sigma, \hat{\mathbf{p}}) (\mathbf{I} - \mathcal{P}_r) \mathbf{f}_0 = 0,$$

which proves the first assertion. Similarly, define $\mathbf{g}_0^T = \mathbf{c}^T \mathcal{C}(\sigma, \hat{\mathbf{p}}) \mathcal{K}(\sigma, \hat{\mathbf{p}})^{-1}$ and observe that $\mathbf{g}_0 \perp \text{Ker}(\mathbf{Q}_r(\sigma, \hat{\mathbf{p}}))$ due to the assumption of the second assumption. Then, one can directly obtain

$$\mathbf{c}^T \mathbf{H}(\sigma, \hat{\mathbf{p}}) - \mathbf{c}^T \mathbf{H}_r(\sigma, \hat{\mathbf{p}}) = \mathbf{g}_0^T (\mathbf{I} - \mathbf{Q}_r) \mathcal{B}(\sigma, \hat{\mathbf{p}}) = 0,$$

which proves the second assertion. \square

Theorem 1.8 extends the interpolation theory for the non-parametric systems to the parameterized ones. It can be easily extended to match higher order derivatives as well. This result is much more general than others in the literature in the sense that we allow transfer function to be in the generalized form; hence not constraining it to be in standard state-space form. Also, the result allows for arbitrary linear and nonlinear dependency in \mathcal{K} , \mathcal{C} , and \mathcal{B} , which themselves can reflect the greater generality of the structured dynamical systems setting considered earlier.

Given $\mathbf{H}(s, p) = \mathcal{C}(s, p)\mathcal{K}(s, p)^{-1}\mathcal{B}(s, p)$ as defined in (43), let us assume that we want to construct a parametric reduced-model that matches $\mathbf{H}(s)$ at frequency points $\{\sigma_i\}_{i=1}^K \in \mathbb{C}$ and $\{\mu_i\}_{i=1}^K \in \mathbb{C}$, the parameter points $\{p^{(j)}\}_{j=1}^L \in \mathbb{C}^q$ along the right tangential directions $\{b_{ij}\}_{i=1, j=1}^{K, L} \in \mathbb{C}^m$ and the right tangential directions $\{c_{ij}\}_{i=1, j=1}^{K, L} \in \mathbb{C}^p$. Define, for $i = 1, \dots, K$ and $j = 1, \dots, L$,

$$\mathbf{v}_{ij} = \mathcal{K}(\sigma_i, p^{(j)})^{-1}\mathcal{B}(\sigma_i, p^{(j)})\mathbf{b}_{i,j} \text{ and } \mathbf{w}_{ij} = \mathcal{K}(\mu_i, p^{(j)})^{-T}\mathcal{C}(\mu_i, p^{(j)})^T\mathbf{c}_{i,j}$$

Then, construct \mathbf{V}_r and \mathbf{W}_r such that

$$\mathbf{V}_r = [\mathbf{v}_{11}, \dots, \mathbf{v}_{1L}, \mathbf{v}_{21}, \dots, \mathbf{v}_{2L}, \dots, \mathbf{v}_{K1}, \dots, \mathbf{v}_{KL}] \in \mathbb{C}^{n \times (KL)}$$

and

$$\mathbf{W}_r = [\mathbf{w}_{11}, \dots, \mathbf{w}_{1L}, \mathbf{w}_{21}, \dots, \mathbf{w}_{2L}, \dots, \mathbf{w}_{K1}, \dots, \mathbf{w}_{KL}] \in \mathbb{C}^{n \times (KL)}.$$

Then, the resulting projection-based reduced-order parametric model of (45) satisfies the interpolation condition stated in Theorem 1.8.

As the discussion above illustrates, *once the parameter points $\{p^{(j)}\}$ are selected*, the machinery for interpolatory model reduction of parametric systems is very similar to that of non-parametric systems. However, in this case one faces the problem choosing not only the frequency points $\{\sigma_i\}$ but also the parameter points $\{p^{(j)}\}$. Hence, the question of effective parameter point selection arises. In some application, the designer specifies important parameter sets to be used in the model reduction. However, the task becomes much harder if one only has the information as the range of parameter space without any knowledge of what parameter sets might be important. In this case, Bui-Thanh et al. [23] proposed the so-called greedy selection algorithm as a possible solution. Even though the method [23] proves to yield high quality approximations, the optimization algorithm that needs to be solved at each step could be computationally expensive. Another strategy for an efficient choice of parameters points in a higher dimensional parameter sparse is the use of sparse grid points [24, 41, 92]. An *optimal* parameter selection strategy was recently proposed by Baur et al. [14] for the special case of the parametric single-input/single-output dynamical systems of the form $\mathbf{H}(s) = (\mathbf{c}_0 + p_1\mathbf{c}_1)^T(s\mathbf{I} - \mathbf{A})^{-1}(\mathbf{b}_0 + p_2\mathbf{b}_1)$ where the scalars p_1 and p_2 are independent parameters and the system dynamic \mathbf{A} is non-parametric. The optimal parameter set is obtained by converting the problem into an equivalent non-parametric optimal \mathcal{H}_2 MIMO approximation.

6.1 Numerical Example

We illustrate the concept of parametric model reduction using a model representing thermal conduction in a semiconductor chip as described in [61]. An efficient model of thermal conduction should allow flexibility in specifying boundary conditions so that designers are able to evaluate the effects of environmental changes on the temperature distribution in the chip. The problem is modeled as homogenous heat diffusion with heat exchange occurring at three device interfaces modeled with convection boundary conditions that introduce film coefficients, p_1 , p_2 , and p_3 , describing the heat exchange on the three device interfaces. Discretization of the underlying partial differential equations leads to a system of ordinary differential equations

$$\mathbf{E}\dot{\mathbf{x}}(t) = (\mathbf{A} + \sum_{i=1}^3 p_i \mathbf{A}_i)\mathbf{x}(t) + \mathbf{B}u(t), \quad \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t),$$

where $\mathbf{E} \in \mathbb{R}^{4257 \times 4257}$ and $\mathbf{A} \in \mathbb{R}^{4257 \times 4257}$ are system matrices, $\mathbf{A}_i \in \mathbb{R}^{4257 \times 4257}$, $i = 1, \dots, 3$, are diagonal matrices arising from the discretization of the convection boundary condition on the i th interface, and $\mathbf{B} \in \mathbb{R}^{4257}$ and $\mathbf{C} \in \mathbb{R}^{7 \times 4257}$. We note even though Theorem 1.8 allows parametric model reduction in a much more general setting, we use a standard state-space model to illustrate the theory.

Each parameter value varies in the range of $[1, 10^4]$. Important parameter set values are listed in [61] of which we use the following two to apply model reduction:

$$p^{(1)} = (10^4, 10^4, 1) \text{ and } p^{(2)} = (1, 1, 1)$$

We follow an \mathcal{H}_2 -inspired approach recently proposed in [14]: Note that Once the parameter set $p^{(i)}$ for $i = 1, 2$ are plugged into the state-space matrices, a non-parametric standard transfer function is obtained, denoted by $\mathbf{H}(s, p^{(i)})$. Then, we apply the optimal \mathcal{H}_2 model reduction technique **IRKA** as outlined in Algorithm 1 to $\mathbf{H}(s, p^{(i)})$ to reduce the order to r_i leading to projection subspaces $\mathbf{V}^{(i)} \in \mathbb{R}^{4257 \times r_i}$ and $\mathbf{W}^{(i)} \in \mathbb{R}^{4257 \times r_i}$ for $i = 1, 2$ with $r_1 = 6$ and $r_2 = 5$. We, then, concatenate these matrices to build the final projection matrices

$$\mathbf{V}_r = [\mathbf{V}^{(1)}, \mathbf{V}^{(2)}] \in \mathbb{R}^{4257 \times 11} \text{ and } \mathbf{W}_r = [\mathbf{W}^{(1)}, \mathbf{W}^{(2)}] \in \mathbb{R}^{4257 \times 11}.$$

We obtain a final parameterized reduced-order model of $\text{orderr} = r_1 + r_2 = 11$:

$$\begin{aligned} \mathbf{W}_r^T \mathbf{E} \mathbf{V}_r \mathbf{x}_r(t) &= (\mathbf{W}_r^T \mathbf{A} \mathbf{V}_r + \sum_{i=1}^3 p_i \mathbf{W}_r^T \mathbf{A}_i \mathbf{V}_r) \mathbf{x}_r(t) + \mathbf{W}_r^T \mathbf{B} u(t), \\ \mathbf{y}_r(t) &= \mathbf{C} \mathbf{V}_r \mathbf{x}_r(t). \end{aligned}$$

A high-quality parameterized reduced-order model must represent the full parameterized model with high fidelity for a wide range of parameter values. To illustrate this is the case, we fix $p_3 = 1$ and vary both p_1 and p_2 over the full parameter range, i.e., between 1 and 10^4 . Then, for each mesh point (i.e. for each triple of parameter values in this range), we compute the corresponding full-order model

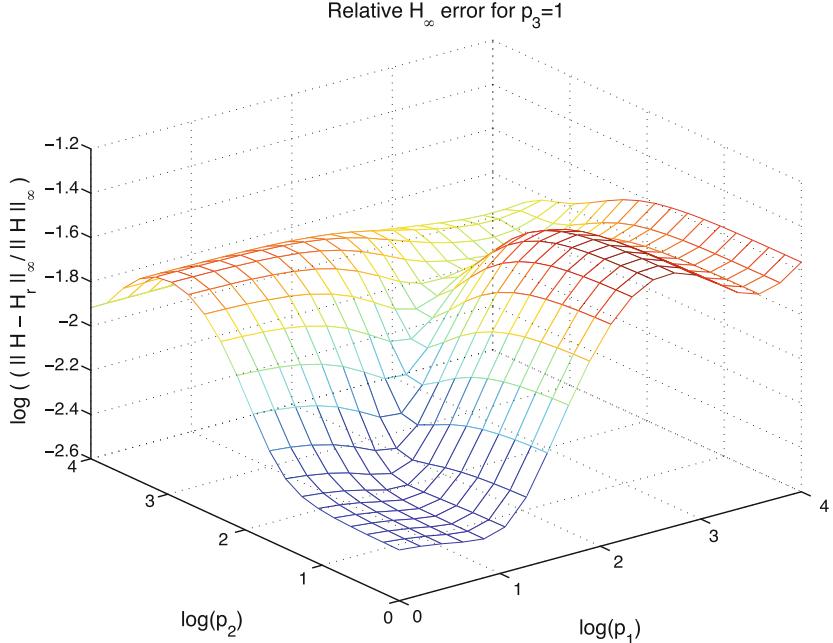


Fig. 4 Relative \mathcal{H}_∞ -error as p_1 and p_2 vary

and the reduced-order model; and compute the corresponding relative \mathcal{H}_∞ -errors. Figure 4 depicts the resulting mesh plot. As the figure illustrates, the reduced-model approximate the full-model uniformly well throughout the full parameter range. The maximum relative \mathcal{H}_∞ -error is 3.50×10^{-2} . Therefore, the parameterized reduced-order model $\mathbf{H}_r(s, p)$ of order $r = 11$ yields a relative accuracy of 10^{-2} for the complete range of p_1 and p_2 .

7 Model Reduction from Measurements

In many instances, input/output measurements replace an explicit model of a to-be-simulated system. In such cases it is of great interest to be able to efficiently construct models and reduced models from the available data. In this section we will address this problem by means of *rational interpolation*. We will show that the natural tool is the *Loewner matrix pencil*, which is closely related to the *Hankel matrix*. For details we refer to [62, 66].

7.1 Motivation: S-Parameters

The growth in communications and networking and the demand for high data bandwidth requires streamlining of the simulation of entire complex systems from chips

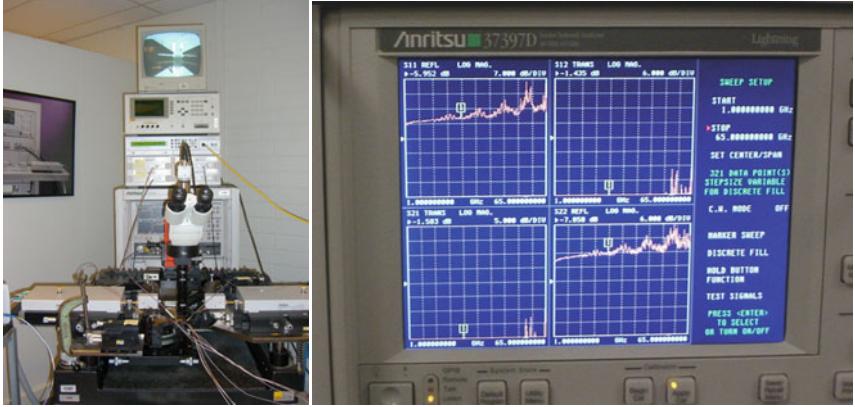


Fig. 5 VNA (Vector Network Analyzer) and VNA screen showing the magnitude of the S -parameters for a two port

to packages to boards. Moreover, in circuit simulation, signal integrity (lack of signal distortion) of high speed electronic components requires that interconnect models be valid over a wide bandwidth.

An important tool for dealing with such problems is the S - or *scattering-parameter* system representation. The S -parameters represent a system as a black box. An important advantage is that these parameters can be measured using VNAs (Vector Network Analyzers) (see Fig. 5). Also, their norm for passive components, is not greater than one, and in high frequencies S -parameters are important because wave phenomena become dominant.

We recall that given a system in input/output representation: $\hat{\mathbf{y}}(s) = \mathbf{H}(s)\hat{\mathbf{u}}(s)$, the associated *S-parameter representation* is

$$\hat{\mathbf{y}}_s = \underbrace{[\mathbf{H}(s) + \mathbf{I}][\mathbf{H}(s) - \mathbf{I}]^{-1}}_{\mathbf{S}(s)} \bar{\mathbf{u}}_s(s),$$

where $\hat{\mathbf{y}}_s = \frac{1}{2}(\hat{\mathbf{y}} + \bar{\mathbf{u}})$ are the *transmitted waves* and, $\bar{\mathbf{u}}_s = \frac{1}{2}(\hat{\mathbf{y}} - \bar{\mathbf{u}})$ are the *reflected waves*. Thus the *S-parameter measurements* $\mathbf{S}(j\omega_k)$, are samples of the frequency response of the S -parameter system representation.

In the sections that follow we will describe a framework which applies also to the efficient modeling from S -parameter measurements. An example is provided in Sect. 7.4.3. For more details and examples see [62].

7.2 The Loewner Matrix Pair and Construction of Interpolants

Suppose we have observed response data as described in Problem 2 in Sect. 2.1. We are given r (right) driving frequencies: $\{\sigma_i\}_{i=1}^r \subset \mathbb{C}$ that use input directions $\{\tilde{\mathbf{b}}_i\}_{i=1}^r \subset \mathbb{C}^m$, to produce system responses, $\{\tilde{\mathbf{y}}_i\}_{i=1}^r \subset \mathbb{C}^p$ and q (left) driving fre-

quencies: $\{\mu_i\}_{i=1}^q \subset \mathbb{C}$ that use dual (left) input directions $\{\tilde{c}_i\}_{i=1}^q \subset \mathbb{C}^p$, to produce dual (left) responses, $\{\tilde{z}_i\}_{i=1}^q \subset \mathbb{C}^m$. We assume that there is an underlying dynamical system as defined in (1) so that

$$\begin{aligned} \tilde{z}_i^T \mathbf{H}(\mu_i) &= \tilde{z}_i^T & \text{and} & & \mathbf{H}(\sigma_j) \tilde{b}_j &= \tilde{y}_j, \\ \text{for } i = 1, \dots, q, & & & & \text{for } j = 1, \dots, r. & \end{aligned}$$

yet we are given access only to the $r + q$ response observations listed above and have no other information about the underlying system $\mathbf{H}(s)$. In this section we will sketch the solution of Problem 2 described in Sect. 2.1. Towards this goal we will introduce the Loewner matrix pair in the tangential interpolation case. The *Loewner matrix* is defined as follows:

$$\mathbb{L} = \begin{bmatrix} \frac{\tilde{z}_1^T \tilde{b}_1 - \tilde{c}_1^T \tilde{y}_1}{\mu_1 - \sigma_1} & \dots & \frac{\tilde{z}_1^T \tilde{b}_r - \tilde{c}_1^T \tilde{y}_r}{\mu_1 - \sigma_r} \\ \vdots & \ddots & \vdots \\ \frac{\tilde{z}_q^T \tilde{b}_1 - \tilde{c}_q^T \tilde{y}_1}{\mu_q - \sigma_1} & \dots & \frac{\tilde{z}_q^T \tilde{b}_r - \tilde{c}_q^T \tilde{y}_r}{\mu_q - \sigma_r} \end{bmatrix} \in \mathbb{C}^{q \times r}.$$

If we define matrices associated with the system observations as:

$$\begin{aligned} \tilde{\mathbf{B}} &= \begin{bmatrix} \vdots & \vdots & \vdots \\ \tilde{b}_1 & \tilde{b}_2 & \dots & \tilde{b}_r \\ \vdots & \vdots & & \vdots \end{bmatrix} & \tilde{\mathbf{Y}} &= \begin{bmatrix} \vdots & \vdots & \vdots \\ \tilde{y}_1 & \tilde{y}_2 & \dots & \tilde{y}_r \\ \vdots & \vdots & & \vdots \end{bmatrix} \\ \tilde{\mathbf{Z}}^T &= \begin{bmatrix} \dots \tilde{z}_1^T \dots \\ \dots \tilde{z}_2^T \dots \\ \vdots \\ \dots \tilde{z}_q^T \dots \end{bmatrix} & \tilde{\mathbf{C}}^T &= \begin{bmatrix} \dots \tilde{c}_1^T \dots \\ \dots \tilde{c}_2^T \dots \\ \vdots \\ \dots \tilde{c}_q^T \dots \end{bmatrix} \end{aligned}$$

\mathbb{L} satisfies the Sylvester equation

$$\mathbb{L}\Sigma - M\mathbb{L} = \tilde{\mathbf{C}}^T \tilde{\mathbf{Y}} - \tilde{\mathbf{Z}}^T \tilde{\mathbf{B}}. \quad (46)$$

where $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r) \in \mathbb{C}^{r \times r}$ and $M = \text{diag}(\mu_1, \mu_2, \dots, \mu_q) \in \mathbb{C}^{q \times q}$. Suppose that state space data $(\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$, of minimal degree n are given such that $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1} \mathbf{B} + \mathbf{D}$. If the generalized eigenvalues of (\mathbf{A}, \mathbf{E}) are distinct from σ_i and μ_j , we define \mathbf{V}_r and \mathbf{W}_q^T as

$$\mathbf{V}_r = [(\sigma_1 \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \tilde{b}_1, \dots, (\sigma_r \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \tilde{b}_r] \in \mathbb{C}^{n \times r} \text{ and}$$

$$\mathbf{W}_q^T = \begin{bmatrix} \mathbf{c}_1^T \mathbf{C}(\mu_1 \mathbf{E} - \mathbf{A})^{-1} \\ \mathbf{c}_2^T \mathbf{C}(\mu_2 \mathbf{E} - \mathbf{A})^{-1} \\ \vdots \\ \mathbf{c}_q^T \mathbf{C}(\mu_q \mathbf{E} - \mathbf{A})^{-1} \end{bmatrix} \in \mathbb{C}^{q \times n}.$$

It follows that

$$\mathbb{L} = -\mathbf{W}_q^T \mathbf{E} \mathbf{V}_r$$

and we call \mathbf{V}_r and \mathbf{W}_q^T *generalized tangential* reachability and observability matrices, respectively.

Next we introduce a new object which is pivotal in our approach. This is the *shifted Loewner matrix*, defined as follows:

$$\mathbb{M} = \begin{bmatrix} \frac{\mu_1 \tilde{\mathbf{z}}_1^T \tilde{\mathbf{b}}_1 - \sigma_1 \tilde{\mathbf{c}}_1^T \tilde{\mathbf{y}}_1}{\mu_1 - \sigma_1} & \dots & \frac{\mu_1 \tilde{\mathbf{z}}_1^T \tilde{\mathbf{b}}_r - \sigma_r \tilde{\mathbf{c}}_1^T \tilde{\mathbf{y}}_r}{\mu_1 - \sigma_r} \\ \vdots & \ddots & \vdots \\ \frac{\mu_q \tilde{\mathbf{z}}_q^T \tilde{\mathbf{b}}_1 - \sigma_1 \tilde{\mathbf{c}}_q^T \tilde{\mathbf{y}}_1}{\mu_q - \sigma_1} & \dots & \frac{\mu_q \tilde{\mathbf{z}}_q^T \tilde{\mathbf{b}}_r - \sigma_r \tilde{\mathbf{c}}_q^T \tilde{\mathbf{y}}_r}{\mu_q - \sigma_r} \end{bmatrix} \in \mathbb{C}^{q \times r}$$

\mathbb{M} satisfies the Sylvester equation

$$\mathbb{M}\Sigma - M\mathbb{M} = \tilde{\mathbf{C}}^T \tilde{\mathbf{Y}}\Sigma - M\tilde{\mathbf{Z}}^T \tilde{\mathbf{B}}. \quad (47)$$

If an interpolant $\mathbf{H}(s)$ is associated with the interpolation data, the shifted Loewner matrix is the Loewner matrix associated to $s\mathbf{H}(s)$. If a state space representation is available, then like for the Loewner matrix, the shifted Loewner matrix can be factored as

$$\mathbb{M} = -\mathbf{W}_q^T \mathbf{A} \mathbf{V}_r.$$

It therefore becomes apparent that \mathbb{L} contains information about \mathbf{E} while \mathbb{M} contains information about \mathbf{A} . These observations are formalized in one of the main results of this section which shows how straightforward the solution of the interpolation problem becomes, in the Loewner matrix framework.

Theorem 1.9. *Assume that $r = q$ and that $\mu_i \neq \sigma_j$ for all $i, j = 1, \dots, r$. Suppose that $\mathbb{M} - s\mathbb{L}$ is invertible for all $s \in \{\sigma_i\} \cup \{\mu_j\}$. Then, with*

$$\mathbf{E}_r = -\mathbb{L}, \quad \mathbf{A}_r = -\mathbb{M}, \quad \mathbf{B}_r = \tilde{\mathbf{Z}}^T, \quad \mathbf{C}_r = \tilde{\mathbf{Y}}, \quad \mathbf{D}_r = 0,$$

$$\mathbf{H}_r(s) = \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{B}_r = \tilde{\mathbf{Z}}^T (\mathbb{M} - s\mathbb{L})^{-1} \tilde{\mathbf{Y}}$$

interpolates the data and furthermore is a minimal realization.

Next we will outline two proofs of this important result. They are both straightforward and hence reveal the main attributes of this approach.

Proof. Multiplying (46) by s and subtracting it from (47) we get

$$(\mathbb{M} - s\mathbb{L})\Sigma - M(\mathbb{M} - s\mathbb{L}) = \tilde{\mathbf{C}}^T \tilde{\mathbf{Y}}(\Sigma - s\mathbf{I}) - (M - s\mathbf{I})\tilde{\mathbf{Z}}^T \tilde{\mathbf{B}}. \quad (48)$$

Multiplying (48) by \mathbf{e}_j (the j th unit vector) on the right and setting $s = \sigma_j$, we obtain

$$\begin{aligned} (\sigma_j \mathbf{I} - M)(\mathbb{M} - \sigma_j \mathbb{L})\mathbf{e}_j &= (\sigma_j \mathbf{I} - M)\tilde{\mathbf{Z}}^T \tilde{\mathbf{b}}_j \Rightarrow \\ (\mathbb{M} - \sigma_j \mathbb{L})\mathbf{e}_j &= \tilde{\mathbf{Z}}^T \tilde{\mathbf{b}}_j \Rightarrow \tilde{\mathbf{Y}}\mathbf{e}_j = \tilde{\mathbf{Y}}(\mathbb{M} - \sigma_j \mathbb{L})^{-1} \tilde{\mathbf{Z}}^T \tilde{\mathbf{b}}_j \end{aligned}$$

Therefore $\mathbf{H}(\sigma_j)\tilde{\mathbf{b}}_j = \tilde{\mathbf{y}}_j$. This proves the right tangential interpolation property. To prove the left tangential interpolation property, we multiply (48) by \mathbf{e}_i^T (the transpose of the i th unit vector) on the left and set $s = \mu_i$:

$$\begin{aligned}\mathbf{e}_i^T(\mathbb{M} - \mu_i \mathbb{L})(\Sigma - \mu_i \mathbf{I}) &= \tilde{\mathbf{c}}_i^T \tilde{\mathbf{Y}} (\Sigma - \mu_i \mathbf{I}) \Rightarrow \\ \mathbf{e}_i^T(\mathbb{M} - \mu_i \mathbb{L}) &= \tilde{\mathbf{c}}_i^T \tilde{\mathbf{Y}} \Rightarrow \mathbf{e}_i^T \tilde{\mathbf{Z}}^T = \tilde{\mathbf{c}}_i^T \tilde{\mathbf{Y}} (\mathbb{M} - \mu_i \mathbb{L})^{-1} \tilde{\mathbf{Z}}^T.\end{aligned}$$

Therefore $\tilde{\mathbf{c}}_i^T \mathbf{H}(\mu_i) = \tilde{\mathbf{z}}_i^T$, which completes the proof. \square

Proof (Alternate version). For this proof we assume that the data have been obtained by sampling a known high order system in state space form $(\mathbf{A}, \mathbf{E}, \mathbf{B}, \mathbf{C})$. Direct calculation verifies that

$$\mathbf{H}(\mu_i) - \mathbf{H}(\sigma_j) = (\sigma_j - \mu_i) \mathbf{C}(\mu_i \mathbf{E} - \mathbf{A})^{-1} \mathbf{E}(\sigma_j \mathbf{E} - \mathbf{A})^{-1} \mathbf{B}.$$

Pre-multiplication of this expression by $\tilde{\mathbf{c}}_i^T$ and post-multiplication by $\tilde{\mathbf{b}}_j$ immediately yields

$$\mathbb{L} = -\mathbf{W}_q^T \mathbf{E} \mathbf{V}_r.$$

One may directly calculate first,

$$\begin{aligned}\sigma_j(\sigma_j \mathbf{E} - \mathbf{A})^{-1} &= \sigma_j(\mu_i \mathbf{E} - \mathbf{A})^{-1}(\mu_i \mathbf{E} - \mathbf{A})(\sigma_j \mathbf{E} - \mathbf{A})^{-1} \\ &= \mu_i \sigma_j(\mu_i \mathbf{E} - \mathbf{A})^{-1} \mathbf{E}(\mu_i \mathbf{E} - \mathbf{A})^{-1} - \sigma_j(\mu_i \mathbf{E} - \mathbf{A})^{-1} \mathbf{A}(\sigma_j \mathbf{E} - \mathbf{A})^{-1}\end{aligned}$$

and

$$\begin{aligned}\mu_i(\mu_i \mathbf{E} - \mathbf{A})^{-1} &= \mu_i(\mu_i \mathbf{E} - \mathbf{A})^{-1}(\sigma_j \mathbf{E} - \mathbf{A})(\sigma_j \mathbf{E} - \mathbf{A})^{-1} \\ &= \mu_i \sigma_j(\mu_i \mathbf{E} - \mathbf{A})^{-1} \mathbf{E}(\mu_i \mathbf{E} - \mathbf{A})^{-1} - \mu_i(\mu_i \mathbf{E} - \mathbf{A})^{-1} \mathbf{A}(\sigma_j \mathbf{E} - \mathbf{A})^{-1}\end{aligned}$$

$$\sigma_j \mathbf{H}(\sigma_j) = \mu_i \sigma_j \mathbf{C}(\mu_i \mathbf{E} - \mathbf{A})^{-1} \mathbf{E}(\sigma_j \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} - \sigma_j \mathbf{C}(\mu_i \mathbf{E} - \mathbf{A})^{-1} \mathbf{A}(\sigma_j \mathbf{E} - \mathbf{A})^{-1} \mathbf{B}$$

$$\mu_i \mathbf{H}(\mu_i) = \mu_i \sigma_j \mathbf{C}(\mu_i \mathbf{E} - \mathbf{A})^{-1} \mathbf{E}(\sigma_j \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} - \mu_i \mathbf{C}(\mu_i \mathbf{E} - \mathbf{A})^{-1} \mathbf{A}(\sigma_j \mathbf{E} - \mathbf{A})^{-1} \mathbf{B}$$

which in turn implies that

$$\mu_i \mathbf{H}(\mu_i) - \sigma_j \mathbf{H}(\sigma_j) = (\sigma_j - \mu_i) \mathbf{C}(\mu_i \mathbf{E} - \mathbf{A})^{-1} \mathbf{A}(\sigma_j \mathbf{E} - \mathbf{A})^{-1} \mathbf{B}.$$

Pre-multiplication of this expression by $\tilde{\mathbf{c}}_i^T$ and post-multiplication by $\tilde{\mathbf{b}}_j$ immediately yields

$$\mathbb{M} = -\mathbf{W}_q^T \mathbf{A} \mathbf{V}_r.$$

Finally, note that $\tilde{\mathbf{Z}}^T = \mathbf{W}_q^T \mathbf{B}$ and $\tilde{\mathbf{Y}} = \mathbf{C} \mathbf{V}_r$. Thus $\mathbf{H}_r(s) = \tilde{\mathbf{Y}}(\mathbb{M} - s \mathbb{L})^{-1} \tilde{\mathbf{Z}}^T$ is a tangential interpolant to $\mathbf{H}(s)$. \square

Remark 1.4. The Loewner matrix was introduced in [5]. As shown therein, its usefulness derives from the fact that its rank is equal to the McMillan degree of the

underlying interpolant. It also turns out that Loewner matrices for confluent interpolation problems (i.e., problems where the values of a certain number of derivatives are provided as well) have Hankel structure. Thus the Loewner matrix provides a direct link with classical realization theory. For an overview of these results we refer to Sects. 4.4 and 4.5 in [2]. In [1], one way of constructing state space realizations based on the Loewner matrix was presented. But the shifted Loewner matrix, first introduced in [66], was missing and consequently the resulting procedure is only applicable to proper rational interpolants.

7.2.1 The General Case

If the assumption of the above theorem is not satisfied, one needs to project onto the column span and onto the row span of a linear combination of the two Loewner matrices. More precisely, let the following assumption be satisfied:

$$\text{rank}(s\mathbb{L} - \mathbb{M}) = \text{rank}(\mathbb{L} \ \mathbb{M}) = \text{rank}\begin{pmatrix} \mathbb{L} \\ \mathbb{M} \end{pmatrix} \geq \rho, \text{ for all } s \in \{\sigma_i\} \cup \{\mu_j\}.$$

ρ is a truncation index that is assumed to be no larger than $\text{rank } s\mathbb{L} - \mathbb{M}$. $\rho \leq q, r$ and choosing ρ to be the *numerical* rank of $s\mathbb{L} - \mathbb{M}$ is convenient. The best tool for determining the numerical rank of $s\mathbb{L} - \mathbb{M}$, is the SVD (Singular Value Decomposition). To that end, suppose

$$s\mathbb{L} - \mathbb{M} = \mathbf{Y}\Theta\mathbf{X}^*,$$

is the SVD of $s\mathbb{L} - \mathbb{M}$ for some choice of $s \in \{\sigma_i\} \cup \{\mu_j\}$ and consider a truncated SVD as $\mathbf{Y}_\rho \in \mathbb{C}^{q \times \rho}$, $\mathbf{X}_\rho \in \mathbb{C}^{r \times \rho}$.

Theorem 1.10. *A realization $[\mathbf{E}_\rho, \mathbf{A}_\rho, \mathbf{B}_\rho, \mathbf{C}_\rho]$, of a minimal solution is given as follows:*

$$\mathbf{E}_\rho = -\mathbf{Y}_\rho^* \mathbb{L} \mathbf{X}_\rho, \quad \mathbf{A}_\rho = -\mathbf{Y}_\rho^* \mathbb{M} \mathbf{X}_\rho, \quad \mathbf{B}_\rho = \mathbf{Y}_\rho^* \tilde{\mathbf{Y}}, \quad \mathbf{C}_\rho = \tilde{\mathbf{Z}}^T \mathbf{X}_\rho.$$

Depending on whether ρ is the exact or approximate rank, we obtain either an interpolant or an approximate interpolant of the data, respectively.

7.3 Loewner and Pick Matrices

The positive real interpolation problem can be formulated as follows. Given triples $(\sigma_i, \tilde{b}_i, \tilde{y}_i)$, $i = 1, \dots, q$, where σ_i are distinct complex numbers in the right-half of the complex plane, \tilde{b}_i and \tilde{y}_i are in \mathbb{C}^r , we seek a rational matrix function $\mathbf{H}(s)$ of size $r \times r$, such that $\mathbf{H}(\sigma_i)\tilde{b}_i = \tilde{y}_i$, $i = 1, \dots, q$, and in addition \mathbf{H} is positive real. This problem does not always have a solution. It is well known that the necessary and sufficient condition for its solution is that the associated *Pick* matrix

$$\Pi = \begin{bmatrix} \frac{\tilde{y}_1^* \tilde{b}_1 + \tilde{b}_1^* \tilde{y}_1}{\bar{\sigma}_1 + \sigma_1} & \dots & \frac{\tilde{y}_1^* \tilde{b}_q + \tilde{b}_1^* \tilde{y}_q}{\bar{\sigma}_1 + \sigma_q} \\ \vdots & \ddots & \vdots \\ \frac{\tilde{y}_q^* \tilde{b}_1 + \tilde{b}_q^* \tilde{y}_1}{\bar{\sigma}_q + \sigma_1} & \dots & \frac{\tilde{y}_q^* \tilde{b}_q + \tilde{b}_q^* \tilde{y}_q}{\bar{\sigma}_q + \sigma_q} \end{bmatrix} \in \mathbb{C}^{q \times q},$$

be positive semi-definite, that is $\Pi = \Pi^* \geq \mathbf{0}$. By comparing Π with the Loewner matrix \mathbb{L} defined in Sect. 7.2, we conclude that if the right (column) array for the former is taken as $(\sigma_i, \tilde{b}_i, \tilde{y}_i)$, $i = 1, \dots, q$, and the left (row) array as $(-\bar{\sigma}_i, \tilde{b}_i^*, -\tilde{y}_i^*)$, $i = 1, \dots, q$, then

$$\Pi = \mathbb{L}.$$

The left array is then called the *mirror-image array*. Thus for this choice of interpolation data the Pick matrix is the same as the Loewner matrix. This shows the importance of the Loewner matrix as a tool for studying rational interpolation.

Remark 1.5. (a) The above considerations provide an algebraization of the positive real interpolation problem. If namely, $\Pi \geq \mathbf{0}$, the minimal-degree rational functions which interpolate *simultaneously* the original array *and* its mirror image array, are automatically *positive real* and hence *stable* as well. The data in the model reduction problem that we studied in Sect. 4, *automatically* satisfy this positive definiteness constraint, and therefore the reduced system is positive real.

(b) It readily follows that interpolants of the original and the mirror-image arrays constructed by means of the Loewner matrix, satisfy

$$[\mathbf{H}(\sigma_i) + \mathbf{H}^T(-\sigma_i)] \tilde{b}_i = \mathbf{0}.$$

In general the zeros σ_i of $\mathbf{H}(s) + \mathbf{H}^T(-s)$ are called *spectral zeros*, and \tilde{b}_i are the corresponding (right) zero directions. Thus the construction of positive real interpolants by means of the Loewner (Pick) matrix, forces these interpolants to have the *given* interpolation points as *spectral zeros*.

(c) The observation that the Pick matrix is a special case of the Loewner matrix, the algebraization discussed above, and passivity preservation by means of spectral zero interpolation, first appeared in [3]. See also [4].

7.4 Examples

7.4.1 A Simple Low-order Example

First we will illustrate the above results by means of a simple example. Consider a 2×2 rational function with minimal realization:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Thus the transfer function is

$$\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} = \begin{bmatrix} s & 1 \\ 1 & \frac{1}{s} \end{bmatrix}.$$

Since $\text{rank } \mathbf{E} = 2$, the McMillan degree of \mathbf{H} is 2. Our goal is to recover this function through interpolation. The data will be chosen in two different ways.

First, we will choose *matrix data*, that is the values of the whole matrix are available at each interpolation point:

$$\sigma_1 = 1, \quad \sigma_2 = 1, \quad \sigma_3 = 2, \quad \sigma_4 = 2,$$

$$\tilde{\mathbf{b}}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \tilde{\mathbf{b}}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \tilde{\mathbf{b}}_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \tilde{\mathbf{b}}_4 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\tilde{\mathbf{y}}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \tilde{\mathbf{y}}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \tilde{\mathbf{y}}_3 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \tilde{\mathbf{y}}_4 = \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}$$

$$\mu_1 = 1, \quad \mu_2 = 1, \quad \mu_3 = 2, \quad \mu_4 = 2,$$

$$\tilde{\mathbf{c}}_1^T = (1, 0), \quad \tilde{\mathbf{c}}_2^T = (0, 1), \quad \tilde{\mathbf{c}}_3^T = (1, 0), \quad \tilde{\mathbf{c}}_4^T = (0, 1)$$

$$\tilde{\mathbf{z}}_1^T = (-1, 1), \quad \tilde{\mathbf{z}}_2^T = (1, -1), \quad \tilde{\mathbf{z}}_3^T = (-2, 1), \quad \tilde{\mathbf{z}}_4^T = (1, -\frac{1}{2})$$

The associated (block) Loewner and shifted Loewner matrices turn out to be:

$$\mathbb{L} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & \frac{1}{2} \\ 1 & 0 & 1 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{4} \end{pmatrix}, \quad \mathbb{M} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

Notice that the rank of both Loewner matrices is 2 while the rank of $x_i \mathbb{L} - \mathbb{M}$ is 3, for all x equal to a σ_i or μ_i . It can be readily verified that the column span of $\sigma_1 \mathbb{L} - \mathbb{M} = \mathbb{L} - \mathbb{M}$ is the same as that of Π , where

$$\Pi = \begin{pmatrix} 1 & 1 & -2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Furthermore the row span of $\mathbb{L} - \mathbb{M}$ is the same as that of Π^* . Thus

$$\hat{\mathbf{A}} = -\Pi^* \mathbb{M} \Pi = \begin{pmatrix} -2 & -3 & 1 \\ -1 & 0 & -4 \\ 1 & 0 & 4 \end{pmatrix}, \quad \hat{\mathbf{E}} = -\Pi^* \mathbb{L} \Pi = \begin{pmatrix} -2 & -2 & \frac{3}{2} \\ -2 & -4 & 4 \\ \frac{3}{2} & 4 & -\frac{17}{4} \end{pmatrix},$$

$$\hat{\mathbf{B}} = \Pi^* \tilde{Z}^T = \begin{pmatrix} 0 & 0 \\ -3 & 2 \\ 3 & -\frac{5}{2} \end{pmatrix}, \quad \hat{\mathbf{C}} = \tilde{Y} \Pi = \begin{pmatrix} 2 & 3 & -1 \\ 2 & 2 & -\frac{3}{2} \end{pmatrix},$$

satisfy $\mathbf{H}(s) = \hat{\mathbf{C}}(s\hat{\mathbf{E}} - \hat{\mathbf{A}})\hat{\mathbf{B}}$, which shows that a (second) minimal realization of \mathbf{H} has been obtained.

The *second* experiment involves tangential data, that is, at each interpolation point only values along certain directions are available.

$$\begin{aligned} \sigma_1 &= 1, & \sigma_2 &= 2, & \sigma_3 &= 3, \\ \tilde{\mathbf{b}}_1 &= \begin{pmatrix} 1 \\ 0 \end{pmatrix}, & \tilde{\mathbf{b}}_2 &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}, & \tilde{\mathbf{b}}_3 &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ \tilde{\mathbf{y}}_1 &= \begin{pmatrix} 1 \\ 1 \end{pmatrix}, & \tilde{\mathbf{y}}_2 &= \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}, & \tilde{\mathbf{y}}_3 &= \begin{pmatrix} 4 \\ \frac{4}{3} \end{pmatrix} \\ \mu_1 &= -1, & \mu_2 &= -2, & \mu_3 &= -3, \\ \tilde{\mathbf{c}}_1^T &= (1, 0), & \tilde{\mathbf{c}}_2^T &= (0, 1), & \tilde{\mathbf{c}}_3^T &= (1, 1) \\ \tilde{\mathbf{z}}_1^T &= (-1, 1), & \tilde{\mathbf{z}}_2^T &= (1, -\frac{1}{2}), & \tilde{\mathbf{z}}_3^T &= (-2, \frac{2}{3}). \end{aligned}$$

Thus the associated Loewner and shifted Loewner matrices are:

$$\mathbb{L} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & \frac{1}{4} & \frac{1}{6} \\ 1 & \frac{1}{6} & \frac{10}{9} \end{bmatrix}, \quad \mathbb{M} = \begin{bmatrix} 0 & 1 & 3 \\ 1 & 0 & 1 \\ -1 & 1 & 2 \end{bmatrix}$$

It readily follows that the conditions of theorem 1.9 are satisfied and hence the quadruple $(-\mathbb{M}, -\mathbb{L}, \tilde{Z}^T, \tilde{Y})$, provides a (third) minimal realization of the original rational function: $\mathbf{H}(s) = -\tilde{Y}[s\mathbb{L} - \mathbb{M}]^{-1}\tilde{Z}^T$.

7.4.2 Coupled Mechanical System

Figure 6 depicts a constrained mechanical system described in [67]. There are g masses in total; the i th mass of weight m_i , is connected to the $(i+1)$ st mass by a spring and a damper with constants k_i and d_i , respectively, and also to the ground by a spring and a damper with constants κ_i and δ_i , respectively. Additionally, the first mass is connected to the last one by a rigid bar (holonomic constraint) and it is influenced by the control $\mathbf{u}(t)$. The vibration of this constrained system is described in generalized state space form as:

$$\mathbf{E}\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \quad \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t),$$

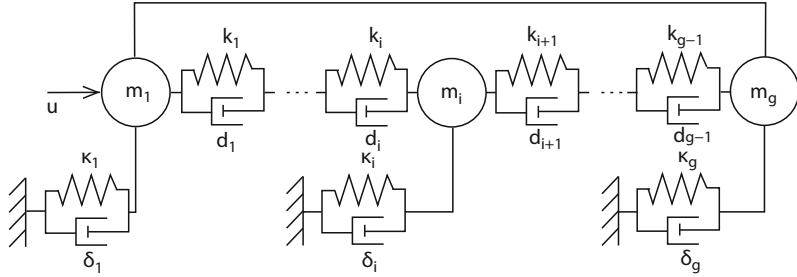


Fig. 6 Constrained mechanical system

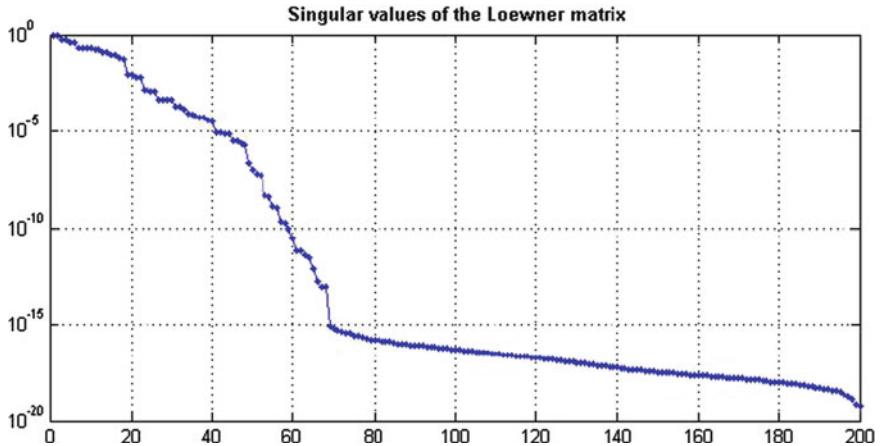


Fig. 7 The singular values of the Loewner matrix

where \mathbf{x} contains the positions and velocities of the masses,

$$\mathbf{E} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{K} & \mathbf{D} & -\mathbf{G}^* \\ \mathbf{G} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{0} \\ \mathbf{B}_2 \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3];$$

furthermore \mathbf{M} is the mass matrix ($g \times g$, diagonal, positive definite), \mathbf{K} is the stiffness matrix ($g \times g$, tri-diagonal), \mathbf{D} is the damping matrix ($g \times g$, tri-diagonal), $\mathbf{G} = [1, 0, \dots, 0, -1]$, is the $1 \times g$ constraint matrix.

In [67], balanced truncation methods for descriptor systems are used to reduce this system. Here we will reduce this system by means of the Loewner framework. Towards this goal, we compute 200 frequency response data, that is $\mathbf{H}(i\omega_i)$, where $\omega_i \in [-2, +2]$. Figure 7 shows the singular values of the Loewner matrix pair, which indicate that a system of order 20 will have an approximate error 10^{-3} (-60 dB).

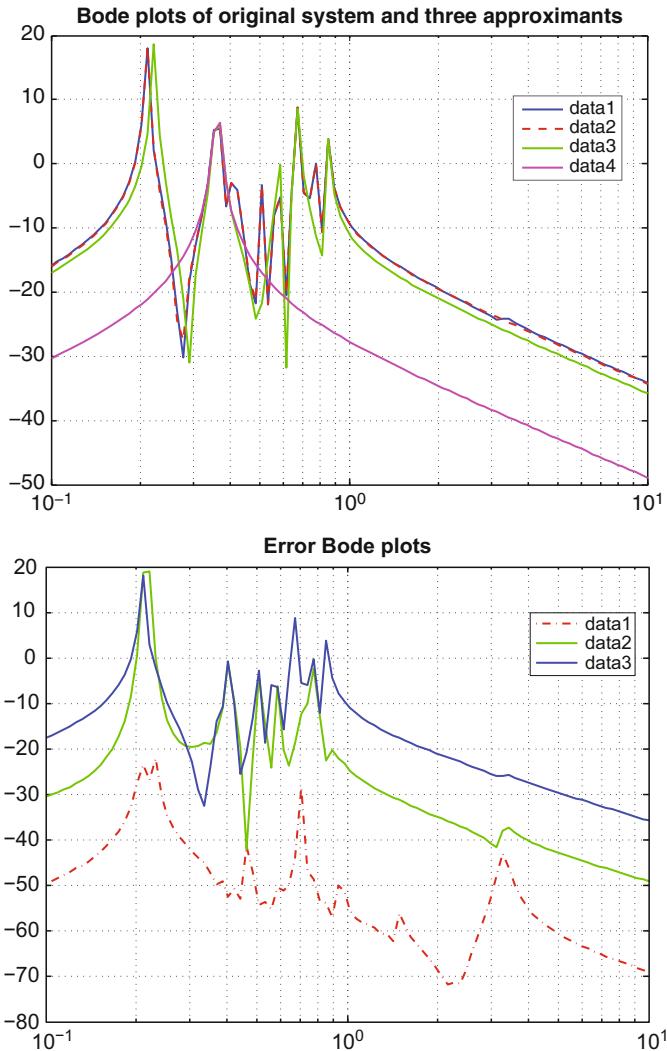


Fig. 8 Upper pane: Frequency responses of original system and approximants (orders 2, 10, 18). Lower pane: Frequency responses of error systems (orders 2,10,18)

Figure 8 shows that (for the chosen values of the parameters) the frequency response has about seven peaks. A second order approximant reproduces (approximately) the highest peak, a tenth order system reproduces (approximately) five peaks, while a system of order 18 provides a good approximation of the whole frequency response (see in particular the error plots – lower pane of Figure 8).

7.4.3 Four-pole Band-pass Filter

In this case 1,000 frequency response measurements are given, of the 2×2 S -parameters of a semi-conductor device which is meant to be a band-pass filter. There is no a priori model available. The range of frequencies is between 40 and 120 GHz; We will use the Loewner matrix procedure applied to the S -parameters. This yields $\mathbb{L}, \mathbb{M} \in \mathbb{C}^{2,000 \times 2,000}$.

In the upper left-hand plot of Fig. 9, the singular values of the Loewner matrix corresponding to the two-port system (upper curve) is compared with the singular values of two one-port subsystems (lower curves). As the decay of all curves is fast, an approximant of order around 20, is expected to provide a good fit. Indeed, as the upper right-hand plot shows, a 21st order approximant provides fits with error less than $-60 dB$. For comparison the fit of a 15th order model is shown in the lower left-hand plot. Sometimes in practical applications, the entries of the two-port S -parameters are modeled separately. In our case 14th order models are sufficient, but the McMillan degree of the two-port is 28 or higher (depending on the symmetries involved, e.g. $S_{11} = S_{22}, S_{12} = S_{21}$).

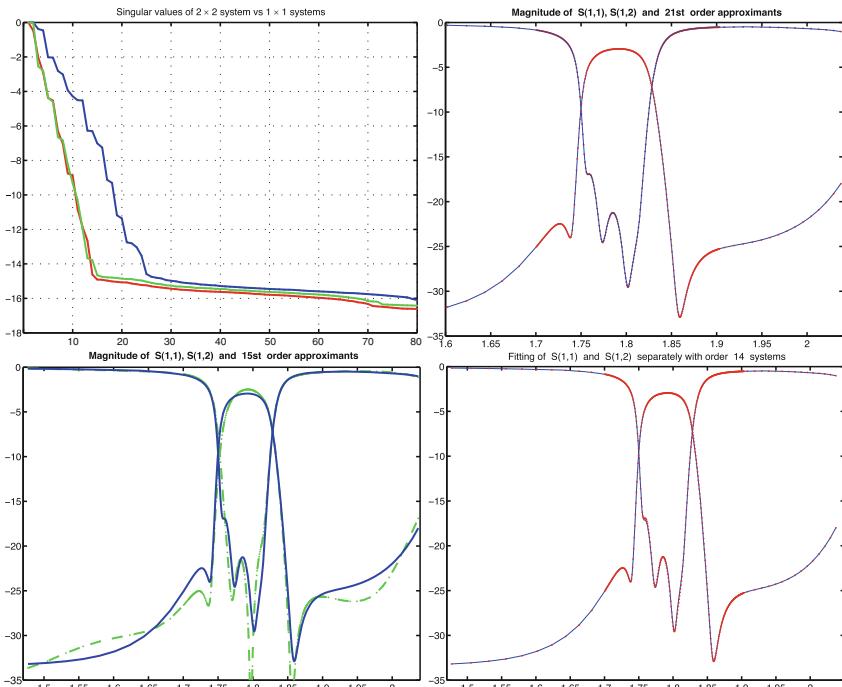


Fig. 9 Upper row, left pane: The singular values of $x\mathbb{L} - \mathbb{M}$, for the two-port and for two one-ports. Upper row, right pane: The $S(1,1)$ and $S(1,2)$ parameter data for a 21st order model. Lower row, left pane: Fitting $S(1,1), S(1,2)$ jointly with a 15th order approximant. Lower row, right pane: Fitting $S(1,1), S(1,2)$ separately with 14th order approximants

8 Conclusions

In this chapter, we have surveyed two projection-based model reduction frameworks both of which make fundamental use of tangential rational interpolation. This approach is extremely flexible in applications; scales well to handle extremely large problems; and is capable of producing very high fidelity reduced order models very efficiently. Throughout, examples are given that illustrate the theoretical concepts discussed.

The first framework we consider assumes the availability of a high-order dynamical system model in state space form (often these are acquired and assembled through the independent discretization and coupling of high resolution distributed parameter models). The retention of high model fidelity is directly recast as the problem of choosing appropriate interpolation points and associated tangent directions. We address this with a detailed discussion of how interpolation points and tangent directions can be chosen to obtain reduced order models which are *optimal* with respect to \mathcal{H}_2 error measures (Sect. 3), or so that reduced order models are obtained which retain the property of passivity (Sect. 4).

The flexibility of our approach is demonstrated in Sects. 5 and 6 where we explore significant generalizations to the basic problem setting. We consider how to preserve second-order system structure; reduction of systems that involve delays or memory terms; and systems having a structured dependence on parameters that must be retained in the reduced systems.

The second major framework we explore allows for the original (high-order) dynamical system model to be *inaccessible*, and assumes that only system response data (e.g. frequency response measurements) are available. We describe an approach using the Loewner matrix pencil and place it in the context of interpolatory model reduction methods. We show that the Loewner matrix pencil constitutes an effective tool in obtaining minimal state-space realizations of reduced order models directly from measured data.

Acknowledgments The work of the A.C. Antoulas was supported in part by the NSF through Grant CCF-0634902. The work of C. Beattie and S. Gugercin has been supported in part by NSF Grants DMS-0505971 and DMS-0645347.

References

1. Anderson, B., Antoulas, A.: Rational interpolation and state variable realizations. In: Decision and Control, 1990. Proceedings of the 29th IEEE Conference on pp. 1865–1870 (1990)
2. Antoulas, A.: Approximation of Large-Scale Dynamical Systems (Advances in Design and Control). Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (2005)
3. Antoulas, A.: A new result on passivity preserving model reduction. *Systems and Control Letters* **54**, 361–374 (2005)
4. Antoulas, A.: On the construction of passive models from frequency response data. *Automatisierungstechnik* **56**, 447–452 (2008)

5. Antoulas, A., Anderson, B.: On the scalar rational interpolation problem. *IMA Journal of Mathematics Control and Information* **3**, 61–68 (1986)
6. Bagley, R., Torvik, P.: A theoretical basis for the application of fractional calculus to viscoelasticity. *Journal of Rheology* **27**, 201–210 (1983)
7. Bai, Z.: Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems. *Applied Numerical Mathematics* **43**(1-2), 9–44 (2002)
8. Bai, Z., Bindel, D., Clark, J., Demmel, J., Pister, K., Zhou, N.: New numerical techniques and tools in SUGAR for 3D MEMS simulation. In: Technical Proceedings of the Fourth International Conference on Modeling and Simulation of Microsystems, pp. 31–34 (2000)
9. Bai, Z., Feldmann, P., Freund, R.: Stable and passive reduced-order models based on partial Padé approximation via the Lanczos process. *Numerical Analysis Manuscript* **97**, 3–10 (1997)
10. Bai, Z., Feldmann, P., Freund, R.: How to make theoretically passive reduced-order models passive in practice. In: Custom Integrated Circuits Conference, 1998. Proceedings of the IEEE 1998, pp. 207–210 (1998)
11. Bai, Z., Freund, R.: A partial Padé-via-Lanczos method for reduced-order modeling. *Linear Algebra and its Applications* **332**(334), 141–166 (2001)
12. Bai, Z., Skoogh, D.: A projection method for model reduction of bilinear dynamical systems. *Linear Algebra and Its Applications* **415**, 406–425 (2006)
13. Bai, Z., Su, Y.: Dimension reduction of second order dynamical systems via a second-order Arnoldi method. *SIAM Journal on Scientific Computing* **5**, 1692–1709 (2005)
14. Baur, U., Beattie, C., Benner, P., Gugercin, S.: Interpolatory projection methods for parameterized model reduction. *Chemnitz Scientific Computing Preprints 09-08*, TU Chemnitz, (ISSN 1864-0087) November 2009
15. Beattie, C., Gugercin, S.: Krylov-based minimization for optimal \mathcal{H}_2 model reduction. In: Proceedings of the 46th IEEE Conference on Decision and Control pp. 4385–4390 (2007)
16. Beattie, C., Gugercin, S.: Interpolatory projection methods for structure-preserving model reduction. *Systems and Control Letters* **58**(3), 225–232 (2009)
17. Beattie, C., Gugercin, S.: A trust region method for optimal \mathcal{H}_2 model reduction. In: Proceedings of the 48th IEEE Conference on Decision and Control (2009)
18. Benner, P.: Solving large-scale control problems. *IEEE Control Systems Magazine* **24**(1), 44–59 (2004)
19. Benner, P., Feng, L.: A robust algorithm for parametric model order reduction based on implicit moment matching. In B. Lohmann and A. Kugi (eds.) Tagungsband GMA-FA 1.30 “Modellbildung, Identifizierung und Simulation in der Automatisierungstechnik”, Workshop in Anif, 26.–28.9.2007, pp. 34–47 (2007)
20. Benner, P., Saak, J.: Efficient numerical solution of the LQR-problem for the heat equation. *Proceedings in Applied Mathematics and Mechanics* **4**(1), 648–649 (2004)
21. Benner, P., Sokolov, V.: Partial realization of descriptor systems. *Systems and Control Letters* **55**(11), 929–938 (2006)
22. Bond, B., Daniel, L.: Parameterized model order reduction of nonlinear dynamical systems. In: IEEE/ACM International Conference on Computer-Aided Design, 2005. ICCAD-2005, pp. 487–494 (2005)
23. Bui-Thanh, T., Willcox, K., Ghattas, O.: Model reduction for large-scale systems with high-dimensional parametric input space. *SIAM Journal on Scientific Computing* **30**(6), 3270–3288 (2008)
24. Bungartz, H.: Dünne Gitter und deren Anwendung bei der adaptiven Lösung der dreidimensionalen Poisson-Gleichung. Dissertation, Institut für Informatik, TU München (1992)
25. Bunse-Gerstner, A., Kubalinska, D., Vossen, G., Wilczek, D.: \mathcal{H}_2 -optimal model reduction for large scale discrete dynamical MIMO systems. *Journal of Computational and Applied Mathematics* (2009). Doi:10.1016/j.cam.2008.12.029
26. Chahlaoui, V., Gallivan, K.A., Vandendorpe, A., Van Dooren, P.: Model reduction of second-order system. In: P. Benner, V. Mehrmann, D. Sorensen (eds.) *Dimension Reduction of Large-Scale Systems, Lecture Notes in Computational Science and Engineering*, vol. 45, pp. 149–172. Springer, Berlin/Heidelberg, Germany (2005)

27. Clark, J.V., Zhou, N., Bindel, D., Schenato, L., Wu, W., Demmel, J., Pister, K.S.J.: 3D MEMS simulation using modified nodal analysis. In: Proceedings of Microscale Systems: Mechanics and Measurements Symposium, p. 6875 (2000)
28. Craig Jr., R.R.: Structural Dynamics: An Introduction to Computer Methods. Wiley, New York (1981)
29. Daniel, L., Siong, O., Chay, L., Lee, K., White, J.: A Multiparameter moment-matching model-reduction approach for generating geometrically parameterized interconnect performance models. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **23**(5), 678–693 (2004)
30. De Villemagne, C., Skelton, R.: Model reductions using a projection formulation. International Journal of Control **46**(6), 2141–2169 (1987)
31. Fanizza, G., Karlsson, J., Lindquist, A., Nagamune, R.: Passivity-preserving model reduction by analytic interpolation. Linear Algebra and Its Applications **425**(2-3), 608–633 (2007)
32. Farle, O., Hill, V., Ingelström, P., Dyczij-Edlinger, R.: Multi-parameter polynomial order reduction of linear finite element models. Mathematical and Computational Modeling of Dynamics Systems **14**(5), 421–434 (2008)
33. Feldmann, P., Freund, R.: Efficient linear circuit analysis by Padé approximation via the Lanczos process. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **14**(5), 639–649 (1995)
34. Feng, L.: Parameter independent model order reduction. Mathematics and Computers in Simulations **68**(3), 221–234 (2005)
35. Feng, L., Benner, P.: A robust algorithm for parametric model order reduction based on implicit moment matching. PAMM Proceedings of the Applied Mathematics and Mechanism **7**(1), 1021,501–1021,502 (2008)
36. Feng, L., Rudnyi, E., Korvink, J.: Preserving the film coefficient as a parameter in the compact thermal model for fast electrothermal simulation. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **24**(12), 1838–1847 (2005)
37. Freund, R., Feldmann, P.: Efficient small-signal circuit analysis and sensitivity computations with the PVL algorithm. In: Proceedings of the 1994 IEEE/ACM International Conference on Computer-aided Design, pp. 404–411. IEEE Computer Society Press Los Alamitos, CA, USA (1994)
38. Freund, R.W.: Padé-type model reduction of second-order and higher-order linear dynamical systems. In: P. Benner, V. Mehrmann, D.C. Sorensen (eds.) Dimension Reduction of Large-Scale Systems, *Lecture Notes in Computational Science and Engineering*, vol. 45, pp. 191–223. Springer, Berlin/Heidelberg (2005)
39. Gallivan, K., Vandendorpe, A., Dooren, P.: Model reduction of MIMO systems via tangential interpolation. SIAM Journal on Matrix Analysis and Applications **26**(2), 328–349 (2005)
40. Glover, K.: All optimal Hankel-norm approximations of linear multivariable systems and their L^∞ -error bounds. International Journal of Control **39**(6), 1115–1193 (1984)
41. Griebel, M.: A parallelizable and vectorizable multi-level algorithm on sparse grids. In: Parallel Algorithms for Partial Differential Equations (Kiel, 1990), *Notes in Numerical Fluid Mechanics*, vol. 31, pp. 94–100. Vieweg, Braunschweig (1991)
42. Grimme, E.: Krylov projection methods for model reduction. Ph.D. thesis, Coordinated-Science Laboratory, University of Illinois at Urbana-Champaign (1997)
43. Grimme, E., Sorensen, D., Dooren, P.: Model reduction of state space systems via an implicitly restarted Lanczos method. Numerical Algorithms **12**(1), 1–31 (1996)
44. Gugercin, S.: Projection methods for model reduction of large-scale dynamical systems. Ph.D. thesis, Ph.D. Dissertation, ECE Department, Rice University, December 2002 (2002)
45. Gugercin, S.: An iterative rational Krylov algorithm (IRKA) for optimal \mathcal{H}_2 model reduction. In: Householder Symposium XVI. Seven Springs Mountain Resort, PA, USA (2005)
46. Gugercin, S., Antoulas, A.: An ϵ error expression for the Lanczos procedure. In: Proceedings of the 42nd IEEE Conference on Decision and Control (2003)
47. Gugercin, S., Antoulas, A.: Model reduction of large-scale systems by least squares. Linear Algebra and its Applications **415**(2-3), 290–321 (2006)

48. Gugercin, S., Antoulas, A., Beattie, C.: A rational Krylov iteration for optimal \mathcal{H}_2 model reduction. In: Proceedings of MTNS, vol. 2006 (2006)
49. Gugercin, S., Antoulas, A., Beattie, C.: \mathcal{H}_2 model reduction for large-scale linear dynamical systems. SIAM Journal on Matrix Analysis and Applications **30**(2), 609–638 (2008)
50. Gugercin, S., Willcox, K.: Krylov projection framework for fourier model reduction. Automatica **44**(1), 209–215 (2008)
51. Gunupudi, P., Khazaka, R., Nakhla, M.: Analysis of transmission line circuits using multi-dimensional model reduction techniques. IEEE Transactions on Advanced Packaging **25**(2), 174–180 (2002)
52. Gunupudi, P., Khazaka, R., Nakhla, M., Smy, T., Celo, D.: Passive parameterized time-domain macromodels for high-speed transmission-line networks. IEEE Transactions on Microwave Theory and Techniques **51**(12), 2347–2354 (2003)
53. Halevi, Y.: Frequency weighted model reduction via optimal projection. IEEE Transactions on Automatic Control **37**(10), 1537–1542 (1992)
54. Hyland, D., Bernstein, D.: The optimal projection equations for model reduction and the relationships among the methods of Wilson, Skelton, and Moore. IEEE Transactions on Automatic Control **30**(12), 1201–1211 (1985)
55. Ionutiu, R., Rommes, J., Antoulas, A.: Passivity preserving model reduction using dominant spectral zero interpolation. IEEE Transactions on CAD (Computer-Aided Design of Integrated Circuits and Systems) **27**, 2250–2263 (2008)
56. Kellems, A., Roos, D., Xiao, N., Cox, S.J.: Low-dimensional morphologically accurate models of subthreshold membrane potential. Journal of Computational Neuroscience, **27**(2), 161–176 (2009)
57. Korvink, J., Rudnyi, E.: Oberwolfach benchmark collection. In: P. Benner, V. Mehrmann, D.C. Sorensen (eds.) Dimension Reduction of Large-Scale Systems, *Lecture Notes in Computational Science and Engineering*, vol. 45, pp. 311–315. Springer, Berlin/Heidelberg, Germany (2005)
58. Krajewski, W., Lepschy, A., Redivo-Zaglia, M., Viaro, U.: A program for solving the L2 reduced-order model problem with fixed denominator degree. Numerical Algorithms **9**(2), 355–377 (1995)
59. Kubalinska, D., Bunse-Gerstner, A., Vossen, G., Wilczek, D.: \mathcal{H}_2 -optimal interpolation based model reduction for large-scale systems. In: Proceedings of the 16th International Conference on System Science, Poland (2007)
60. Kunkel, P.: Differential-Algebraic Equations: Analysis and Numerical Solution. European Mathematical Society, Amsterdam (2006)
61. Lasance, C.: Two benchmarks to facilitate the study of compact thermal modelingphenomena. IEEE Transactions on Components and Packaging Technologies [see also IEEE Transactions on Packaging and Manufacturing Technology, Part A: Packaging Technologies] **24**(4), 559–565 (2001)
62. Lefteriu, S., Antoulas, A.: A new approach to modeling multi-port systems from frequency domain data. IEEE Transactions on CAD (Computer-Aided Design of Integrated Circuits and Systems) **29**, 14–27 (2010)
63. Leitman, M., Fisher, G.: The linear theory of viscoelasticity. Handbuch der Physik **6**, 10–31 (1973)
64. Leung, A.M., Khazaka, R.: Parametric model order reduction technique for design optimization. In: IEEE International Symposium on Circuits and Systems, 2005. ISCAS 2005, vol. 2, pp. 1290–1293 (2005)
65. Ma, M., Leung, A.M., Khazaka, R.: Sparse Macromodels for Parametric Networks. In: Proceedings of the IEEE International Symposium on Circuits and Systems (2006)
66. Mayo, A., Antoulas, A.: A framework for the solution of the generalized realization problem. Linear Algebra and Its Applications **425**(2-3), 634–662 (2007)
67. Mehrmann, V., Stykel, T.: Balanced truncation model reduction for large-scale systems in descriptor form. In: P. Benner, V. Mehrmann, D. Sorensen (eds.) Dimension Reduction of Large-Scale Systems, pp. 83–115. Springer, Belin (2005)

68. Meier III, L., Luenberger, D.: Approximation of linear constant systems. *IEEE Transactions on Automatic Control* **12**(5), 585–588 (1967)
69. Meyer, D., Srinivasan, S.: Balancing and model reduction for second-order form linear systems. *IEEE Transactions on Automatic Control* **41**(11), 1632–1644 (1996)
70. Moore, B.: Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control* **26**(1), 17–32 (1981)
71. Moosmann, K., Korvink, J.: Automatic parametric mor for mems design. In: B. Lohmann, A. Kugi (eds.) Tagungsband GMA-FA 1.30 “Modellbildung, Identifikation und Simulation in der Automatisierungstechnik”, Workshop am Bostalsee, 27.–29.9.2006, pp. 89–99 (2006)
72. Mullis, C., Roberts, R.: Synthesis of minimum roundoff noise fixed point digital filters. *IEEE Transactions on Circuits and Systems* **23**(9), 551–562 (1976)
73. Odabasioglu, A., et al.: PRIMA. *IEEE Transaction on Computer-Aided Design of Integrated Circuits and Systems* **17**(8) (1998)
74. Pillage, L., Rohrer, R.: Asymptotic waveform evaluation for timing analysis. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **9**(4), 352–366 (1990)
75. Preumont, A.: Vibration Control of Active Structures: An Introduction. Springer, Berlin (2002)
76. Raghavan, V., Rohrer, R., Pillage, L., Lee, J., Bracken, J., Alaybeyi, M.: AWE-inspired. In: Custom Integrated Circuits Conference, 1993. Proceedings of the IEEE 1993 (1993)
77. Rommes, J., Martins, M.: Efficient computation of multivariable transfer function dominant poles using subspace acceleration. *IEEE Transactions on Power Systems* **21**, 1471–1483 (2006)
78. Rudnyi, E., Moosmann, C., Greiner, A., Bechtold, T., Korvink, J.: Parameter Preserving Model Reduction for MEMS System-level Simulation and Design. In: Fifth MATHMOD Proceedings **1** (2006)
79. Ruhe, A.: Rational Krylov algorithms for nonsymmetric eigenvalue problems. II: matrix pair. *Linear Algebra and Its Applications Applications*, pp. 282–295 (1984)
80. Sorensen, D.: Passivity preserving model reduction via interpolation of spectral zeros. *Systems and Control Letters* **54**, 354–360 (2005)
81. Spanos, J., Milman, M., Mingori, D.: A new algorithm for L^2 optimal model reduction. *Automatica (Journal of IFAC)* **28**(5), 897–909 (1992)
82. Su, T.J., Jr., R.C.: Model reduction and control of flexible structures using Krylov vectors. *Journal of Guid. Control Dyn.* **14**, 260–267 (1991)
83. Van Dooren, P., Gallivan, K., Absil, P.: \mathcal{H}_2 -optimal model reduction of MIMO systems. *Applied Mathematics Letters* **21**(12), 1267–1273 (2008)
84. Weaver, W., Johnston, P.: Structural dynamics by finite elements. Prentice-Hall, Upper Saddle River (1987)
85. Weile, D., Michielssen, E., Grimme, E., Gallivan, K.: A method for generating rational interpolant reduced order models of two-parameter linear systems. *Applied Mathematics Letters* **12**(5), 93–102 (1999)
86. Weile, D.S., Michielssen, E., Grimme, E., Gallivan, K.: A method for generating rational interpolant reduced order models of two-parameter linear systems. *Applied Mathematics Letters* **12**(5), 93–102 (1999)
87. Willcox, K., Megretski, A.: Fourier series for accurate, stable, reduced-order models in large-scale applications. *SIAM Journal of Scientific Computing* **26**(3), 944–962 (2005)
88. Wilson, D.: Optimum solution of model-reduction problem. *Proceedings of IEE* **117**(6), 1161–1165 (1970)
89. Yan, W., Lam, J.: An approximate approach to h^2 optimal model reduction. *IEEE Transactions on Automatic Control* **44**(7), 1341–1358 (1999)
90. Yousuff, A., Skelton, R.: Covariance equivalent realizations with applications to model reduction of large-scale systems. *Control and Dynamic Systems* **22**, 273–348 (1985)
91. Yousuff, A., Wagie, D., Skelton, R.: Linear system approximation via covariance equivalent realizations. *Journal of Mathematical Analysis and Applications* **106**(1), 91–115 (1985)

92. Zenger, C.: Sparse grids. In: Parallel Algorithms for Partial Differential Equations (Kiel, 1990), *Notes Numerical Fluid Mechanics*, vol. 31, pp. 241–251. Vieweg, Braunschweig (1991)
93. Zhaojun, B., Qiang, Y.: Error estimation of the Pade approximation of transfer functions via the Lanczos process. *Electronic Transactions on Numerical Analysis* **7**, 1–17 (1998)
94. Zhou, K., Doyle, J., Glover, K.: Robust and Optimal Control. Prentice-Hall, Upper Saddle River, NJ (1996)
95. Zicic, D., Watson, L., Beattie, C.: Contragredient transformations applied to the optimal projection equations. *Linear Algebra and Its Applications* **188**, 665–676 (1993)

Efficient Model Reduction for the Control of Large-Scale Systems

Richard Colgren

1 Introduction

When analyzing and controlling large-scale systems, it is extremely important to develop efficient modeling processes. The key dynamic elements must be identified and spurious dynamic elements eliminated. This allows the controls engineer to implement the optimal control strategy for the problem at hand. Model reduction techniques provide an extremely effective way to address this requirement.

In this chapter the evolution of model reduction techniques for designing control systems for large-scale systems is summarized. These start with simple approaches such as spectral decomposition and simultaneous gradient error reduction and then progresses through a variety of balanced and related model reduction approaches. Motivations for the use of these methods are given. Emphasis is given to approaches which are easily applied to the large, generalized models which are created by Computer Aided Design (CAD) tools. A U-2S example application is provided and described.

The provided example of a large-scale system of very high order is a model of an air vehicle with significant aeroservoelastic coupling. Future unstable vehicles, or those employing features such as relaxed static stability, can display aeroservoelastic coupling. Evaluation of the interactions between the dynamic modes must be accomplished using an efficient, integrated modeling approach. The extensive use of composite materials has also resulted in greater aeroservoelastic coupling. Finite element models of complex systems are sparse and are also intrinsically of very high order. Such large-scale systems must be subjected to comprehensive analysis.

The methods discussed in this chapter preserve the frequency response characteristics of the system model being examined while reducing its size to one practical for direct controls design. They support a variety of optimal and robust control system

R. Colgren
Vice President and Chief Scientist, Viking Aerospace LLC, 100 Riverfront Road, Suite B,
Lawrence, KS 66044, USA
e-mail: rcolgren@gmail.com

design approaches. Some control system design methods require the use of all of the system's states. Without the use of model reduction techniques the full state control system design would be too large to practically implement. States that would be included within the controller might also be outside the bandwidth of the servos or actuators. This might generate an ill-conditioned problem. The use of model reduction methods provides a way to generate full state controller solutions of reduced size, and to further simplify these full state feedback solutions once they have been generated.

2 Spectral Decomposition

This method is an efficient way to generate a reduced order model of a large-scale system when all of its subsystems are decoupled or are at most weakly coupled. It was initially developed by Chiodi and Davis at the Lockheed Corporation [1]. It depends on systems having distinct eigenvalues within well separated frequency responses as shown in Fig. 1. These modal groupings are then decoupled using the eigenvector matrix.

The starting point for this method is the standard n th order state space representation given in (1).

$$dx/dt = Ax + Bu; y = Cx \quad (1)$$

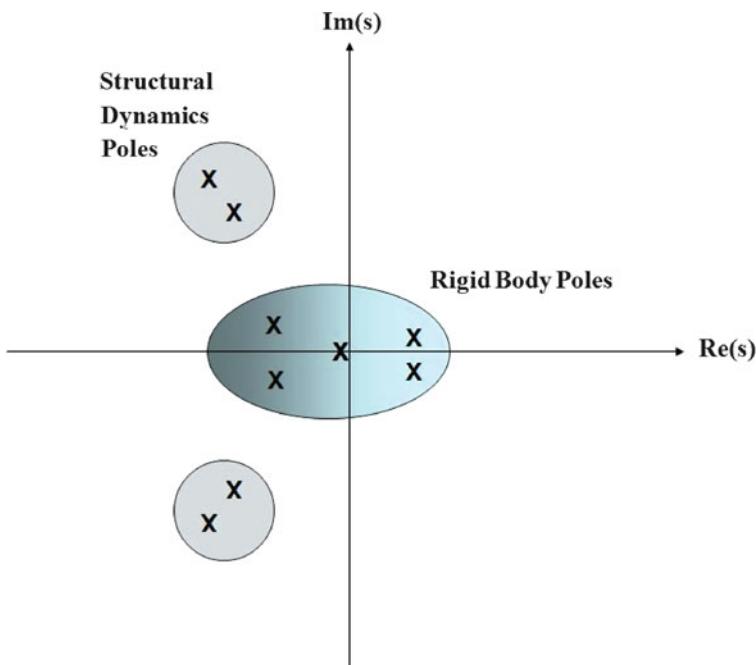


Fig. 1 Distinct frequency range groupings of a model's eigenvalues

The system matrix A is then represented using the system's eigenvalue and eigenvector matrices as shown in (2).

$$dx/dt = V\Lambda V^{-1}x + Bu; y = Cx \quad (2)$$

This system description can then be expanded into the following series representation as provided in (3). Note that the sum is initially computed using the full number of states within the model, but that for control system design it is computed using the eigenvalues to be obtained. Also, please note that $V_iV_i^{-1}$ is the residue of the eigenvalue λ_i in $(sI - A)^{-1}$.

$$dx/dt = [\sum V_iV_i^{-1}\lambda_i]x + Bu; y = Cx \quad (3)$$

From this representation the system can be reduced into its individual elements as is done in (4)

$$dx_i/dt = \underline{A}_i x + \underline{B}_i u \quad (4)$$

The spectral decomposition process is described in Fig. 2. The system eigenvalues and eigenvectors are next calculated. For each eigenvalue to be retained within the reduced order model the eigenvector and its inverse are multiplied to generate $V_iV_i^{-1}$. Each matrix element is multiplied by its corresponding eigenvalue λ_i . The transformed state matrix is then constructed by solving for the sum as given in (5) for all the retained eigenvalues.

$$\underline{A} = \sum V_iV_i^{-1}\lambda_i \quad (5)$$

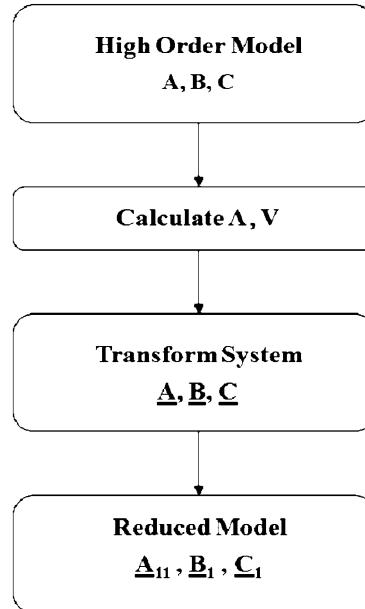


Fig. 2 Spectral decomposition process

To calculate $\underline{B} = TB$ and $\underline{C} = CT$ the transformation matrix T must also be calculated. This transformation is provided as in (6).

$$T = \sum V_i V_i^{-1} \quad (6)$$

After spectral decomposition the system is transformed into that provided as in (7)–(9).

$$d\underline{x}_1/dt = \underline{A}_{11}\underline{x}_1 + \underline{A}_{12}\underline{x}_2 + \underline{B}_1 u \quad (7)$$

$$d\underline{x}_2/dt = \underline{A}_{21}\underline{x}_1 + \underline{A}_{22}\underline{x}_2 + \underline{B}_2 u \quad (8)$$

$$y = \underline{C}_1\underline{x}_1 + \underline{C}_2\underline{x}_2 \quad (9)$$

The reduced system model as described by the states to be used for the design of the control system is given as in (10) and (11).

$$d\underline{x}_1/dt = \underline{A}_{11}\underline{x}_1 + \underline{B}_1 u \quad (10)$$

$$y = \underline{C}_1\underline{x}_1 \quad (11)$$

3 Simultaneous Gradient Error Reduction

This method provides an efficient way to generate a reduced order model of a large-scale system whether all of its subsystems and their outputs are decoupled or not. It does require that all of its inputs are no more than weakly coupled to each other. For flying qualities prediction and for aeroservodynamic modeling [2,3] the author originally developed this method for simultaneously fitting several frequency responses depicting high order systems to low order transfer functions including time delays [4,5]. This method has been applied to the design the control systems for several large-scale aerospace systems.

Simultaneous gradient error reduction is applied over a finite bandwidth. Controls analysis and design are always accomplished over a finite bandwidth. This allows modes outside of the control bandwidth to be neglected without impacting the responsiveness or the robustness of the closed loop system. As an example, for representative handling qualities it is believed that the low and the high order system models must have sufficiently similar frequency responses between approximately 1 to 10 rps. This means that, in the case of such a fit, that the pilot will judge the dynamics to be equivalent between the high order and the reduced order system models.

This procedure uses a conjugate gradient search routine to adjust the parameters of transfer functions including pure time delays with the desired form until their frequency responses are as close as possible to the frequency responses from the high order dynamics model. This is a multi-variable approach whereby several output frequency responses to the same input can be analyzed simultaneously. The

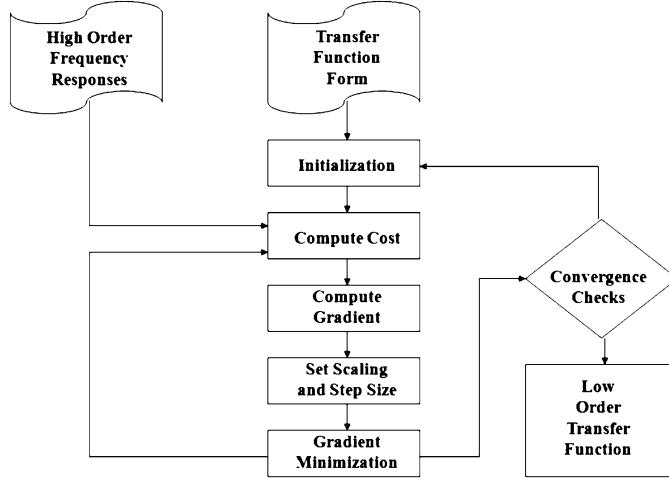


Fig. 3 Simultaneous gradient error reduction

user is allowed to fix, free, or simultaneously fit the various modes, including time delays, after selecting the order of the transfer function [5]. This procedure is outlined in Fig. 3.

The desired result is to match a transfer function to each of the given plant's magnitude $A(\omega)$ and phase frequency response $\Theta\omega$ data over the control bandwidth. Expressed in the form of a complex number $A(\omega)e^{j\Theta(\omega)}$, these plant transfer functions should respond in an equivalent manner to the full order plant model over the desired frequency range for the given pair of inputs and outputs. The reduced order transfer function representation is given in factored form and incorporates a pure time delay. All of the output responses to a single large-scale system input can be simultaneously fit using this approach. The output response models to the other large-scale system's inputs are generated in subsequent analysis.

The cost function J represents the fit error between the plant data and the reduced order model. It is formulated as the integral of the absolute value of the error squared between the data and the reduced order model as shown in (12)

$$J = \int_0^\infty |G(j\omega) - A(\omega)e^{j\Theta(\omega)}|^2 d\omega \quad (12)$$

The fit error is weighted over each infinitesimal frequency interval. This fit error can be assigned a relative importance as a function of frequency. A conjugate gradient routine is used for error reduction because of its simplicity. Parameters are constrained to maintain the same sign during minimization. This is done to maintain stability characteristics according to the Nyquist Stability Criteria. The gradient ΔJ can be expressed analytically by differentiating the cost function given in (13).

$$\Delta J = \operatorname{Re} \sum [G(j\omega_i) - A(\omega_i)e^{j\Theta(\omega_i)}] \Delta G(j\omega) d\omega \quad (13)$$

$G(j\omega)$ is defined in (14). In this equation $Z(j\omega)$ is composed of the partial derivatives of $G(j\omega)$ divided by their corresponding factors.

$$\Delta G(j\omega) = G(j\omega)Z(j\omega) \quad (14)$$

Gradient search techniques generally require parameter scaling to efficiently converge. An ideal scaling uses a cost function equally sensitive to each parameter. The law to accomplish this scales each parameter by its own magnitude as in (15). Re-scaling is done after completing each separate iteration.

$$P_i = b_i / |b_i| \quad (15)$$

The elements in the scaled gradient J are formed using (16).

$$\delta J / \delta p_i = \delta b_i / \delta p_i * \delta J / \delta b_i = |b_i| * \delta J / \delta b_i \quad (16)$$

Simultaneously fit parameters are constrained within bounds derived by initially allowing each transfer function to converge to an independent solution. An unweighted average is then computed, and is recomputed after each search. This is continued until all reduced order transfer functions have converged, the simultaneously fit parameters have converged, or the maximum number of iterations have been exceeded. In studies it was shown that the use of weighted averages did not improve the cost function of the final result. Its use also created convergence stability problems. If the problem as formulated does not have a true joint minimum, the gradient search routine will give the best estimated joint minimum.

4 Balancing

This method offers an efficient way to generate a reduced order model of a large-scale system whether all of its subsystems and their inputs and outputs are decoupled or not. This method for reducing the order of large-scale system models was originally developed by Enns while he was at Stanford [6]. An internally balanced system representation has input and output grammians that are equal and diagonal [7]. The magnitude of each diagonal element provides a measure of the controllability and observability of the corresponding state [8]. This is used as a guide in selecting those states which are the least controllable or observable for elimination [9]. The sum of the diagonal elements of the grammian corresponding to the states eliminated provides an error bound between the high order representation and the reduced order model.

Balancing relies on the notion of a system as a mapping from the inputs to the servos or actuators and then to the sensor outputs [10]. This mapping is viewed as a combination of the reachability mapping from the actuator input signals to the state vector and an observability mapping from the state to the sensor output signals. It follows from this mapping concept that the state is an intermediate quantity between

the inputs in the past and the sensor outputs in the future. The minimum amount of control energy required to reach the desired state vector is inversely proportional to the reachability grammian. Similarly, the amount of sensor output energy generated by the state vector is proportional to the observability grammian. Balancing examines the amount each state component participates in the mapping from the input to the output. The model reduction problem is now reduced into one of truncating small terms from the partial fraction decomposition.

The following section pertains to a system again formulated as in (1), where A and B are controllable and A and C are observable. Given these assumptions, the controllability grammian U and the observability grammian Y are the solutions to the Lyapunov equations given as in (17) and (18)

$$\underline{dx}/dt = \underline{Ax} + \underline{Bu} = TAT^{-1}x + T^{-1}Bu \quad (17)$$

$$y = \underline{Cx} = CTx \quad (18)$$

The transformation matrix T also relates the grammians through equations (19) and (20) using an algorithm first developed by Laub [11, 12]. These contragedient transformations (where both U and Y are diagonal) can be calculated either as the best conditioned contragedient transformation or as an internally balanced transformation.

$$\underline{U} = T^{-1}UT^{-1} \quad (19)$$

$$\underline{Y} = T'YT \quad (20)$$

For an internally balanced transformation in (21) applies. In this representation the columns of T are the eigenvectors of the product UY .

$$UY = T^{-1}\Lambda T \quad (21)$$

To compute the balanced representation of the original large-scale system model, it first must be decoupled into stable and unstable subsystem models. Balancing is used for this step to generate a transformation based on the system's eigenvalues. For simplicity, neutrally stable modes can be slightly perturbed to make them marginally stable. If the conditionality of the resulting stable projection is a concern, all neutrally stable modes can also be preserved as a part of the unstable system. This can lead to a higher order final reduced system model. The resulting system is given in (22). In this equation $G_+(s)$ is the stable subsystem and $G_-(s)$ is the unstable subsystem

$$G(s) = C(sI - A)^{-1}B = G_+(s) + G_-(s) \quad (22)$$

Without input or output weighting the stable subsystem is transformed into a real Schur form. Next the Cholesky factors L_u and L_y of U and Y are computed. A singular value decomposition of the product L_yL_u is then performed. The singular values and the corresponding vectors are arranged in order of decreasing singular values. The transformation T is then computed in a final Schur transformation.

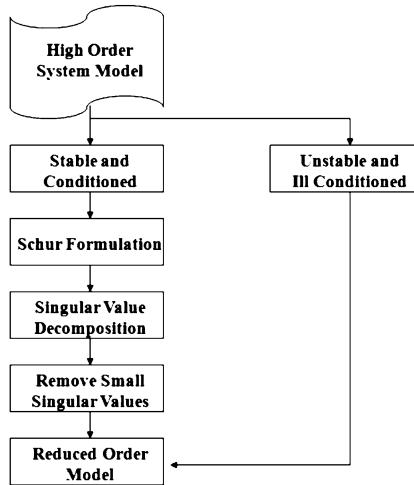


Fig. 4 Balancing process

States which have relatively small singular values and thus have a relatively small effect on the response of the large-scale system are then truncated from the model of the stable subsystem. The unstable and ill conditioned subsystem is completely preserved throughout this process. It is next recombined with the reduced order representation of the stable subsystem. This process is shown in Fig. 4.

4.1 Techniques Not Requiring Balancing

Other model reduction methods with an equivalent range of applicability have been designed to avoid the balancing process. One such method uses the Hankel Minimum Degree Approximate algorithm as described in [12]. This method is similar to balanced additive model reduction routines but can produce a reduced order model more reliably when the desired reduced model has nearly controllable and/or observable states [13]. These conditions are equivalent to having Hankel singular values very close to the machine accuracy.

For a stable system the Hankel singular values indicate the relative energy of each state with respect to the state energy of the entire system [14]. The reduced order system is directly determined by examining the system's Hankel singular values. An optimal reduced order system is selected using this algorithm to satisfy the error bound criterion regardless of the order selected at the beginning of the process [12]. Given the state space representation of the system as shown in (23), along with the value selected for k (the desired reduced order), the reduced order model is generated.

$$dx/dt = Ax + Bu; y = Cx \quad (23)$$

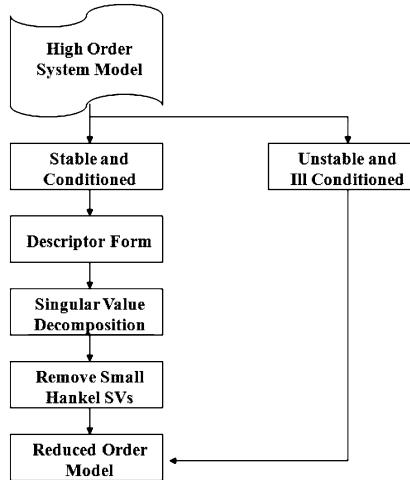


Fig. 5 Hankel minimum degree approximate process

The following steps, as shown in Fig. 5, produce a similarity transformation to truncate the original state space system into the k^{th} order reduced model. Weights on the original model's inputs and/or its outputs can make the model reduction algorithm focus on the specific frequency range of interest. These weights are required to be stable, minimum phase, and invertible. Note that as in the previous section the unstable subsystem must be combined with the reduced order stable subsystem to create the final reduced order model.

In more detail the steps to accomplish model reduction without balancing using the Hankel Minimum Degree Approximate algorithm are given as follows. First, the original large-scale system model must be decoupled into stable and unstable subsystem models. Next, the controllability and observability grammians P and Q must be generated. Using these grammians the descriptor given in (24) is generated

$$E = QP(j\omega) - \rho^2 I \quad (24)$$

where $\sigma_k > \rho \geq \sigma_{k+1}$.

A singular value decomposition is then accomplished on the descriptor. The system is next transformed to generate the system representation given in (25) and (26).

$$\frac{dx}{dt} = U'(\rho^2 A' + QAP)Vx + U'(QB - C')u \quad (25)$$

$$y = [CP - \rho B']Vu \quad (26)$$

The resulting system is then partitioned and truncated to become a k th order system. The final k th order Hankel Minimum Degree Approximate is the stable part of the state space realization. Its unstable part must be recombined with the reduced order model of the stable part.

The nature of the resulting error between the original system $G(j\omega)$ and the final reduced order model is discussed in [14]. This error is described by an all-pass function. A detailed description of the Hankel Minimum Degree Approximate algorithm can be found in [13].

4.2 Balancing Over A Disk

This procedure is a further development of the balancing technique. It was originally developed by Jonckheere at the University of Southern California [15, 16]. Its advantages are that it gives precise frequency response bounds over a desired bandwidth. Elimination of eigenvalues by inspection makes the balancing better conditioned and computationally efficient as well as further decreasing the final size of the reduced order system.

As was done using the previous two approaches, the unstable subsystem is removed and only the stable subsystem is processed and reduced. Note that with proper sign changes the unstable subsystem could be balanced over a disk. Because of the Nyquist Stability Criteria the number of open loop poles should not be reduced. In actual application work the number of unstable poles represents a very small percentage of the total number of poles. Thus reducing the number of unstable poles would have little effect on the dimension of the final reduced order system.

Standard balancing has infinite bandwidth. It thus does not necessarily provide the smallest possible H_∞ error. What is desired is the smallest frequency response error over the selected bandwidth. An infinite bandwidth is not required. From the Hankel Singular Values for each pole the error bound is generated as in (27).

$$\sup |G(s) - \underline{G}(s)| = \Sigma \delta \quad (27)$$

To develop a reduced order model optimized over a finite bandwidth and to improve the conditionality of the stable subsystem to be balanced, the subsystem is balanced over a disk. This is accomplished over a region as shown in Fig. 6. The disk is placed in the complex plane based on two considerations. First, the full order model $G(s)$ must be analytic in the disk [10, 17]. For the targeted error bound the disk must be placed to exclude any eigenvalues of $G(j\omega)$. Second, the disk should cover the interval of real frequencies ($-\Omega < \omega < \Omega$). This bandwidth limit Ω is usually based on the response limitations imposed by the control servos or actuators. This guarantees that the error bound only includes this critical frequency range. A good rule is to use $\alpha = \frac{1}{2}$ (the largest real part of the poles of $G(j\omega)$). This is done using a bilinear mapping as shown in Fig. 7. The error bound becomes that given in (28)

$$\sup |G(s) - \underline{G}_{DISK(s)}| = \Sigma \delta_{DISK} \quad (28)$$

Note that $\delta_{DISK} \leq \delta$. Those eigenvalues that have small singular values are deleted from the stable subsystem. The reduced order stable subsystem is recombined with the unstable subsystem. The overall process for reducing the order of a large-scale system over a disk is presented in Fig. 8.

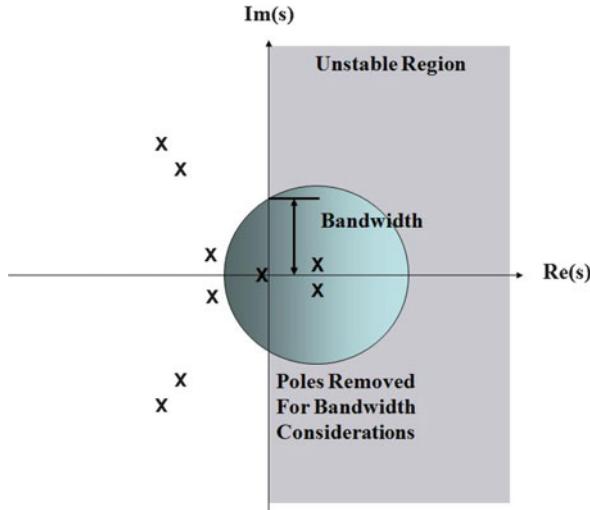


Fig. 6 Balancing over a Disk

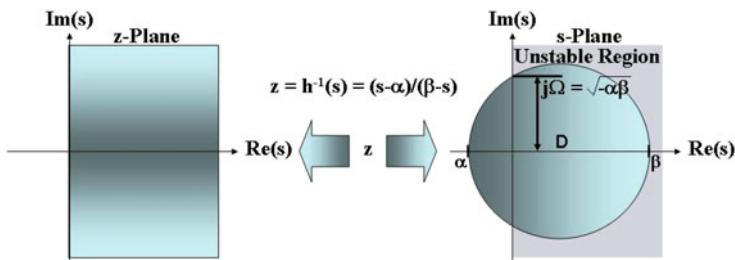


Fig. 7 Bilinear mapping

5 Example: Large-Scale System Application

The model reduction techniques discussed in the previous sections are based on minimizing the error between the high order and the low order models as measured within the frequency domain. Using these approaches, a large-scale 140th order multiple input, multiple output full dynamics model of the U-2S including rigid body and structural dynamics modes was reduced in order [4,18]. Within this section 90th, 80th and 40th order reduced models are documented. The response of the 90th order model is, by visual inspection, identical to the 140th order large-scale system's response. The response of the 80th order model begins to show differences from that of the large-scale system. The response shown in Fig. 9 is that of the roll rate due to a rudder input. This difference results from the migration of two extremely high frequency, lightly damped zeros from the left half into the right half of the complex plane. The large-scale system has ten non-minimum phase zeros while the 80th order model has 12 non-minimum phase zeros. Although the frequency response

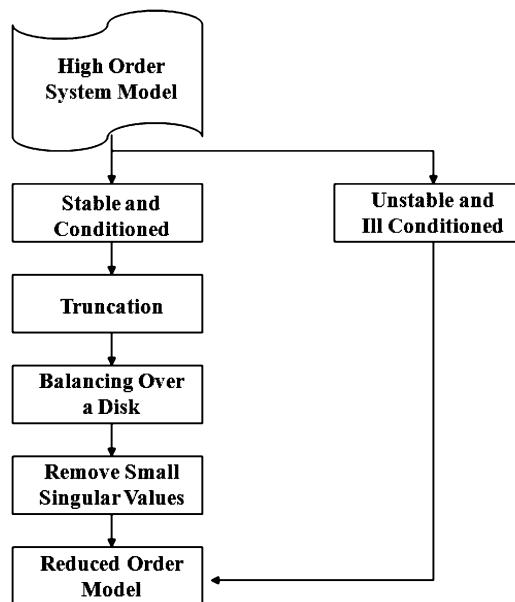


Fig. 8 Asymptotic Balancing Process

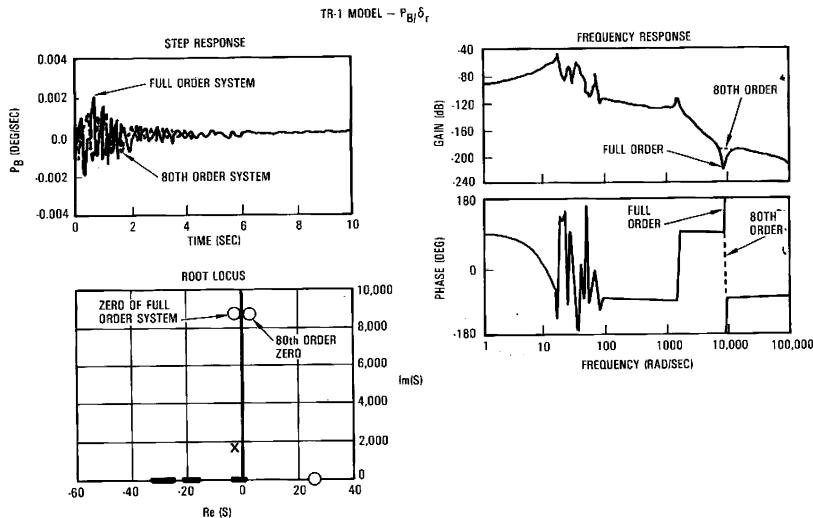


Fig. 9 Effect of non-minimum phase zeros

match shows very little variation between these two model's responses up to the frequency of this complex zero, the difference between the transient responses must be noted. In practice this difference is so far outside the control bandwidth of the servos that it will not have an impact on the design of the control laws.

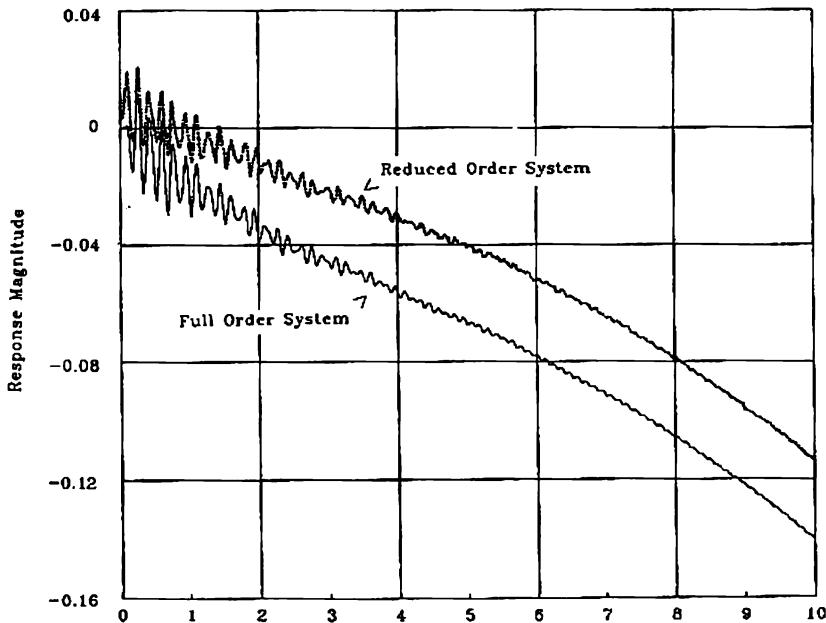


Fig. 10 Effect of Residualization

When the model is further reduced to 40th order a steady state offset error appears in the pitch rate response due to an elevator input. This offset is seen in Fig. 10. Residualization is easily used to shift the low order response to nearly overlay the 140th order large-scale system's equivalent response. Only minor differences in the peaks of the first 5 oscillations are seen after residualization. Besides this minor error the residualized 40th order system's transient response overlays that of the 140th order large-scale system. This residualized system's time response is omitted from Fig. 10 for clarity.

References

1. Davis W. J. and Chiodi O. A. A method for the decoupling of equations by spectral decomposition. *Lockheed Report LR 27479*, Dec 1975
2. Socolinsky D. A., Wolff L. B., Neuheisel J. D., and Eveland C. K. Background information and user guide for mil-f-8785c, military specification, flying qualities of piloted airplanes. *AFWAL-TR-81-3109*, Sept 1981
3. Seidel R. C. Transfer-function-parameter estimation from frequency response data - a fortran program. *NASA TM X-3286*, Sept 1975
4. Colgren R. D. Methods for model reduction. *AIAA-88-4144*, pages 15–17, Aug 1998
5. Colgren R. D. *Simultaneous Fit Equivalent Systems Program – User's Guide*. Lockheed California Company, Burbank, CA, 1985
6. Enns D. *Model Reduction for Control System Design*. PhD thesis, Department of Aeronautics and Astronautics Stanford University, 1984

7. Laub A. J., Heath M. T., Paige C. C., and Ward R. C. Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms. *IEEE Transactions on Automatic Control*, Feb 1987
8. Kailath T. *Linear Systems*. Prentice-Hall, Upper Saddle River, 1980
9. Safonov M. G. and Chiang R. Y. A schur method for balanced model reduction. *IEEE Trans. on Automat. Contr.*, 34(7):729–733, july 1989
10. Colgren R. D. and Jonckheere E. A. Balanced model reduction for large aircraft models. *MTNS91*, June 1991
11. Anderson E. Bai Z., Bischof C., Blackford S., Demmel J., Dongarra J., Du Croz J., Greenbaum A., Hammarling S., McKenney A., and Sorensen D. Lapack User's Guide, 1999
12. Balas G., Chiang R., Packard A., and Safonov M. Robust control toolbox 3 user's guide. *Revision for Version 3.3.3 (Release 2009a)*, Natick, MA, March 2009., March 2009
13. Safonov M. G., Chiang R. Y., and Limebeer D. J. N. Optimal hankel model reduction for nonminimal systems. *IEEE Transactions on Automatic and Control*, 35(4):496–502, April 1990
14. Glover K. All optimal hankel norm approximation of linear multivariable systems, and their 1 infinity – error bounds. *International Journal of Control*, 39(6):1145–1193, 1984
15. Jonckheere, E. A. and Silverman, L. M. A new set of invariants for linear systems – application to reduced order controller design. *IEEE Transactions on Automatic Control*, AC-28:953–964, 1983
16. Jonckheere, E. A. and Silverman, L. M. A new set of invariants for linear systems – application and approximation. In *International Symposium on Theory, Networks and Systems*, Santa Monica, CA, 1981
17. Colgren and R. D. Finite bandwidth model reduction applied to the advanced supersonic transport. *COMCON3*, Victoria, BC, Canada, Oct 1991
18. Dudinski R. J., and Colgren R. D. Time domain effects of model order reduction. In *3rd IEEE International Symposium on Intelligent Control*, pp. 24–26, Arlington, VA, Aug 1988

Dynamics of Tensegrity Systems

Maurício C. de Oliveira and Anders S. Wroldsen

1 Introduction and Motivation

Buckminster Fuller [1] coined the word *tensegrity* as a conjunction of the two words *tension* and *integrity* [3, 6, 10]. The artist Kenneth Snelson was the first to build a three-dimensional tensegrity structure [9], which he perfected throughout the years, such as the sculptures in Fig. 1. Tensegrity components are very simple elements, often just sticks and strings. Our interest is in engineering structures. The structures of interest have a large number of compressive parts (rods) and tensile parts (strings). On a *class 1 tensegrity system*, such as the sculptures in Fig. 1, there is no contact between the rigid bodies (rods). Integrity, or as we would say, stability, is provided by the network of tensile members (strings). No individual member is subject to torques even when the complete structure bends. In a *class k tensegrity system*, we allow as many as k rigid bodies (rods) to be in contact through *ball joints*. This last requirement preserves the ability of class k tensegrity systems to bend without any member experiencing torque.

The purpose of this chapter is to derive dynamic models for Tensegrity Systems. We will produce models that can cope with the large number of members often present in Tensegrity Systems. The main concerns for the mathematical models are that they should produce efficient computer simulations and that they may be used as a basis for control design. In this chapter we focus primarily on the first aspect,

M.C. de Oliveira

Department of Mechanical and Aerospace Engineering, University of California San Diego,
9500 Gilman Dr., La Jolla CA, 92093-0411, USA

e-mail: mauricio@ucsd.edu

A.S. Wroldsen

Centre for Ships and Ocean Structures, Norwegian University of Science and Technology,
Trondheim, Norway

e-mail: wroldsen@ntnu.no

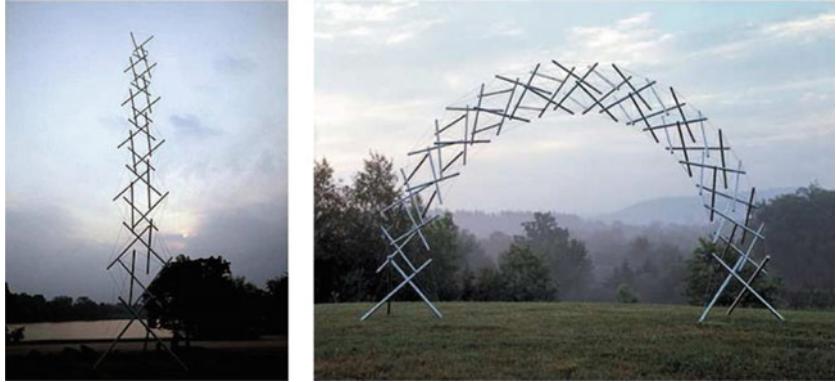


Fig. 1 Class 1 tensegrity towers by Kenneth Snelson

dynamic modeling. The interested reader is referred to [12] for an application of such models to control of Tensegrity Systems and to the book [7] as a general reference on tensegrity systems.

We will construct dynamic models for Tensegrity Systems made of sticks and strings in the form of ordinary differential equations under the following assumptions: (a) rods are rigid, thin and long so rotational motion about the longitudinal axis can be neglected; (b) strings are massless elastic elements with Hookean (linear) behavior only when in tension; (c) the connectivity of the structure is fixed.

These assumptions reflect tensegrity structures where the rods are massive and stiff, here approximated as rigid, as compared with a network of lightweight, elastic strings. Herein we are motivated by the network approach in [8]. We first study the dynamics of a single rod, in Sect. 2. These will be used to derive dynamic models for class 1 tensegrity systems in Sect. 3. These models can be applied to any type of class 1 tensegrity system. In Sect. 4 we turn our attention to the much more complex problem of deriving dynamic models for class k tensegrity systems. Instead of providing a general solution, we study a particular example of a class 2 tensegrity system, a simple cable, which will highlight the inherent difficulties to obtain dynamic models for class k tensegrity systems and illustrate the benefits of our proposed approach. The chapter is concluded with some general remarks on the implementation of these models in a computer simulation environment.

2 Dynamics of a Single Rigid Rod

We start by defining some important quantities associated with the dynamics of the single rigid rod in three-dimensional space as illustrated in Fig. 2. This rod has mass $m > 0$ and length $\ell > 0$ with extreme points $\mathbf{n}_j, \mathbf{n}_i \in \mathbb{R}^3$, hence $\|\mathbf{n}_j - \mathbf{n}_i\| = \ell$. We often make use of the normalized rod vector

$$\mathbf{b} = \ell^{-1}(\mathbf{n}_j - \mathbf{n}_i), \quad \|\mathbf{b}\| = 1. \quad (1)$$

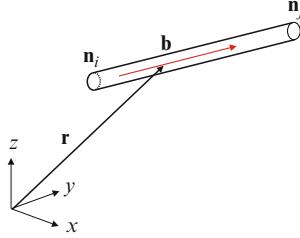


Fig. 2 Illustration of a rigid rod with and its configuration components (\mathbf{r}, \mathbf{b}) in \mathbb{R}^3

Any point in the rod can be located by the vector

$$\mathbf{v}(\mu) = \mu \mathbf{n}_j + (1 - \mu) \mathbf{n}_i,$$

where $\mu \in [0, 1]$. Let $\rho(\mu) \geq 0$ be a density function defined on the interval $\mu \in [0, 1]$ which describe the mass density along the rod, that is

$$m = \int_0^1 \rho(\mu) d\mu > 0.$$

We describe the position of the rod by means of the *configuration matrix*

$$\mathbf{Q} = [\mathbf{r} \ \mathbf{b}] \in \mathbb{R}^{3 \times 2} \quad (2)$$

where $\mathbf{r} = \mathbf{v}(\sigma)$, $\sigma \in [0, 1]$, is any fixed point in the rod. Whenever possible \mathbf{r} will be made to coincide with the center of mass of the rod. Any point in the rod can be equivalently described as a linear function of the configuration vector

$$\mathbf{v}(\eta) = \mathbf{Q} \begin{bmatrix} 1 \\ \eta \end{bmatrix}, \quad \eta \in [-\sigma\ell, (1 - \sigma)\ell].$$

Note that μ and η are related by $\mu = \sigma + \eta/\ell$. It is using η that we can compute higher order mass moments around \mathbf{r} , the next two of which are

$$f(\sigma) = m^1(\sigma), \quad J(\sigma) = m^2(\sigma), \quad m^k(\sigma) := \frac{1}{\ell} \int_{-\sigma\ell}^{(1-\sigma)\ell} \rho(\sigma + \eta/\ell) \eta^k d\eta,$$

Such moments are associated with an important quantity, the *kinetic energy* of the rod which is given by the formula

$$\begin{aligned} T &= \frac{1}{2\ell} \int_{-\sigma\ell}^{(1-\sigma)\ell} \rho(\sigma + \eta/\ell) \dot{\mathbf{v}}(\eta)^T \dot{\mathbf{v}}(\eta) d\eta \\ &= \frac{1}{2\ell} \text{trace} \left(\int_{-\sigma\ell}^{(1-\sigma)\ell} \rho(\sigma + \eta/\ell) \begin{bmatrix} 1 & \eta \\ \eta & \eta^2 \end{bmatrix} d\eta \dot{\mathbf{Q}}^T \dot{\mathbf{Q}} \right) = \frac{1}{2} \text{trace} (\mathbf{J}(\sigma) \dot{\mathbf{Q}}^T \dot{\mathbf{Q}}), \quad (3) \end{aligned}$$

where $\mathbf{J}(\sigma)$ is the positive semidefinite matrix of moments

$$\mathbf{J}(\sigma) = \begin{bmatrix} m & f(\sigma) \\ f(\sigma) & J(\sigma) \end{bmatrix} \succeq 0.$$

Note that if we choose $\sigma = \int_0^1 \rho(\mu) \mu d\mu$ so that \mathbf{r} coincides with the center of mass of the rod then $f(\sigma) = 0$. This leads to the well known decoupling of the kinetic energy in translational and rotational components in a rigid body described by its center of mass. One should choose to describe a rod by its center of mass whenever possible, with the main exception being the case when constraints are present in points of the rod other than the center of mass. We will illustrate this case later in Sect. 4. In the present chapter we confine our discussion to the particular case of uniform mass distributions (see [7] for a discussion on some nonuniform mass distributions). For rods with mass uniformly distributed along the rod we have $\rho(\mu) = m\ell^{-1}$. In this case the mass moments f and J are

$$f(\sigma) = \frac{1}{2}m\ell(1 - 2\sigma), \quad J(\sigma) = \frac{1}{3}m\ell^2(1 - 3\sigma + 3\sigma^2)$$

which are functions of σ , hence depends on the choice of the fixed point \mathbf{r} . Indeed, when the center of mass is the center of the bar, i.e. $\sigma = 1/2$, f and J are the familiar

$$f(1/2) = 0, \quad J(1/2) = \frac{1}{12}m\ell^2.$$

Another familiar choice is when \mathbf{r} coincides with one of the extreme points of the rod, say $\mathbf{r} = \mathbf{n}_i$ ($\sigma = 0$) so that

$$f(0) = \frac{1}{2}m\ell, \quad J(0) = \frac{1}{3}m\ell^2.$$

Interestingly, for any rod with uniform mass distribution, matrix $\mathbf{J}(\sigma)$ is positive definite, i.e. $\mathbf{J}(\sigma) \succ 0$, regardless of σ (see [7]).

2.1 Nodes as Functions of the Configuration

In dynamics, the node vectors must be expressed as a function of the configuration, in our case the matrix \mathbf{Q} . One of the major advantages of our approach is that the relationship between the configuration matrix and nodes is linear, as illustrated next.

Example 3.1. Consider the single rod pinned at one end and with three strings on the other end as illustrated in Fig. 3. Let \mathbf{Q} be as in (2). The end nodes of the rod can be computed as

$$\mathbf{n}_i = \mathbf{Q} \begin{bmatrix} 1 \\ \eta_i \end{bmatrix}, \quad \eta_1 = 0, \quad \eta_2 = \ell.$$

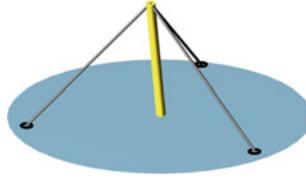


Fig. 3 A single rod with three strings

so that

$$\mathbf{N} = [\mathbf{n}_1 \cdots \mathbf{n}_5] = \mathbf{Q} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & \ell & 0 & 0 & 0 \end{bmatrix} + [\mathbf{0} \ \mathbf{0} \ \mathbf{n}_3 \ \mathbf{n}_4 \ \mathbf{n}_5],$$

where nodes \mathbf{n}_i , $i = 3, \dots, 5$ are the attachment points of the three strings that are not on the rod.

In general we should have

$$\mathbf{N} = \mathbf{Q} \Psi^T + \mathbf{Y}, \quad \mathbf{N}, \mathbf{Y} \in \mathbb{R}^{3 \times n}, \quad \Psi \in \mathbb{R}^{n \times 2}. \quad (4)$$

where $\Psi \in \mathbb{R}^{n \times 2}$ and $\mathbf{Y} \in \mathbb{R}^{3 \times n}$ are constant. The above expression is valid even when more than one rod is considered (see Sect. 3).

2.2 String Forces

In structural tensegrity systems, forces on the rod are due to the elongation of strings and ground reactions. See Sect. 4 for a treatment of gravitational forces. For simplicity, we assume that the strings are Hookean and that they are firmly attached to nodes on the rods or on fixed space coordinates (see [7] for more details). That is, strings are linear force elements with rest-length l_i^0 and stiffness k_i . The force vector of the i th string is

$$\mathbf{t}_i := \begin{cases} 0, & \|\mathbf{s}_i\| < l_i^0, \\ -\kappa_i(\|\mathbf{s}_i\| - l_i^0)\frac{\mathbf{s}_i}{\|\mathbf{s}_i\|}, & \|\mathbf{s}_i\| \geq l_i^0. \end{cases}$$

where \mathbf{s}_i is a vector in the direction of the i th string. String vectors are linear functions of the nodes of the structure. Assembling a matrix of string vectors and nodes

$$\mathbf{S} = [\mathbf{s}_1 \cdots \mathbf{s}_m] \in \mathbb{R}^{3 \times m}, \quad \mathbf{N} = [\mathbf{n}_1 \cdots \mathbf{n}_n] \in \mathbb{R}^{3 \times n}, \quad \mathbf{S} = \mathbf{N} \mathbf{C}_S^T,$$

the vector \mathbf{n}_k denotes the k th node in the structure and matrix $\mathbf{C}_S \in \mathbb{R}^{m \times n}$ is the string connectivity matrix. The connectivity matrix is made of ones, zeros and

minus ones representing the string/rod connections within the structure (see [7] and the next illustrative example). It follows that

$$\mathbf{T} = -\mathbf{S}\Gamma, \quad \mathbf{F} = [\mathbf{f}_1 \cdots \mathbf{f}_n] = \mathbf{T}\mathbf{C}_S = -\mathbf{N}\mathbf{C}_S^T\Gamma\mathbf{C}_S$$

where we made use of the diagonal matrix Γ which contains the *force densities*

$$\gamma_i := \max\{0, -\kappa_i(\|\mathbf{s}_i\| - l_i^0)\}/\|\mathbf{s}_i\|\}$$

on its diagonal. The matrix \mathbf{F} is the matrix of nodal forces.

2.3 Generalized Forces and Torques

Equations of motion must be written in terms of the configuration matrix, whereas the forces in the previous section are expressed at the nodes. Hence, we need to express forces in term of the configuration matrix coordinates. That is, to compute *generalized forces*. Because of linearity of (4) the matrix of generalized forces is computed as (see [7] for details)

$$\mathbf{F}_Q = -(\mathbf{Q}\Psi^T + \mathbf{Y})\mathbf{C}_S^T\Gamma\mathbf{C}_S\Psi. \quad (5)$$

A closer look at (5) reveals that

$$\mathbf{F}_Q = [\mathbf{f}_r \mathbf{f}_b], \quad \mathbf{f}_r = \sum_{i=1}^n \mathbf{f}_i, \quad \mathbf{f}_b = \sum_{i=1}^n \eta_i \mathbf{f}_i.$$

where \mathbf{f}_r is simply the sum of all forces applied to the rods and \mathbf{f}_b is related to the sum of the torques on the rod [7].

2.4 Equations of Motion

In Sect. 4 we will deal with tensegrity structures in which some or all the rods may have kinematic constraints. In such structures it may be advantageous to make \mathbf{r} to not coincide with the location of the center of mass of the rod. For this reason we derive the equations of motion using Lagrangian methods, whose full potential we explore in Sect. 4 (see [7] for an alternative Newtonian derivation). In this section we consider the simpler case in which \mathbf{r} is assumed to coincide with the center of mass. This assumption will be removed later.

Consider the Lagrangian function

$$L = T - V - \frac{J\xi}{2} (\mathbf{b}^T \mathbf{b} - 1), \quad (6)$$

where ξ is the Lagrange multiplier responsible for enforcing the constraint that vector \mathbf{b} must remain unitary (1) and where V is some appropriately defined potential

function. Assume that \mathbf{r} coincides with the location of the center of mass of the rod, i.e. $f = 0$. Following standard derivations as shown in [7] we arrive at the equations of motion

$$m\ddot{\mathbf{r}} = \mathbf{f}_r, \quad J\ddot{\mathbf{b}} = \mathbf{f}_b - J\xi\mathbf{b}, \quad \mathbf{b}^T\mathbf{b} - 1 = 0. \quad (7)$$

where \mathbf{f}_r and \mathbf{f}_b are the vector of generalized forces acting on the rod written in the coordinates \mathbf{q} (see Sect. 2.3). We can avoid the explicit calculation of the Lagrange multiplier ξ by using the constraint (1). As shown in [7],

$$\xi = (\|\dot{\mathbf{b}}\|/\|\mathbf{b}\|)^2 + J^{-1}\mathbf{b}^T\mathbf{f}_b/\|\mathbf{b}\|^2.$$

Substituting ξ on (7) produces the rotational equations of motion

$$\ddot{\mathbf{b}} = J^{-1}(\mathbf{I} - (\mathbf{b}\mathbf{b}^T)/\|\mathbf{b}\|^2)\mathbf{f}_b - (\|\dot{\mathbf{b}}\|/\|\mathbf{b}\|)^2\mathbf{b}. \quad (8)$$

These equations are the basis for the dynamics of class 1 tensegrity systems to be discussed in Sect. 3.

3 Class 1 Tensegrity Systems

A tensegrity system where no rigid rod is in contact with any other rod is said to be a *Class 1 Tensegrity System*. Such systems are the ones often identified with the concept of tensegrity as originally defined by Kenneth Snelson and Buckminster Fuller [3, 6]. The equations of motion developed in the previous section can be extended to cope with general class 1 tensegrity systems in a straightforward way. Instead of presenting a lengthy derivation of the equations of motion for general class 1 tensegrity systems, we shall limit ourselves to indicate what are the steps needed to undertake such generalization based on the material presented so far.

Because in class 1 tensegrity systems no rods touch each other, there are no extra constraints to be taken into consideration beyond the ones already dealt with in Sect. 2. In fact, using the Lagrangian approach, all that is needed to derive equations of motion for a class 1 tensegrity System with N rods is to define the combined Lagrangian

$$L = \sum_{j=1}^N L_j \quad (9)$$

where each L_j is a Lagrangian function, as in (6), written for each rod $j = 1, \dots, N$, and following the procedure outlined in Sect. 2.4 for enforcing the individual rod constraints and deriving the equations of motion. With that in mind define the configuration matrix

$$\mathbf{Q} = [\mathbf{R} \ \mathbf{B}] \in \mathbb{R}^{3 \times 2N} \quad (10)$$

where

$$\mathbf{R} = [\mathbf{r}_1 \cdots \mathbf{r}_N], \quad \mathbf{B} = [\mathbf{b}_1 \cdots \mathbf{b}_N] \in \mathbb{R}^{3 \times N}. \quad (11)$$

Note that, in the absence of constraints, the (4) is still valid provided an appropriate matrix $\Psi \in \mathbb{R}^{n \times 2N}$ is constructed. Likewise, generalized forces are easily computed using (5). A surprisingly compact matrix expression for the resulting equations of motion is possible by combining (7) and (8) as in the following theorem. See [7] for a detailed proof.

Theorem 3.1. Consider a Class 1 Tensegrity System with N rigid fixed length rods. Define the configuration matrix (10)

$$\mathbf{Q} = [\mathbf{R} \ \mathbf{B}] \in \mathbb{R}^{3 \times 2N}$$

where the columns of \mathbf{R} describe the center of mass of the N rods and the columns of \mathbf{B} describe the rod vectors. Let $\Psi \in \mathbb{R}^{n \times 2N}$ and $\mathbf{Y} \in \mathbb{R}^{3 \times n}$ be constant matrices that relate the $n \geq 2N$ nodes of the structure with the configuration matrix through (4)

$$\mathbf{N} = \mathbf{Q} \Psi^T + \mathbf{Y}, \quad \mathbf{N}, \mathbf{Y} \in \mathbb{R}^{3 \times n}, \quad \Psi \in \mathbb{R}^{n \times 2}.$$

The dynamics of such class 1 tensegrity system satisfy

$$(\ddot{\mathbf{Q}} + \mathbf{Q} \Xi) \mathbf{M} = \mathbf{F}_{\mathbf{Q}} \quad (12)$$

where

$$\mathbf{M} = \text{diag}[m_1, \dots, m_N, J_1, \dots, J_N], \quad \Xi = \text{diag}[0, \dots, 0, \xi_1, \dots, \xi_N].$$

The Lagrange multipliers ξ_i , $i = 1, \dots, N$ are computed by

$$\xi_i = (\|\mathbf{b}_i\| / \|\mathbf{b}_i\|)^2 + J_i^{-1} \mathbf{b}_i^T \mathbf{f}_{\mathbf{b}_i} / \|\mathbf{b}_i\|^2 \quad (13)$$

where $\mathbf{f}_{\mathbf{b}_i}$ are columns of the matrix $\mathbf{F}_{\mathbf{B}}$ which is part of the matrix of generalized forces

$$\mathbf{F}_{\mathbf{Q}} = [\mathbf{F}_{\mathbf{R}} \ \mathbf{F}_{\mathbf{B}}] \in \mathbb{R}^{3 \times 2N}$$

which is computed by (5)

$$\mathbf{F}_{\mathbf{Q}} = [\mathbf{W} - (\mathbf{Q} \Psi^T + \mathbf{Y}) \mathbf{C}_S^T \Gamma \mathbf{C}_S] \Psi$$

where $\mathbf{C}_S \in \mathbb{R}^{m \times n}$ is the string connectivity matrix, and the columns of matrix

$$\mathbf{W} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \cdots \ \mathbf{w}_{2N}]$$

are external forces where \mathbf{w}_i acts on node \mathbf{n}_i .

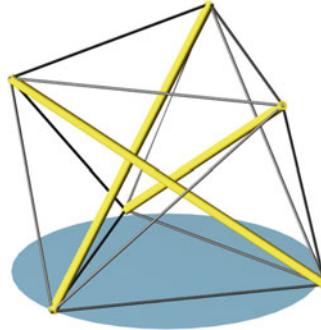


Fig. 4 A class 1 tensegrity prism with three rods and 12 strings

The following example illustrates the use of the expressions introduced in the theorem above.

Example 3.2. Consider the tensegrity prism depicted in Fig. 4. This structure has six nodes, three rods and 12 strings. Let the node matrix be

$$\mathbf{N} = [\mathbf{n}_1 \ \mathbf{n}_2 \ \mathbf{n}_3 \ \mathbf{n}_4 \ \mathbf{n}_5 \ \mathbf{n}_6],$$

where each pair of nodes is a pair of bottom and top nodes on a rod. That is

$$\mathbf{B} = [\ell_1^{-1}(\mathbf{n}_1 - \mathbf{n}_2) \ \ell_2^{-1}(\mathbf{n}_3 - \mathbf{n}_4) \ \ell_3^{-1}(\mathbf{n}_5 - \mathbf{n}_6)].$$

Assuming that the mass m_j of the j th rods is uniformly distributed then the center of each rods is its center of mass

$$\mathbf{R} = \frac{1}{2} [\mathbf{n}_1 + \mathbf{n}_2 \ \mathbf{n}_3 + \mathbf{n}_4 \ \mathbf{n}_5 + \mathbf{n}_6].$$

The nodes can be retrieved from the configuration matrix $\mathbf{Q} = [\mathbf{R} \ \mathbf{B}]$ through (4) with

$$\Psi = \frac{1}{2} \begin{bmatrix} 2 & 0 & 0 & \ell_1 & 0 & 0 \\ 2 & 0 & 0 & -\ell_1 & 0 & 0 \\ 0 & 2 & 0 & 0 & \ell_2 & 0 \\ 0 & 2 & 0 & 0 & -\ell_2 & 0 \\ 0 & 0 & 2 & 0 & 0 & \ell_3 \\ 0 & 0 & 2 & 0 & 0 & -\ell_3 \end{bmatrix}$$

and, because of the uniform mass distribution and the choice of \mathbf{R} , we have that $f_j = 0$ and

$$J_j = \frac{1}{12} m_j \ell_j^2, \quad j = \{1, 2, 3\}.$$

The string connectivity is

$$\mathbf{C}_S = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \end{bmatrix}.$$

With this information one can write the equations of motion (12).

4 Class k Tensegrity Systems

So far we have focused on class 1 tensegrity systems, where the rigid rods are never in contact. Other systems of rods and strings in which the rods may be in contact are also tensegrity systems (more generally, a system of rigid bodies and tensile elements). A tensegrity system in which as many as $k \geq 1$ rigid rods are in contact is a *class k tensegrity system*. See [7] for a detailed discussion. Dynamic models for class k tensegrity systems are a lot more complicated than the ones we presented for Class 1 tensegrity Systems. Yet, such systems still present relevant structure that may help simplify the derivation and computation of the equations of motion. The main difficulty arises now from the fact that rods may be in contact (through ball joints) and kinematic constraints other than the preservation of length of the bar vector have to be enforced on the equations of motion. These constraints will vary from system to system, which makes difficult the obtention of a general formula. Instead, we shall discuss perhaps the simplest class k tensegrity system, a class 2 cable as illustrated in Fig. 5. The intent is to illustrate the difficulties of such models and the benefits that can be obtained by exploring the problem structure by using the approach of the previous sections.

4.1 A Class 2 Tensegrity Cable Model

Our class 2 tensegrity cable is a chain system composed of N interconnected rigid rods used to model the behavior of a long and heavy cable. We assume, as before for simplicity, that the mass of each rigid rod is distributed uniformly. For reasons that will become clear in the sequel, we choose the coordinate \mathbf{r}_i to be the near end

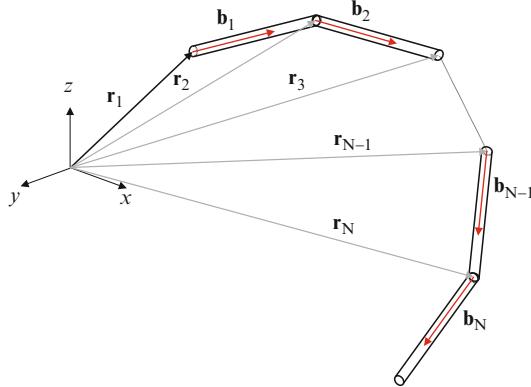


Fig. 5 Illustration of a chain system of N interconnected rigid rods

of each rod, which implies $f_i = m_i \ell_i / 2$, $J_i = m_i \ell_i^2 / 3$, for $i = 1, \dots, N$. This setup is illustrated in Fig. 5. The vector \mathbf{r}_1 describes the motion at the first node of the first rod, that is the starting end of the cable. In this section we assume that the acceleration of such node is to be prescribed. That is, we assume that \mathbf{r}_1 and all its derivatives are known *a priori* and are independent of the motion of the cable. The cable can then be thought as being *suspended* by \mathbf{r}_1 , the problem of interest being then to describe the dynamics of the remaining elements of the cable. We assume that the cable is immersed in a constant gravitational field. This setup reflects, for instance, a cable that is being towed by a massive boat in a marine application or by an excitation system in the laboratory [11].

With the above discussion in mind we choose as configuration the matrix

$$\mathbf{Q} = [\mathbf{b}_1 \cdots \mathbf{b}_N]. \quad (14)$$

Vector \mathbf{r}_1 is not present because it is assumed to be known *a priori*. This assumption can be removed without difficulty. Clearly the challenge in modeling this class 2 tensegrity system is taking into account the constraints that the rods are connected (by ball joints in our present model). Because of our choice of configuration \mathbf{Q} , such constraints are linear. Enforcing the constraint that the far and near ends of successive rods are connected amounts at enforcing

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \ell_i \mathbf{b}_i, \quad i = 1, \dots, N-1,$$

which are clearly linear functions of \mathbf{Q} . Furthermore, these equations can be solved recursively yielding

$$\mathbf{r}_i = \mathbf{r}_1 + \sum_{j=1}^{i-1} \ell_j \mathbf{b}_j, \quad i = 1, \dots, N. \quad (15)$$

For this reason it will be possible to explicitly enforce these constraints when writing the Lagrangian function, as opposed to using Lagrange multipliers. This is the route taken for the rest of this session. In systems where loops (closed kinematic chain) are

present the situation is more complicated because it may not be possible to explicitly compute a reduced configuration vector as done above.

In order to derive the equations of motion we assemble a combined Lagrangian function as in (9) where each L_i is given by (6). The terms of L_i are computed by adding the kinetic energy of each rigid rod, from (3),

$$\begin{aligned} T_i &= \frac{1}{2}m_i \dot{\mathbf{r}}_i^T \dot{\mathbf{r}}_i + f_i \dot{\mathbf{r}}_i^T \dot{\mathbf{b}}_i + \frac{1}{2}J_i \dot{\mathbf{b}}_i^T \dot{\mathbf{b}}_i, \\ &= \frac{1}{2}m_i \left(\dot{\mathbf{r}}_1 + \sum_{j=1}^{i-1} \ell_j \dot{\mathbf{b}}_j \right)^T \left(\dot{\mathbf{r}}_1 + \sum_{j=1}^{i-1} \ell_j \dot{\mathbf{b}}_j \right) + \left[f_i \left(\dot{\mathbf{r}}_1 + \sum_{j=1}^{i-1} \ell_j \dot{\mathbf{b}}_j \right) + \frac{1}{2}J_i \dot{\mathbf{b}}_i \right]^T \dot{\mathbf{b}}_i, \end{aligned}$$

where we made use of (15) explicitly, as anticipated. In this derivation we chose to leave the expression T_i in terms of the individual vector components as opposed to the configuration matrix \mathbf{Q} . This will be reversed later.

In order to account for the constant gravitational field in which the cable is immersed, the potential energy of the i th rod is computed as

$$\begin{aligned} V_i &= g \int_0^1 \rho_i(\mu) \mathbf{v}_i(\mu)^T \mathbf{e}_z d\mu \\ &= m_i g / \ell_i \int_0^{\ell_i} (\mathbf{r}_i + \eta \mathbf{b}_i)^T \mathbf{e}_z d\eta = m_i g \left(\mathbf{r}_1 + \sum_{j=1}^{i-1} \ell_j \mathbf{b}_j + \frac{1}{2} \ell_i \mathbf{b}_i \right)^T \mathbf{e}_z, \quad (16) \end{aligned}$$

where $\mathbf{e}_z \in \mathbb{R}^3$ is a unit vector in the positive z -direction. The last two expressions follow from the assumption of uniform mass distribution and the choice of configuration. The complete Lagrangian function is then assembled

$$L = \sum_{i=1}^N \left(T_i - V_i - \frac{J_i \xi_i}{2} (\mathbf{b}_i^T \mathbf{b}_i - 1) \right)$$

and the equations of motion are derived by computing the Euler–Lagrange equation which, for convenience, will have the left and right hand sides split as

$$\ddot{\mathbf{Q}} \mathbf{M} + \ddot{\mathbf{r}}_1 \mathbf{h}^T = \frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{Q}}} = \frac{\partial L}{\partial \mathbf{Q}} = \mathbf{F}_{\mathbf{Q}} - \mathbf{Q} \boldsymbol{\Xi} \mathbf{J}, \quad (17)$$

where

$$\mathbf{J} = \text{diag}[J_1, \dots, J_N], \quad \mathbf{F}_{\mathbf{Q}} := \sum_{j=1}^N -\frac{\partial V_j}{\partial \mathbf{Q}} = \mathbf{e}_z \mathbf{g}^T.$$

As in class 1 tensegrity systems, the *mass matrix* \mathbf{M} is a constant $N \times N$ matrix, with entries to be defined in (20). Vectors \mathbf{g} and \mathbf{h} are N -dimensional with entries to be defined in (22). Compare how (17) is similar to (12). The main differences are that \mathbf{M} is now a full matrix, as opposed to diagonal, and, of course, we should expect that

the calculation of Ξ will be much more complex than (13). It is to the calculation of Ξ that we now turn our attention.

In order to compute Ξ , using the properties of Kronecker¹ products [2], let us temporarily rearrange the equations of motion (17) in vector form

$$\begin{bmatrix} \tilde{\mathbf{M}} & \mathbf{Y} \\ \mathbf{Y}^T & \mathbf{0} \end{bmatrix} \begin{pmatrix} \ddot{\mathbf{q}} \\ \xi \end{pmatrix} = \begin{pmatrix} \mathbf{u} \\ -\mathbf{y} \end{pmatrix}, \quad (18)$$

where

$$\mathbf{q} = \text{vec}(\mathbf{Q}), \quad \tilde{\mathbf{M}} = \mathbf{M} \otimes \mathbf{I}_3, \quad \mathbf{u} = \text{vec}(\mathbf{U}), \quad \mathbf{U} := \mathbf{F}_{\mathbf{Q}} - \ddot{\mathbf{r}}_1 \mathbf{h}^T = \mathbf{e}_z \mathbf{g}^T - \ddot{\mathbf{r}}_1 \mathbf{h}^T$$

and

$$\mathbf{Y} = [J_1 \mathbf{e}_1 \otimes \mathbf{b}_1 \cdots J_N \mathbf{e}_N \otimes \mathbf{b}_N], \quad \mathbf{y} = \begin{pmatrix} \dot{\mathbf{b}}_1^T \mathbf{b}_1 \\ \vdots \\ \dot{\mathbf{b}}_N^T \mathbf{b}_N \end{pmatrix}.$$

The last N equations of (18) come from differentiating the bar length constraint $\|\mathbf{b}_i\|^2 = 1$ twice with respect to time. As in the case of Class 1 Tensegrity systems, we can solve for ξ explicitly as follows

$$\xi = (\mathbf{Y}^T \tilde{\mathbf{M}}^{-1} \mathbf{Y})^{-1} (\mathbf{Y}^T \tilde{\mathbf{M}}^{-1} \mathbf{u} + \mathbf{y}).$$

Online computation of ξ is then dominated by the factorization of the $N \times N$ positive definite matrix $\mathbf{Y}^T \tilde{\mathbf{M}}^{-1} \mathbf{Y}$, because \mathbf{M} (hence $\tilde{\mathbf{M}}$) is constant and can be factored offline. Evaluation of all terms involved in this computation can be further sped up by exploring the structure of the involved matrices. Indeed, compute

$$\mathbf{V} = \mathbf{U} \mathbf{M}^{-1} = \mathbf{e}_z (\mathbf{M}^{-1} \mathbf{g})^T - \ddot{\mathbf{r}}_1 (\mathbf{M}^{-1} \mathbf{h})^T, \quad \mathbf{Z} = \mathbf{Q} \mathbf{J} = [J_1 \mathbf{b}_1 \cdots J_N \mathbf{b}_N]$$

One can verify that

$$\tilde{\mathbf{M}}^{-1} \mathbf{u} = (\mathbf{M}^{-1} \otimes \mathbf{I}_3) \mathbf{u} = \text{vec}(\mathbf{v}), \quad \mathbf{Y}^T \tilde{\mathbf{M}}^{-1} \mathbf{u} = \begin{pmatrix} \mathbf{z}_1^T \mathbf{v}_1 \\ \vdots \\ \mathbf{z}_N^T \mathbf{v}_N \end{pmatrix}$$

where \mathbf{v}_i (or \mathbf{z}_i) denotes the i th column of matrix \mathbf{V} (or \mathbf{Z}). By denoting by $\mathbf{1}$ a vector with all entries equal to 1 and the entrywise product of two matrices by the symbol ‘ \circ ’ (known as the Hadamard product) we have that

$$\mathbf{Y}^T \tilde{\mathbf{M}}^{-1} \mathbf{u} + \mathbf{y} = (\mathbf{Z} \circ \mathbf{V})^T \mathbf{1} + \mathbf{y}.$$

¹ The Kronecker product of two $n \times n$ matrices \mathbf{A} and \mathbf{B} , i.e. $\mathbf{A} \otimes \mathbf{B}$, is an $n^2 \times n^2$ matrix composed of $n \times n$ blocks of matrices of the type $A_{ij}\mathbf{B}$.

Using similar ideas one can show that

$$\mathbf{Y}^T \tilde{\mathbf{M}}^{-1} \mathbf{Y} = \mathbf{M}^{-1} \circ (\mathbf{Z}^T \mathbf{Z}).$$

The final result is summarized in the following theorem.

Theorem 3.2. Consider the Class 2 Tensegrity Systems Cable Model with N rigid fixed length rods connected as in Figure 5, immersed in a constant gravitational field in the \mathbf{e}_z direction. Let the trajectory of the node \mathbf{r}_1 be prescribed a priori, independently of the rest of the system. Define the configuration matrix $\mathbf{Q} = \mathbf{B} \in \mathbb{R}^{3 \times N}$, as in (14), where the columns of \mathbf{B} describe the rod vectors.

The dynamics of this class 2 tensegrity systems cable model satisfy

$$\ddot{\mathbf{Q}} \mathbf{M} + \mathbf{Q} \Xi \mathbf{J} = \mathbf{F}_{\mathbf{Q}} - \ddot{\mathbf{r}}_1 \mathbf{h}^T, \quad (19)$$

where $\mathbf{M} = (\mathbf{M}_{ij})$ is a constant $N \times N$ matrix with entries

$$\mathbf{M}_{ij} = \begin{cases} \ell_j(f_i + \ell_i \sum_{k=i+1}^N m_k), & i > j, \\ J_i + \ell_i^2 \sum_{k=i+1}^N m_k, & i = j, \\ \mathbf{M}_{ji}, & i < j. \end{cases} \quad (20)$$

The generalized force $\mathbf{F}_{\mathbf{Q}}$ is given by

$$\mathbf{F}_{\mathbf{Q}} := \sum_{j=1}^N -\frac{\partial V_j}{\partial \mathbf{Q}} = \mathbf{e}_z \mathbf{g}^T. \quad (21)$$

The entries of the N -dimensional vectors $\mathbf{g} = (\mathbf{g}_i)$ and $\mathbf{h} = (\mathbf{h}_i)$ are

$$\mathbf{g}_i = -g \ell_i \left(\frac{1}{2} m_i + \sum_{j=i+1}^N m_j \right), \quad \mathbf{h}_i := f_i + \ell_i \sum_{k=i+1}^N m_k. \quad (22)$$

and \mathbf{J} and Ξ are the diagonal matrices

$$\mathbf{J} = \text{diag}[J_1, \dots, J_N], \quad \Xi = \text{diag}[\xi_1, \dots, \xi_N].$$

The Lagrange multipliers ξ_i , $i = 1, \dots, N$ are computed by

$$\xi = [\mathbf{M}^{-1} \circ (\mathbf{Z}^T \mathbf{Z})]^{-1} [(\mathbf{Z} \circ \mathbf{V})^T \mathbf{1} + \mathbf{y}], \quad (23)$$

where the matrices \mathbf{V} and \mathbf{Z} are given by

$$\mathbf{V} = \mathbf{e}_z (\mathbf{M}^{-1} \mathbf{g})^T - \ddot{\mathbf{r}}_1 (\mathbf{M}^{-1} \mathbf{h})^T, \quad \mathbf{Z} = \mathbf{Q} \mathbf{J} \quad (24)$$

and $\mathbf{y} = (\mathbf{y}_i)$ is a N -dimensional vector with entries $\mathbf{y}_i := \|\dot{\mathbf{b}}_i\|^2$.

The next example illustrates the theorem in the case of a “double” pendulum.

Example 3.3. Let $N = 2$, $m_1 = m_2 = 1$ and $\ell_1 = \ell_2 = \ell$. Hence for uniformly distributed mass rods $f_1 = f_2 = \ell/2$ and $J_1 = J_2 = \ell^2/3$. Then

$$\mathbf{M} = \ell^2 \begin{bmatrix} 4 & 1 \\ \frac{3}{2} & \frac{2}{3} \\ 1 & 1 \\ \frac{1}{2} & \frac{3}{2} \end{bmatrix}, \quad \mathbf{g} = g\ell \begin{pmatrix} \frac{3}{2} \\ 1 \\ \frac{1}{2} \end{pmatrix}, \quad \mathbf{h} = \ell \begin{pmatrix} \frac{3}{2} \\ 1 \\ \frac{1}{2} \end{pmatrix}.$$

5 Concluding Remarks

The dynamic models developed in this chapter, i.e. (12) and (19), have remarkable structure. First they have compact matrix forms that facilitate computation, without the need to recourse to Kronecker products, as in traditional dynamic models. Second, they are ordinary differential equations even though the systems are not described in minimal coordinates. By parametrizing the configuration matrix directly in terms of the components of the rod vectors, the usual transcendental nonlinearities involved with the use of angles, angular velocities and coordinate transformations are avoided. Indeed, the absence of trigonometric functions in this formulation leads to a simplicity in the analytical form of the dynamics. The actual number of degrees of freedom for each rod is 5, whereas, the model (12) has as many equations as required for 6 degrees of freedom for each rod. That is, the equations are a non-minimal realization of the dynamics. The *mathematical structure* of the equations are simple, however. This allows much easier integration of structure and control design, since the control variables (string force densities) appear linearly, and the simple structure of the nonlinearities can be exploited when controlling the system. The interested reader is referred to [7, 12] for further discussions on these topics.

The proposed structure also facilitates computation. Indeed, the differential equations (12) and (19) can be integrated using standard explicit methods, e.g. Runge–Kutta. In [4] we provide a detailed analysis in which the numerical stability of such methods is investigated in the case of a single rod. In a nut shell, the conclusion is that with a minor modification, namely the periodic normalization of the unitary rod vectors in \mathbf{B} , the numerical integration is very reliable and accurate. Indeed, we have very successfully implemented numerical code that can integrate tensegrity systems with tens of thousands of rods. This code has been used, for instance to integrate a Class 1 Tensegrity model of the dynamic behavior of the cytoskeleton of erythrocytes or red-blood cells (see [5] for details).

References

1. Fuller, R.B.: US patent 3 063 521 Tensile Integrity Structures. United States Patent Office (1959)
2. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge, UK (1985)

3. Alvani, H.: Origins of tensegrity: views of Emmerich, Fuller and Snellson. *International Journal of Space Structures* **11**, 27–55 (1996)
4. de Oliveira, M.C.: Dynamics of systems with rods. In: Proceedings of the 45th IEEE Conference on Decision and Control Conference, pp. 2326–2331. San Diego, CA (2006)
5. de Oliveira, M.C., Vera, C., Valdez, P., Sharma, Y., Skelton, R.E., Sung, L.A.: Network nanomechanics of the erythrocyte membrane skeleton in equibiaxial deformation. To appear in the *Biophysical Journal*
6. Sadao, S.: Fuller on tensegrity. *International Journal of Space Structures* **11**, 37–42 (1996)
7. Skelton, R.E., de Oliveira, M.C.: Tensegrity Systems. Springer, Berlin (2009). ISBN: 978-0387742410
8. Skelton, R.E., Pinaud, J.P., Mingori, D.L.: Dynamics of the shell class of tensegrity structures. *Journal of The Franklin Institute-Engineering And Applied Mathematics* **338**(2-3), 255–320 (2001)
9. Snellson, K.: US patent 3 169 611 Continuous Tension Discontinuous Compression Structures. United States Patent Office (1965)
10. Snellson, K.: Letter to R. Motro. originally published in International Journal of Space Structures (1990). URL <http://www.grunch.net/snellen/rmoto.html>
11. Wroldsen, A.S.: Modelling and control of tensegrity structures. Ph.D. thesis, Department of Marine Technology, Norwegian University of Science and Technology (2007)
12. Wroldsen, A.S., de Oliveira, M.C., Skelton, R.E.: Modelling and control of non-minimal non-linear realisations of tensegrity systems. *International Journal of Control* **82**(3), 389–407 (2009). DOI {10.1080/00207170801953094}

Modeling a Complex Aero-Engine Using Reduced Order Models

Xuewu Dai, Timofei Breikin, Hong Wang, Gennady Kulikov, and Valentin Arkov

1 Introduction

Gas turbine engines are widely used in many industrial applications and engine condition monitoring is a vital issue for the aircraft in-service use and flight safety. From the variety of condition monitoring methods, the model-based approach is perhaps the most promising for real-time condition monitoring. This approach can predict the engine characteristics at the expense of “algorithmic redundancy” and requires real-time simulation. The main obstacles for using full thermodynamic models in the engine condition monitoring schemes are high computing load, and inability to incorporate unforeseen changes.

It is clear that in the on-board condition monitoring of gas turbine engines, a reduced order model is required due to the limited on-board computation resources. This model should be individual for each particular engine. This type of models cannot be obtained based only on full thermodynamic model of the engine. Identification of each individual engine model parameters based on experimental data is required. Simplification of the model, the model order reduction, introduces additional

X. Dai

School of Electronic and Information Engineering, Southwest University, Chongqing 400715, China

and

He was with Control Systems Centre, School of Electrical and Electronic Engineering, University of Manchester, Manchester M60 1QD, UK

e-mail: dxw.dai@gmail.com

T. Breikin and H. Wang

Control Systems Centre, School of Electrical and Electronic Engineering, University of Manchester, Manchester, UK, M60 1QD

e-mail: t.breikin@manchester.ac.uk; hong.wang@manchester.ac.uk

G. Kulikov and V. Arkov

Department of Automated Control Systems, Ufa State Aviation Technical University, K. Marx Street 12, Ufa 450000, Russia

e-mail: kulikov@asu.ugatu.ac.ru; arkov@asu.ugatu.ac.ru

problems during the model parameters identification from real engine data. Furthermore, in order to retain the fault information in the residual, the model's long term prediction (simulation) performance is of the main interest, rather than the one-step a-head prediction. It is particularly important for detecting incipient faults. Although Output Error (OE) model shows better simulation performance than other models (e.g., ARX, ARMAX, etc.), the dependency within the output errors presents a number of challenges. This dependency makes LSE method biased leading to a poor long-term prediction. As described literature, even if the model itself is linear, the objective function of OE model is highly nonlinear and some kind of iterative nonlinear optimization is needed [15].

This chapter describes the modeling phase of the engine model-based condition monitoring. The aim is to find a fast OE model identification algorithm suitable for the limited on-board computation facilities of the on-board condition monitoring system. The chapter is organized as follows. Section 2 provides an introduction to Gas Turbine System. Section 3 discusses the performance criterion and structure selection during data driven reduced order modeling. Section 4 defines the objective function for OE modeling and presents the development of DNLS (Dynamic Nonlinear Least Squares) identification algorithm. In order to accelerate the identification speed, two techniques are employed. The first one is Iterative calculation of the gradient (the Jacobian). The dependency within output errors is taken into account by adding an weighted sum of past gradients to the current Jacobian matrix. Secondly Hessian approximation is used. In order to further accelerate the convergence speed at a relatively low computation cost, the second order Hessian matrix is used and approximated by the first order information. Finally, the potential of the proposed DNLS for reduced order modeling of a gas turbine engine is illustrated in Sect. 5. Data gathered at an aero engine test-bed serves as the test vehicle to demonstrate the improvements in convergence speed and the reduction of computation cost.

2 Gas Turbine System

A generalized schematic of a typical reheat bypass turbojet engine is shown in Fig. 1, where the main units of the flowing part of the engine are as follows: I – fan or low pressure (LP) compressor; II – high-pressure (HP) compressor; III – bypass duct; IV – main combustion chamber; V – mixing chamber; VI – afterburner; VII – variable jet nozzle.

The possible factors shown in Fig. 1 that affects the engine operation are: 1 – variable stator vanes of fan ($\alpha_{SV,f}$); 2 – air bypass from first stages of compressor into bypass duct (m); 3 – variable stator vanes of compressor ($\alpha_{SV,c}$); 4 – variable radial clearance in final stages of compressor (Δr_c); 5 – fuel supply in combustion chamber (W_f); 6 – fuel supply in combustion chamber of bypass duct ($W_{f,II}$); 7 – variable radial clearance in HP turbine ($\Delta r_{t,HP}$); 8 – variable stator vanes of fan turbine ($\alpha_{SV,ft}$); 9 – variable radial clearance in fan turbine (Δr_{ft}); 10 – mixing area

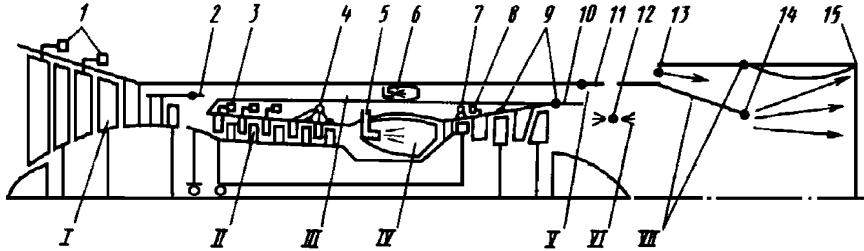


Fig. 1 Generalized schematic of reheat bypass twin-shaft turbo jet with variable working process

(A_{mix}); 11 – system of reverse thrust (F_{rev}); 12 – fuel supply in afterburner ($W_{f.ab}$); 13 – air supply in ejector nozzle (n_{ej}); 14 – critical nozzle area (A_{cr}); 15 – output nozzle area (A_n). The main control factors here are fuel flow into the main combustion chamber W_f and afterburner $W_{f.ab}$, critical nozzle area A_{cr} and output nozzle area A_n . In the following subsections nonlinear static and dynamic models of the engine are presented to demonstrate their complexity. Detailed description of the models parameters and variables as well as the nomenclature used can be found in [11].

2.1 Nonlinear Static Model of Gas Turbine

Under some assumptions [11], nonlinear equations describing operation of a gas turbine engine can be presented in the following form

$$\left. \begin{array}{l} T_H = 288 - 0.0065H \\ p_H = p_0(T_H/T_0)^{5.25} \end{array} \right\}, \text{ if } H < 11,000 \text{ m}$$

$$\left. \begin{array}{l} T_H = 216.5K \\ p_H = p_{11,000} \exp\left(\frac{11,000-H}{6,318}\right) \end{array} \right\}, \text{ if } H > 11,000 \text{ m}$$

$$\begin{aligned} T_{in}^* &= T_H(1 + 0.2M^2) \\ p_{in}^* &= \sigma_{in} p_H (1 + 0.2M^2)^{3.5} \\ n_{red} &= n \sqrt{288/T_H^*} \\ p_c^* &= p_g^*/\sigma_{comb} \end{aligned}$$

$$\begin{aligned} \pi_c^* &= p_c^* / p_{in}^* \\ W_{a.red} &= f(\pi_c^*, n_{red}) \\ W_a &= f(W_{a.red}, T_{in}^*, p_{in}^*) \\ \eta_c^* &= f(\pi_c^*, \eta_{red}) \end{aligned}$$

$$\begin{aligned} T_c^* &= f(\eta_c^*, \pi_c^*, T_{in}^*) \\ p_t^* &= p_{noz}^* / \sigma_{noz} \\ \pi_t^* &= p_g^* / p_t^* \\ \eta_{t.red} &= n / \sqrt{288/T_g^*} \end{aligned}$$

$$\begin{aligned} W_{g,red} &= f(\pi_t^*, n_{r,red}) \\ W_g &= f(W_{g,red}, T_g^*, p_g^*) \\ \eta_t^* &= f(\pi_t^*, n_{t,red}) \\ T_t^* &= f(T_g^*, \pi_t^*, \eta_t^*) \end{aligned}$$

$$\begin{aligned} W_n &= f(A_n, p_n^*, T_n^*) \\ N_t &= f(W_g, T_g^*, T_t^*) \\ N_c &= f(W_a, T_{in}, T_c^*) \end{aligned}$$

In addition, the following balance relationships, representing air/gas flow balance, thermal balance in the combustion chamber and jet nozzle and power balance between the compressor and turbine, should be taken into account.

$$\begin{aligned} W_g - (W_a + W_f) &= 0 \\ c_{p_a} T_c^* W_a + H_u \eta_g W_f - c_{p_g} T_g^* W_g &= 0 \\ W_n - W_g &= 0 \\ c_{p_g} T_c^* W_g - c_{p_g} T_n^* W_n &= 0 \\ \eta_{mech} N_t - N_c &= 0 \end{aligned}$$

2.2 Nonlinear Dynamic Model of Gas Turbine

For dynamic modeling, the system given in Sect. 2.1 should be complemented by the differential equation describing dynamics of the shaft rotation and differential equations accounting the accumulation of the gas energy and mass within major volumes of the flowing part of the engine [11]:

$$\begin{aligned} \dot{p}_g^* &= \frac{p_g^*}{T_g^*} \dot{T}_g^* + \frac{RT_g^*}{V_{comb}} (W_a + W_t - W_g) \\ \dot{T}_g^* &= \frac{1}{c_v m_c} [(c_{p_a} T_c^* W_a + H_u \eta_g W_f - c_{p_g} T_t^* W_g) - c_v T_g^* (W_a + W_t - W_g)] \\ \dot{p}_n^* &= \frac{p_n^*}{T_n^*} \dot{T}_n^* + \frac{RT_n^*}{V_n} (W_g - W_n) \\ \dot{T}_n^* &= \frac{1}{c_v m_n} [(c_{p_g} T_t^* W_g - c_{p_g} T_n^* W_n) - c_v T_n^* (W_g - W_n)] \\ \dot{n} &= \frac{\eta_{mech} N_t - N_c}{Jn(\pi/30)^2} \end{aligned} \quad (1)$$

Thus, the dynamic model of a gas turbine engine can be presented in state space form:

$$\begin{aligned} \dot{\mathbf{X}} &= \mathbf{F}_x(\mathbf{X}, \mathbf{U}, \mathbf{V}) \\ \dot{\mathbf{Y}} &= \mathbf{F}_y(\mathbf{X}, \mathbf{U}, \mathbf{V}) \end{aligned} \quad (2)$$

where \mathbf{X} is the state vector, \mathbf{U} is control vector, \mathbf{V} is vector of flight conditions, \mathbf{Y} is vector of output observable coordinates, \mathbf{F}_x and \mathbf{F}_y are nonlinear operators. This model is a *detailed Nonlinear Dynamic Model* making it possible to study the dynamic properties of a turbine engine.

The programs for gas turbine engine control are determined with respect to the aircraft requirements. For example, a controller may use programs for maintenance of maximum maneuverability of the aircraft, or for maximum efficiency, or specific programs for take-off and landing. Each program is determined via optimization of some criterion or a group of criteria. To implement model-based condition monitoring during these programs execution either very computationally extensive model (which still would not account specific characteristics of the engine) should be employed or alternatively on-line identification of the simplified reduced order models (specific for the engine and its operating conditions) can be used. The later one has obvious advantageous accounting specific characteristics of an individual engine, degradation of the engine parameters, and other uncertainties as external operating conditions, fuel quality, etc.

3 Problem Formulation for Reduced Order Data Driven Modeling

Although the detailed nonlinear dynamic model (1) is the best mathematical model to represent the gas turbine engines, it involves nonlinear equations and it is impractical to implement it in on-board condition monitoring due to the limited computation resources [11]. In addition, the model should be individual for each particular engine.

When the engine works around some operating point, the nozzle area and inlet guide vanes are fixed and the engine is controlled by the fuel flow. In this context, the reheat system is assumed inoperative, the compressor bleed valve closed. Due to the very small time constants in the thermodynamic processes and relatively large time constants in shaft speeds, the higher order nonlinear model can be reduced to a state model with the states of shaft speeds [5]. Hence, for on-board condition monitoring, the shaft speeds are regarded as the primary outputs, from which the internal engine pressures and thrust can be calculated. Because the engine performance is linked with the shaft speeds closely, in condition monitoring, much attention is paid to modeling the dynamic relationship between these shaft speeds and the fuel flow.

Furthermore, the reduced order model can be mathematically described by a set of SISO (Single-Input/Single-Output) transfer functions relating the shaft speeds n to fuel flow W_f . In the discrete time domain, the corresponding discrete-time specification of such a SISO transfer function is the difference equation:

$$y^*(t) = f(\theta^*, y^*(t-1), y^*(t-2), \dots, y^*(t-n), u^*(t-1), \dots, u^*(t-m)). \quad (3)$$

where $u^*(t)$ and $y^*(t)$ are the unmeasurable actual system input and output, respectively. And θ^* is the actual parameter vector associated with $f(\cdot)$.

Consider a system corrupted by input noise $d_u(t)$ and output noise $d_y(t)$, as shown in Fig. 2 where $d_u(t)$ and $d_y(t)$ denote either the measurement noise or the external disturbance, the measured input $u(t)$ and output $y(t)$ can be expressed as

$$u(t) = u^*(t) + d_u(t) \quad (4)$$

$$y(t) = y^*(t) + d_y(t) \quad (5)$$

In the case of modeling aero engines, $u(t)$ and $y(t)$ denotes the fuel flow W_f and the shaft speeds n of the engine model (1), respectively. Note that function $f(\theta^*)$ is a mathematical description of how the engine's fuel flow and shaft speed variables relate to each other.

A model is used to approximate the function $f(\theta^*)$ by $\hat{f}(\hat{\theta})$ and estimate the future output according the observed data sequence $[u(t), y(t)]$ up to time t . The system identification is such a technique to build the function structure of $\hat{f}(\hat{\theta})$ and estimate the parameters $\hat{\theta}$.

Not that the criterion on what is a good model is highly problem-dependent. It is useful to define and clarify what is a good model for condition monitoring, and select the right model structure for parameter identification. In terms of condition monitoring, the most important objective is to find a “good” reduced order model whose output error is robust to the disturbances d_u , d_y and sensitive to the faults f_a , f_s . Here, d_u and d_y denotes the input and output disturbances(noises), respectively. f_a denotes the actuator fault, and f_s denotes the sensor faults. They can be seen in Fig. 2.

3.1 Criterion Selection

In the initial stage, it was recognized that in the context of condition monitoring, the long-term prediction performance is more of interest than the one-step-ahead prediction. It is particularly true for detecting incipient faults. This section discusses why the long-term prediction performance is selected as the main criterion for condition monitoring.

Basically, there are two connection modes used for condition monitoring [8], [2]: (1) parallel connection for long-term prediction (which corresponds to the “Output Error” model or “Infinite Impulse Response” filter); (2) parallel-series connection for one-step ahead prediction (which corresponds to the “Equation Error” model).

In the parallel connection mode, as shown in Fig. 2a, the model runs in parallel to the system (thereby, forming the so-called *parallel model* or *IIR filter*) and the error is termed as *long-term prediction error*, because the prediction $\hat{y}(t)$ is calculated from the previous input u alone and has no direct link to the previous system output y . This error is also called *simulation error* [13], as the model is often used to simulate the actual plant.

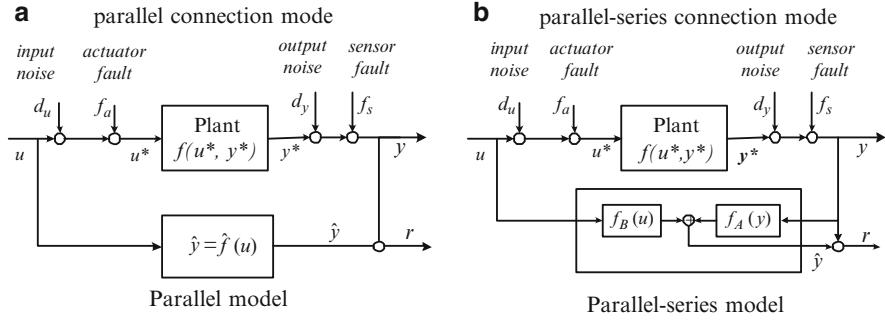


Fig. 2 Comparison of long-step prediction and one-step-ahead prediction

Figure 2b shows a block diagram of the parallel-series connection mode, where the model consists of two FIR (Finite Impulse Response) filters f_A and f_B . f_A is the *series part* running in series to the plant, and f_B the *parallel part* running in parallel to the plant. Compared to the parallel connection mode, the obvious difference is the feeding back of actual system output $y(t)$ to the model. Because $\hat{y}(t)$ depends on the last system outputs $y(t-1), y(t-2), \dots$ up to time $t-1$, it is called one-step-ahead prediction and the corresponding error is termed as *one-step-ahead prediction error*, or simply *prediction error* [13].

Although the parallel-series model is able to represent the dynamic characteristics of the system to some extent, it is not an exact duplication of the plant, because the plant output $y(t-i)$ is used to correct the model prediction $\hat{y}(t)$. Indeed, the parallel-series model is designed for fitting the output rather than simulating the plant.

On the other hand, the parallel model repeats the dynamic behavior of the plant in the same manner as the plant (that is to simulate the plant). Hence, it is easy to understand that the one-step-ahead predictor may not give the best fault detection performance, and the long-term prediction works better in terms of fault detection.

For illustrating this, both models are used to detect the actuator faults happening at 10 s. As shown in Fig. 3a1, b1, the abrupt and incipient faults are simulated and added to the input signal, respectively. Figure 3a2, b2 shows the residuals given by the parallel-series model, where no obvious changes can be seen when faults happening. However, the magnitude changes can be easily detected from the residuals of the parallel model, as shown in Fig. 3a3, b3. Moreover, the shapes of the residuals also reflect the pattern of faults, which may be useful for fault estimation

These phenomena can be interpreted as that, in the parallel-series connection, the system output is fed into the model to correct the model prediction. When some fault emerges in the system, the faulty information contained in the system output is also passed to the model, and the model prediction is adjusted to match the faulty output. Thus, the effects of faults are cancelled in the residual due to the minus operation. Hence, it fails to detect the fault by checking the residual. The residuals in the parallel connection, however, depend on the input alone and are unaffected by the faulty output. Thus the fault information remains in the residual.

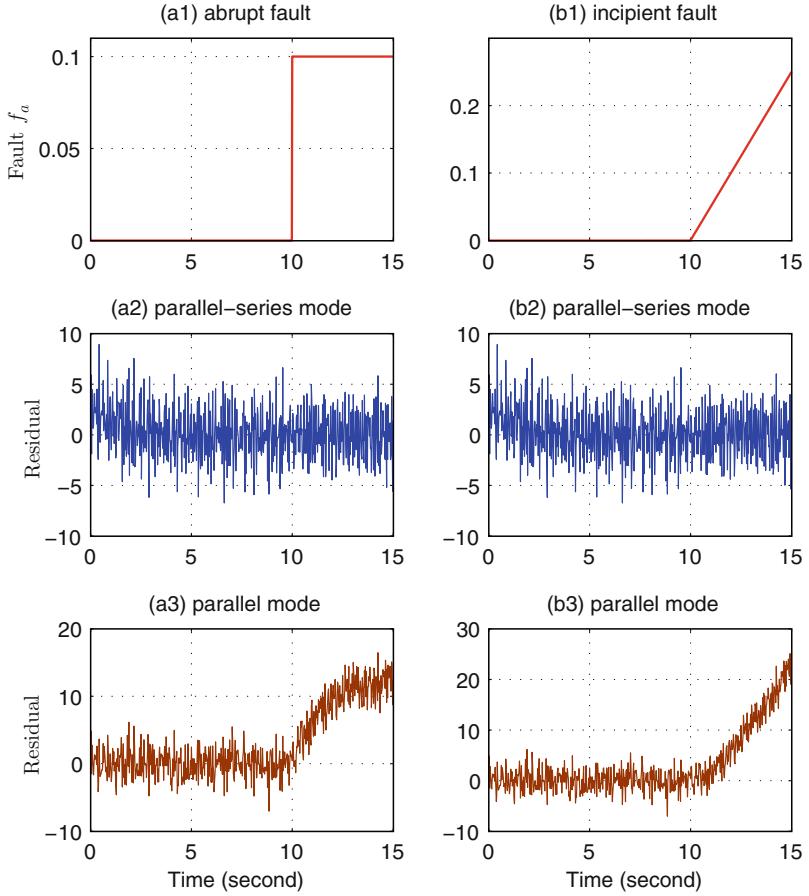


Fig. 3 Residuals of long-term prediction and one-step ahead prediction

Therefore, in the context of condition monitoring, the parallel model is preferred. As a consequence, the long-term prediction performance of the parallel model is selected as the main criterion in this study.

3.2 Model Selection: EE vs. OE

From the viewpoint of system identification [2], there are two basic error concepts: (1) output error and (2) equation error. Figure 4 gives an interpretation of these concepts in terms of a block diagram.

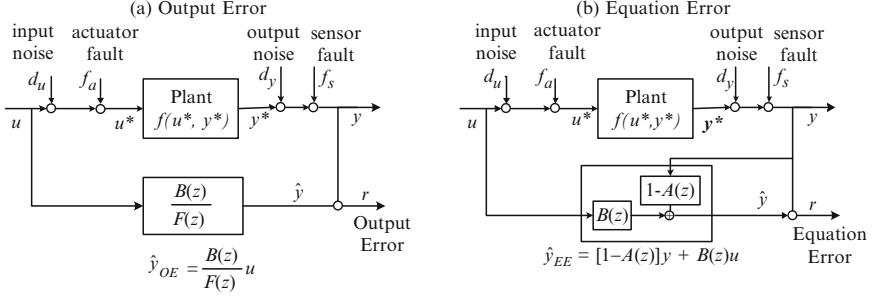


Fig. 4 Definitions of output error and equation error

3.2.1 Equation Error (EE) Model

The most simple way to describe the input–output relation $f(\cdot)$ is to represent it as a difference equation in the discrete time domain:

$$y(t) = \varphi^T(t) \cdot \theta + e(t) \quad (6)$$

where

$$\theta = [a_1 \dots a_{n_a} b_1 \dots b_{n_b}]^T \quad (7)$$

is a parameter vector, and

$$\varphi(t) = [y(t-1) \dots y(t-n_a) u(t-1) \dots u(t-n_b)]^T \quad (8)$$

an observation vector at time t , and $e(t)$ the noise term acting as a direct error in the difference equation. The model (6) is often called as *equation error model*.

By introducing the backward shift operator z^{-1} , the model prediction can be written in transfer function form:

$$\hat{y}_{EE}(t) = [1 - A(z)]y(t) + B(z)u(t) \quad (9)$$

where

$$A(z) = 1 + a_1 z^{-1} + \dots + a_{n_a} z^{-n_a} \quad (10)$$

and

$$B(z) = b_1 z^{-1} + \dots + b_{n_b} z^{-n_b}. \quad (11)$$

Alternatively, it can be written in the vector form

$$\hat{y}_{EE}(t) = \varphi^T(t) \cdot \theta \quad (12)$$

And the quantities

$$\begin{aligned} r_{EE}(t) &= y(t) - \hat{y}_{EE}(t) \\ &= y(t) - \varphi^T(t) \cdot \theta \end{aligned} \quad (13)$$

are called *equation errors*.

Recalling the parallel-series model (see Fig. 2), one can see that the EE model in fact is some kind of parallel-series model. $1 - A(z)$ corresponds to the series part f_A , $B(z)$ (11) is the counterpart of f_B and $\hat{y}_{EE}(t)$ in (9) is the one-step ahead prediction. The model (6) is also called an ARX (Auto-Regressive eXogeneous) model, where AR refers to the autoregressive part $A(z)y(t)$, X to the extra input $B(z)u(t)$, and $\phi(t)$ to the *regression vector* at time t .

3.2.2 Output Error (OE) Model

The EE model separates the relation between the input and output into two transfer functions $1 - A(z)$ (relating \hat{y} to y) and $B(z)$ (relating \hat{y} to u). From a physical point of view, it may seem more natural to present $f()$ in one transfer function. An immediate way of presenting a transfer function is to parameterize $f()$ as a rational function. Thus the model is given as

$$\hat{y}_{OE}(t) = \frac{B(z)}{F(z)}u(t) \quad (14)$$

with

$$B(z) = b_1z^{-1} + \dots + a_{n_b}z^{-n_b} \quad (15)$$

and

$$F(z) = 1 + f_1z^{-1} + \dots + f_{n_f}z^{-n_f} \quad (16)$$

The parameter vector to be identified is

$$\theta = [f_1 \ f_2 \ \dots \ f_{n_f} \ b_1 \dots b_{n_b}]^T \quad (17)$$

By multiplying both sides of (14) with $F(z)$ and moving terms $f_i z^{-i} \hat{y}(t)$, ($i = 1, \dots, n_f$) to the right side, the prediction $\hat{y}_{OE}(t)$ can be rewritten in the vector form:

$$\hat{y}_{OE}(t) = \hat{\phi}^T(t) \cdot \theta \quad (18)$$

with a *pseudo-regression vector* [13]

$$\hat{\phi}(t) = [\hat{y}(t-1) \ \dots \ \hat{y}(t-n_f) \ u(t-1) \ \dots \ u(t-n_b)]^T \quad (19)$$

The prediction error of (18)

$$\begin{aligned} r_{OE}(t) &= y(t) - \hat{y}_{OE}(t) \\ &= y(t) - \hat{\phi}^T(t) \cdot \theta \end{aligned} \quad (20)$$

is called *Output Error*.

It is worth noting that the pseudo-regression vector (19) in OE model differs from the regressor (8) in EE model. Compared to $y(t-j)$, $j = 1, 2, \dots, n_a$ in the regressor

(8) (which are the measured plant outputs), $\hat{y}(t - i)$ in the pseudo-regression vector (19) are not observed. Instead, $\hat{y}(t - i)$ are the previous model predictions.

One can find that the model connections are closely linked with the error concepts. The parallel connection corresponds to the output error, and the parallel-series connection is associated with the equation error.

As discussed earlier, a good model for condition monitoring is the parallel model, and a good criterion is the long-term prediction performance. It follows that OE model is better than EE model in terms of condition monitoring, and Sect. 4 will solve the problem of identifying the parameters of OE model.

4 NLS for OE Parameter Identification

The *least-squares estimate (LSE)* algorithm has been proved to be very effective for the ARX model identification. The objective of LSE is to minimize the mean squared equation error (MSEE):

$$V_{EE}(\theta) = \frac{1}{N} \sum_{t=1}^N \frac{1}{2} [y(t) - \varphi^T(t)\theta]^2 \quad (21)$$

Here, $V_{EE}(\theta)$ is a quadratic function of the parameter θ . It has been proved in literatures that the LSE can be successfully applied to the ARX model and gives the best parameters estimation in terms of minimizing $V_{EE}(\theta)$. However, the estimates of $\{a_i\}$ given by LSE may be biased if the residuals are correlated ([2, 12], p. 207, p. 256 in [7, 13]).

Due to the dependence within the residual, the LSE method failing to give unbiased parameter estimates results in a relatively poor long-term prediction performance. It is of interest to note that: the biased parameter estimates benefit to the smaller one-step prediction errors. This agrees with the point that the aim of LSE is to minimize the one-step-ahead prediction errors (21).

On the other hand, the objective of identifying OE model is to minimize the mean squared output error (MSOE):

$$V_{OE}(\theta) = \frac{1}{N} \sum_{t=1}^N \frac{1}{2} [y(t) - \hat{\varphi}^T(t)\theta]^2 \quad (22)$$

Note that, although they have the similar form in appearance, $V_{OE}(\theta)$ differs $V_{EE}(\theta)$ a lot due to the difference between $\varphi(t)$ (8) and $\hat{\varphi}(t)$ (19). Hence, (21) is a quadratic function, but (22) is highly nonlinear with respect to the parameter θ [1, 14].

Because of the high nonlinearity of OE objective function, the OE model identification is not an easy task, even if the model is linear. Some iterative optimization algorithm is inevitable, and the identification algorithm is generally expressed as

$$\theta_{k+1} = \theta_k - \eta \cdot [R(k)]^{-1} \cdot \frac{\partial V(\theta_k)}{\partial \theta} \quad (23)$$

where the subscript k gives the number of the iteration, η is a series of positive scalars (in term of step size) tending to zero or a small value, and $R(k)$ a n -by- n ($n = \dim \theta$) positive definite matrix to modify the search direction $-\frac{\partial V(\theta_k)}{\partial \theta}$. Here, $\frac{\partial V(\theta_k)}{\partial \theta}$ denotes the derivatives of $V(\theta)$ with respect to θ at k th iteration.

The following sections will give the technical details of the proposed fast identification algorithm. For ease of notation, the subscript OE is omitted and the terms $r(t)$, $V(\theta)$ will denote the output error $r_{OE}(t)$, the OE objective function $V_{OE}(\theta)$, respectively, in the following sections otherwise specified.

4.1 Calculation of $\partial V(\theta_k)/\partial \theta$ and the Jacobian

It is important to keep in mind that $\hat{\phi}(t)$ in (22) involves model prediction $\hat{y}(t-i)$, $i = 1 \dots n_f$, and $\hat{y}(t-i)$ is a function of parameters θ . It yields that $\hat{\phi}(t)$ depends on θ too. For explicitly expressing the links between \hat{y} and θ , $\hat{\phi}(t, \theta_k)$ is adopted to denote $\hat{\phi}(t)$ at k th optimization step:

$$\hat{\phi}(t, \theta_k) = [\hat{y}(t-1, \theta_k) \dots \hat{y}(t-n_f, \theta_k) \ u(t-1) \dots u(t-n_b)]^T \quad (24)$$

Let $g_k(t)$ denote the individual local gradient information at time t during the k th iteration, that is

$$g_k(t) = \frac{\partial r(t)}{\partial \theta}, \quad (25)$$

followed by

$$g_k(t) = \frac{\partial(y(t) - \hat{y}(t, \theta_k))}{\partial \theta} = -\frac{\partial(\hat{y}(t, \theta_k))}{\partial \theta}. \quad (26)$$

Since

$$\begin{aligned} \frac{\partial(\hat{y}(t, \theta_k))}{\partial \theta} &= \frac{\partial(\hat{\phi}^T(t, \theta_k) \cdot \theta_k)}{\partial \theta} \\ &= (\hat{\phi}(t, \theta_k) + \frac{\partial \hat{\phi}(t, \theta_k)}{\partial \theta} \cdot \theta_k), \end{aligned} \quad (27)$$

it follows that

$$g_k(t) = -[\hat{\phi}(t, \theta_k) + \frac{\partial \hat{\phi}(t, \theta_k)}{\partial \theta} \cdot \theta_k]. \quad (28)$$

According to the definition of $\hat{\phi}(t, \theta_k)$ and (26), $\frac{\partial \hat{\phi}(t, \theta_k)}{\partial \theta}$ is

$$\begin{aligned} \frac{\partial \hat{\phi}(t, \theta_k)}{\partial \theta} &= \left[\frac{\partial \hat{y}(t-1, \theta_k)}{\partial \theta} \dots \frac{\partial \hat{y}(t-n_f, \theta_k)}{\partial \theta} \ \frac{\partial u(t-1)}{\partial \theta} \dots \frac{\partial u(t-n_b)}{\partial \theta} \right] \\ &= -[g_k(t-1) \dots g_k(t-n_f) \ 0 \dots 0] \end{aligned} \quad (29)$$

Substituting (29) into (28) gives

$$g_k(t) = -[\hat{\varphi}(t, \theta_k) - [g_k(t-1) \cdots g_k(t-n_f) \ 0 \cdots 0] \cdot \theta_k] \quad (30)$$

Equation (30) indicates that $g_k(t)$ at time instant t not only depends on current observation vector $\hat{\varphi}(t, \theta)$ but also depends on the previous gradients $\{g_k(\tau)\}$, $(\tau < t)$. Therefore, the iterative calculation of $g_k(t)$ (30) represents the dependency within prediction errors $\{r(t)\}$.

The gradient on all the observed data is calculated by summing all values of local gradients.

$$\frac{\partial V(\theta_k)}{\partial \theta} = \frac{1}{2} \sum_{t=1}^N \frac{\partial r^2(t)}{\partial \theta} = \sum_{t=1}^N r(t) \cdot g_k(t) \quad (31)$$

Compared with ARX model (where the gradient is $\varphi(t)$ alone), the calculation of gradient as (30) differs at the additional sum terms $[g_k(t-1) \cdots g_k(t-n_f) \ 0 \cdots 0] \cdot \theta_k$. The added terms enable the derivative of $V(\theta)$ reflect the actual gradient of objective function (22) correctly.

Note that $g_k(t)$ is calculated at every time t . As a result, the Jacobian of objective function (22) is achieved simply by transforming the sequence $\{g_k(t)\}$ into a matrix form:

$$\mathbf{J} = [g_k(1) \ g_k(2) \cdots g_k(N)]^T \quad (32)$$

where \mathbf{J} is N -by- $(n_f + n_b)$ matrix.

Let a column vector Γ to denote the long-term prediction error sequence $\Gamma = [r(1) \ r(2) \cdots r(N)]^T$, then (31) can be rewritten as

$$\frac{\partial V(\theta_k)}{\partial \theta} = \mathbf{J}^T \cdot \Gamma \quad (33)$$

4.2 Approximation of $R(k)$ and the Hessian

As shown in many literatures, for minimizing a nonlinear function, it is may be inefficient to use the first-order alone as the search direction [6, 10]. In most cases, the inverse of gradient does not point to the minimum point straightaway. It is particularly true when the objective function surface has a valley, e.g., the Rosenbrock function. Thus a matrix $R(k)$ is employed to adjust the optimization direction from $\frac{\partial V(\theta)}{\partial \theta}$. With the aid of Taylor expansion, it has been shown that using a quadratic model to approximate the high nonlinear function (22) is beneficial for improving the optimization convergence. A good selection of $R(k)$ in (23) is the Hessian matrix [10]. However, the calculation of the Hessian matrix is computation-consuming. In order to reduce the computation costs for the on-board application, the Hessian needs to be approximated properly at a lower cost.

One of the solution is to make use of the Jacobian to approximate the Hessian. Consider the second-order Taylor expansion of function (22) around some parameter θ^*

$$\begin{aligned} V(\theta) \approx V(\theta^*) &+ \left[\frac{\partial V(\theta)}{\partial \theta} \Big|_{\theta=\theta^*} \right]^T (\theta - \theta^*) \\ &+ \frac{1}{2} (\theta - \theta^*)^T \mathbf{H}(\theta^*) (\theta - \theta^*) \end{aligned} \quad (34)$$

where $\mathbf{H}(\theta^*)$ is the Hessian matrix evaluated at θ^* . Note that this approximation is valid when θ is in the neighborhood of θ^* .

For compact notation, $\frac{\partial V(\theta)}{\partial \theta} \Big|_{\theta=\theta^*}$ is replaced by $\frac{\partial V(\theta^*)}{\partial \theta}$. From expansion (34), the condition of minimum of $V(\theta)$ can be expressed as

$$\frac{\partial V(\theta^*)}{\partial \theta} + \mathbf{H}(\theta^*) (\theta - \theta^*) \approx 0 \quad (35)$$

followed by

$$\theta \approx \theta^* - \mathbf{H}(\theta^*)^{-1} \cdot \frac{\partial V(\theta^*)}{\partial \theta} \quad (36)$$

Comparing (36) with (23), one can easily find that the search direction should be modified by the inverse Hessian matrix. It follows that the $\mathbf{R}(k)$ can be replaced by the Hessian matrix.

The problem turns into an approximation of the Hessian matrix. According to Gauss-Newton methods, an easy way to form Hessian estimate is to make use of the first derivative information, as shown below.

$$\begin{aligned} \mathbf{H}(\theta) &\triangleq \frac{\partial^2 V(\theta)}{\partial \theta \partial \theta^T} \\ &= \sum_{t=1}^N \frac{\partial^2 (r(t))}{\partial \theta \partial \theta^T} \\ &= \sum_{t=1}^N \left[\frac{\partial r(t)}{\partial \theta} \cdot \frac{\partial r(t)}{\partial \theta^T} + r(t) \frac{\partial^2 r(t)}{\partial \theta \partial \theta^T} \right] \\ &= \mathbf{J}^T \cdot \mathbf{J} + r(t) \frac{\partial^2 r(t)}{\partial \theta \partial \theta^T} \end{aligned} \quad (37)$$

It is therefore convenient to ignore the second term on the right-hand side and approximate $\mathbf{H}(\theta)$ by $\mathbf{J}^T \cdot \mathbf{J}$. It follows that

$$\mathbf{R}(k) = \mathbf{J}^T \cdot \mathbf{J} \quad (38)$$

In terms of the identification convergence, this approximation approach potentially offer the best trade-off of two worlds: First, by using the Hessian matrix as $R(k)$ (23), the search direction is modified from inverse gradient direction to point

toward to the minimum more straightaway. Secondly, it only needs to compute the first-order derivatives that have been available in the computation of local gradients $g_k(t)$. Thus, the calculation of the Hessian does not introduce too much more extra computation burden, and results a better parameter updating direction.

As a result, the improved NLS algorithm for reduced order OE model is named dynamic nonlinear least squares (DNLS) and given as follows:

Dynamic Nonlinear Least Squares Algorithm

- (1) Let $k = 0$, $g_k(1) \dots g_k(n_f) = 0$, and set the initial values of θ_k according to a priori knowledge
- (2) At each time instant t , ($t = n_f + 1, \dots, N$), compute the local gradient $g_k(t)$ by $g_k(t) = -[\dot{\phi}(t, \theta_k) - [g_k(t-1) \dots g_k(t-n_f) 0 \dots 0] \cdot \theta_k]$
- (3) At time instant N , rearrange $\{g_k(t)\}$ to form the Jacobian $\mathbf{J} = [g_k(1) \ g_k(2) \dots g_k(N)]^T$
- (4) Update the parameter by $\theta_{k+1} = \theta_k - \eta \cdot (\mathbf{J}^T \cdot \mathbf{J})^{-1} \cdot \mathbf{J}^T \cdot \Gamma$, where $\Gamma = [r(1) \ r(2) \dots r(N)]^T$ and η is a fixed or adjustable step-size.
- 5) Stop condition check: check whether the maximum iteration has been achieved or the updating of θ_k has been very small. If not, set $k = k + 1$ and go back Step 2.

Remark 4.1. Since the Hessian matrix is approximated by $\mathbf{J}^T \cdot \mathbf{J}$, the approximation is only accurate enough when the parameters θ is close enough to the minimum, as shown in the second-order Taylor expansion (34). In this case, the step size η is equal to 1. However, in the earlier stage of search where θ may be far from the minimum, η is not equal to 1 and a line search method is adopted to select the optimal value of η in each search iteration.

Remark 4.2. The dependency within long-term prediction errors is solved by computing the gradient in an iterative manner, as shown in (30), leading to a better gradient calculation.

Remark 4.3. In the proposed DNLS, the objective function is approximated by a quadratic function that is more accurate than the first-order approximation, thus the search direction is better than those common deepest descent methods. Furthermore, the Jacobian \mathbf{J} is just a rearrangement of sequence of gradients $g(t)$ ($t = 1 \dots N$) without additional computation and the approximation of the Hessian does not need more computation than common deepest descent methods. Therefore, one of the benefits is that the improvement on performance is only paid by a small increase of computation expense, which make this method suitable for on-board modeling and condition monitoring.

5 Application and Results

In this section, the proposed DNLS algorithm is employed to identify the parameters of the reduced order OE model of a two shaft gas turbine engine. Real engine data gathered from normal engine operation at the engine test-bed are used [3]. In the duration of this test, the angle of the VGVs (Variable Guide Vanes) of the low pressure compressor and the reheat nozzle area were fixed to their low speed positions, the engine fuel flow $W_f(t)$ is the control input and the high pressure shaft speed $N_{HP}(t)$ is the primary output.

Because the engine runs at the low speed operating point, the input/output data are first preprocessed by subtracting the physical equilibrium. Therefore, the data used indeed for dynamic modeling are $\Delta W_f(t)$ and $\Delta N_{HP}(t)$. For simplicity of notation, let $u(t)$, $y(t)$ denote the input $\Delta W_f(t)$ and output $\Delta N_{HP}(t)$, respectively. 1,500 data pairs were collected in total. The first 750 pairs compose the training data set for parameters estimation and the remaining 750 pairs make up the validation data set for model validation.

Figure 5 illustrates the distribution of the training data set and validation data set, where ‘*’ represents the training data and ‘o’ the test data. It is obvious that these two data sets cover a little different dynamics of the gas turbine, although they are overlapped around the center which represents the current operating point.

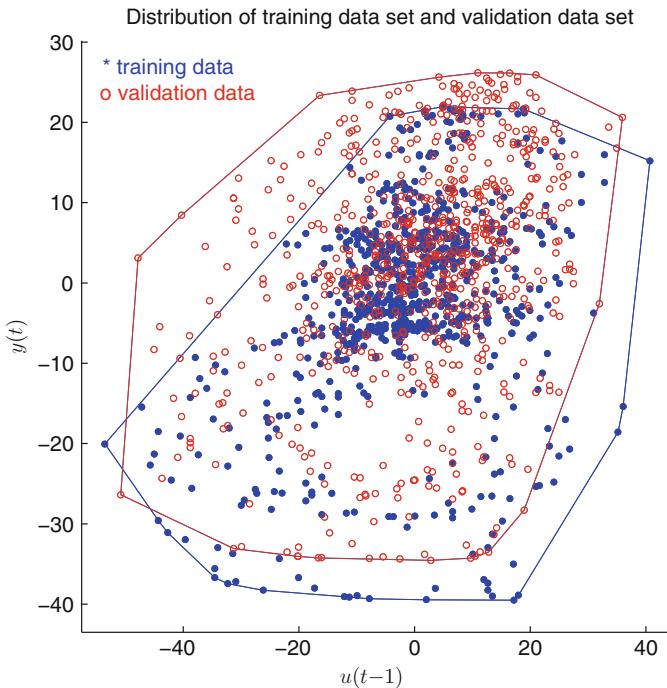


Fig. 5 Scatter plot of the output $y(t)$ versus the input $u(t-1)$

The aim of dynamic modeling is to obtain a reduced order model whose long-term prediction errors are minimized. To measure the algorithm performances, two criteria are examined: (1) prediction accuracy and (2) computation costs. The prediction accuracy is measured by MSOE (Mean Square Output Error)

$$MSOE = \frac{1}{N} \sum_{t=1}^N \frac{1}{2} [y(t) - \hat{y}_{OE}(t)]^2 \quad (39)$$

where $N = 750$ is the length of the training or validation data set. Note that some methods (e.g., LSE) are for minimizing one-step ahead prediction errors. For such algorithms, the MSEE (Mean Square Equation Error) on training data set is used for comparison.

$$MSEE = \frac{1}{N} \sum_{t=1}^N \frac{1}{2} [y(t) - \hat{y}_{EE}(t)]^2. \quad (40)$$

The standard deviation of the errors is also adopted

$$STD = \sqrt{\frac{1}{N-1} \sum_{t=1}^N (r(t) - \mathbf{E}(r(t)))^2} \quad (41)$$

where $r(t)$ is either the OE errors or EE errors, E denotes the mathematical expectation.

The computation cost is measured by how many evaluations of the objective function are carried out during the identification, because the main computation burden in most iterative optimization algorithms is the calculation of the long-term prediction sequence $\hat{y}_{OE}(t), t = 1 \dots N$.

5.1 First-Order Model

Here, the first order model of the gas turbine engine is considered, where $\hat{\phi}(t) = [\hat{y}(t-1) \ u(t-1)]^T$ and $\theta = [f_1 \ b_1]^T$. In these experiments, although the model is a linear model, the objective function (22) is a high order nonlinear surface with a narrow valley around the minimum point as shown in Fig. 6.

For comparison purposes, five different algorithms are presented here. The LSE and ARX approaches provided by *System Identification toolbox* do not contain iterative search, hence their computation costs are set 1. All the rest algorithms are iterative search methods. The standard OE method and the RIV (Refined Instrumental Variable) [16, 17] are also included for comparison. A fuzzy neural networks ANFIS (Adaptive Networks-based Fuzzy Inference System) [9] is also used. The results show that such neural networks are not suitable for long-term prediction.

The exhaustive search method examines each possible parameter value in order to find the ‘possible’ global solution. In our experiments, it runs in two main steps and the total number of objective function evaluation is 12000.

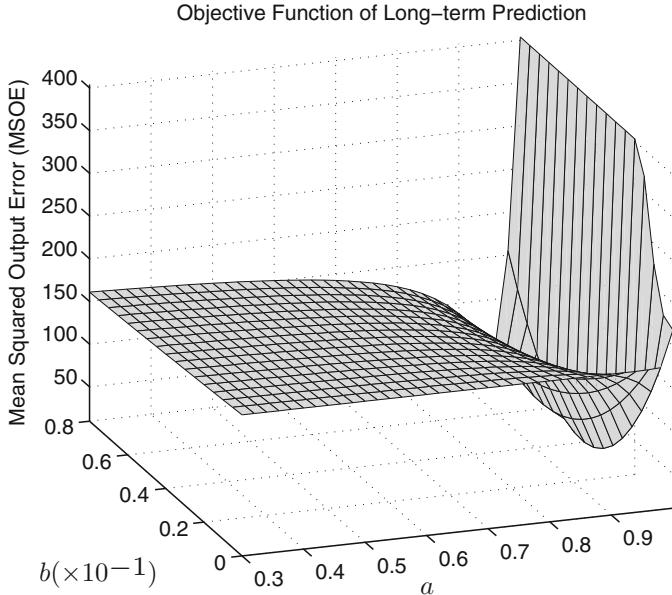


Fig. 6 Objective function of the first-order model

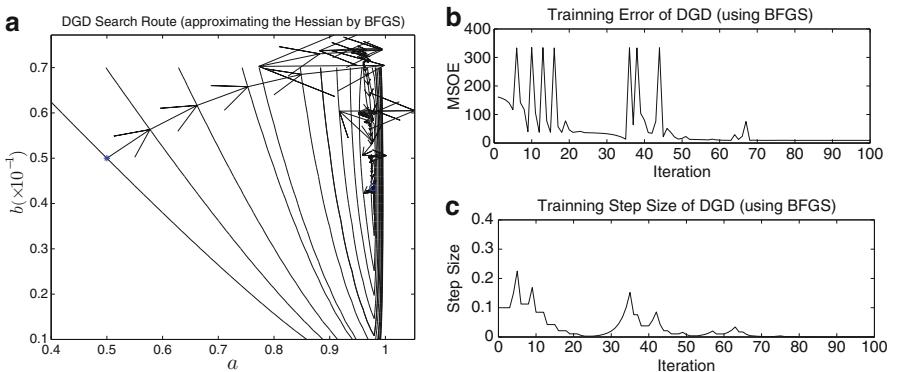


Fig. 7 The DGD algorithm: (a) Searching route of DGD, (b) Training errors and (c) step size

In all the gradient-based iterative algorithms, $[f_1, b_1] = [0.5 \ 0.05]$ is set as the default initial values of parameters and the initial step size is 0.1. The gradient search method uses the inverse gradient as search direction and set $R(k)$ in (23) as an identity matrix. In DGD (Dynamic Gradient Descent) approach [4], the BFGS (Broyden–Fletcher–Goldfarb–Shanno) [13], [6] are adopted to modify the search direction from inverse gradient direction. The detailed search progress of DGD is presented in Fig. 7. The DGD algorithm needs about 100 objective function

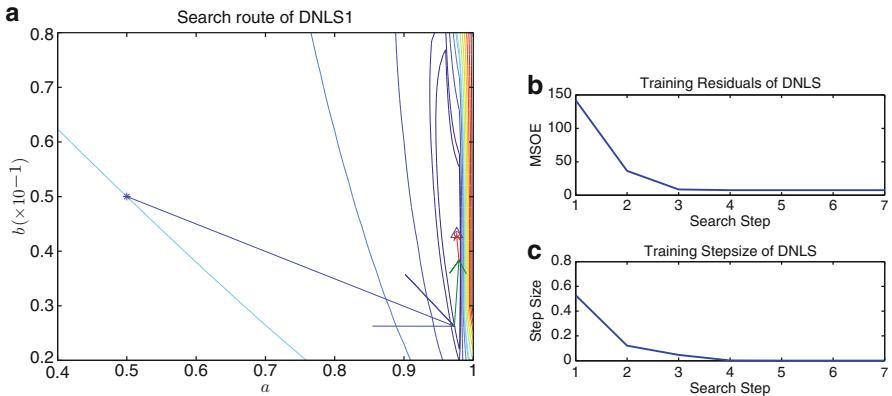


Fig. 8 The DNLS algorithm: (a) Searching route, (b) training errors and (c) training stepsize

evaluations to approach to the optimum point, and a distance from the minimum still can be seen. One reason is that the search direction does not point to the minimum at the beginning stage. Further reason is that the search route continually zigzags from one side of the valley to another after entering the valley as shown in Fig. 7.

In the proposed DNLS approaches, because of line search, two different stop conditions are tested in DNLS1 and DNLS2, respectively. The search process of DNLS1 is shown in Fig. 8. Apparently, the solution point given by the DNLS is closer to the minimum and the convergence speed is accelerated, as depicted in Fig. 8. The search direction of DNLS points a more straightforward way to the minimum. Even in the valley, it still looks better. More clearly, from Fig. 8b, c, it can be seen that only 7 steps are involved to arrive the optimum point and the value of objective function drops dramatically and steadily. Compared to DGD, this improvement is benefited from a better Hessian approximation in (38). Figure 9 shows the DNLS results of long-term prediction on validation data with MSOE of 9.131786.

The comparison of different methods is shown in Table 1 (on training data) and Table 2 (on validation data). It can be concluded from these tables that: (1) In terms of one-step-ahead prediction (equation error), LSE, ANFIS, ARX achieved slightly smaller equation errors than those methods for long-term prediction. However, LSE, ANFIS, ARX failed on long-term prediction on both training data set and validation data set. (2) In terms of long-term prediction (output error), the methods OE, RIV, exhaustive search, Gradient Descent, DGD and DNLS obtain similar results. Their performances on long-term prediction are better than those methods designed for one-step-prediction. (3) In terms of computation costs, the DNLS is advantageous over the exhaustive search, gradient descent, DGD approaches. The main contribution of DNLS is the reduction of computation costs that makes this approach better for on-board identification.

One thing interesting is that the neural network ANFIS achieves the smallest MSEE 0.18728 on training data, but it gives the worst MSOE performance on both training data and validation data. Although static feed-forward neural networks

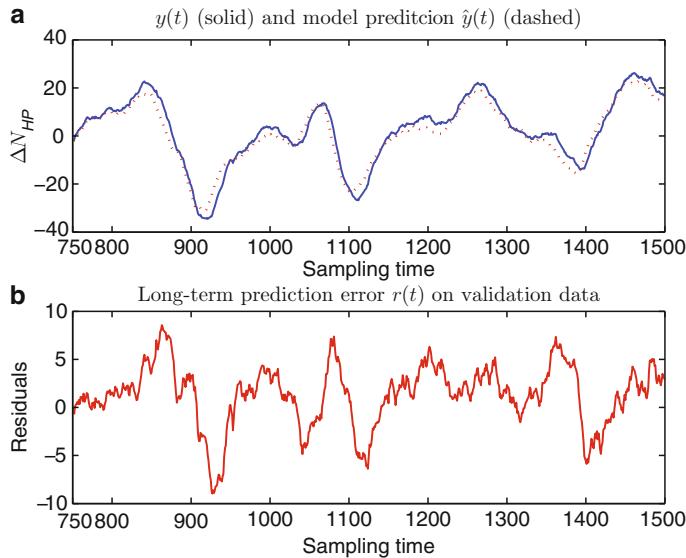


Fig. 9 Long-term prediction of DNLS for the first-order model on the validation data

Table 1 Comparison on training data set (first order model)

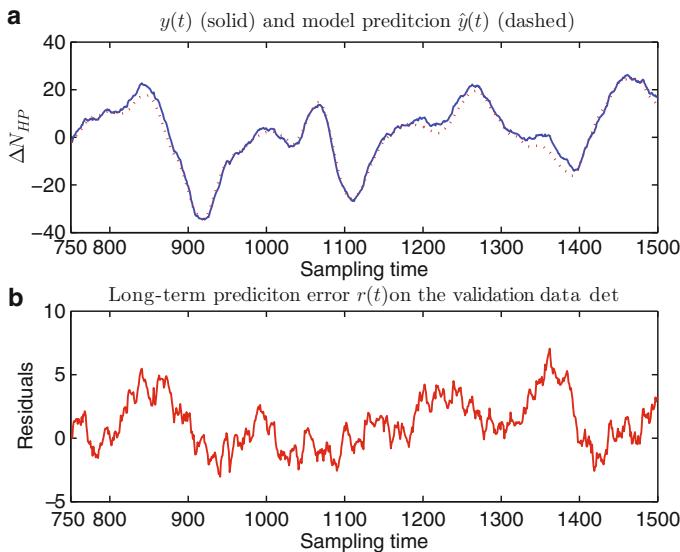
Methods	Computation Costs	1-Step-ahead prediction		Long-term prediction	
		STD	MSEE	STD	MSOE
LSE	1	0.46133	0.21284	4.4273	19.6359
ARX	1	0.46169	0.21322	5.1073	26.1860
OE	N/A	0.54868	0.30120	2.8465	8.12804
RIV	N/A	0.54713	0.29965	2.7479	7.6628
ANFIS	250	0.43276	0.18728	5.2745	27.8212
Exhaustive search	12,000	0.48241	0.23280	2.7470	7.63754
Gradient descent ^a	1,000	0.48270	0.23312	2.7471	7.63788
DGD ^b	101	0.48239	0.23281	2.7470	7.63754
DNLS1 ^c	59	0.48238	0.23280	2.7470	7.63755
DNLS2 ^c	80	0.48237	0.23280	2.7470	7.63754

^a The gradient search approach uses inverse gradient as search direction ^bThe DGD approach uses BFGS algorithm to adjust search direction ^cDNLS1 and DNLS2 stops after 59 and 80 objective function evaluations respectively because of different stop conditions

(such as ANFIS) have good ability in approximating static mathematical functions, they may be not suitable for dynamic system simulation (in terms of long-term prediction). The smallest MSEE given by the ANFIS on the training data set may be caused by the overfit during the training stage. Hence, the ANFIS interpolates the training data set very well, but it fails to extrapolate the validation data.

Table 2 Comparison on validation data set (first order model)

Methods	STD	MSOE
LSE	3. 7796	16. 8985
ARX	4. 0895	19. 5996
OE	2. 9301	10. 2658
RIV	3. 2490	12. 2963
ANFIS	3. 0258×10^3	1. 1908×10^7
Exhaustive search	3. 1807	11. 8369
Gradient descent	8. 1806	11. 8311
DGD	3. 1807	11. 8369
DNLS1	3. 1811	11. 8398
DNLS2	3. 1807	11. 8368

**Fig. 10** Long-term prediction of DNLS for the second-order model on the validation data

5.2 Second-Order Model

The second-order model also is identified to test the ability of DNLS, where $\hat{\phi}(t) = [\hat{y}(t-1) \ \hat{y}(t-2) \ u(t-1)u(t-2)]^T$ and $\theta = [f_1 \ f_2 \ b_1 \ b_2]^T$. The MSE drops to 3.31867176 and the computation cost is 60, as shown in Fig. 10. The parameters of such a model are $\theta = [1.8604765 \ -0.8641699 \ 0.070453335 \ -0.0074745994]^T$.

The comparison against other approaches is listed in Table 3. It can be seen that the computation cost of the proposed DNLS is just 103.

Table 3 Comparison for second order model

Methods	Computation Costs	on Training data		on Validation data	
		STD	MSOE	STD	MSOE
LSE	1	3.41291	1.16613 ^a	2.87529	10.3296
ARX	1	3.93632	1.55375 ^a	3.05683	11.5071
OE	N/A	1.32140	1.89871	1.85980	4.76809
RIV	N/A	1.25340	1.81345	1.97333	5.19088
Exhaustive search	10^6	1.25624	1.78773	1.95562	5.16392
Gradient descent	20,000	2.30235	5.39615	2.74971	9.22707
DGD	1,000	1.25625	1.78773	1.95560	5.16380
DNLS	103	1.25624	1.78773	1.95563	5.16391

^a These two are MSEEs, because LSE, ARX are for EE model only

6 Summary

In this chapter, by a comparative study, the OE model is selected as the model for modeling gas turbine engines. The discussion on parallel/serial model connection and the corresponding equationg/output error shed light on the way to unbiased parameter identification. It has been proved that LSE gives biased parameter estimation in the context of reduced order modeling. In the identification algorithm design, the nonlinear least-squares method is used and problem of correlated prediction errors in the OE model is resolved by calculating the gradient in an iterative manner. In order to accelerate the identification speed and meet the limitation on the on-board computation resources, the second-order Hessian matrix is approximated by the first-order Jacobian matrix.

The main contribution of this study is to propose an iterative calculation of the gradient to deal with the correlated prediction errors. Furthermore, by approximating the Hessian with the the Jacobian, the identification speed is accelerated with a minor increase of computation costs. The experiment results on modeling a gas turbine engine have shown that the proposed approach achieves a faster convergence speed at relatively lower computational costs.

Acknowledgement This work was supported by the UK Leverhulme Trust (F/00 120/BC) and the China National Science Foundation under Grants (60828007 and 60974029), 973 National Basic Research Program of China under Grant (2009CB320601), the 111 project under (B08015) and the Fundamental Research Funds for the Central Universities (XDKJ2009C024).

References

1. B. Anderson and C. Johnson, Jr. On reduced-order adaptive output error identification and adaptive IIR filtering. *IEEE Transactions on Automatic Control*, 27(4):927–933, 1982
2. K. J. Astrom and P. Eykhoff. System identification – a survey. *Automatica*, 7(2):123–162, March 1971

3. T.V. Breikin, G. G. Kulikov, V. Y. Arkov, and Peter. J. Fleming. Dynamic modelling for condition monitoring of gas turbines: Genetic algorithms approach. In *Proceedings of 16th IFAC World Congress*, 2005
4. X. Dai, T. Breikin, and H. Wang. An algorithm for identification of reduced-order dynamic models of gas turbines. In *Proceedings of 1st International Conference on Innovative Computing, Information and Control*, volume 1, pages 134–137, 2006
5. C. Evans. Testing and modelling aircraft gas turbines: An introduction and overview. In *UKACC International Conference on Control'98*, pages 1361–1366, 1998
6. R. Fletcher. A new approach to variable metric algorithms. *The Computer Journal*, 13: 317–322, 1970
7. M. Hong, T. Soderstrom, and W. X. Zheng. A simplified form of the bias-eliminating least squares method for errors-in-variables identification. *IEEE Transactions on Automatic Control*, 52(9):1754–1756, September 2007
8. R. Isermann. Model-based fault detection and diagnosis – status and applications. In *Proceeding of 16th IFAC Symposium on Automatic Control in Aerospace*, 2004
9. J.-S.R. Jang, C.-T. Sun, and E. Mizutani. *Neuro-Fuzzy and Soft Computing*. Prentice-Hall, Englewood Cliffs, 1997
10. D. Kahaner, C. B. Moler, and S. Nash. *Numerical Methods and Software*. Prentice-Hall, Englewood Cliffs, 1989
11. G. G. Kulikov and H. A. Thompson. *Dynamic Modelling of Gas Turbines: Identification, Simulation, Condition Monitoring and Optimal Control*. Springer, London, 2004
12. I. Landau. Unbiased recursive identification using model reference adaptive techniques. *IEEE Transactions on Automatic Control*, 21(2):194–202, April 1976
13. L. Ljung. *System Identification: Theory for the User*. Prentice Hall, London, 2nd edition, 1999
14. S. L. Netto, P. S. R. Diniz, and P. Agathoklis. Adaptive IIR filtering algorithms for system identification: A general framework. *IEEE Transactions on Education*, 38(1):54–66, February 1995
15. A. Wills and B. Ninness. On gradient-based search for multivariable system estimates. *IEEE Transactions on Automatic Control*, 53:298–306, 2008
16. P. C. Young. *Recursive Estimation and Time-Series Analysis*. Springer, Berlin, 1984
17. P. C. Young, H. Garnier, and M. Gilson. An optimal instrumental variable approach for identifying hybrid Box-Jenkins models. In *14th IFAC Symposium on System Identification SYSID06*, pages 225–230, Newcastle, NSW, Australia, 2006

Part II

**Large-Scale Systems Control
and Applications**

Robust Control of Large-Scale Systems: Efficient Selection of Inputs and Outputs

Javad Lavaei, Somayeh Sojoudi, and Amir G. Aghdam

1 Introduction

There has been a growing interest in recent years in robust control of systems with parametric uncertainty [4, 9, 13, 16, 18]. The dynamic behavior of this type of systems is typically governed by a set of differential equations whose coefficients belong to fairly-known uncertainty regions. Although there are several methods to capture the uncertain nature of a real-world system (e.g., by modeling it as a structured or unstructured uncertainty [6]), it turns out that the most realistic means of describing uncertainty is to parameterize it and then specify its domain of variation.

Robust stability is an important requirement in the control of a system with parametric uncertainty. This problem has been extensively studied in the case of linear time-invariant (LTI) control systems with specific types of uncertainty regions. For instance, sum-of-squares (SOS) relaxations are numerically efficient techniques introduced in [18] and [4] for checking the robust stability of polynomially uncertain systems. Moreover, a necessary and sufficient condition is proposed in [13] for the robust stability verification of this class of uncertain systems, by solving a hierarchy of semi-definite programming (SDP) problems.

The concepts of controllability and observability were introduced in the literature, and it was shown that they play a key role in various feedback control analysis and design problems such as model reduction, optimal control, state estimation, etc. [6]. Several techniques are provided in the literature to verify the controllability

J. Lavaei

Control and Dynamical Systems, California Institute of Technology, Pasadena, USA

e-mail: lavaei@cds.caltech.edu

S. Sojoudi

Control and Dynamical Systems, California Institute of Technology, Pasadena, USA

e-mail: sojoudi@cds.caltech.edu

A.G. Aghdam

Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada

e-mail: aghdam@ece.concordia.ca

and/or observability of a system. However, in many applications it is important to know how much controllable or observable a system is. Gramian matrices were introduced to address this issue by providing a quantitative measure for controllability and observability [6]. While these notions were originally introduced for fixed known systems, they have been investigated thoroughly in the past two decades for the case of uncertain systems [21, 22, 27].

On the other hand, real-world systems are often composed of multiple interacting components, and hence possess sophisticated structures. They are typically modeled as large-scale interconnected systems, for which classical control analysis and design techniques are usually inefficient. Several results are reported in the literature for structurally constrained control of large-scale systems in the contexts of decentralized and overlapping control, to address the shortcomings of the traditional control techniques [5, 11, 12, 24, 25].

This work aims to measure the minimum of the smallest singular value for the controllability and observability Gramians of parametric systems, over a given uncertainty region. Given a polynomially uncertain LTI system with uncertain parameters defined on a semi-algebraic set, it is asserted that the controllability (observability) Gramian is a rational matrix in the corresponding parameters. It is desired to attain the minimum singular value of this matrix over the uncertainty region, but due to the rational structure of the matrix one cannot take advantage of the efficient techniques such as SOS tools. To bypass this obstacle, it is shown that this rational matrix can be replaced by a polynomial approximation which satisfies an important relation. An SOS formula is then obtained to find the underlying infimum. The special case of a polytopic uncertainty region is also investigated, due to its importance in practice. An alternative approach is proposed for this special case, with a substantially reduced computational burden.

Two primary applications of this work are as follows:

- To achieve robust closed-loop performance for a system subject to perturbation, it is very important to know the minimum energy required to control the system for any possible values of the parameters in the uncertainty region. This energy is known to be proportional to the inverse of the infimum of the singular values sought in this work. For instance, this infimum would determine if a system which is controllable for the nominal parameters, is also controllable with a sufficiently safe margin in a practical environment, where the parameters are subject to variation around the nominal values.
- In a real-world system with several interacting subsystems (which may be geographically distributed), it may not be feasible to establish information flow between all control agents. In other words, for such systems it is more desirable to have some form of decentralization, where each control input is constructed in terms of only those outputs which are available to the corresponding local controller due to the communication and computation limitations. To determine which inputs (or outputs) are most effective in the overall control operation and which ones are negligible, one can obtain the minimum input energy for different information flow structures, and choose the best structure by comparing

the resultant values [28]. The results of the present work can be used to measure the required energy for different control structures systematically (this will be clarified in a numerical example).

This chapter is organized as follows. The problem is formulated in Sect. 2, where some important background results are provided. The main results of the chapter are developed in Sect. 3 for systems with polynomial uncertainty, and the special case of a polytopic region is also addressed in detail. The results are illustrated in Sect. 4 through a numerical example, and finally the concluding remarks are summarized in Sect. 5.

2 Preliminaries and Problem Formulation

Consider an uncertain large-scale interconnected system with the following representation:

$$\begin{aligned}\dot{x}(t) &= A(\alpha)x(t) + B(\alpha)u(t) \\ y(t) &= C(\alpha)x(t) + D(\alpha)u(t)\end{aligned}\tag{1}$$

where

- $x(t) \in \mathbf{R}^n$, $u(t) \in \mathbf{R}^m$ and $y(t) \in \mathbf{R}^r$ are the state, input and output of the system, respectively.
- $\alpha := [\alpha_1, \alpha_2, \dots, \alpha_k]$ denotes the vector of unknown, fixed uncertain parameters of the system.
- $A(\alpha), B(\alpha), C(\alpha)$ and $D(\alpha)$ are matrix polynomials in the variable α .

The system (1) is referred to as a *polynomially uncertain system*. Assume that the uncertainty vector α belongs to a given semi-algebraic set \mathcal{D} characterized as follows:

$$\mathcal{D} = \{\alpha \in \mathbf{R}^k \mid f_1(\alpha) \geq 0, \dots, f_z(\alpha) \geq 0\}\tag{2}$$

where $f_1(\alpha), \dots, f_z(\alpha)$ are known scalar polynomials. Note that many practical uncertainty regions can be expressed either exactly or approximately in the above form.

Due to the large-scale nature of the system, the vectors $u(t)$ and $y(t)$ may contain several entries, and this has important implications in controller design, in general. Hence, assume that only a subset of the output vector, denoted by $\tilde{y}(t)$, and a subset of the input vector, denoted by $\tilde{u}(t)$, are desired to be used in the control structure, for the sake of simplifying the control operation. Let $\mathcal{S}(\alpha)$ be the system with the reduced input and output size, and denote its state-space representation as:

$$\begin{aligned}\dot{x}(t) &= A(\alpha)x(t) + \tilde{B}(\alpha)\tilde{u}(t) \\ \tilde{y}(t) &= \tilde{C}(\alpha)x(t) + \tilde{D}(\alpha)\tilde{u}(t)\end{aligned}\tag{3}$$

The objective is to evaluate the controllability/observability degradation resulted from reducing the size of the input and output of the system (1). In other words, it is

intended to measure the controllability/observability degree of the system $\mathcal{S}(\alpha)$ in comparison to that of the original system (1). Addressing the above point is central to this chapter.

Suppose that the open-loop system $\mathcal{S}(\alpha)$ is robustly stable over the region \mathcal{D} , i.e., all eigenvalues of the matrix $A(\alpha)$ lie in the open left-half s -plane for every $\alpha \in \mathcal{D}$. This assumption is required for defining the infinite-horizon controllability Gramian of the system. It is noteworthy that the verification of the open-loop robust stability of the system $\mathcal{S}(\alpha)$ can be carried out systematically, using existing methods in the literature such as the SOS technique given in [13] or the SDP method proposed in [4]. Denote the controllability and observability Gramians of the system $\mathcal{S}(\alpha)$ with $W_c(\alpha)$ and $W_o(\alpha)$, respectively, which are defined as follows:

$$W_c(\alpha) = \int_0^\infty e^{A(\alpha)t} \tilde{B}(\alpha) \tilde{B}(\alpha)^T e^{A(\alpha)^T t} dt \quad (4a)$$

$$W_o(\alpha) = \int_0^\infty e^{A(\alpha)^T t} \tilde{C}(\alpha)^T \tilde{C}(\alpha) e^{A(\alpha)t} dt \quad (4b)$$

In this chapter, the system $\mathcal{S}(\alpha)$ is said to be *robustly controllable (observable)* if it is controllable (observable) for all $\alpha \in \mathcal{D}$. Note that the robust controllability of the system $\mathcal{S}(\alpha)$ is equivalent to the positive-definiteness of the matrix $W_c(\alpha)$ for all $\alpha \in \mathcal{D}$. The Gramian matrix $W_c(\alpha)$ provides a measure for the degree of controllability (rather than just controllability/uncontrollability). Indeed, it is known that the input energy required for controlling the system is, roughly speaking, proportional to the inverse of the matrix $W_c(\alpha)$, and more specifically, proportional to the inverse of its smallest singular value (Proposition 4.5 in [6]). Hence, the robust controllability degree of the system $\mathcal{S}(\alpha)$ can be assessed in terms the minimum of the smallest singular value of the matrix $W_c(\alpha)$ over the region \mathcal{D} . This chapter is concerned with the computation of this minimum using an efficient method. Note that although this work focuses on the robust controllability problem, the results can be easily applied to the robust observability problem, as well (due to the existence of a natural duality between the two problems). Since the main development of this chapter is contingent upon the notion of SOS, some relevant background material is provided in Sect. 2.1

2.1 Background on Sum-of-Squares

This section aims to present important results in the area of sum-of-squares (SOS). A scalar polynomial $p(\alpha)$ is said to be SOS if there exists a set of scalar polynomials $p_1(\alpha), p_2(\alpha), \dots, p_q(\alpha)$ such that:

$$p(\alpha) = p_1(\alpha)^2 + p_2(\alpha)^2 + \cdots + p_q(\alpha)^2, \quad \forall \alpha \in \mathbf{R}^k \quad (5)$$

It is evident that an SOS polynomial is always nonnegative, but the converse statement is not necessarily true. Indeed, a nonnegative polynomial cannot always be

written as a sum of squared polynomials [1]. It is to be noted that checking whether a given polynomial is SOS amounts to solving a convex optimization problem [19]. Sparked by Polya's work, a great deal of effort has been made in the literature to make a connection between positive (nonnegative) polynomials and SOS polynomials. For instance, assume that $p(\alpha)$ is a homogeneous scalar polynomial which is strictly positive over the region $\mathbf{R}^k \setminus \{0\}$. Polya's theorem states that there exists a natural number ζ such that the coefficients of the polynomial $(\alpha_1 + \cdots + \alpha_k)^\zeta p(\alpha)$ are all nonnegative [7]. Hence, it is easy to show that $(\alpha_1^2 + \cdots + \alpha_k^2)^\zeta p(\alpha^2)$ is SOS, where $\alpha^2 := [\alpha_1^2 \cdots \alpha_k^2]$. Now, assume that $p(\alpha)$ is an arbitrary polynomial (not necessarily a homogeneous one). Obviously, if there exists a set of SOS polynomials $p_0(\alpha), p_1(\alpha), \dots, p_k(\alpha)$ such that:

$$p(\alpha) = p_0(\alpha) + p_1(\alpha)f_1(\alpha) + \cdots + p_k(\alpha)f_k(\alpha) \quad (6)$$

then the polynomial $p(\alpha)$ is nonnegative over the region \mathcal{D} (due to the fact that $p_0(\alpha), p_1(\alpha), \dots, p_k(\alpha)$ are always nonnegative, and $f_1(\alpha), \dots, f_k(\alpha)$ are also non-negative over the region \mathcal{D}). The question arises as to whether the converse statement is true: given a polynomial $p(\alpha)$ that is nonnegative over the region \mathcal{D} , can it be written in the form (6) for some SOS polynomials $p_0(\alpha), p_1(\alpha), \dots, p_k(\alpha)$? The answer to this question is negative. However, Putinar's theorem states that if the polynomial $p(\alpha)$ is strictly positive over the region \mathcal{D} , then it can be expressed as (6), provided the region \mathcal{D} satisfies some mild conditions [20].

The above results will be used in the present chapter, but in a more general case when $p(\alpha)$ is a matrix polynomial.

3 Robust Controllability Degree

The following lemma is a generalized form of Lemma 1 in [4], and presents some interesting properties of the controllability Gramian.

Lemma 1 *There exist a matrix polynomial $H(\alpha)$ and a scalar polynomial $h(\alpha)$ such that the Gramian $W_c(\alpha)$ can be written as $\frac{H(\alpha)}{h(\alpha)}$, where:*

- $H(\alpha)$ is positive semi-definite for all $\alpha \in \mathcal{D}$.
- $h(\alpha)$ is strictly positive for all $\alpha \in \mathcal{D}$.

Proof. It is known that the Gramian matrix $W_c(\alpha)$, $\alpha \in \mathcal{D}$, is the unique solution of the following continuous-time Lyapunov equation:

$$A(\alpha)W_c(\alpha) + W_c(\alpha)A(\alpha)^T = -\tilde{B}(\alpha)\tilde{B}(\alpha)^T \quad (7)$$

Hence, one can write:

$$[I \otimes A(\alpha) + A(\alpha) \otimes I] \text{vec}\{W_c(\alpha)\} = \text{vec}\{-\tilde{B}(\alpha)\tilde{B}(\alpha)^T\} \quad (8)$$

where \otimes denotes the Kronecker product, and $\text{vec}\{\cdot\}$ is an operator which takes a matrix and converts it to a vector by stacking its columns on top of one another. Define now:

$$\begin{aligned} h(\alpha) &:= (-1)^n \det\{I \otimes A(\alpha) + A(\alpha) \otimes I\}, \quad \forall \alpha \in \mathbf{R}^k \\ H(\alpha) &:= h(\alpha)W_c(\alpha), \quad \forall \alpha \in \mathcal{D} \end{aligned} \quad (9)$$

Note that although $h(\alpha)$ is defined over the entire k -dimensional space, $H(\alpha)$ is defined only over the uncertainty region due to the fact that the Gramian $W_c(\alpha)$ used in its definition may not be well-defined outside the uncertainty region. It can be concluded from (8) and (9) that $H(\alpha)$ and $h(\alpha)$ are matrix and scalar polynomials, respectively, over the region \mathcal{D} . The domain of definition of $H(\alpha)$ can be easily extended to the whole space \mathbf{R}^k . Therefore, assume that $H(\alpha)$ and $h(\alpha)$ are two polynomials defined over the entire space, which satisfy the relations in (9) for every $\alpha \in \mathcal{D}$. To prove that $h(\alpha)$ is strictly positive over the uncertainty region, first define $\lambda_1(\alpha), \lambda_2(\alpha), \dots, \lambda_n(\alpha)$ as the eigenvalues of the matrix $A(\alpha)$. Due to a well-known property of the Kronecker product, one can write:

$$h(\alpha) = (-1)^n \prod_{i=1}^n \prod_{j=1}^n (\lambda_i(\alpha) + \lambda_j(\alpha)) \quad (10)$$

Given a fixed $\alpha \in \mathcal{D}$, assume with no loss of generality that $\lambda_1(\alpha), \dots, \lambda_q(\alpha)$ are real numbers and that the remaining eigenvalues $\lambda_{q+1}(\alpha), \dots, \lambda_n(\alpha)$ are non-real complex numbers. A few observations can be made as follows:

- For every $i, j \in \{1, 2, \dots, n\}$ such that either $i > q$ or $j > q$, both of the terms $\lambda_i(\alpha) + \lambda_j(\alpha)$ and $\lambda_i(\alpha)^* + \lambda_j(\alpha)^*$ appear in the product given in the right side of (10), which makes the product strictly positive.
- For every $i, j \in \{1, 2, \dots, q\}$, the term $\lambda_i(\alpha) + \lambda_j(\alpha)$ is strictly negative (as the system is robustly stable). Moreover, there are q^2 of such terms in (10).
- Since the real-valued matrix $A(\alpha)$ has an even number of non-real complex eigenvalues, $n - q$ is an even number.

These facts lead to the conclusion that the scalar polynomial $h(\alpha)$ is strictly positive for all $\alpha \in \mathcal{D}$. On the other hand, the definition of the Gramian $W_c(\alpha)$ implies that it is positive-semidefinite over the uncertainty region. As a result, the definition of $H(\alpha)$ in (9) yields that this matrix is positive semi-definite for all $\alpha \in \mathcal{D}$. \square

In light of Lemma 1, the Gramian matrix $W_c(\alpha)$ is normally a non-polynomial rational function. This property impedes the use of the available SOS techniques, as they merely deal with polynomials. Hence, it is desirable to approximate $W_c(\alpha)$ by a polynomial. One naive way to do so is to replace the exponential terms in (4a) with some truncated Taylor series. However, the resultant polynomial approximation would not necessarily satisfy any important properties (namely the ones that will be introduced in Theorem 1 and are essential to the development of this chapter). Hence, a more advanced approximation technique will be provided in the sequel.

Notation 1 Given a matrix M , $\underline{\sigma}\{M\}$ denotes its minimum singular value.

Notation 2 Given a vector $\beta = [\beta_1, \beta_2, \dots, \beta_k]$, let β^2 be defined as $\beta^2 = [\beta_1^2, \beta_2^2, \dots, \beta_k^2]$.

Assumption 1 The set \mathcal{D} is compact, and there exist SOS scalar polynomials $w_0(\alpha), w_1(\alpha), \dots, w_z(\alpha)$ such that all vectors α satisfying the inequality:

$$w_0(\alpha) + w_1(\alpha)f_1(\alpha) + \dots + w_z(\alpha)f_z(\alpha) \geq 0 \quad (11)$$

form a compact set.

There are two important points concerning Assumption 1. First, the validity of this assumption can be checked by solving a proper SOS problem. Furthermore, if the assumption does not hold, then the results of this chapter will become only sufficient, as opposed to both necessary and sufficient.

Consider $H(\alpha)$ and $h(\alpha)$ introduced in Lemma 1. It follows from the positiveness of $h(\alpha)$ and the compactness of \mathcal{D} that there exist reals μ_1 and μ_2 such that:

$$0 < \mu_1 < h(\alpha) < \mu_2, \quad \forall \alpha \in \mathcal{D} \quad (12)$$

Definition 1 Define $P_i(\alpha)$, $\alpha \in \mathcal{D}$, to be:

$$P_i(\alpha) := W_c(\alpha) \times \left(1 - \left(1 - \frac{h(\alpha)}{\mu_2} \right)^{2i} \right), \quad i \in \mathbb{N} \quad (13)$$

The next theorem provides a means to approximate the rational controllability Gramian by a polynomial with any arbitrary precision.

Theorem 1 The following statements are true:

(i) For every $i \in \mathbb{N}$, $P_i(\alpha)$ is a positive semi-definite matrix polynomial over the region \mathcal{D} that satisfies the matrix inequality:

$$A(\alpha)P_i(\alpha) + P_i(\alpha)A(\alpha)^T + \tilde{B}(\alpha)\tilde{B}(\alpha)^T \geq 0 \quad (14)$$

(ii) Given $\alpha \in \mathcal{D}$, the sequence $\{P_i(\alpha)\}_1^\infty$ converges to $W_c(\alpha)$ from below monotonically, i.e. $P_1(\alpha) \leq P_2(\alpha) \leq P_3(\alpha) \leq \dots$ and $\lim_{i \rightarrow \infty} \|P_i(\alpha) - W_c(\alpha)\| = 0$. Moreover:

$$P_i(\alpha) \leq W_c(\alpha) \leq \left(1 - \left(1 - \frac{\mu_1}{\mu_2} \right)^{2i} \right)^{-1} P_i(\alpha) \quad (15)$$

Proof of Part (i). The function $P_i(\alpha)$ being matrix polynomial is a consequence of the following facts:

- In light of Lemma 1, $W_c(\alpha)$ can be written as $\frac{H(\alpha)}{h(\alpha)}$.

- The polynomial:

$$\left(1 - \left(1 - \frac{h(\alpha)}{\mu_2}\right)^{2i}\right) \quad (16)$$

is divisible by $1 - (1 - \frac{h(\alpha)}{\mu_2})$, and hence is divisible by $h(\alpha)$ as well.

On the other hand, one can write:

$$\begin{aligned} A(\alpha)P_i(\alpha) + P_i(\alpha)A(\alpha)^T + \tilde{B}(\alpha)\tilde{B}(\alpha)^T &= \left(1 - \left(1 - \frac{h(\alpha)}{\mu_2}\right)^{2i}\right) \\ &\times (A(\alpha)W_c(\alpha) + W_c(\alpha)A(\alpha)^T) + \tilde{B}(\alpha)\tilde{B}(\alpha)^T \\ &= \tilde{B}(\alpha)\tilde{B}(\alpha)^T \left(1 - \frac{h(\alpha)}{\mu_2}\right)^{2i} \geq 0 \end{aligned} \quad (17)$$

This completes the proof.

Proof of Part (ii). The proof of this part follows directly from Definition 1 and the inequality (12). \square

Corollary 1 *The minimum singular values of $W_c(\alpha)$ and $P_i(\alpha)$ are related to each other by the following inequalities:*

$$\left(1 - \left(1 - \frac{\mu_1}{\mu_2}\right)^{2i}\right) \min_{\alpha \in \mathcal{D}} \underline{\sigma}\{W_c(\alpha)\} \leq \min_{\alpha \in \mathcal{D}} \underline{\sigma}\{P_i(\alpha)\} \leq \min_{\alpha \in \mathcal{D}} \underline{\sigma}\{W_c(\alpha)\} \quad (18)$$

where $i = 1, 2, \dots$

Proof. The proof is an immediate consequence of Theorem 1. \square

Theorem 1 shows that $W_c(\alpha)$ can be approximated, with any arbitrary precision, by a matrix polynomial satisfying a certain matrix inequality. Moreover, implicit bounds on the smallest singular value of the controllability matrix are provided in Corollary 1. In order to obtain a proper range of values for the degree of a polynomial which can approximate $W_c(\alpha)$ satisfactorily, it is required to know the ratio $\frac{\mu_1}{\mu_2}$ *a priori*. Fortunately, the relation (10) can be used to obtain this quantity, as it indicates that this ratio is, roughly speaking, related to the minimum and maximum eigenvalues of $A(\alpha)$ over the region \mathcal{D} . Note that the ratio $\frac{\mu_1}{\mu_2}$ quantifies the uncertainty degree of the open-loop system (i.e. the matrix $A(\alpha)$) in terms of the location of the eigenvalues. Let an optimization problem be introduced in the sequel.

Optimization 1 *Given the system $S(\alpha)$ and the uncertainty region \mathcal{D} , maximize the real-valued scalar variable μ subject to the constraint that there exist a symmetric matrix polynomial $P(\alpha)$ and SOS matrix polynomials $S_0(\alpha), \dots, S_z(\alpha), \tilde{S}_0(\alpha), \dots, \tilde{S}_z(\alpha)$ (all $n \times n$) such that:*

$$A(\alpha)P(\alpha) + P(\alpha)A(\alpha)^T + \tilde{B}(\alpha)\tilde{B}(\alpha)^T = S_0(\alpha) + \sum_{i=1}^z S_i(\alpha)f_i(\alpha) \quad (19a)$$

$$P(\alpha) = \mu I_n + \tilde{S}_0(\alpha) + \sum_{i=1}^z \tilde{S}_i(\alpha)f_i(\alpha) \quad (19b)$$

where I_n is the $n \times n$ identity matrix. Denote the solution of this optimization problem with μ^* .

Theorem 2 The quantity $\min_{\alpha \in \mathcal{D}} \underline{\sigma}\{W_c(\alpha)\}$ is equal to μ^* .

Proof. Let $P(\alpha)$ be a matrix polynomial for which there exist a real μ and SOS matrix polynomials $S_0(\alpha), \dots, S_z(\alpha), \tilde{S}_0(\alpha), \dots, \tilde{S}_z(\alpha)$ satisfying the equalities given in (19). One can write:

$$A(\alpha)P(\alpha) + P(\alpha)A(\alpha)^T + \tilde{B}(\alpha)\tilde{B}(\alpha)^T \geq 0 \quad (20a)$$

$$P(\alpha) \geq \mu I_n \quad (20b)$$

for all $\alpha \in \mathcal{D}$. It follows from (7) and (20a) that:

$$A(\alpha)(W_c(\alpha) - P(\alpha)) + (W_c(\alpha) - P(\alpha))A(\alpha)^T \leq 0, \quad \forall \alpha \in \mathcal{D} \quad (21)$$

Therefore, $P(\alpha) \leq W_c(\alpha)$, $\forall \alpha \in \mathcal{D}$. This, together with the inequality (20b), yields:

$$\mu \leq \min_{\alpha \in \mathcal{D}} \underline{\sigma}\{W_c(\alpha)\} \quad (22)$$

As a result:

$$\mu^* \leq \min_{\alpha \in \mathcal{D}} \underline{\sigma}\{W_c(\alpha)\} \quad (23)$$

On the other hand, notice that:

$$P_j(\alpha) > (\min_{\alpha \in \mathcal{D}} \underline{\sigma}\{P_j(\alpha)\} - \varepsilon)I_n, \quad \forall \alpha \in \mathcal{D} \quad (24)$$

for any $j \in \mathbb{N}$ and $\varepsilon \in \mathbf{R}^+$ (note that \mathbf{R}^+ represents the set of positive real numbers). Therefore, it can be inferred from Theorem 1 and Assumption 1 of the present work, and Theorem 2 in [23] (i.e. the general matrix form of Putinar's theorem given in the previous section) that there exist SOS matrix polynomials $\tilde{S}_0(\alpha), \dots, \tilde{S}_z(\alpha)$ such that:

$$P_j(\alpha) = (\min_{\alpha \in \mathcal{D}} \underline{\sigma}\{P_j(\alpha)\} - \varepsilon)I_n + \tilde{S}_0(\alpha) + \sum_{i=1}^z \tilde{S}_i(\alpha)f_i(\alpha) \quad (25)$$

Now, recall from Theorem 1 that:

$$A(\alpha)P_j(\alpha) + P_j(\alpha)A(\alpha)^T + \tilde{B}(\alpha)\tilde{B}(\alpha)^T \geq 0 \quad (26)$$

Since the above inequality cannot be simply relaxed to strict inequality, Putinar's theorem cannot be used directly (as explained in Sect. 2.1). To bypass this obstacle, one can use the result of Theorem 2 in [23] that there exist SOS matrix polynomials $\bar{S}_0(\alpha), \dots, \bar{S}_z(\alpha)$ such that:

$$\left(1 - \frac{h(\alpha)}{\mu_2}\right)^{2j} = \bar{S}_0(\alpha) + \sum_{i=1}^z \bar{S}_i(\alpha)f_i(\alpha) \quad (27)$$

because the left side of the above equation is always strictly positive over the region \mathcal{D} . Hence, it is concluded from (17) that:

$$\begin{aligned}
 A(\alpha)P_j(\alpha) + P_j(\alpha)A(\alpha)^T + \tilde{B}(\alpha)\tilde{B}(\alpha)^T &= \tilde{B}(\alpha)\tilde{B}(\alpha)^T \left(1 - \frac{h(\alpha)}{\mu_2}\right)^{2j} \\
 &= \bar{S}_0(\alpha)\tilde{B}(\alpha)\tilde{B}(\alpha)^T \\
 &\quad + \sum_{i=1}^z \bar{S}_i(\alpha)\tilde{B}(\alpha)\tilde{B}(\alpha)^T f_i(\alpha) \\
 &= S_0(\alpha) + \sum_{i=1}^z S_i(\alpha)f_i(\alpha)
 \end{aligned} \tag{28}$$

where:

$$S_i(\alpha) := \bar{S}_i(\alpha)\tilde{B}(\alpha)\tilde{B}(\alpha)^T, \quad i = 0, 1, \dots, k \tag{29}$$

are SOS matrix polynomials. It results from (25) and (28) that:

$$\min_{\alpha \in \mathcal{D}} \underline{\sigma}\{P_j(\alpha)\} - \varepsilon \leq \mu^* \tag{30}$$

The proof is completed by taking the inequalities (23) and (30) into consideration and letting i and ε go to infinity and zero, respectively, and by using Corollary 1. \square

Theorem 2 provides a methodology for finding the quantity $\min_{\alpha \in \mathcal{D}} \underline{\sigma}\{W_c(\alpha)\}$ indirectly via solving Optimization 1. This optimization problem is in the form of SOS, which can be handled by a semi-definite program. For this purpose, one can use a proper software tool such as YALMIP or SOSTOOLS [14, 15]. Nevertheless, it is first required to consider some upper bounds *a priori* on the degrees of the polynomials involved in the corresponding optimization problem, from which a *lower bound* on the solution of Optimization 1 can be found. In other words, this optimization problem can be formulated as a hierarchy of SDP problems, whose solutions converge asymptotically to the quantity of interest, i.e. $\min_{\alpha \in \mathcal{D}} \underline{\sigma}\{W_c(\alpha)\}$, from below.

Remark 1 *Despite the fact that a high-order rational matrix ($W_c(\alpha)$ in the present problem) cannot, in general, be approximated satisfactorily by a low-order polynomial matrix $P(\alpha)$, it will be illustrated later in an example that a relatively low-order polynomial typically works well here. This is due to the fact that these two functions need to be sufficiently close only at a critical point corresponding to the solution of the optimization problem (as opposed to everywhere in the region \mathcal{D}).*

3.1 Special Case: A Polytopic Region

Although Theorem 2 provides a numerically tractable method for measuring the robust controllability of a system, the proposed optimization problem can be simplified

significantly for special cases of interest. For instance, assume that \mathcal{D} is a polytopic region \mathcal{P} given by:

$$\mathcal{P} = \{\alpha | \alpha_1 + \cdots + \alpha_k = 1, \alpha_1, \dots, \alpha_k \geq 0\} \quad (31)$$

This type of uncertainty region is of particular interest, due to its important applications.

Assumption 2 Assume that $A(\alpha)$ and $\tilde{B}(\alpha)$ are homogeneous matrix polynomials, and let their degrees be denoted by ζ_1 and ζ_2 , respectively.

Note that Assumption 2 holds automatically for polytopic systems, with $\zeta_1 = \zeta_2 = 1$.

Theorem 3 The quantity $\min_{\alpha \in \mathcal{P}} \underline{\sigma} W_c(\alpha)$ is equal to the maximum value of μ for which there exists a homogeneous matrix polynomial $\tilde{P}(\alpha)$ satisfying the following inequalities for all $\alpha \in \mathbf{R}^k$:

$$\tilde{P}(\alpha^2) \geq 0, \quad (32a)$$

$$\begin{aligned} & \left[A(\alpha^2) \left(\tilde{P}(\alpha^2) + \mu(\alpha\alpha^T)^{\zeta_3} I_n \right) + \left(\tilde{P}(\alpha^2) + \mu(\alpha\alpha^T)^{\zeta_3} I_n \right) A^T(\alpha^2) \right] \\ & \times (\alpha\alpha^T)^{\max(0, 2\zeta_2 - \zeta_1 - \zeta_3)} + \tilde{B}(\alpha^2) \tilde{B}(\alpha^2)^T (\alpha\alpha^T)^{\max(0, \zeta_3 + \zeta_1 - 2\zeta_2)} \geq 0 \end{aligned} \quad (32b)$$

where ζ_3 denotes the degree of the polynomial $\tilde{P}(\alpha)$.

Proof. The proof follows by refining the proofs of Theorems 1 and 2 in such a way that the homogeneity of the system matrices as well as the polytopic structure of the uncertainty region are both taken into account. To this end, the homogenization technique given in [4] is adopted. First of all, notice that the equation (8) yields:

$$\text{vec}\{W_c(\alpha)\} = -[I \otimes A(\alpha) + A(\alpha) \otimes I]^{-1} \text{vec}\{\tilde{B}(\alpha)\tilde{B}(\alpha)^T\} \quad (33)$$

Note also that $I \otimes A(\alpha) + A(\alpha) \otimes I$ and $\tilde{B}(\alpha)\tilde{B}(\alpha)^T$ are both homogeneous polynomials. Therefore, one can conclude that the polynomials $H(\alpha)$ and $h(\alpha)$ introduced in Lemma 1 can both be assumed to be homogeneous. Now, let $P_i(\alpha)$ be defined as:

$$P_i(\alpha) := W_c(\alpha) \times \left(\left(\sum_{j=1}^k \alpha_j \right)^{2i\zeta_4} - \left(\left(\sum_{j=1}^k \alpha_j \right)^{\zeta_4} - \frac{h(\alpha)}{\mu_2} \right)^{2i} \right) \quad (34)$$

in lieu of the one defined in (13), where ζ_4 denotes the degree of $h(\alpha)$. Note that $P_i(\alpha)$ given above is identical to the one defined in (13) over the region \mathcal{P} , and its subtle difference is that it is homogeneous. In other words, $W_c(\alpha)$ is approximated by a homogeneous polynomial here. It is easy to show that Theorem 1 holds for the

polynomial $P_i(\alpha)$ defined in (34). On the other hand, due to the polytopic nature of the set \mathcal{P} in (31), the inequality (14) in Theorem 1 can be rewritten as:

$$\begin{aligned} & (A(\alpha)P_i(\alpha) + P_i(\alpha)A(\alpha)^T) \left(\sum_{j=1}^k \alpha_j \right)^{\max(0, 2\zeta_2 - \zeta_1 - \zeta_3)} \\ & + \tilde{B}(\alpha)\tilde{B}(\alpha)^T \left(\sum_{j=1}^k \alpha_j \right)^{\max(0, \zeta_1 + \zeta_3 - 2\zeta_2)} \geq 0, \quad \forall \alpha \in \mathcal{P} \end{aligned} \quad (35)$$

where ζ_3^i denotes the degree of $P_i(\alpha)$. One can write:

$$\begin{aligned} \tilde{P}_i(\alpha) &:= P_i(\alpha) - \min_{\alpha \in \mathcal{P}} \{P_i(\alpha)\} \left(\sum_{j=1}^k \alpha_j \right)^{\zeta_3^i} I_n \\ &= P_i(\alpha) - \min_{\alpha \in \mathcal{P}} \{P_i(\alpha)\} I_n \geq 0, \quad \forall \alpha \in \mathcal{P} \end{aligned} \quad (36)$$

Hence:

$$\tilde{P}_i(\alpha) \geq 0, \quad \forall \alpha \in \mathcal{P} \quad (37)$$

Notice now the following facts:

- A matrix polynomial $M(\alpha)$ is nonnegative over \mathcal{P} if and only if $M(\alpha^2)$ is non-negative over the unit sphere $\tilde{\mathcal{D}}$ defined as follows:

$$\tilde{\mathcal{D}} = \{\alpha | \alpha_1^2 + \cdots + \alpha_k^2 = 1\} \quad (38)$$

- Assume that $N(\alpha)$ is a homogeneous matrix polynomial of degree γ . One can write:

$$N(\beta) = \|\beta\|^\gamma \times N\left(\frac{\beta}{\|\beta\|}\right), \quad \forall \beta \neq 0 \quad (39)$$

Since $\frac{\beta}{\|\beta\|} \in \tilde{\mathcal{D}}$, one can conclude that $N(\alpha)$ is nonnegative over $\tilde{\mathcal{D}}$ if and only if it is nonnegative over the whole space \mathbf{R}^k .

Thus, the proof is completed by pursuing the argument given in the proof of Theorem 2, and using the inequalities (35) and (37) after replacing α with α^2 and removing the constraint $\alpha \in \mathcal{P}$ (note that the polynomial in the left side of the inequality (35) is homogeneous). \square

Optimization 2 Maximize μ subject to the constraint that there exist a homogeneous matrix polynomial $P(\alpha)$ and SOS matrix polynomials $S_1(\alpha)$ and $S_2(\alpha)$ (all $n \times n$) such that:

$$P(\alpha^2) = S_1(\alpha), \quad (40a)$$

$$\begin{aligned} & \left[A(\alpha^2) \left(P(\alpha^2) + \mu(\alpha\alpha^T)^{\zeta_3} I_n \right) + \left(P(\alpha^2) + \mu(\alpha\alpha^T)^{\zeta_3} I_n \right) A^T(\alpha^2) \right] \\ & \times (\alpha\alpha^T)^{\max(0, 2\zeta_2 - \zeta_1 - \zeta_3)} + \tilde{B}(\alpha^2)\tilde{B}(\alpha^2)^T (\alpha\alpha^T)^{\max(0, \zeta_3 + \zeta_1 - 2\zeta_2)} = S_2(\alpha) \end{aligned} \quad (40b)$$

where ζ_3 denotes the degree of the polynomial $P(\alpha)$. Denote the solution of this optimization problem with $\tilde{\mu}^*$.

The following lemma is required in order to delve into the properties of the optimal parameter $\tilde{\mu}^*$ defined above.

Lemma 2 *Let $M(\alpha)$ be a homogeneous matrix polynomial with the property that $M(\alpha^2)$ is positive definite for every $\alpha \in \mathbf{R}^k \setminus \{0\}$. There exists a natural number c so that $(\alpha\alpha^T)^c M(\alpha^2)$ is SOS.*

Proof. The proof of this lemma relies heavily on the extension of Polya's theorem [7] to the matrix case, as carried out in [23]. More precisely, since $M(\alpha)$ is positive definite over the polytope \mathcal{P} , it follows from Theorem 3 in [23] that there exists a natural number c such that $(\alpha_1 + \alpha_2 + \dots + \alpha_k)^c M(\alpha)$ has only positive semi-definite matrix coefficients. This implies that the coefficients of $(\alpha\alpha^T)^c M(\alpha^2)$ are all positive semi-definite, and in addition, its monomials are squared terms. As a result, $(\alpha\alpha^T)^c M(\alpha^2)$ is SOS. \square

Theorem 4 *The quantity $\min_{\alpha \in \mathcal{P}} \{W_c(\alpha)\}$ is equal to $\tilde{\mu}^*$.*

Proof. The proof will be performed in two steps. First, observe that if (40) holds for some matrices $P(\alpha)$, $S_1(\alpha)$ and $S_2(\alpha)$, then (32) is satisfied for $\tilde{P}(\alpha) = P(\alpha)$.

Conversely, assume that a matrix polynomial $\tilde{P}(\alpha)$ satisfies the inequalities given in (32), with the non-strict inequalities replaced by strict inequalities for every $\alpha \in \mathbf{R}^k \setminus \{0\}$ (such strict inequalities correspond to the case when the term *maximum* is substituted by *supremum*). Note that for a strict inequality in (32a), one would need to replace $\tilde{P}(\alpha)$ and μ with $\tilde{P}(\alpha) - \varepsilon(\alpha\alpha^T)^{\zeta_3}$ and $\mu + \varepsilon$, respectively, for a positive infinitesimal number ε . Now, one can apply Lemma 2 to these inequalities to conclude that there exists a natural number c such that if the expressions in the left sides of the inequalities (32a) and (32b) are multiplied by $(\alpha\alpha^T)^c$, then they become SOS matrix polynomials. It is enough to choose $P(\alpha)$ as $\tilde{P}(\alpha)(\alpha\alpha^T)^c$ for the inequalities given in (40) to hold (for some appropriate matrices $S_1(\alpha)$ and $S_2(\alpha)$). \square

Remark 2 *Optimization 2 is obtained from Theorem 3 by replacing the nonnegativity constraints with SOS conditions. Although one may argue that this can potentially introduce some conservatism due to the known gap between the set of SOS polynomials and the set of nonnegative polynomials, Theorem 4 shows that this is not the case. In other words, the proposed replacement does not make the resultant conditions conservative at all.*

Note that Optimization 2 can be handled using proper software as noted in Remark 1, provided some *a priori* upper bounds are set on the degrees of the relevant polynomials being sought. The question arises as to whether choosing higher degree polynomials would result in a tighter (less conservative) solution. This question is addressed in the next corollary.

Corollary 2 *The solution of Optimization 2 is a monotone nondecreasing function with respect to ζ_3 (the degree of the polynomial $\tilde{P}(\alpha)$).*

Proof. The proof is a direct consequence of the fact that if $\tilde{P}(\alpha)$ satisfies the constraints of Optimization 2 for some μ , then $\tilde{P}(\alpha)(\alpha_1 + \dots + \alpha_k)$ also satisfies them for the same μ , but for some other suitable matrices $S_1(\alpha)$ and $S_2(\alpha)$. \square

3.2 Comparison with Existing Results

It follows immediately from the result of [2] that the Gramian matrix can be approximated by a polynomial. Nonetheless, the relationship between the order of the corresponding polynomial and the approximation error is not investigated in [2]. In contrast, it is shown in Theorem 1 of the present chapter that the approximation error reduces exponentially with respect to the order of the corresponding polynomial. In particular, given a desired error bound, one can find the required degree of the approximating polynomial by computing μ_1 and μ_2 which depend on the smallest and largest eigenvalues of $A(\alpha)$ over the region \mathcal{P} . It is worth noting that the exponential reduction of the error is only for the approximation of the Gramian rational matrix with a polynomial matrix. Hence, for any fixed order of the polynomial approximation, Optimization 1 is to be solved in which the degrees of the polynomials $S_0(\alpha), \dots, S_z(\alpha), \tilde{S}_0(\alpha), \dots, \tilde{S}_z(\alpha)$ may be undesirably large. This issue arises in almost all optimization problems derived based on Polya's or Putinar's theorems.

As far as the complexity is concerned, it is easy to verify that Optimization 2 introduced in the present work is basically as complex as the optimization problem tackled in [4]. This is partly due to the fact that there are an SOS homogeneous polynomial and two SOS constraints of a particular form in both approaches. For a detailed comparison between the complexities of the results given in [4], [18] and [13], the interested reader may refer to [13].

A recent paper [17] deals with the robust analysis and synthesis of linear uncertain systems. Although this paper presents useful results in the LMI framework, Optimization 2 is a much simpler problem compared to the ones proposed in [17]. It is noteworthy that all of the papers surveyed above deal with an inequality of the form:

$$A(\alpha)P(\alpha) + P(\alpha)A(\alpha)^T + \tilde{B}(\alpha)\tilde{B}(\alpha)^T \leq 0 \quad (41)$$

whereas this work studies the opposite inequality (see Theorem 1):

$$A(\alpha)P(\alpha) + P(\alpha)A(\alpha)^T + \tilde{B}(\alpha)\tilde{B}(\alpha)^T \geq 0 \quad (42)$$

This is a consequence of the fact that it is desired to *maximize* the *minimum* singular value. In other words, due to the special structure of the optimization considered here, the existing results cannot be directly applied to the underlying problem.

4 Numerical Example

Example 1

A simple example is given here to profoundly illustrate the results of this work. Consider an LTI uncertain fourth-order system with the following state-space matrices:

$$\begin{aligned} A(\alpha) &= \begin{bmatrix} -0.65 & -0.6 & -1 & 1.5 \\ -0.05 & -0.95 & -1.85 & -0.4 \\ -0.7 & 1.6 & -3.05 & 0.2 \\ -1.5 & -0.1 & -2.85 & -0.35 \end{bmatrix} \alpha_1 + \begin{bmatrix} -0.75 & 1.4 & 0 & 2.3 \\ -1.35 & -1.75 & -0.65 & -1.1 \\ 0.8 & 2.2 & -3.25 & -2.5 \\ -0.5 & 1.7 & 0.15 & 0.45 \end{bmatrix} \alpha_2, \\ B(\alpha) &= \begin{bmatrix} -1.2 & -0.3 & -0.85 & 1.55 \\ 0.35 & -1.05 & -1.3 & -1.8 \\ 0.25 & 1.2 & -2.5 & -0.8 \\ -0.3 & 1.45 & -1 & -0.3 \end{bmatrix} \alpha_1 + \begin{bmatrix} 0.45 & 1.7 & 0.85 & 0.75 \\ -1.7 & -0.7 & 0.65 & 0.7 \\ 0.55 & 1 & -0.75 & -1.7 \\ -0.2 & 0.25 & 1.15 & 0.75 \end{bmatrix} \alpha_2 \end{aligned} \quad (43)$$

where α_1 and α_2 are the uncertain parameters of the system, which belong to the polytope $\mathcal{P} = \{(\alpha_1, \alpha_2) | \alpha_1 + \alpha_2 = 1, \alpha_1, \alpha_2 \geq 0\}$. Regard (43) as an interconnected system with four inputs. It is desired to determine which inputs contribute weakly to the control of the system, and hence can be ignored for the sake of cost reduction. In other words, the objective is to find out which inputs play a vital role in controlling the system. To this end, for any given set $g \subset \{1, 2, 3, 4\}$ let $\mathcal{S}^g(\alpha)$ represent the system $\mathcal{S}(\alpha)$ after ignoring those inputs whose indices belong to g . Denote the controllability Gramian of this system with $W_c^g(\alpha)$.

Given $\alpha \in \mathcal{P}$ and a final state x_0 of unit norm, consider the problem of finding an input $u(t)$ over the time interval $(-\infty, 0]$ with minimum L_2 norm such that it drives the state of the system $\mathcal{S}^g(\alpha)$ from $x(-\infty) = 0$ to $x(0) = x_0$. As discussed in Sect. 4.3 of [6], one possible solution is given by:

$$u_{opt}(t) = B^g(\alpha)^T e^{-A(\alpha)^T t} W_c^g(\alpha)^{-1} x_0, \quad t \leq 0 \quad (44)$$

where $B^g(\alpha)$ is the B -matrix of the system $\mathcal{S}^g(\alpha)$. The optimal input energy can be computed as:

$$\|u_{opt}\|^2 = x_0^T W_c^g(\alpha)^{-1} x_0 \quad (45)$$

Define $v(g)$ to be the maximum value of this optimal input energy over all final states x_0 of unit norm and all $\alpha \in \mathcal{P}$. The idea behind this definition is that $v(g)$ provides an upper bound for the input energy required to drive the state of the uncertain system $\mathcal{S}^g(\alpha)$, $\alpha \in \mathcal{P}$, from the origin at time $t = -\infty$ to any arbitrary point in the unit ball at time $t = 0$. Note that if $v(g)$ corresponds to a final state x_0^* and an uncertain parameter α^* , then x_0^* must be a unit eigenvector of $W_c^g(\alpha^*)$ associated with its smallest eigenvalue (singular value). This relationship can be expressed by:

$$\frac{1}{v(g)} = \min_{\alpha \in \mathcal{P}} \underline{\sigma}\{W_c^g(\alpha)\} \quad (46)$$

The objective is to evaluate $v(g)$ for different choices of g . To this end, four cases are considered as follows:

- *Case 1:* $g = \{2, 3\}$.
- *Case 2:* $g = \{1\}$.
- *Case 3:* $g = \{4\}$.
- *Case 4:* $g = \{\}$.

To obtain $v(g)$ for any of the above cases, it suffices to solve Optimization 2 with the appropriate matrix $B^g(\alpha)$. For this purpose, the order of the polynomial $P(\alpha)$ being sought should be chosen *a priori*. This optimization is treated using YALMIP on a Dell laptop with a 1.6 GHz processor and 512 MB memory, and the results are given in Table 1. The last column of this table gives the points (α_1^*, α_2^*) for which the largest optimal input energy $v(g)$ is required. These points are computed by gridding the polytope properly, and performing an exhaustive search. Therefore, the entries of the last column are computed using a brute force technique, which will be exploited to verify the results obtained by solving Optimization 2. The second column of the table gives the solution of Optimization 2 at the second relaxation, i.e., when $P(\alpha)$ is assumed to be a homogeneous polynomial of order 2 (with the monomials $\alpha_1^2, \alpha_2^2, \alpha_1 \alpha_2$). It can be verified that the solution obtained for any of the cases 1, 2 or 3 corresponds to the minimum singular value of the Gramian matrix evaluated at the optimal point given in the last column of the table. This proves that the relaxation arrives at the correct solution for cases 1, 2, and 3. For case 4, Optimization 2 is also solved at the fourth relaxation (by considering the monomials $\alpha_1^4, \alpha_1^3 \alpha_2, \alpha_1^2 \alpha_2^2, \alpha_1^1 \alpha_2^3, \alpha_2^4$ for $P(\alpha)$). The corresponding solution is given in the third column, which is, in fact, the exact optimal value. The CPU time consumed for solving Optimization 2 at the second relaxation is given in the fourth column for each case; these values show that the problem is solved very fast. Using the results in columns 2 and 3 of the table as well as the equation (46), the quantity $v(g)$ is calculated and provided in column 5. It is to be noted that the proposed approach requires that the matrix $A(\alpha)$ be stable (Hurwitz) over the polytope. This can be verified using the method developed in [4]. It is worth mentioning that the ratio $\frac{\mu_1}{\mu_2}$ is equal to 0.075 in this example.

The values given in Table 1 (case 2) imply that the first input of the system is fairly important and ignoring it in the controller design would substantially increase the control energy required for shifting the state vector from certain points in the state-space. In contrast, the last input may be neglected, because its contribution is not significant as reflected by the small minimum singular value (case 3). However,

Table 1 Numerical results for Example 1

Case	Second relaxation	Fourth relaxation	CPU time for second relaxation	$v(g)$	(α_1^*, α_2^*)
1	0.0068	*	0.88 sec	147.06	(0.820, 0.180)
2	0.0744	*	0.82 sec	13.44	(0.240, 0.760)
3	0.1920	*	0.78 sec	5.21	(0.225, 0.775)
4	0.2516	0.2531	0.85 sec	3.97	(0.234, 0.766)

if the second and third inputs are ignored concurrently (case 1), although the system remains robustly controllable by the remaining inputs, the required control energy would be huge.

For each of the above-mentioned cases, let the optimal input corresponding to the worst-case scenario (i.e. the final state x_0^* and the uncertain variable α^*) be applied to the system. Note that this input is given by (44), with $\alpha = \alpha^*$ (provided in Table 1) and $x_0 = x_0^*$ as defined earlier. The resulting input and state of the system are plotted in Figs. 1–4. Notice that although these signals extend from $t = -\infty$ to $t = 0$, they are sketched only on the interval $t \in [-15, 0]$. These figures confirm the theoretical

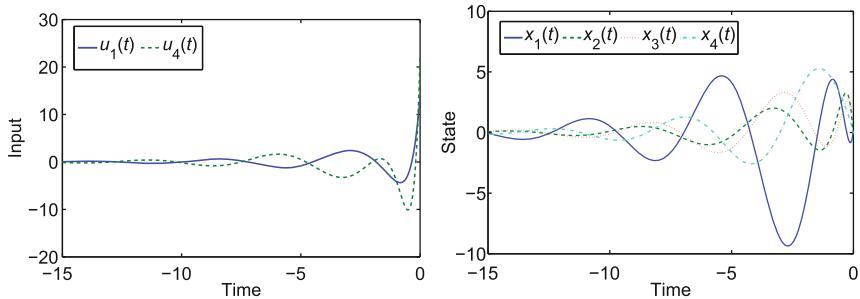


Fig. 1 The input and state of the system in case 1 (i.e., when inputs 2 and 3 are blocked)

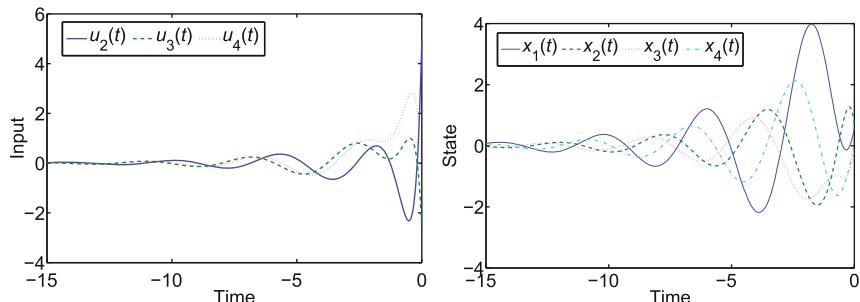


Fig. 2 The input and state of the system in case 2 (i.e., when input 1 is blocked)

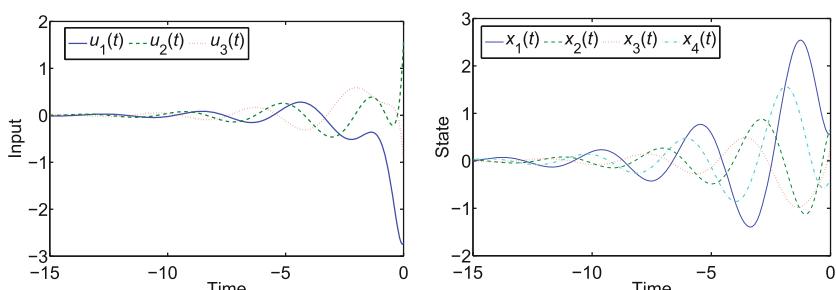


Fig. 3 The input and state of the system in case 3 (i.e., when input 4 is blocked)

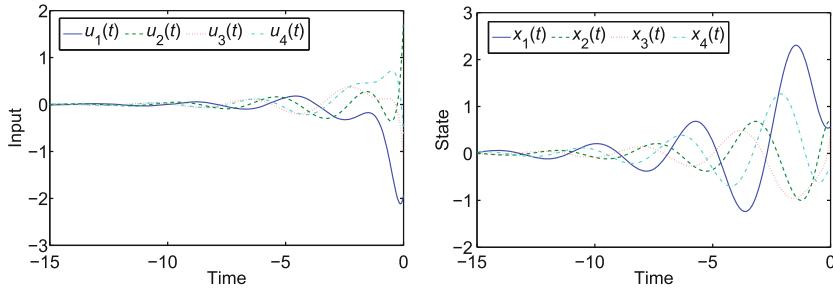


Fig. 4 The input and state of the system in case 4 (i.e., when none of the inputs is blocked)

results obtained in this work. For instance, one can observe that in the case when both inputs $u_2(t)$ and $u_3(t)$ are blocked, the worst-case optimal input of the system has a large overshoot occurring at $t = 0$ (about 22 in magnitude).

5 Summary

Given a continuous-time linear time-invariant (LTI) system which is polynomially uncertain on a semi-algebraic region, in this chapter we obtain the minimum of the smallest singular value of its controllability (observability) Gramian matrix. For this purpose, it is first shown that the Gramian is a rational function which can be approximated by a matrix polynomial (with any arbitrary precision) that satisfies an important relation. A sum-of-squares (SOS) formula is then derived for solving the underlying problem, which can be efficiently handled using proper software. An alternative SOS method is subsequently obtained for the case when the uncertainty region is a polytope. This allows one to measure the robust controllability (observability) degree of the system, when its parameters are subject to perturbation. The method proposed here can be used to find a dominant subset of inputs and outputs for any given large-scale system, by determining the effectiveness of each input and output in the overall operation of the control system. Simulations demonstrate the efficacy of the proposed results.

Acknowledgement The authors would like to acknowledge that this chapter is written based on their recent work [26].

References

1. Blekherman, G.: There are significantly more nonnegative polynomials than sums of squares. *Israel Journal of Mathematics* **153**, 355–380 (2006)
2. Bliman, P. A., Oliveira, R. C. L. F., Montagner, V. F., Peres, P. L. D.: Existence of homogeneous polynomial solutions for parameter-dependent linear matrix inequalities with parameters in the simplex. *Proceedings of 45th IEEE Conference on Decision and Control*. San Diego, USA. 1486–1491 (2006)

3. Chesi, G.: On the non-conservatism of a novel LMI relaxation for robust analysis of polytopic systems. *Automatica* **44**, 2973–2976 (2008)
4. Chesi, G., Garulli, A., Tesi, A., Vicino, A.: Polynomially parameter-dependent Lyapunov functions for robust stability of polytopic systems: an LMI approach. *IEEE Transactions on Automatic Control* **50**, 365–370 (2005)
5. Davison, E. J., Chang, T. N.: Decentralized stabilization and pole assignment for general proper systems. *IEEE Transactions Automatic Control* **35**, 652–664 (1990)
6. Dullerud, G. E., Paganini, F.: A course in robust control theory: a convex approach. *Texts in Applied Mathematics*, Springer, Berlin (2005)
7. Hardy, G. H., Littlewood, J. E., Polya, G.: *Inequalities*. Cambridge University Press, Cambridge, UK, Second edition (1952)
8. Hillar, C. J., Nie, J.: An elementary and constructive solution to Hilbert's 17th Problem for matrices. *Proceedings of the American Mathematical Society* **136**, 73–76 (2008)
9. Kau, S., Liu, Y., Hong, L., Lee, C., Fang, C., Lee, L.: A new LMI condition for robust stability of discrete-time uncertain systems. *Systems and Control Letters* **54**, 1195–1203 (2005)
10. Lavaei, J., Aghdam, A. G.: Optimal periodic feedback design for continuous-time LTI systems with constrained control structure. *International Journal of Control* **80**, 220–230 (2007)
11. Lavaei, J., Aghdam, A. G.: A graph theoretic method to find decentralized fixed modes of LTI systems. *Automatica* **43**, 2129–2133 (2007)
12. Lavaei, J., Aghdam, A. G.: Control of continuous-time LTI systems by means of structurally constrained controllers. *Automatica* **44**, 141–148 (2008)
13. Lavaei, J., Aghdam, A. G.: Robust stability of LTI systems over semi-algebraic sets using sum-of-squares matrix polynomials. *IEEE Transactions on Automatic Control* **53**, 417–423 (2008)
14. Lofberg, J.: A toolbox for modeling and optimization in MATLAB. Proc. of the CACSD Conference. Taipei, Taiwan (2004) Available via <http://control.ee.ethz.ch/~joloef/yalmip.php>
15. Prajna, S., Papachristodoulou, A., Seiler, P., Parrilo, P. A.: SOSTOOLS sum of squares optimization toolbox for MATLAB. Users guide (2004) Available via <http://www.cds.caltech.edu/sostools>
16. Oliveira, M. C., Geromel, J. C.: A class of robust stability conditions where linear parameter dependence of the Lyapunov function is a necessary condition for arbitrary parameter dependence. *Systems and Control Letters* **54**, 1131–1134 (2005)
17. Oliveira, R. C. L. F., Oliveira, M. C., Peres, P. L. D.: Convergent LMI relaxations for robust analysis of uncertain linear systems using lifted polynomial parameter-dependent Lyapunov functions. *Systems and Control Letters* **57**, 680–689 (2008)
18. Oliveira, R. C. L. F., Peres, P. L. D.: LMI conditions for robust stability analysis based on polynomially parameter-dependent Lyapunov functions. *Systems and Control Letters* **55**, 52–61 (2006)
19. Parrilo, P. A.: Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Ph.D. dissertation, California Institute of Technology (2000)
20. Putinar, M.: Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal* **42**, 969–984 (1993)
21. Sastry, S. S., Desoer, C. A.: The robustness of controllability and observability of linear time-varying systems. *IEEE Transactions on Automatic Control* **27**, 933–939 (1982)
22. Savkin, A. V., Petersen, I. R.: Weak robust controllability and observability of uncertain linear systems. *IEEE Transactions on Automatic Control* **44**, 1037–1041 (1999)
23. Scherer, C. W., Hol, C. W. J.: Matrix sum-of-squares relaxations for robust semi-definite programs. *Mathematical Programming* **107**, 189–211 (2006)
24. Siljak, D. D.: *Decentralized control of complex systems*. Academic, Cambridge (1991)
25. Sojoudi, S., Aghdam, A. G.: Characterizing all classes of LTI stabilizing structurally constrained controllers by means of combinatorics. *Proceedings 46th IEEE Conference on Decision and Control*. New Orleans, USA, 4415–4420 (2007)

26. Sojoudi, S., Lavaei, J., Aghdam, A. G.: Robust controllability and observability degrees of polynomially uncertain systems. *Automatica* **45**(11), 2640–2645, Nov. (2009)
27. Ugrinovskii, V. A.: Robust controllability of linear stochastic uncertain systems. *Automatica* **41**, 807–813 (2005)
28. Wang, S. H., Davison, E. J.: On the stabilization of decentralized control systems. *IEEE Transactions on Automatic Control* **18**, 473–478 (1973)

Decentralized Output-Feedback Control of Large-Scale Interconnected Systems via Dynamic High-Gain Scaling

P. Krishnamurthy and F. Khorrami

1 Introduction

Large-scale systems occurring in several application domains (including, as a very short representative list, power systems, multi-agent systems, communication and transportation networks, supply chains, etc.) can be profitably viewed as interconnections of multiple subsystems. In this general context, the development of control algorithms for interconnected large-scale systems has attracted considerable research interest. Interest in decentralized control designs has been significantly renewed in recent years due to noteworthy extensions promised by the application of new results emerging in nonlinear robust and adaptive output-feedback control. In this vein, this chapter addresses the design of decentralized output-feedback controllers for a class of nonlinear interconnected large-scale systems based on our recent results on the dynamic high-gain scaling control design technique. The results here follow the general direction in the decentralized literature of attempting to generalize the form of the dynamics of subsystems and simultaneously weaken the assumptions on subsystem interconnections. We consider a class of interconnected large-scale systems with each subsystem being of the form

$$\begin{aligned}\dot{z}_{(i,m)} &= q_{(i,m)}(z, x, u, t, \boldsymbol{\varpi}), \quad i = 1, \dots, s_m + 1 \\ \dot{x}_{(i,m)} &= \phi_{(i,m)}(z, x, u, t, \boldsymbol{\varpi}) + \phi_{(i,i+1,m)}(x_{(1,m)})x_{(i+1,m)}, \quad i = 1, \dots, s_m - 1 \\ \dot{x}_{(i,m)} &= \phi_{(i,m)}(z, x, u, t, \boldsymbol{\varpi}) + \phi_{(i,i+1,m)}(x_{(1,m)})x_{(i+1,m)} \\ &\quad + \mu_{(i-s_m,m)}(x_{(1,m)})u_m, \quad i = s_m, \dots, n_m - 1\end{aligned}$$

P. Krishnamurthy

Control/Robotics Research Laboratory (CRRL), Department of Electrical
and Computer Engineering, Polytechnic Institute of NYU, Brooklyn, NY 11201, USA
e-mail: pk@crrl.poly.edu

F. Khorrami

Control/Robotics Research Laboratory (CRRL), Department of Electrical and Computer
Engineering, Polytechnic Institute of NYU, Brooklyn, NY 11201, USA
e-mail: khorrami@smart.poly.edu

$$\begin{aligned}\dot{x}_{(n_m,m)} &= \phi_{(n_m,m)}(z, x, u, t, \varpi) + \mu_{(n_m-s_m,m)}(x_{(1,m)})u_m \\ y_m &= x_{(1,m)}\end{aligned}\tag{1}$$

where $x_m = [x_{(1,m)}, \dots, x_{(n_m,m)}]^T \in \mathcal{R}^{n_m}$ is the state, $y_m = x_{(1,m)} \in \mathcal{R}$ the output, $u_m \in \mathcal{R}$ the input, and $z_m = [z_{(1,m)}^T, \dots, z_{(s_m+1,m)}^T]^T \in \mathcal{R}^{n_z(1,m)+\dots+n_z(s_m+1,m)}$ the state of the appended dynamics of the m^{th} subsystem. M is the number of subsystems, $x = [x_1^T, \dots, x_M^T]^T$, $u = [u_1, \dots, u_M]^T$, and $z = [z_1^T, \dots, z_M^T]^T$. $\phi_{(i,i+1,m)}$, $i = 1, \dots, n_m - 1$, and $\mu_{(i,m)}$, $i = 0, \dots, n_m - s_m$, are known scalar real-valued continuous functions. $q_{(i,m)}$, $i = 1, \dots, s_m + 1$, and $\phi_{(i,m)}$, $i = 1, \dots, n_m$, are continuous scalar real-valued uncertain functions. s_m is the relative degree of the m^{th} subsystem. $\varpi \in \mathcal{R}^{n_\varpi}$ is the exogenous disturbance input.

Early results in decentralized control focused on linear systems [1, 5] and linearly bounded interconnections [8, 22]. In [25], higher order (i.e., polynomial type) interconnections were considered for large-scale systems assuming matching conditions. Backstepping-based robust decentralized controllers were designed in [3, 4] for systems of output-feedback canonical form including uncertain parameters and polynomially bounded uncertainties. Using the Cascading Upper Diagonal Dominance (CUDD) based technique in [19], a decentralized output-feedback disturbance attenuation scheme was proposed in [9] for interconnected large-scale systems with each subsystem being in the generalized output-feedback canonical form [19] and with nonlinear appended dynamics.

In a recent sequence of papers, we have proposed a new dynamic high-gain scaling-based control design approach with a central ingredient of the technique being solutions of coupled Lyapunov inequalities [10, 11]. Unlike previous control design approaches such as backstepping [21], forwarding [24], and older high-gain techniques [2, 7, 23], the new dynamic scaling-based approach provides a unified framework applicable to both state-feedback and output-feedback problems for both strict-feedback [11] and feedforward [12] systems. Furthermore, the new dynamic scaling-based technique provides strong robustness properties and allows appended Input-to-State Stable (ISS) dynamics driven by all states¹ [11] and also cross-products of unknown parameters and unmeasured states [13]. In this chapter, we show that the dynamic high-gain scaling-based technique can be applied to the decentralized output-feedback stabilization and disturbance attenuation problems, thus yielding decentralized results for a wider class of large-scale interconnected systems than available from prior results. The output-feedback decentralized control design based on the adaptive dual dynamic high-gain observer/controller methodology is described in Sect. 2. Thereafter, the application of the generalized scaling technique to eliminate the cascading dominance assumption on upper diagonal terms is described in Sect. 3.

¹ Previous results required the ISS appended dynamics to have nonzero gain only from the output y .

2 Decentralized Control Based on The Adaptive Dual Dynamic High-Gain Scaling Paradigm

2.1 Assumptions

The design is carried out under the Assumptions A1–A5, each of which is required to hold for all $m \in \{1, \dots, M\}$.

Assumption A1. A constant $\sigma_m > 0$ exists such that

$$|\phi_{(i,i+1,m)}(x_{(1,m)})| \geq \sigma_m, 1 \leq i \leq n_m - 1 \quad (2)$$

$$|\mu_{(0,m)}(x_{(1,m)})| \geq \sigma_m \quad (3)$$

for all $x_{(1,m)} \in \mathcal{R}$. Furthermore, the sign of each $\phi_{(i,i+1,m)}$, $i = 1, \dots, n_m - 1$, is independent of its argument.

Assumption A2. The inverse dynamics of (1)

$$\begin{aligned} \dot{Y}_{(i,m)} = & \phi_{(s_m+i,m)}(z, x, v - \tilde{u}, t, \varpi) + \phi_{(s_m+i, s_m+i+1, m)}(x_{(1,m)}) Y_{(i+1,m)} \\ & + \mu_{(i,m)}(x_{(1,m)})(v_m - \tilde{u}_m), 1 \leq i \leq n_m - s_m \end{aligned} \quad (4)$$

are Input-to-State practically Stable (ISpS) with state $Y_m = [Y_{(1,m)}, \dots, Y_{(n_m-s_m,m)}]^T = [x_{(s_m+1,m)}, \dots, x_{(n_m,m)}]^T$ and input $[z, \tilde{x}_m, v, t, \varpi]^T$ where $Y_{(n_m-s_m+1,m)} \equiv 0$ is a dummy variable, \tilde{x}_m is the vector comprised of the entire state x except $(x_{(s_m+1,m)}, \dots, x_{(n_m,m)})$, $v_m = u_m + \tilde{u}_m$, $\tilde{u}_m = \frac{\phi_{(s_m,s_m+1,m)}(x_{(1,m)})}{\mu_{(0,m)}(x_{(1,m)})} Y_{(1,m)}$, $v = [v_1, \dots, v_M]^T$, and $\tilde{u} = [\tilde{u}_1, \dots, \tilde{u}_M]^T$.

Furthermore, an ISpS Lyapunov function $V_{Y_m}(Y_m)$ exists which satisfies²

$$\begin{aligned} \dot{V}_{Y_m} \leq & -\alpha_{Y_m}(|Y_m|) + \beta_{Y_m}(|x_{(1,m)}|) \left[v_m^2 + \sum_{i=2}^{s_m} x_{(i,m)}^2 + \sum_{i=2}^{s_m} \Lambda_{(i,m)}^2(|z_{(i,m)}|) \right. \\ & \left. + \theta_m \sum_{k=1}^M \Lambda_{(1,m,k)}^2(|z_{(1,k)}|) \right] + \theta_m \sum_{k=1}^M \beta_{(Y_m,k)}(x_{(1,k)}) x_{(1,k)}^2 + \beta_{(Y_m,\varpi)}(|\varpi|) \end{aligned} \quad (5)$$

where θ_m is an unknown non-negative constant. α_{Y_m} is a known class K_∞ function. β_{Y_m} , $\Lambda_{(i,m)}$, $i = 2, \dots, s_m$, $\Lambda_{(1,m,k)}$ and $\beta_{(Y_m,k)}$, $k = 1, \dots, M$, and $\beta_{(Y_m,\varpi)}$ are known continuous non-negative functions. A class K_∞ function \bar{V}_{Y_m} exists such that $V_{Y_m}(Y_m) \leq \bar{V}_{Y_m}(|Y_m|)$ for all $Y_m \in \mathcal{R}^{n_m-s_m}$. Non-negative constants κ_{Y_m} and $\tilde{\kappa}_{Y_m}$ exist such that $|Y_m|^2 \leq \kappa_{Y_m} \alpha_{Y_m}(|Y_m|)$ and $|Y_m|^2 \leq \tilde{\kappa}_{Y_m} V_{Y_m}$ for all $Y_m \in \mathcal{R}^{n_m-s_m}$.

Assumption A3. The uncertain functions $\phi_{(i,m)}$ satisfy inequalities (6) for $1 \leq i \leq s_m$ and (7) for $s_m + 1 \leq i \leq n_m$

² To minimize notation, instead of introducing an additional positive constant in \dot{V}_{Y_m} as in the standard ISpS Lyapunov inequality [6], we have subsumed the effect of an additive positive constant into $\beta_{(Y_m,\varpi)}$ by not requiring $\beta_{(Y_m,\varpi)}$ to vanish at the origin.

$$|\phi_{(i,m)}| \leq \Gamma_m(x_{(1,m)}) \left[\sum_{j=2}^i |x_{(j,m)}| + \sum_{j=2}^i \tilde{\Lambda}_{(j,m)}(|z_{(j,m)}|) + \theta_m \sum_{k=1}^M \tilde{\Lambda}_{(1,m,k)}(|z_{(1,k)}|) \right] \\ + \theta_m \sum_{k=1}^M \Gamma_{(m,k)}(x_{(1,k)}) |x_{(1,k)}| + \Gamma_{(m,\varpi)}(|\varpi|) \quad (6)$$

$$|\phi_{(i,m)}| \leq \Gamma_m(x_{(1,m)}) \left[\sum_{j=2}^{n_m} |x_{(j,m)}| + \sum_{j=2}^{s_m+1} \tilde{\Lambda}_{(j,m)}(|z_{(j,m)}|) \right. \\ \left. + \theta_m \sum_{k=1}^M \tilde{\Lambda}_{(1,m,k)}(|z_{(1,k)}|) + \theta_m |u_m| \right] \\ + \theta_m \sum_{k=1}^M \Gamma_{(m,k)}(x_{(1,k)}) |x_{(1,k)}| + \Gamma_{(m,\varpi)}(|\varpi|) \quad (7)$$

for all $t \geq 0$, $x_m \in \mathcal{R}^{n_m}$, $m = 1, \dots, M$, $z_{(i,m)} \in \mathcal{R}^{n_{z(i,m)}}$, $i = 1, \dots, s_m + 1$, $m = 1, \dots, M$, $u \in \mathcal{R}^M$, and $\varpi \in \mathcal{R}^{n_\varpi}$, with θ_m being an unknown non-negative constant and Γ_m , $\tilde{\Lambda}_{(j,m)}$, $j = 2, \dots, s_m + 1$, $\tilde{\Lambda}_{(1,m,k)}$, $k = 1, \dots, M$, $\Gamma_{(m,k)}$, $k = 1, \dots, M$, and $\Gamma_{(m,\varpi)}$ being known continuous non-negative functions.

Assumption A4. Positive constants $\bar{\rho}_{(i,m)}$ and $\underline{\rho}_{(i,m)}$ exist such that

$$|\phi_{(i,i+1,m)}(x_{(1,m)})| \geq \bar{\rho}_{(i,m)} |\phi_{(i-1,i,m)}(x_{(1,m)})|, i = 3, \dots, s_m - 1 \quad (8)$$

$$|\phi_{(i,i+1,m)}(x_{(1,m)})| \leq \underline{\rho}_{(i,m)} |\phi_{(i-1,i,m)}(x_{(1,m)})|, i = 3, \dots, n_m - 1 \quad (9)$$

for all $x_{(1,m)} \in \mathcal{R}$.

Assumption A5. The $z_{(i,m)}$, $i = 1, \dots, s_m + 1$, subsystems are ISpS with ISpS Lyapunov functions $V_{z_{(i,m)}}$ satisfying

$$\dot{V}_{z_{(1,m)}} \leq -\alpha_{z_{(1,m)}}(|z_{(1,m)}|) + \theta_m \sum_{k=1}^M \beta_{(z_{(1,m)},k)}(|x_{(1,k)}|) + \beta_{(z_{(1,m)},\varpi)}(|\varpi|) \quad (10)$$

$$\dot{V}_{z_{(i,m)}} \leq -\alpha_{z_{(i,m)}}(|z_{(i,m)}|) + \theta_m \sum_{k=1}^M \beta_{(z_{(i,m)},k)}(|x_{(1,k)}|) x_{(1,k)}^2 \\ + \beta_{z_{(i,m)}}(|x_{(1,m)}|) \sum_{j=2}^i x_{(j,m)}^2 + \beta_{(z_{(i,m)},\varpi)}(|\varpi|), i = 2, \dots, s_m \quad (11)$$

$$\dot{V}_{z_{(s_m+1,m)}} \leq -\alpha_{z_{(s_m+1,m)}}(|z_{(s_m+1,m)}|) + \theta_m \sum_{k=1}^M \beta_{(z_{(s_m+1,m)},k)}(|x_{(1,k)}|) x_{(1,k)}^2 \\ + \beta_{z_{(s_m+1,m)}}(|x_{(1,m)}|) \left[\sum_{j=2}^{n_m} x_{(j,m)}^2 + \theta_m u_m^2 \right] + \beta_{(z_{(s_m+1,m)},\varpi)}(|\varpi|) \quad (12)$$

where θ_m is an unknown non-negative constant, $\alpha_{z_{(i,m)}}$, $i = 1, \dots, s_m + 1$, are known class K_∞ functions, and $\beta_{z_{(i,m)}}$, $i = 2, \dots, s_m + 1$, $\beta_{(z_{(i,m)},k)}$, $i = 1, \dots, s_m + 1$, $k =$

$1, \dots, M$, and $\beta_{(z_{(i,m)}, \varpi)}$, $i = 1, \dots, s_m + 1$, are known continuous non-negative functions. A class K_∞ function $\bar{V}_{z_{(1,m)}}$, exists such that $V_{z_{(1,m)}}(z_{(1,m)}) \leq \bar{V}_{z_{(1,m)}}(|z_{(1,m)}|)$ for all $z_{(1,m)} \in \mathcal{R}^{n_{z_{(1,m)}}}$. Also, positive constants $\bar{V}_{z_{(i,m)}}$, $i = 2, \dots, s_m + 1$, exist such that $V_{z_{(i,m)}}(z_{(i,m)}) \leq \bar{V}_{z_{(i,m)}} \alpha_{z_{(i,m)}}(|z_{(i,m)}|)$ for all $z_{(i,m)} \in \mathcal{R}^{n_{z_{(i,m)}}}$. Furthermore, non-negative constants $\kappa_{z_{(i,m)}}$ and $\tilde{\kappa}_{z_{(i,m)}}$ exist such that³

$$\Lambda_{(i,m)}^2(|z_{(i,m)}|) + \tilde{\Lambda}_{(i,m)}^2(|z_{(i,m)}|) \leq \kappa_{z_{(i,m)}} \alpha_{z_{(i,m)}}(|z_{(i,m)}|) \quad (13)$$

$$\Lambda_{(i,m)}^2(|z_{(i,m)}|) + \tilde{\Lambda}_{(i,m)}^2(|z_{(i,m)}|) \leq \tilde{\kappa}_{z_{(i,m)}} V_{z_{(i,m)}} \quad (14)$$

for all $z_{(i,m)} \in \mathcal{R}^{n_{z_{(i,m)}}}$, $i = 2, \dots, s_m + 1$. The following local order estimates hold as $\pi \rightarrow 0^+$:

- (a) $\sum_{k=1}^M [\Lambda_{(1,k,m)}^2(\pi) + \tilde{\Lambda}_{(1,k,m)}^2(\pi)] = O[\alpha_{z_{(1,m)}}(\pi)]$
- (b) $\sum_{k=1}^M [\Lambda_{(1,k,m)}(\pi) + \tilde{\Lambda}_{(1,k,m)}(\pi)] = O[\pi]$
- (c) $\sum_{k=1}^M \beta_{(z_{(1,k)}, m)}(\pi) = O[\pi^2]$.

Remark 1. Under Assumptions A1 and A4, symmetric positive-definite matrices P_{o_m} and P_{c_m} , positive constants v_{o_m} , \tilde{v}_{o_m} , v_{c_m} , \bar{v}_{o_m} , v_{c_m} , \underline{v}_{c_m} , and \bar{v}_{c_m} , and continuous functions $g_{(2,m)}, \dots, g_{(n_m,m)}$, $k_{(2,m)}, \dots, k_{(s_m,m)}$ can be found [10] such that for all $x_{(1,m)} \in \mathcal{R}$,

$$\left. \begin{aligned} P_{o_m} A_{o_m}(x_{(1,m)}) + A_{o_m}^T(x_{(1,m)}) P_{o_m} &\leq -v_{o_m} I_{n_m-1} - \tilde{v}_{o_m} |\phi_{(2,3,m)}(x_{(1,m)})| C_m^T C_m \\ \underline{v}_{o_m} I_{n_m-1} &\leq P_{o_m} \tilde{D}_{o_m} + \tilde{D}_{o_m} P_{o_m} \leq \bar{v}_{o_m} I_{n_m-1} \end{aligned} \right\} \quad (15)$$

$$\left. \begin{aligned} P_{c_m} A_{c_m}(x_{(1,m)}) + A_{c_m}^T(x_{(1,m)}) P_{c_m} &\leq -v_{c_m} |\phi_{(2,3,m)}(x_{(1,m)})| I_{s_m-1} \\ \underline{v}_{c_m} I_{s_m-1} &\leq P_{c_m} \tilde{D}_{c_m} + \tilde{D}_{c_m} P_{c_m} \leq \bar{v}_{c_m} I_{s_m-1} \end{aligned} \right\} \quad (16)$$

where⁴

$$A_{o_m} = \text{upperdiag}(\phi_{(2,3,m)}, \dots, \phi_{(n_m-1,n_m,m)}) - [g_{(2,m)}, \dots, g_{(n_m,m)}]^T C_m \quad (17)$$

$$\tilde{D}_{o_m} = D_{o_m} - \frac{1}{2} I_{n_m-1}; D_{o_m} = \text{diag}(1, 2, \dots, n_m - 1) \quad (18)$$

$$A_{c_m} = \text{upperdiag}(\phi_{(2,3,m)}, \dots, \phi_{(s_m-1,s_m,m)}) - B_m[k_{(2,m)}, \dots, k_{(s_m,m)}] \quad (19)$$

$$\tilde{D}_{c_m} = D_{c_m} - \frac{1}{2} I_{s_m-1}; D_{c_m} = \text{diag}(1, 2, \dots, s_m - 1) \quad (20)$$

where $C_m = [1, 0, \dots, 0]$ is an $(n_m - 1)$ dimensional row vector and $B_m = [0, \dots, 0, 1]^T$ is an $(s_m - 1)$ dimensional column vector. Furthermore, by Theorem A1 in [11], a positive constant \bar{G}_m exists such that $(\sum_{i=2}^{s_m} g_{(i,m)}^2)^{\frac{1}{2}} \leq \bar{G}_m |\phi_{(2,3,m)}(x_{(1,m)})|$.

³ $\Lambda_{(i,m)}$ for $i = s_m + 1$ is a dummy variable identically equal to zero.

⁴ For notational clarity, we drop the arguments of functions when no confusion will result.

2.2 Observer and Controller Designs

A reduced-order observer with states $\hat{x}_m = [\hat{x}_{(2,m)}, \dots, \hat{x}_{(n_m,m)}]^T$ is designed as

$$\begin{aligned}\dot{\hat{x}}_{(i,m)} &= \phi_{(i,i+1,m)}(x_{(1,m)})[\hat{x}_{(i+1,m)} + r_m^i f_{i+1}(x_{(1,m)})] + \mu_{(i-s_m,m)}(x_{(1,m)})u_m \\ &\quad - r_m^{i-1} g_{(i,m)}(x_{(1,m)})[\hat{x}_{(2,m)} + r_m f_2(x_{(1,m)})] \\ &\quad - (i-1) \dot{r}_m r_m^{i-2} f_{(i,m)}(x_{(1,m)}), \quad 2 \leq i \leq n_m\end{aligned}\quad (21)$$

where r_m is the high-gain scaling parameter introduced in the dynamic controller for the m^{th} subsystem and

$$f_{(i,m)}(x_{(1,m)}) = \int_0^{x_{(1,m)}} \frac{g_{(i,m)}(\pi)}{\phi_{(1,2,m)}(\pi)} d\pi \quad (22)$$

for $2 \leq i \leq n_m$. The observer errors $e_{(2,m)}, \dots, e_{(n_m,m)}$ and scaled observer errors $\epsilon_{(2,m)}, \dots, \epsilon_{(n_m,m)}$ are defined as

$$e_{(i,m)} = \hat{x}_{(i,m)} + r_m^{i-1} f_{(i,m)}(x_{(1,m)}) - x_{(i,m)} \quad (23)$$

$$\epsilon_{(i,m)} = \frac{e_{(i,m)}}{r_m^{i-1}}, \quad i = 2, \dots, n_m. \quad (24)$$

The dynamics of $\epsilon_m = [\epsilon_{(2,m)}, \dots, \epsilon_{(n_m,m)}]^T$ are

$$\dot{\epsilon}_m = r_m A_{o_m} \epsilon_m - \frac{\dot{r}_m}{r_m} D_{o_m} \epsilon_m + \bar{\Phi}_m \quad (25)$$

where $\bar{\Phi}_m = [\bar{\Phi}_{(2,m)}, \dots, \bar{\Phi}_{(n_m,m)}]^T$ with

$$\bar{\Phi}_{(i,m)} = -\frac{\phi_{(i,m)}(z, x, u, t, \varpi)}{r_m^{i-1}} + g_{(i,m)}(x_{(1,m)}) \frac{\phi_{(1,m)}(z, x, u, t, \varpi)}{\phi_{(1,2,m)}(x_{(1,m)})}. \quad (26)$$

Introduce $\eta_m = [\eta_{(2,m)}, \dots, \eta_{(s_m,m)}]^T$ with

$$\begin{aligned}\eta_{(2,m)} &= \frac{\hat{x}_{(2,m)} + r_m f_{(2,m)}(x_{(1,m)}) + \zeta_m(x_{(1,m)}, \hat{\theta}_m)}{r_m} \\ \eta_{(i,m)} &= \frac{\hat{x}_{(i,m)} + r_m^{i-1} f_{(i,m)}(x_{(1,m)})}{r_m^{i-1}}, \quad i = 3, \dots, s_m,\end{aligned}\quad (27)$$

$\hat{\theta}_m$ being a dynamic adaptation parameter and

$$\zeta_m(x_{(1,m)}, \hat{\theta}_m) = (1 + \hat{\theta}_m)x_{(1,m)}\zeta_{(1,m)}(x_{(1,m)}) \quad (28)$$

with $\zeta_{(1,m)}$ being a design freedom to be chosen later. The control law is designed as

$$u_m = \frac{r_m^{s_m}}{\mu_{(0,m)}(x_{(1,m)})} \left[-\phi_{(s_m, s_m+1, m)}(x_{(1,m)}) \left(\frac{\hat{x}_{(s_m+1, m)}}{r_m^{s_m}} + f_{(s_m+1, m)}(x_{(1,m)}) \right) - \sum_{i=2}^{s_m} k_{(i,m)}(x_{(1,m)}) \eta_{(i,m)} \right]. \quad (29)$$

The dynamics of η_m are given by

$$\begin{aligned} \dot{\eta}_m &= r_m A_{c_m} \eta_m - \frac{\dot{r}_m}{r_m} D_{c_m} \eta_m + \Phi_m - r_m G_m \varepsilon_{(2,m)} \\ &\quad + H_m (\eta_{(2,m)} - \varepsilon_{(2,m)}) + \Xi_m \end{aligned} \quad (30)$$

$$G_m = [g_{(2,m)}, \dots, g_{(s_m,m)}]^T ; \quad \Phi_m = G_m \frac{\phi_{(1,m)}}{\phi_{(1,2,m)}} \quad (31)$$

$$H_m = \left[(1 + \hat{\theta}_m) \{ \zeta'_{(1,m)} x_{(1,m)} + \zeta_{(1,m)} \} \phi_{(1,2,m)}, 0, \dots, 0 \right]^T \quad (32)$$

$$\begin{aligned} \Xi_m &= \frac{1}{r_m} \left[\dot{\hat{\theta}}_m \zeta_{(1,m)} x_{(1,m)} + (1 + \hat{\theta}_m) [\zeta'_{(1,m)} x_{(1,m)} + \zeta_{(1,m)}] \right. \\ &\quad \times \left. [\phi_{(1,m)} - (1 + \hat{\theta}_m) \zeta_{(1,m)} x_{(1,m)} \phi_{(1,2,m)}], 0, \dots, 0 \right]^T \end{aligned} \quad (33)$$

where $\zeta'_{(1,m)}(x_{(1,m)})$ denotes $\frac{d\zeta_{(1,m)}(\pi)}{d\pi} \Big|_{\pi=x_{(1,m)}}$.

2.3 Stability Analysis

Noting that the inverse dynamics with state $\Upsilon_m = [x_{(s_m+1,m)}, \dots, x_{(n_m,m)}]^T$ are of the form (4) with

$$v_m = \frac{r_m^{s_m}}{\mu_{(0,m)}(x_{(1,m)})} \left[-\phi_{(s_m, s_m+1, m)}(x_{(1,m)}) \varepsilon_{(s_m+1, m)} - K_m^T(x_{(1,m)}) \eta_m \right] \quad (34)$$

where $K_m(x_{(1,m)}) = [k_{(2,m)}(x_{(1,m)}), \dots, k_{(s_m,m)}(x_{(1,m)})]^T$, and using Assumption A2,

$$\begin{aligned} \frac{d}{dt} \left(\frac{V_{\Upsilon_m}}{r_m^{2s_m-1}} \right) &\leq -\frac{\alpha r_m (|\Upsilon_m|)}{r_m^{2s_m-1}} - (2s_m - 1) \frac{\dot{r}_m}{r_m^{2s_m}} V_{\Upsilon_m} + \frac{3}{r_m} \beta_{\Upsilon_m} [1 + \hat{\theta}_m]^2 \zeta_{(1,m)}^2 x_{(1,m)}^2 \\ &\quad + \frac{1}{r_m^2} \beta_{\Upsilon_m} \sum_{i=2}^{s_m} \frac{\Lambda_{(i,m)}^2 (|z_{(i,m)}|)}{r_m^{2i-3}} + \theta_m \beta_{\Upsilon_m} \sum_{k=1}^M \Lambda_{(1,m,k)}^2 (|z_{(1,k)}|) \\ &\quad + r_m \beta_{\Upsilon_m} \left[3 + \frac{2}{\mu_{(0,m)}^2} \phi_{(s_m, s_m+1, m)}^2 + \frac{2}{\mu_{(0,m)}^2} |K_m|^2 \right] (|\varepsilon_m|^2 + |\eta_m|^2) \\ &\quad + \frac{\theta_m}{r_m} \sum_{k=1}^M \beta_{(\Upsilon_m, k)} x_{(1,k)}^2 + \frac{1}{r_m} \beta_{(\Upsilon_m, \varpi)}. \end{aligned} \quad (35)$$

Consider the observer and controller Lyapunov functions defined as

$$V_{(o,m)} = r_m \boldsymbol{\varepsilon}_m^T P_{o_m} \boldsymbol{\varepsilon}_m \quad (36)$$

$$V_{(c,m)} = r_m \boldsymbol{\eta}_m^T P_{c_m} \boldsymbol{\eta}_m + \frac{1}{2} \left(1 + \frac{1}{r_m} \right) x_{(1,m)}^2 \quad (37)$$

where P_{o_m} and P_{c_m} are picked as in Remark 1. The dynamics of the high-gain parameter r_m and the adaptation parameter $\hat{\theta}_m$ will be designed such that r_m is larger than 1, $\hat{\theta}_m$ is positive, and r_m and $\hat{\theta}_m$ are monotonically non-decreasing. $\zeta_{(1,m)}$ will be designed such that $\phi_{(1,2,m)} \zeta_{(1,m)}$ (and hence $x_{(1,m)} \phi_{(1,2,m)} \zeta_m$) is positive. Differentiating $V_{(o,m)}$ and $V_{(c,m)}$ and using (15) and (16),

$$\dot{V}_{(o,m)} \leq -r_m^2 v_{o_m} |\boldsymbol{\varepsilon}_m|^2 - r_m^2 \tilde{v}_{o_m} |\phi_{(2,3,m)}| \boldsymbol{\varepsilon}_{(2,m)}^2 - \dot{r}_m v_{o_m} |\boldsymbol{\varepsilon}_m|^2 + 2r_m \boldsymbol{\varepsilon}_m^T P_{o_m} \bar{\Phi}_m \quad (38)$$

$$\begin{aligned} \dot{V}_{(c,m)} &\leq -r_m^2 v_{c_m} |\phi_{(2,3,m)}| |\boldsymbol{\eta}_m|^2 - \dot{r}_m v_{c_m} |\boldsymbol{\eta}_m|^2 - x_{(1,m)} \phi_{(1,2,m)} \zeta_m - \frac{1}{2} \frac{\dot{r}_m}{r_m^2} x_{(1,m)}^2 \\ &\quad + 2r_m \boldsymbol{\eta}_m^T P_{c_m} G_m \frac{\phi_{(1,m)}}{\phi_{(1,2,m)}} - 2r_m^2 \boldsymbol{\eta}_m^T P_{c_m} G_m \boldsymbol{\varepsilon}_{(2,m)} \\ &\quad + 2r_m \boldsymbol{\eta}_m^T P_{c_m} H_m (\boldsymbol{\eta}_{(2,m)} - \boldsymbol{\varepsilon}_{(2,m)}) + 2r_m \boldsymbol{\eta}_m^T P_{c_m} \Xi_m \\ &\quad + 2|x_{(1,m)} \phi_{(1,m)}| + 2|x_{(1,m)} (r_m \boldsymbol{\eta}_{(2,m)} - r_m \boldsymbol{\varepsilon}_{(2,m)}) \phi_{(1,2,m)}|. \end{aligned} \quad (39)$$

A composite Lyapunov function for the x_m component of the m^{th} subsystem is defined as

$$V_{x_m} = c_m V_{(o,m)} + V_{(c,m)} + (c_m \kappa_{Y_m} + 1) \frac{V_{Y_m}}{r_m^{2s_m-1}} \quad (40)$$

where c_m is any positive constant larger than⁵ $[8\lambda_{max}^2(P_{c_m}) \bar{G}_m^2] / [\tilde{v}_{o_m} v_{c_m}]$. Using (35), (38), and (39), and forming upper bounds for the terms in (38) and (39) along the lines in [11],

$$\begin{aligned} \dot{V}_{x_m} &\leq -r_m^2 \frac{c_m v_{o_m}}{4} |\boldsymbol{\varepsilon}_m|^2 - r_m^2 \frac{v_{c_m}}{4} |\phi_{(2,3,m)}| |\boldsymbol{\eta}_m|^2 - \dot{r}_m c_m v_{o_m} |\boldsymbol{\varepsilon}_m|^2 - \dot{r}_m v_{c_m} |\boldsymbol{\eta}_m|^2 \\ &\quad - (1 + \hat{\theta}_m) x_{(1,m)}^2 \phi_{(1,2,m)} \zeta_{(1,m)} - \frac{1}{2} \frac{\dot{r}_m}{r_m^2} x_{(1,m)}^2 - \frac{\alpha_{Y_m} (|Y_m|)}{r_m^{2s_m-1}} \\ &\quad - (2s_m - 1)(c_m \kappa_{Y_m} + 1) \frac{\dot{r}_m}{r_m^{2s_m}} V_{Y_m} + q_{(1,m)}(x_{(1,m)}) x_{(1,m)}^2 \\ &\quad + \theta_m^* \sum_{k=1}^M q_{(2,m,k)}(x_{(1,k)}) x_{(1,k)}^2 + \frac{1}{r_m} q_{(3,m)}(x_{(1,m)}, \hat{\theta}_m) \zeta_{(1,m)}^2 x_{(1,m)}^2 \\ &\quad + r_m^{\frac{3}{2}} w_m(x_{(1,m)}, \hat{\theta}_m, \dot{\hat{\theta}}_m) [|\boldsymbol{\varepsilon}_m|^2 + |\boldsymbol{\eta}_m|^2] \end{aligned}$$

⁵ $\lambda_{max}(P)$ with P being a square symmetric matrix denotes the maximum eigenvalue of P .

$$\begin{aligned}
& + G_m^* \theta_m^* h_m(x_{(1,m)}) \sum_{k=1}^M \left[\Lambda_{(1,m,k)}^2(|z_{(1,k)}|) + \tilde{\Lambda}_{(1,m,k)}^2(|z_{(1,k)}|) \right] \\
& + \left[\frac{c_m}{r_m^{\frac{1}{2}}} + (c_m \kappa_{Y_m} + 1) \frac{\beta_{Y_m}(|x_{(1,m)}|)}{r_m^2} \right] \sum_{i=2}^{s_m+1} \frac{[\Lambda_{(i,m)}^2(|z_{(i,m)}|) + \tilde{\Lambda}_{(i,m)}^2(|z_{(i,m)}|)]}{r_m^{2i-3}} \\
& + \theta_m^* \frac{8c_m n_m}{v_{o_m}} \lambda_{max}^2(P_{o_m}) \frac{\Gamma_m^2(x_{(1,m)})}{\mu_{(0,m)}^2(x_{(1,m)})} \\
& \times \left[\phi_{(s_m, s_m+1, m)}(x_{(1,m)}) \left[\frac{\hat{x}_{(s_m+1, m)}}{r_m^{s_m}} + f_{(s_m+1, m)}(x_{(1,m)}) \right] + K_m^T(x_{(1,m)}) \eta_m \right]^2 \\
& + \beta_{(m, \varpi)}(|\varpi|)
\end{aligned} \tag{41}$$

where

$$\theta_m^* = 1 + \theta_m + \theta_m^2 \tag{42}$$

$$G_m^* = c_m \kappa_{Y_m} + 4 + c_m + \frac{16n_m c_m}{v_{o_m}} \lambda_{max}^2(P_{o_m}) \bar{G}_m^2 + \frac{16}{\sigma_m v_{c_m}} \lambda_{max}^2(P_{c_m}) \bar{G}_m^2 \tag{43}$$

and the functions $q_{(1,m)}(x_{(1,m)})$, $q_{(2,m,k)}(x_{(1,k)})$, $q_{(3,m)}(x_{(1,m)}, \hat{\theta}_m)$, $w_m(x_{(1,m)}, \hat{\theta}_m, \dot{\hat{\theta}}_m)$, $h_m(x_{(1,m)})$, and $\beta_{(m, \varpi)}(|\varpi|)$ are given by

$$q_{(1,m)} = 3 + 2c_m + \left[\frac{2}{c_m v_{o_m}} + \frac{2}{v_{c_m} |\phi_{(2,3,m)}|} \right] \phi_{(1,2,m)}^2 \tag{44}$$

$$\begin{aligned}
q_{(2,m,m)} & = 2c_m(1 + \Gamma_{(m,m)}^2) + 3\Gamma_{(m,m)}^2 + \frac{16n_m c_m}{v_{o_m}} \lambda_{max}^2(P_{o_m}) \bar{G}_m^2 \frac{\phi_{(2,3,m)}^2}{\phi_{(1,2,m)}^2} \Gamma_m^2 \\
& + \frac{16}{v_{c_m}} \lambda_{max}^2(P_{c_m}) \bar{G}_m^2 \frac{|\phi_{(2,3,m)}|}{\phi_{(1,2,m)}^2} \Gamma_m^2 + (c_m \kappa_{Y_m} + 1) \beta_{(Y_m, m)}
\end{aligned} \tag{45}$$

$$q_{(2,m,k)} = 2c_m(1 + \Gamma_{(m,k)}^2) + 3\Gamma_{(m,k)}^2 + (c_m \kappa_{Y_m} + 1) \beta_{(Y_m, k)} \text{ for } k \neq m \tag{46}$$

$$q_{(3,m)} = 3(c_m \kappa_{Y_m} + 1)(1 + \hat{\theta}_m)^2 \beta_{Y_m} \tag{47}$$

$$\begin{aligned}
w_m & = (c_m \kappa_{Y_m} + 1) \beta_{Y_m} \left[3 + \frac{2}{\mu_{(0,m)}^2} \phi_{(s_m, s_m+1, m)}^2 + \frac{2}{\mu_{(0,m)}^2} |K_m|^2 \right] \\
& + c_m n_m \lambda_{max}^2(P_{o_m}) \Gamma_m^2 + c_m n_m \lambda_{max}^2(P_{o_m}) (1 + \hat{\theta}_m)^2 \Gamma_m^2 [1 + |\zeta_{(1,m)}|]^2
\end{aligned}$$

$$\begin{aligned}
& + 3c_m n_m \lambda_{max}(P_{o_m}) \Gamma_m + 2c_m n_m^2 \lambda_{max}^2(P_{o_m}) \Gamma_m^2 \\
& + 3\lambda_{max}(P_{c_m})(1 + \hat{\theta}_m) |\zeta'_{(1,m)} x_{(1,m)} + \zeta_{(1,m)}| |\phi_{(1,2,m)}| \\
& + \lambda_{max}^2(P_{c_m}) \left\{ 2\dot{\theta}_m^2 \zeta_{(1,m)}^2 + 2(1 + \hat{\theta}_m)^4 [\zeta'_{(1,m)} x_{(1,m)} + \zeta_{(1,m)}]^2 \phi_{(1,2,m)}^2 \zeta_{(1,m)}^2 \right. \\
& \left. + (1 + \hat{\theta}_m)^2 [\zeta'_{(1,m)} x_{(1,m)} + \zeta_{(1,m)}]^2 \right\} \tag{48}
\end{aligned}$$

$$h_m = 1 + \Gamma_m^2 + \beta_{Y_m} + \frac{\phi_{(2,3,m)}^2}{\phi_{(1,2,m)}^2} \Gamma_m^2 \tag{49}$$

$$\beta_{(m,\varpi)} = \beta_{(Y_m,\varpi)} + \left(\frac{1}{16v_{c_m}\sigma} + \frac{c_m}{16v_{o_m}} \right) \Gamma_{(m,\varpi)}^2. \tag{50}$$

The terms involving $z_{(1,k)}$, $k = 1, \dots, M$ in (41) can be upper bounded as

$$\begin{aligned}
h_m(x_{(1,m)}) [\Lambda_{(1,m,k)}^2(|z_{(1,k)}|) + \tilde{\Lambda}_{(1,m,k)}^2(|z_{(1,k)}|)] & \leq \Delta_{(1,m,k)}(|z_{(1,k)}|) \\
& + \Delta_{(2,m,k)}(x_{(1,m)}) \tag{51}
\end{aligned}$$

where

$$\Delta_{(1,m,k)}(|z_{(1,k)}|) = [\Lambda_{(1,m,k)}^2(|z_{(1,k)}|) + \tilde{\Lambda}_{(1,m,k)}^2(|z_{(1,k)}|)] \sup_{|\pi| \leq |z_{(1,k)}|} h_m(\pi) \tag{52}$$

$$\Delta_{(2,m,k)}(x_{(1,m)}) = \sup_{0 \leq \pi \leq |x_{(1,m)}|} [\Lambda_{(1,m,k)}^2(\pi) + \tilde{\Lambda}_{(1,m,k)}^2(\pi)] h_m(x_{(1,m)}). \tag{53}$$

By Assumption A5,

$$\Lambda_{(1,m,k)}^2(\pi) + \tilde{\Lambda}_{(1,m,k)}^2(\pi) = O[\alpha_{z_{(1,k)}}(\pi)] \tag{54}$$

$$\Lambda_{(1,m,k)}(\pi) + \tilde{\Lambda}_{(1,m,k)}(\pi) = O[\pi] \tag{55}$$

as $\pi \rightarrow 0^+$. Hence, $\Delta_{(1,m,k)}(\pi) = O[\alpha_{z_{(1,k)}}(\pi)]$, $\Delta_{(1,m,k)}(\pi) = O[\pi^2]$, and $\Delta_{(2,m,k)}(\pi) = O[\pi^2]$ as $\pi \rightarrow 0^+$. Hence, a continuous non-negative function $\overline{\Delta}_{(2,m,k)}$ exists such that

$$\Delta_{(2,m,k)}(x_{(1,m)}) \leq x_{(1,m)}^2 \overline{\Delta}_{(2,m,k)}(x_{(1,m)}). \tag{56}$$

Using a reasoning similar to that used in the proof of Theorem 2 in [26], it is seen that the local order estimate $\Delta_{(1,k,m)}(\pi) = O[\alpha_{z_{(1,m)}}(\pi)]$ as $\pi \rightarrow 0^+$ implies the existence of a new Lyapunov function $\tilde{V}_{z_{(1,m)}}$, class K_∞ functions $\tilde{\alpha}_{z_{(1,m)}}$ and $\tilde{\beta}_{(z_{(1,m)},k)}$, and continuous non-negative functions $\alpha_{\theta m}$ and $\tilde{\beta}_{(z_{(1,m)},\varpi)}$ such that

$$\dot{\tilde{V}}_{z_{(1,m)}} \leq -\tilde{\alpha}_{z_{(1,m)}}(|z_{(1,m)}|) + \alpha_{\theta m}(\theta_m) \sum_{k=1}^M \tilde{\beta}_{(z_{(1,m)},k)}(|x_{(1,k)}|) + \tilde{\beta}_{(z_{(1,m)},\varpi)}(|\varpi|) \tag{57}$$

with $\tilde{\alpha}_{z_{(1,m)}}(\pi) = O[\alpha_{z_{(1,m)}}(\pi)]$ as $\pi \rightarrow 0^+$, $\tilde{\beta}_{(z_{(1,m)},k)}$ independent of θ_m , $\tilde{\beta}_{(z_{(1,m)},k)}(\pi) = O[\beta_{(z_{(1,m)},k)}(\pi)]$ as $\pi \rightarrow 0^+$, and with $\tilde{\alpha}_{z_{(1,m)}}$ satisfying the inequality $\tilde{\alpha}_{z_{(1,m)}}(|z_{(1,m)}|) \geq \sum_{k=1}^M \Delta_{(1,k,m)}(|z_{(1,m)}|) \forall z_{(1,m)} \in \mathcal{R}^{n_{z_{(1,m)}}}$. Hence, a continuous non-negative function $\bar{\beta}_{(z_{(1,m)},k)}$ exists such that

$$\tilde{\beta}_{(z_{(1,m)},k)}(|x_{(1,k)}|) \leq x_{(1,k)}^2 \bar{\beta}_{(z_{(1,m)},k)}(x_{(1,k)}). \quad (58)$$

Furthermore, it can be shown that a class K_∞ function $\bar{V}_{z_{(1,m)}}$ exists such that $\tilde{V}_{z_{(1,m)}}(z_{(1,m)}) \leq \bar{V}_{z_{(1,m)}}(|z_{(1,m)}|)$ for all $z_{(1,m)} \in \mathcal{R}^{n_{z_{(1,m)}}}$.

Using Assumption A5,

$$\begin{aligned} \frac{d}{dt} \left(\frac{V_{z_{(i,m)}}}{r_m^{2i-\frac{5}{2}}} \right) &\leq -\frac{\alpha_{z_{(i,m)}}(|z_{(i,m)}|)}{r_m^{2i-\frac{5}{2}}} + \frac{1}{r_m} \theta_m \sum_{k=1}^M \beta_{(z_{(i,m)},k)}(|x_{(1,k)}|) x_{(1,k)}^2 \\ &\quad + \frac{3}{r_m} \beta_{z_{(i,m)}}(|x_{(1,m)}|) (1 + \hat{\theta}_m)^2 \zeta_{(1,m)}^2 x_{(1,m)}^2 \\ &\quad + 3r_m \beta_{z_{(i,m)}}(|x_{(1,m)}|) [|\eta_m|^2 + |\varepsilon_m|^2] \\ &\quad + \frac{1}{r_m} \beta_{(z_{(i,m)},\varpi)}(|\varpi|) - \left(2i - \frac{5}{2}\right) \frac{\dot{r}_m}{r_m^{2i-\frac{3}{2}}} V_{z_{(i,m)}} \end{aligned} \quad (59)$$

for $i = 2, \dots, s_m$. Similarly,

$$\begin{aligned} \frac{d}{dt} \left(\frac{V_{z_{(s_m+1,m)}}}{r_m^{2s_m-\frac{1}{2}}} \right) &\leq -\frac{\alpha_{z_{(s_m+1,m)}}(|z_{(s_m+1,m)}|)}{r_m^{2s_m-\frac{1}{2}}} + \frac{1}{r_m} \theta_m \sum_{k=1}^M \beta_{(z_{(s_m+1,m)},k)}(|x_{(1,k)}|) x_{(1,k)}^2 \\ &\quad + \frac{3}{r_m} \beta_{z_{(s_m+1,m)}}(|x_{(1,m)}|) (1 + \hat{\theta}_m)^2 \zeta_{(1,m)}^2 x_{(1,m)}^2 \\ &\quad + 3r_m \beta_{z_{(s_m+1,m)}}(|x_{(1,m)}|) [|\eta_m|^2 + |\varepsilon_m|^2] \\ &\quad + \beta_{z_{(s_m+1,m)}}(|x_{(1,m)}|) \frac{|\Upsilon_m|^2}{r_m^{2s_m-\frac{1}{2}}} \\ &\quad + \theta_m r_m^{\frac{1}{2}} \frac{\beta_{z_{(s_m+1,m)}}(|x_{(1,m)}|)}{\mu_{(0,m)}^2(x_{(1,m)})} \left[\phi_{(s_m, s_m+1, m)}(x_{(1,m)}) \left[\frac{\hat{x}_{(s_m+1,m)}}{r_m^{s_m}} \right. \right. \\ &\quad \left. \left. + f_{(s_m+1,m)}(x_{(1,m)}) \right] + K_m^T(x_{(1,m)}) \eta_m \right]^2 \\ &\quad + \frac{1}{r_m} \beta_{(z_{(s_m+1,m)},\varpi)}(|\varpi|) - \left(2s_m - \frac{1}{2}\right) \frac{\dot{r}_m}{r_m^{2s_m+\frac{1}{2}}} V_{z_{(s_m+1,m)}}. \end{aligned} \quad (60)$$

The composite Lyapunov function for the overall system is defined as

$$V = \sum_{m=1}^M V_m \quad (61)$$

with

$$V_m = V_{x_m} + (c_m + 1) \sum_{i=2}^{s_m+1} \frac{\kappa_{z(i,m)} V_{z(i,m)}}{r_m^{2i-\frac{5}{2}}} + (1 + G_m^* \theta_m^*) \tilde{V}_{z(1,m)} + \frac{1}{2} (\hat{\theta}_m - \bar{\theta})^2 \quad (62)$$

where

$$\bar{\theta} = \sum_{m=1}^M \{(1 + \theta_m^*)[1 + \alpha_{\theta m}(\theta_m)]\} = \sum_{m=1}^M \{(2 + \theta_m + \theta_m^2)[1 + \alpha_{\theta m}(\theta_m)]\}. \quad (63)$$

The parameter estimator dynamics are designed as

$$\begin{aligned} \dot{\hat{\theta}}_m &= -\gamma_{\theta m} \hat{\theta}_m + \sum_{k=1}^M \left[q_{(2,k,m)}(x_{(1,m)}) x_{(1,m)}^2 + G_k^* \Delta_{(2,m,k)}(x_{(1,m)}) \right. \\ &\quad \left. + (1 + G_k^*) \tilde{\beta}_{(z(1,k),m)}(|x_{(1,m)}|) + (c_k + 1) x_{(1,m)}^2 \sum_{i=2}^{s_k+1} \kappa_{z(i,k)} \beta_{(z(i,k),m)}(|x_{(1,m)}|) \right] \\ &\quad + \left\{ \frac{8c_m n_m}{v_{o_m} \mu_{(0,m)}^2(x_{(1,m)})} \lambda_{max}^2(P_{o_m}) \Gamma_m^2(x_{(1,m)}) \right. \\ &\quad \left. + (c_m + 1) \frac{\kappa_{z(s_m+1,m)} r_m^{\frac{1}{2}}}{\mu_{(0,m)}^2(x_{(1,m)})} \beta_{z(s_m+1,m)}(|x_{(1,m)}|) \right\} \left[\phi_{(s_m, s_m+1, m)}(x_{(1,m)}) \left[\frac{\hat{x}_{(s_m+1,m)}}{r_m^{s_m}} \right. \right. \\ &\quad \left. \left. + f_{(s_m+1,m)}(x_{(1,m)}) \right] + K_m^T(x_{(1,m)}) \eta_m \right]^2 \end{aligned} \quad (64)$$

where $\gamma_{\theta m}$ is a nonnegative design parameter. $\hat{\theta}_m$ is initialized to be positive. By (64), $\dot{\hat{\theta}}_m \geq 0$ so that $\hat{\theta}_m$ remains positive for all time. Using (41), (57), (59), and (64),

$$\dot{V} \leq \sum_{m=1}^M \chi_m \quad (65)$$

where

$$\begin{aligned} \chi_m &= - \left[r_m^2 \frac{c_m v_{o_m}}{4} |\epsilon_m|^2 + r_m^2 \frac{v_{c_m}}{4} |\phi_{(2,3,m)}(x_{(1,m)})| |\eta_m|^2 \right. \\ &\quad \left. + \dot{r}_m c_m v_{o_m} |\epsilon_m|^2 + \dot{r}_m v_{c_m} |\eta_m|^2 \right. \\ &\quad \left. + (1 + \hat{\theta}_m) x_{(1,m)}^2 \phi_{(1,2,m)}(x_{(1,m)}) \zeta_{(1,m)}(x_{(1,m)}) + \frac{1}{2} \frac{\dot{r}_m}{r_m^2} x_{(1,m)}^2 \right. \\ &\quad \left. + \frac{\alpha_{Y_m}(|Y_m|)}{r_m^{2s_m-1}} + \frac{\gamma_{\theta m}}{2} (\hat{\theta}_m - \bar{\theta})^2 + (2s_m - 1)(c_m \kappa_{Y_m} + 1) \frac{\dot{r}_m}{r_m^{2s_m}} V_{Y_m} \right. \\ &\quad \left. + \tilde{\alpha}_{z(1,m)}(|z_{(1,m)}|) + \sum_{i=2}^{s_m+1} \frac{\kappa_{z(i,m)} \alpha_{z(i,m)}(|z_{(i,m)}|)}{r_m^{2i-\frac{5}{2}}} \right] \end{aligned}$$

$$\begin{aligned}
& + (c_m + 1) \sum_{i=2}^{s_m+1} \frac{(2i - \frac{5}{2}) \kappa_{z(i,m)} V_{z(i,m)} \dot{r}_m}{r_m^{2i-\frac{3}{2}}} \\
& + \left[q_{(1,m)}(x_{(1,m)}) x_{(1,m)}^2 + \sum_{k=1}^M \hat{\theta}_m \bar{q}_{(2,k,m)}(x_{(1,m)}) x_{(1,m)}^2 \right. \\
& + \frac{1}{r_m} \bar{q}_{(3,m)}(x_{(1,m)}, \hat{\theta}_m) \zeta_{(1,m)}^2(x_{(1,m)}) x_{(1,m)}^2 + q_{(4,m)}(x_{(1,m)}, \hat{\theta}_m) \frac{|\gamma_m|^2}{r_m^{2s_m-\frac{1}{2}}} + \frac{\gamma_{\theta m}}{2} \bar{\theta}^2 \\
& + q_{(5,m)}(x_{(1,m)}) \sum_{i=2}^{s_m+1} \frac{\Lambda_{(i,m)}^2(|z_{(i,m)}|) + \tilde{\Lambda}_{(i,m)}^2(|z_{(i,m)}|)}{r_m^{2i-1}} \\
& \left. + r_m^{\frac{3}{2}} \bar{w}_m(x_{(1,m)}, \hat{\theta}_m, \dot{\hat{\theta}}_m) [|\varepsilon_m|^2 + |\eta_m|^2] + \bar{\beta}_{(m,\varpi)}(|\varpi|) \right] \quad (66)
\end{aligned}$$

where the functions $\bar{q}_{(2,m,k)}(x_{(1,k)})$, $\bar{q}_{(3,m)}(x_{(1,m)}, \hat{\theta}_m)$, $q_{(4,m)}(x_{(1,m)}, \hat{\theta}_m)$, $q_{(5,m)}(x_{(1,m)})$, $\bar{w}_m(x_{(1,m)}, \hat{\theta}_m, \dot{\hat{\theta}}_m)$, and $\bar{\beta}_{(m,\varpi)}(|\varpi|)$ are given by

$$\begin{aligned}
\bar{q}_{(2,m,k)} & = q_{(2,m,k)}(x_{(1,k)}) + G_m^* \bar{\Delta}_{(2,k,m)}(x_{(1,k)}) + (1 + G_m^*) \bar{\beta}_{(z_{(1,m)}, k)}(|x_{(1,k)}|) \\
& + (c_m + 1) \sum_{i=2}^{s_m+1} \kappa_{z(i,m)} \beta_{(z(i,m), k)}(|x_{(1,k)}|) \quad (67)
\end{aligned}$$

$$\bar{q}_{(3,m)} = q_{(3,m)}(x_{(1,m)}, \hat{\theta}_m) + 3(1 + \hat{\theta}_m)^2 (c_m + 1) \sum_{i=2}^{s_m+1} \kappa_{z(i,m)} \beta_{(z(i,m))}(|x_{(1,m)}|) \quad (68)$$

$$\begin{aligned}
q_{(4,m)} & = 3\hat{\theta}_m \left\{ \frac{8c_m n_m}{v_{o_m} \mu_{(0,m)}^2} \lambda_{max}^2(P_{o_m}) \Gamma_m^2 \right. \\
& \left. + (c_m + 1) \kappa_{z(s_m+1,m)} \frac{\beta_{z(s_m+1,m)}(|x_{(1,m)}|)}{\mu_{(0,m)}^2} \right\} \phi_{(s_m, s_m+1, m)}^2 \\
& + (c_m + 1) \kappa_{z(s_m+1,m)} \beta_{z(s_m+1,m)}(|x_{(1,m)}|) \quad (69)
\end{aligned}$$

$$q_{(5,m)} = (c_m \kappa_{\gamma_m} + 1) \beta_{\gamma_m}(|x_{(1,m)}|) \quad (70)$$

$$\begin{aligned}
\bar{w}_m & = w_m(x_{(1,m)}, \hat{\theta}_m, \dot{\hat{\theta}}_m) + 3(c_m + 1) \sum_{i=2}^{s_m+1} \kappa_{z(i,m)} \beta_{(z(i,m))}(|x_{(1,m)}|) \\
& + 3\hat{\theta}_m \left\{ \frac{8c_m n_m}{v_{o_m} \mu_{(0,m)}^2} \lambda_{max}^2(P_{o_m}) \Gamma_m^2 + (c_m + 1) \kappa_{z(s_m+1,m)} \frac{\beta_{z(s_m+1,m)}(|x_{(1,m)}|)}{\mu_{(0,m)}^2} \right\} \\
& \times \left[|K_m(x_{(1,m)})| + |\phi_{(s_m, s_m+1, m)}| \right]^2 \quad (71)
\end{aligned}$$

$$\begin{aligned}
\bar{\beta}_{(m,\varpi)} & = \beta_{(m,\varpi)}(|\varpi|) + (c_m + 1) \sum_{i=2}^{s_m+1} \kappa_{z(i,m)} \beta_{(z(i,m), \varpi)}(|\varpi|) \\
& + (1 + G_m^* \theta_m^*) \tilde{\beta}_{(z_{(1,m)}, \varpi)}(|\varpi|). \quad (72)
\end{aligned}$$

The design freedom $\zeta_{(1,m)}$ is picked to dominate the functions $q_{(1,m)}$ and $\bar{q}_{(2,k,m)}$ by defining

$$\begin{aligned}\zeta_{(1,m)}(x_{(1,m)}) &= \frac{1}{\sigma_m} \text{sign}(\phi_{(1,2,m)}(0)) \left[q_{(1,m)}(x_{(1,m)}) \right. \\ &\quad \left. + \sum_{k=1}^M \bar{q}_{(2,k,m)}(x_{(1,m)}) + \zeta_m^*(x_{(1,m)}) \right]\end{aligned}\quad (73)$$

with ζ_m^* being a positive function of $x_{(1,m)}$ lower bounded by a positive constant $\underline{\zeta}_m^*$. As in [11], the dynamics of the high-gain scaling parameters $r_m, m = 1 \dots, M$, are designed as

$$\dot{r}_m = \lambda_m \left(R_m(x_{(1,m)}, \hat{\theta}_m, \dot{\hat{\theta}}_m) - r_m \right) \Omega_m(r_m, x_{(1,m)}, \hat{\theta}_m, \dot{\hat{\theta}}_m) \quad (74)$$

with λ_m being any continuous non-negative function such that $\lambda_m(\pi) = 1$ for $\pi \geq 0$ and $\lambda_m(\pi) = 0$ for $\pi \leq -\varepsilon_{rm}$ with ε_{rm} being any positive constant. By (74), r_m is monotonically non-decreasing so that by initializing $r_m(0) \geq 1$, we have $r_m(t) \geq 1$ for all time. R_m is chosen such that if $r_m \geq R_m$, then, for any nonnegative value of \dot{r}_m , the negative terms in (66) dominate over all the positive terms in (66) (except for $\bar{\beta}_{(m,\varpi)}(|\varpi|)$ and $\frac{\gamma_{\theta_m}}{2} \bar{\theta}^2$). The function Ω_m is chosen such that if $\dot{r}_m \geq \Omega_m$, then, for any value of r_m not smaller than 1, the negative terms in (66) dominate over all the positive terms in (66) (again, except for $\bar{\beta}_{(m,\varpi)}(|\varpi|)$ and $\frac{\gamma_{\theta_m}}{2} \bar{\theta}^2$). Functions $R_m(x_{(1,m)}, \hat{\theta}_m, \dot{\hat{\theta}}_m)$ and $\Omega_m(r_m, x_{(1,m)}, \hat{\theta}_m, \dot{\hat{\theta}}_m)$ to satisfy these considerations are given by

$$R_m = \max \left\{ \frac{2\bar{q}_{(3,m)}(x_{(1,m)}, \hat{\theta}_m) \zeta_{(1,m)}^2(x_{(1,m)})}{\zeta_m^*(x_{(1,m)})}, \left(\frac{8\bar{w}_m(x_{(1,m)}, \hat{\theta}_m, \dot{\hat{\theta}}_m)}{\min(c_m V_{o_m}, \sigma_m V_{c_m})} \right)^2, \right. \\ \left. [2q_{(4,m)}(x_{(1,m)}, \hat{\theta}_m) \kappa_{Y_m}]^2, 2q_{(5,m)}(x_{(1,m)}) \right\} \quad (75)$$

$$\Omega_m = \max \left\{ 2r_m \bar{q}_{(3,m)}(x_{(1,m)}, \hat{\theta}_m) \zeta_{(1,m)}^2(x_{(1,m)}), \right. \\ \frac{q_{(5,m)}(x_{(1,m)})}{r_m^{\frac{1}{2}}(c_m + 1)} \max \left\{ \frac{\tilde{\kappa}_{z_{(i,m)}}}{\kappa_{z_{(i,m)}}(2i - \frac{5}{2})} \mid i = 2, \dots, s_m + 1 \right\}, \\ \left. \frac{r_m^{\frac{1}{2}} \tilde{\kappa}_{Y_m} q_{(4,m)}(x_{(1,m)}, \hat{\theta}_m)}{(2s_m - 1)(c_m \kappa_{Y_m} + 1)}, \frac{r_m^{\frac{3}{2}} \bar{w}_m(x_{(1,m)}, \hat{\theta}_m, \dot{\hat{\theta}}_m)}{\min(c_m V_{o_m}, \underline{V}_{c_m})} \right\}. \quad (76)$$

From the dynamics (74), it is seen that r_m remains bounded if $x_{(1,m)}, \hat{\theta}_m$, and $\dot{\hat{\theta}}_m$ remain bounded. From (64), $\dot{\hat{\theta}}_m$ is a function of $r_m, x_{(1,m)}, \hat{x}_{(s_m+1,m)}$, and η_m that is guaranteed to be bounded if $x_{(1,m)}, r_m^{\frac{1}{2}} |\eta_m|^2, x_{(s_m+1,m)}^2/r_m^{2s_m-\frac{1}{2}}$, and $r_m^{\frac{1}{2}} \epsilon_{s_m+1}^2$ are bounded. Hence, boundedness of $\dot{\hat{\theta}}_m$ follows from boundedness of V_m implying that

boundedness of r_m follows from boundedness of V_m . For each $m \in \{1, \dots, M\}$ at each time t , one of the following two cases must hold:

- Case \mathcal{A} : $r_m > R_m(x_{(1,m)}, \hat{\theta}_m, \dot{\hat{\theta}}_m)$ (which implies that $\dot{r}_m = \Omega_m(r_m, x_{(1,m)}, \hat{\theta}_m, \dot{\hat{\theta}}_m)$)
Case \mathcal{B} : $r_m \leq R_m(x_{(1,m)}, \hat{\theta}_m, \dot{\hat{\theta}}_m)$.

From (66), (73), (75), and (76), it is seen that in both cases, an upper bound for χ_m can be obtained as

$$\begin{aligned} \chi_m \leq & - \left[r_m^2 \frac{c_m v_{o_m}}{8} |\varepsilon_m|^2 + r_m^2 \frac{v_{c_m}}{8} |\phi_{(2,3,m)}(x_{(1,m)})||\eta_m|^2 + \frac{1}{2} x_{(1,m)}^2 \zeta_m^*(x_{(1,m)}) \right. \\ & + \frac{\alpha_{Y_m}(|Y_m|)}{2r_m^{2s_m-1}} + \tilde{\alpha}_{z_{(1,m)}}(|z_{(1,m)}|) + \sum_{i=2}^{s_m+1} \frac{\kappa_{z_{(i,m)}} \alpha_{z_{(i,m)}}(|z_{(i,m)}|)}{2r_m^{2i-\frac{5}{2}}} + \frac{\gamma_{\theta m}}{2} (\hat{\theta}_m - \bar{\theta})^2 \\ & \left. + \left[\bar{\beta}_{(m,\varpi)}(|\varpi|) + \frac{\gamma_{\theta m}}{2} \bar{\theta}^2 \right] \right]. \end{aligned} \quad (77)$$

Hence, for any combination of the occurrences of the Cases \mathcal{A} and \mathcal{B} for the M subsystems, the following inequality is satisfied:

$$\begin{aligned} \dot{V} \leq & - \sum_{m=1}^M \left[r_m^2 \frac{c_m v_{o_m}}{8} |\varepsilon_m|^2 + r_m^2 \frac{v_{c_m}}{8} |\phi_{(2,3,m)}(x_{(1,m)})||\eta_m|^2 + \frac{1}{2} x_{(1,m)}^2 \zeta_m^*(x_{(1,m)}) \right. \\ & + \frac{\alpha_{Y_m}(|Y_m|)}{2r_m^{2s_m-1}} + \tilde{\alpha}_{z_{(1,m)}}(|z_{(1,m)}|) + \sum_{i=2}^{s_m+1} \frac{\kappa_{z_{(i,m)}} \alpha_{z_{(i,m)}}(|z_{(i,m)}|)}{2r_m^{2i-\frac{5}{2}}} + \frac{\gamma_{\theta m}}{2} (\hat{\theta}_m - \bar{\theta})^2 \\ & \left. + \sum_{m=1}^M \left[\bar{\beta}_{(m,\varpi)}(|\varpi|) + \frac{\gamma_{\theta m}}{2} \bar{\theta}^2 \right] \right]. \end{aligned} \quad (78)$$

Using standard Lyapunov arguments, Input-to-Output practical Stability (IOpS) and integral Input-to-Output practical Stability (iIOpS) properties can be inferred from the Lyapunov inequality (78). Analogous to the analysis in [9], it is inferred from (78) that the inequality

$$\dot{V} \leq -\underline{\mathcal{H}}(V) + \bar{\chi}_1 + \bar{\chi}_2 \quad (79)$$

holds with $\bar{\chi}_1 = \sum_{m=1}^M 0.5 \gamma_{\theta m} \bar{\theta}^2$ being an uncertain constant, $\bar{\chi}_2 = \sum_{m=1}^M \bar{\beta}_{(m,\varpi)}(|\varpi|)$, and with $\underline{\mathcal{H}}$ being a class K_∞ function given by

$$\begin{aligned} \underline{\mathcal{H}}(V) = & \min \left(\min \left\{ \alpha_{Y_m} \circ \bar{V}_{Y_m}^{-1} \left(\frac{V}{(2M+1)(c_m \kappa_{Y_m} + 1)} \right) \middle| m = 1, \dots, M \right\}, \right. \\ & \min \left\{ \tilde{\alpha}_{z_{(1,m)}} \circ \bar{V}_{z_{(1,m)}}^{-1} \left(\frac{V}{(2M+1)(1 + G_m^* \theta_m^*)} \right) \middle| m = 1, \dots, M \right\}, \\ & \left. \frac{c_{\mathcal{H}}}{2M+1} V \right) \end{aligned} \quad (80)$$

where

$$c_{\mathcal{H}} = \min \bigcup_{m=1}^M \left\{ \frac{v_{o_m}}{8\lambda_{\max}(P_{o_m})}, \frac{v_{c_m}\sigma_m}{8\lambda_{\max}(P_{c_m})}, \frac{\zeta_m^*}{2}, \frac{1}{2(c_m+1)\bar{V}_{z_{(i,m)}}}, \gamma_{\theta_m} \right\}. \quad (81)$$

As in [9], (79) can be used to infer that as $t \rightarrow \infty$, all solutions tend to a compact set in which $V \leq \underline{\mathcal{H}}^{-1}(\limsup_{t \rightarrow \infty} [\bar{\chi}_1 + \bar{\chi}_2])$; in this compact set, the inequality $\sum_{m=1}^M y_m^2 \leq 2\underline{\mathcal{H}}^{-1}(\limsup_{t \rightarrow \infty} [\bar{\chi}_1 + \bar{\chi}_2])$ is satisfied. Also, it can be inferred from (79) that the following inequalities hold:

$$\begin{aligned} \frac{1}{2} \sum_{m=1}^M y_m^2(t) &\leq V(t_0) + \int_{t_0}^t \sum_{m=1}^M \bar{\beta}_{(m,\varpi)}(|\varpi(\pi)|) d\pi + \bar{\chi}_1(t - t_0) \\ \frac{1}{2} \sum_{m=1}^M \zeta_m^* \int_{t_0}^t y_m^2(\pi) d\pi &\leq V(t_0) + \int_{t_0}^t \sum_{m=1}^M \bar{\beta}_{(m,\varpi)}(|\varpi(\pi)|) d\pi + \bar{\chi}_1(t - t_0) \end{aligned} \quad (82)$$

where t and t_0 are any time instants with $t > t_0$.

The closed-loop stability properties are summarized in Theorems 1–3. Also, note that from (50) and (72), $\bar{\beta}_{(m,\varpi)}$ is a linear combination of $\beta_{(r_m,\varpi)}$, $\Gamma_{(m,\varpi)}^2$, $\tilde{\beta}_{(z_{(1,m)},\varpi)}$, and $\beta_{(z_{(i,m)},\varpi)}$, $i = 2, \dots, s_m + 1$. This implies that if ϖ enters into each subsystem in a linear fashion, i.e., if $\Gamma_{(m,\varpi)}(|\varpi|)$ is linear in $|\varpi|$ and $\beta_{(r_m,\varpi)}(|\varpi|)$ and $\beta_{(z_{(i,m)},\varpi)}(|\varpi|)$, $i = 1, \dots, s_m + 1$ are quadratic in $|\varpi|$, then it can be inferred from (78) that \mathcal{L}_2 disturbance attenuation can be achieved by picking controller parameters appropriately.

Theorem 1. Under Assumptions A1–A5, the designed dynamic compensator given by (21), (29), (64), and (74) achieves Bounded-Input-Bounded-State (BIBS) stability and IOpS of the closed-loop system with $(x, z, r_1, \dots, r_M, \hat{\theta}_1, \dots, \hat{\theta}_M, \hat{x}_1, \dots, \hat{x}_M)$ considered to be the state of the closed-loop system, ϖ the input, and $(x, z, \hat{x}_1, \dots, \hat{x}_M)$ the output. Furthermore, practical regulation of $x_{(1,m)}$, $m = 1, \dots, m$ to zero is achieved in the presence of bounded disturbances, i.e., $\sum_{m=1}^M |x_{(1,m)}(t)|$ can be asymptotically regulated to within as small a value as desired by appropriately tuning the controller parameters.

Theorem 2. Under Assumptions A1–A5, given any initial conditions $(x(0), z(0))$ for the overall plant state and $(r_m(0), \hat{\theta}_m(0), \hat{x}_m(0))$, $m = 1, \dots, M$, for the controller states with $r_m(0) \geq 1$, $m = 1, \dots, M$, if the disturbance input terms go to zero asymptotically, i.e., if

$$\sum_{m=1}^M \left[\Gamma_{(m,\varpi)}(\varpi(t)) + \beta_{(r_m,\varpi)}(\varpi(t)) + \sum_{i=1}^{s_m+1} \beta_{(z_{(i,m)},\varpi)}(\varpi(t)) \right] \rightarrow 0$$

as $t \rightarrow \infty$, then the signals $x(t), z(t), \hat{x}_1(t), \dots, \hat{x}_M(t)$ go to zero asymptotically as $t \rightarrow \infty$ if the controller parameters γ_{θ_m} , $m = 1, \dots, M$, are picked to be zero.

Theorem 3. Under Assumptions A1–A5 and the additional Assumption A6 below, given any values of the initial conditions $(x(0), z(0))$ for the overall plant state and

$(r_m(0), \hat{\theta}_m(0), \hat{x}_m(0)), m = 1, \dots, M$, for the controller states with $r_m(0) \geq 1, m = 1, \dots, M$, the designed dynamic controller achieves boundedness of all closed-loop states.

Assumption A6. The values of $\int_0^\infty \Gamma_{(m,\varpi)}^2(|\varpi(t)|)dt$, $\int_0^\infty \beta_{(r_m,\varpi)}(|\varpi(t)|)dt$, and $\int_0^\infty \beta_{(z_{(i,m)},\varpi)}(|\varpi(t)|)dt, i = 1, \dots, s_m + 1$ are finite for all $m = 1, \dots, M$.

3 Generalized Scaling: Application to Decentralized Control

In Sect. 2, the control design utilized, as a central assumption (see Assumption A4), bidirectional cascading dominance of the upper diagonal terms $\phi_{(i,i+1,m)}$ of each subsystem. This assumption can be eliminated using the generalized scaling technique [14, 17] as shown in this section. In this technique, the observer-context cascading dominance is induced by choosing the high-gain scaling powers appropriately and the controller-context cascading dominance is rendered irrelevant by resorting to a backstepping-based controller instead of a high-gain scaling-based controller. The price paid, however, due to the use of a backstepping-based controller is that the full generality of the structure of the system uncertainties and the coupling with the appended dynamics cannot be attained in this design, as described further in Remark 2. For instance, it is not possible to handle the triangular structure of state-coupled appended dynamics as in (1); instead, the appended dynamics z_m can only be materially driven (in a Lyapunov inequality sense – see Assumption A4') by the subsystem outputs.

The decentralized control design in this section addresses a class of interconnected large-scale systems wherein each subsystem is of the form:

$$\begin{aligned} \dot{z}_m &= q_m(z, x, u, t, \varpi) \\ \dot{x}_{(i,m)} &= \phi_{(i,m)}(z, x, u, t, \varpi) + \phi_{(i,i+1,m)}(x_{(1,m)})x_{(i+1,m)}, i = 1, \dots, s_m - 1 \\ \dot{x}_{(i,m)} &= \phi_{(i,m)}(z, x, u, t, \varpi) + \phi_{(i,i+1,m)}(x_{(1,m)})x_{(i+1,m)} \\ &\quad + \mu_{(i-s_m,m)}(x_{(1,m)})u_m, i = s_m, \dots, n_m \\ &\quad \vdots \\ \dot{x}_{(n_m,m)} &= \phi_{(n_m,m)}(z, x, u, t, \varpi) + \mu_{(n_m-s_m,m)}(x_{(1,m)})u_m \\ y_m &= x_{(1,m)} \end{aligned} \tag{83}$$

where, as in Sect. 2, $x_m = [x_{(1,m)}, \dots, x_{(n_m,m)}]^T \in \mathcal{R}^{n_m}$ is the state, $u_m \in \mathcal{R}$ is the input, $y_m \in \mathcal{R}$ is the output, and $z_m \in \mathcal{R}^{n_z}$ is the state of the appended dynamics of the m^{th} subsystem. M is the number of subsystems, $x = [x_1^T, \dots, x_M^T]^T$, $u = [u_1, \dots, u_M]^T$, and $z = [z_1^T, \dots, z_M^T]^T$. $\phi_{(i,i+1,m)}, i = 1, \dots, n_m - 1$ and $\mu_{(i,m)}$ are known continuous scalar real-valued functions. s_m is the relative degree of the m^{th} subsystem. $\varpi \in \mathcal{R}^{n_\varpi}$ is the exogenous disturbance input. $\phi_{(i,m)}, i = 1, \dots, n_m$ and q_m are continuous scalar real-valued uncertain functions.

The design is fundamentally based on our earlier result in [18] where a single subsystem of form (83) (i.e., $M = m = 1$), but without appended dynamics z_1 and with slightly stronger assumptions than used here on bounds on functions $\phi_{(i,1)}$, was considered and an output-feedback controller was proposed. The design in [18] was based on the dynamic high-gain scaling paradigm [11, 20, 23] but introduced a multiple time scaling through the use of arbitrary (not necessarily successive) powers of a dynamic high-gain scaling parameter r enabled through a new result on coupled parameter-dependent Lyapunov inequalities [16–18]. The utilization of non-successive powers of the dynamic high-gain scaling parameter in [18] allowed the removal of the cascading dominance assumption on upper diagonal terms (i.e., Assumption A4 that the ratios $\phi_{(i,i+1,1)}/\phi_{(i-1,i,1)}$ and $\phi_{(i-1,i,1)}/\phi_{(i,i+1,1)}$ for $i = 2, \dots, n_1 - 1$ are bounded) which was central in the earlier results [11, 15, 20, 23]. The construction in [18] resulted in the cascading dominance property being induced in the scaled system when r was of an appropriate size; the dynamics of r were then designed to achieve the required properties of the signal $r(t)$. In contrast with [18], we consider here an appended dynamics z_m , uncertain parameters in the bounds on $\phi_{(i,m)}$, a disturbance input ϖ , and introduction of multiple subsystems with nonlinear interconnections. The decentralized extension of the technique from [18] proves particularly challenging due to the fact that the observer gains in this approach are designed as functions of the high-gain scaling parameter and thus tend to amplify subsystem cross-coupling arising from $\phi_{(1,m)}$ as seen in the stability analysis.

3.1 Assumptions

The design is carried out under the Assumptions A1 and A2'–A4', each of which is required to hold for all $m \in \{1, \dots, M\}$. Assumption A1 is identically as in Sect. 2 and is not repeated here. Assumptions A2'–A4' are as follows.

Assumption A2'. The *inverse dynamics* of (83) satisfies the Bounded-Input-Bounded-State (BIBS) condition that the system given by $\dot{Y}_m = \Omega_m(x_{(1,m)})Y_m + v_m$ is BIBS stable with $[x_{(1,m)}, v_m^T]^T \in \mathcal{R}^{n_m - s_m + 1}$ considered the input and $Y_m \in \mathcal{R}^{n_m - s_m}$ being the state where the $(i,j)^{th}$ element of the $(n_m - s_m) \times (n_m - s_m)$ matrix $\Omega_m(x_{(1,m)})$ is defined as

$$\begin{aligned}\Omega_{m(i,i+1)}(x_{(1,m)}) &= \phi_{(s_m+i,s_m+i+1,m)}(x_{(1,m)}) , \quad i = 1, \dots, n_m - s_m - 1 \\ \Omega_{m(i,1)}(x_{(1,m)}) &= -\frac{\mu_{(i,m)}(x_{(1,m)})}{\mu_{(0,m)}(x_{(1,m)})} \phi_{(s_m,s_m+1,m)}(x_{(1,m)}) , \quad i = 1, \dots, n_m - s_m\end{aligned}\quad (84)$$

with zeros elsewhere.

Assumption A3'. Continuous functions $\hat{\phi}_{(i,m)}(t, x_{(1,m)}, \dots, x_{(i,m)})$ and nonnegative functions $\phi_{(i,j,m)}$, $\tilde{\Lambda}_{(m,k)}$, $k = 1, \dots, M$, $\Gamma_{(m,k)}$, $k = 1, \dots, M$, and $\Gamma_{(m,\varpi)}$ are known such that

$$|\phi_{(1,m)}(z, x, u, t, \varpi)| \leq \theta_m \sum_{k=1}^M \left[\Gamma_{(m,k)}(x_{(1,k)}) |x_{(1,k)}| + \Lambda_{(m,k)}(|z_k|) \right] + \Gamma_{(m,\varpi)}(|\varpi|) \quad (85)$$

$$\begin{aligned} & |\hat{\phi}_{(i,m)}(t, x_{(1,m)}, \hat{x}_{(2,m)}, \dots, \hat{x}_{(i,m)}) \\ & - \phi_{(i,m)}(z, x, u, t, \varpi)| \leq \theta_m \sum_{k=1}^M \left[\Gamma_{(m,k)}(x_{(1,k)}) |x_{(1,k)}| + \Lambda_{(m,k)}(|z_k|) \right] \\ & + \sum_{j=2}^i \phi_{(i,j,m)}(x_{(1,m)}) |\hat{x}_{(j,m)} - x_{(j,m)}| \\ & + \Gamma_{(m,\varpi)}(|\varpi|), \quad 2 \leq i \leq n_m \end{aligned} \quad (86)$$

for all $t \geq 0$, $x_m \in \mathcal{R}^{n_m}$, $m = 1, \dots, M$, $z_m \in \mathcal{R}^{n_{z_m}}$, $m = 1, \dots, M$, $u \in \mathcal{R}^M$, and $\varpi \in \mathcal{R}^{n_\varpi}$, with θ_m being an unknown non-negative constant.

Assumption A4'. The z_m subsystem is ISpS with ISpS Lyapunov function V_{z_m} satisfying

$$\dot{V}_{z_m} \leq -\alpha_{z_m}(|z_m|) + \theta_m \sum_{k=1}^M \beta_{(z_m,k)}(|x_{(1,k)}|) + \beta_{(z_m,\varpi)}(|\varpi|) \quad (87)$$

where θ_m is an unknown non-negative constant, α_{z_m} is a known class K_∞ function, and $\beta_{(z_m,k)}$, $k = 1, \dots, M$ and $\beta_{(z_m,\varpi)}$ are known continuous non-negative functions. A class K_∞ function \bar{V}_{z_m} exists such that $V_{z_m}(z_m) \leq \bar{V}_{z_m}(|z_m|)$ for all $z_m \in \mathcal{R}^{n_{z_m}}$. The following local order estimates hold as $\pi \rightarrow 0^+$:

- (a) $\sum_{k=1}^M \Lambda_{(k,m)}^2(\pi) = O[\alpha_{z_m}(\pi)]$,
- (b) $\sum_{k=1}^M \Lambda_{(k,m)}(\pi) = O[\pi]$,
- (c) $\sum_{k=1}^M \beta_{(z_k,m)}(\pi) = O[\pi^2]$.

Remark 2. Unlike in Sect. 2, the control design in this section does not require the cascading dominance assumption on upper diagonal terms (Assumption A4). Instead, the observer-context cascading dominance will be induced through a generalized scaling. The price, however, that one must pay to relax the cascading dominance assumption is that the assumption on the functions $\phi_{(i,m)}$ needs to be stronger than in Sect. 2 since a high-gain controller cannot be used due to the fact that the cascading dominance of upper diagonal terms required in observer and controller contexts are dual, i.e., the observer-context cascading dominance condition requires ratios $|\phi_{(i,i+1,m)}|/|\phi_{(i-1,i,m)}|$ to be upper bounded while the controller-context cascading dominance condition requires ratios $|\phi_{(i-1,i,m)}|/|\phi_{(i,i+1,m)}|$ to be upper bounded. Hence, the high-gain observer design requires upper diagonal terms nearer to the output to be larger while the high-gain controller design requires upper diagonal terms closer to the input to be larger. Therefore, it is not, in general, possible to design a high-gain observer and high-gain controller using the generalized scaling technique since either observer-context or controller-context cascading dominance can be induced by the scaling, but not both. The output-feedback design in

this section uses the generalized scaling technique for the observer which is then coupled with a backstepping controller. This constrains the functions $\phi_{(i,m)}$ to be incrementally linear in unmeasured states and prevents them from having the more general bound which can be handled in Sect. 2. Specifically, compare Assumption A3 with Assumption A3'. Also, compare Assumption A2 with Assumption A2' and Assumption A5 with Assumption A4'.

3.2 Observer Design

A reduced-order observer for the m^{th} subsystem of the interconnected large-scale system (83) is given by⁶

$$\begin{aligned}\dot{\hat{x}}_{(i,m)} &= \hat{\phi}_{(i,m)}(t, x_{(1,m)}, \hat{x}_{(2,m)} + f_{(2,m)}(r_m, x_{(1,m)}), \dots, \hat{x}_{(i,m)} + f_{(i,m)}(r_m, x_{(1,m)})) \\ &\quad + \phi_{(i,i+1,m)}(x_{(1,m)})[\hat{x}_{(i+1,m)} + f_{(i+1,m)}(r_m, x_{(1,m)})] \\ &\quad + g_{(i,m)}(r_m, x_{(1,m)})[\hat{x}_{(2,m)} + f_{(2,m)}(r_m, x_{(1,m)})] \\ &\quad + \mu_{(i-s_m,m)}(x_{(1,m)})u_m - \dot{r}_m h_{(i,m)}(r_m, x_{(1,m)}), 2 \leq i \leq n_m\end{aligned}\quad (88)$$

where r_m is a dynamic high-gain scaling parameter, $f_{(i,m)}(r_m, x_{(1,m)})$ are design functions of r_m and $x_{(1,m)}$ which will be picked during the stability analysis, and

$$\begin{aligned}g_{(i,m)}(r_m, x_{(1,m)}) &= -\phi_{(1,2,m)}(x_{(1,m)}) \frac{\partial f_{(i,m)}(r_m, x_{(1,m)})}{\partial x_{(1,m)}} \\ h_{(i,m)}(r_m, x_{(1,m)}) &= \frac{\partial f_{(i,m)}(r_m, x_{(1,m)})}{\partial r_m}.\end{aligned}\quad (89)$$

The dynamics of the high-gain scaling parameter r_m will be designed to be of the form $\dot{r}_m = w_m(r_m, x_{(1,m)})$ with w_m being $(s_m - 2)$ -times continuously differentiable. r_m is initialized greater than 1. The dynamics of r_m designed during the stability analysis will ensure that r_m is non-decreasing. Defining the observer errors

$$e_{(i,m)} = \hat{x}_{(i,m)} + f_{(i,m)}(r_m, x_{(1,m)}) - x_{(i,m)}, 2 \leq i \leq n_m, \quad (90)$$

the observer error dynamics are, $2 \leq i \leq n_m$,

$$\begin{aligned}\dot{e}_{(i,m)} &= \tilde{\phi}_{(i,m)} - \phi_{(i,m)} + \phi_{(i,i+1,m)}(x_{(1,m)})e_{(i+1,m)} \\ &\quad - g_{(i,m)}(r_m, x_{(1,m)}) \frac{\phi_{(1,m)}}{\phi_{(1,2,m)}(x_{(1,m)})} + g_{(i,m)}(r_m, x_{(1,m)})e_{(2,m)}\end{aligned}\quad (91)$$

with $e_{(n_m+1,m)} = 0$ being a dummy variable where, for notational convenience, we have introduced

⁶ For simplicity of notation, we introduce the dummy variables $\phi_{(n_m,n_m+1,m)} = \hat{x}_{(n_m+1,m)} = f_{(n_m+1,m)} = g_{(n_m+1,m)} = 0$ and $\mu_{(i,m)} \equiv 0$ for $i < 0$.

$$\tilde{\phi}_{(i,m)} = \hat{\phi}_{(i,m)}(t, x_{(1,m)}, \hat{x}_{(2,m)} + f_{(2,m)}(r_m, x_{(1,m)}), \dots, \hat{x}_{(i,m)} + f_{(i,m)}(r_m, x_{(1,m)})), \\ \text{for } i = 2, \dots, n_m. \quad (92)$$

Hence, the dynamics of $e_m = [e_{(2,m)}, \dots, e_{(n_m,m)}]^T$ are

$$\dot{e}_m = \tilde{\Phi}_m + [A_{(o,m)} + G_m C_m] e_m \quad (93)$$

where

$$C_m = [1, 0, \dots, 0] \quad (94)$$

$$G_m(r_m, x_{(1,m)}) = [g_{(2,m)}(r_m, x_{(1,m)}), \dots, g_{(n_m,m)}(r_m, x_{(1,m)})]^T \quad (95)$$

$$\tilde{\Phi}_m = [\tilde{\Phi}_{(2,m)}, \dots, \tilde{\Phi}_{(n_m,m)}]^T \quad (96)$$

$$\tilde{\Phi}_{(i,m)} = \tilde{\phi}_{(i,m)} - \phi_{(i,m)} - g_{(i,m)}(r_m, x_{(1,m)}) \frac{\phi_{(1,m)}}{\phi_{(1,2,m)}}, \quad i = 2, \dots, n_m \quad (97)$$

$$A_{(o,m)} = \begin{bmatrix} 0 & \phi_{(2,3,m)} & 0 & \dots & 0 \\ 0 & 0 & \phi_{(3,4,m)} & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & & & & \phi_{(n_m-1,n_m,m)} \\ 0 & 0 & \dots & & 0 \end{bmatrix}.$$

3.3 Controller Design

Define

$$\xi_{(i,m)} = \hat{x}_{(i,m)} + f_{(i,m)}(r_m, x_{(1,m)}), \quad i = 2, \dots, s_m. \quad (98)$$

The controller for the m th subsystem is designed through backstepping [21] using the subsystem with states $(x_{(1,m)}, \xi_{(2,m)}, \dots, \xi_{(s_m,m)})$ whose dynamics are

$$\begin{aligned} \dot{x}_{(1,m)} &= -\phi_{(1,2,m)}(x_{(1,m)})e_{(2,m)} + \phi_{(1,m)} + \phi_{(1,2,m)}(x_{(1,m)})\xi_{(2,m)} \\ \dot{\xi}_{(i,m)} &= g_{(i,m)}(r_m, x_{(1,m)})e_{(2,m)} + \hat{\phi}_{(i,m)}(t, x_{(1,m)}, \xi_{(2,m)}, \dots, \xi_{(i,m)}) \\ &\quad - g_{(i,m)}(r_m, x_{(1,m)}) \frac{\phi_{(1,m)}}{\phi_{(1,2,m)}(x_{(1,m)})} \\ &\quad + \phi_{(i,i+1,m)}(x_{(1,m)})\xi_{(i+1,m)}, \quad i = 2, \dots, s_m - 1 \\ \dot{\xi}_{(s_m,m)} &= g_{(s_m,m)}(r_m, x_{(1,m)})e_{(2,m)} + \hat{\phi}_{(s_m,m)}(t, x_{(1,m)}, \xi_{(2,m)}, \dots, \xi_{(s_m,m)}) \\ &\quad - g_{(s_m,m)}(r_m, x_{(1,m)}) \frac{\phi_{(1,m)}}{\phi_{(1,2,m)}(x_{(1,m)})} \\ &\quad + \phi_{(s_m,s_m+1,m)}(x_{(1,m)})[\hat{x}_{(s_m+1,m)} + f_{(s_m+1,m)}(r_m, x_{(1,m)})] \\ &\quad + \mu_{(0,m)}(x_{(1,m)})u_m. \end{aligned} \quad (99)$$

The backstepping-based controller design follows along similar lines as in [18] except for the introduction of the adaptation parameter $\hat{\theta}_m$, and more importantly, the introduction of a parameter $\bar{\theta}_m$ used at a key point in the stability analysis.

Step 1: The backstepping is commenced using the Lyapunov function $V_{(1,m)} = \frac{1}{2}\eta_{(1,m)}^2$ with $\eta_{(1,m)} = x_{(1,m)}$ yielding

$$\begin{aligned} \dot{V}_{(1,m)} &= -x_{(1,m)}\phi_{(1,2,m)}[e_{(2,m)} - \xi_{(2,m)}] + x_{(1,m)}\phi_{(1,m)} \\ &\leq -\alpha_m(r_m, x_{(1,m)})x_{(1,m)}^2 + \phi_{(1,2,m)}\eta_{(1,m)}\eta_{(2,m)} + \frac{e_{(2,m)}^2}{4r_m^2\bar{\theta}_m} + \frac{1}{4\bar{\theta}_m}\phi_{(1,m)}^2 \\ &\quad + \psi_{(1,m)}(\dot{\hat{\theta}}_m + \gamma_{2,m}\hat{\theta}_m) + (\bar{\theta}_m - \hat{\theta}_m - \gamma_{(1,m)}\psi_{(1,m)})\tau_{(1,m)} \end{aligned} \quad (100)$$

where $\gamma_{(1,m)}$ is an arbitrary positive constant, $\gamma_{(2,m)}$ is an arbitrary nonnegative constant, α_m is a smooth nonnegative function to be picked during stability analysis, $\hat{\theta}_m$ is an adaptation parameter, $\bar{\theta}_m$ is an (unknown) positive constant (depending on θ_m) which will be specified during stability analysis, and

$$\eta_{(2,m)} = \xi_{(2,m)} - \xi_{(2,m)}^*(r_m, x_{(1,m)}) \quad (101)$$

$$\begin{aligned} \xi_{(2,m)}^*(r_m, x_{(1,m)}) &= -\frac{1}{\phi_{(1,2,m)}(x_{(1,m)})} \left[\hat{\theta}_m r_m^2 x_{(1,m)} \phi_{(1,2,m)}^2(x_{(1,m)}) \right. \\ &\quad \left. + \hat{\theta}_m x_{(1,m)} + \alpha_m(r_m, x_{(1,m)})x_{(1,m)} \right] \end{aligned} \quad (102)$$

$$\psi_{(1,m)} = 0 \quad (103)$$

$$\tau_{(1,m)} = x_{(1,m)}^2 + r_m^2 x_{(1,m)}^2 \phi_{(1,2,m)}^2 \quad (104)$$

Step i ($2 \leq i \leq s_m - 1$): Assume that at step $(i-1)$, a Lyapunov function $V_{(i-1,m)}$ has been designed such that

$$\begin{aligned} \dot{V}_{(i-1,m)} &\leq -\alpha_m(r_m, x_{(1,m)})x_{(1,m)}^2 - \sum_{j=2}^{i-1} \zeta_{(j,m)}\eta_{(j,m)}^2 + \phi_{(i-1,i,m)}\eta_{(i-1,m)}\eta_{(i,m)} \\ &\quad + \frac{(i-1)e_{(2,m)}^2}{4r_m^2\bar{\theta}_m} + \frac{i-1}{4\bar{\theta}_m}\phi_{(1,m)}^2 + (\bar{\theta}_m - \hat{\theta}_m - \gamma_{(1,m)}\psi_{(i-1,m)})\tau_{(i-1,m)} \\ &\quad + \psi_{(i-1,m)}[\dot{\hat{\theta}}_m + \gamma_{(2,m)}\hat{\theta}_m] \end{aligned} \quad (105)$$

where, for $j = 3, \dots, i$,

$$\eta_{(j,m)} = \xi_{(j,m)} - \xi_{(j,m)}^*(t, r_m, \hat{\theta}_m, x_{(1,m)}, \xi_{(2,m)}, \dots, \xi_{(j-1,m)}),$$

with $\xi_{(j,m)}^*$ being functions designed in the previous steps of backstepping. Defining

$$V_{(i,m)} = V_{(i-1,m)} + \frac{1}{2}\eta_{(i,m)}^2, \quad (106)$$

and differentiating,

$$\begin{aligned}\dot{V}_{(i,m)} &\leq -\alpha_m(r_m, x_{(1,m)})x_{(1,m)}^2 - \sum_{j=2}^i \zeta_{(j,m)} \eta_{(j,m)}^2 + \phi_{(i,i+1,m)} \eta_{(i,m)} \eta_{(i+1,m)} + \frac{i e_{(2,m)}^2}{4r_m^2 \theta_m} \\ &+ \frac{i}{4\theta_m} \phi_{(1,m)}^2 + (\bar{\theta}_m - \hat{\theta}_m - \gamma_{(1,m)} \psi_{(i,m)}) \tau_{(i,m)} + \psi_{(i,m)} [\hat{\theta}_m + \gamma_{(2,m)} \hat{\theta}_m]\end{aligned}$$

where

$$\eta_{(i+1,m)} = \xi_{(i+1,m)} - \xi_{(i+1,m)}^*(t, r_m, \hat{\theta}_m, x_{(1,m)}, \xi_{(2,m)}, \dots, \xi_{(i,m)}) \quad (107)$$

$$\begin{aligned}\xi_{(i+1,m)}^* &= -\frac{1}{\phi_{(i,i+1,m)}(x_{(1,m)})} \left\{ \zeta_{(i,m)} \eta_{(i,m)} + \phi_{(i-1,i,m)}(x_{(1,m)}) \eta_{(i-1,m)} \right. \\ &+ \hat{\phi}_{(i,m)}(t, x_{(1,m)}, \xi_{(2,m)}, \dots, \xi_{(i,m)}) - \frac{\partial \xi_{(i,m)}^*}{\partial t} - \frac{\partial \xi_{(i,m)}^*}{\partial r_m} w_m(r_m, x_{(1,m)}) \\ &- \frac{\partial \xi_{(i,m)}^*}{\partial x_{(1,m)}} \phi_{(1,2,m)}(x_{(1,m)}) \xi_{(2,m)} \\ &- \sum_{j=2}^{i-1} \frac{\partial \xi_{(i,m)}^*}{\partial \xi_{(j,m)}} \left[\hat{\phi}_{(j,m)}(t, x_{(1,m)}, \xi_{(2,m)}, \dots, \xi_{(j,m)}) + \phi_{(j,j+1,m)}(x_{(1,m)}) \xi_{(j+1,m)} \right] \\ &+ (\hat{\theta}_m + \gamma_{(1,m)} \psi_{(i,m)}) \left(r_m^2 + \frac{1}{\phi_{(1,2,m)}^2(x_{(1,m)})} \right) \eta_{(i,m)} \left[g_{(i,m)}(r_m, x_{(1,m)}) \right. \\ &+ \frac{\partial \xi_{(i,m)}^*}{\partial x_{(1,m)}} \phi_{(1,2,m)}(x_{(1,m)}) - \sum_{j=2}^{i-1} \frac{\partial \xi_{(i,m)}^*}{\partial \xi_{(j,m)}} g_{(j,m)}(r_m, x_{(1,m)}) \left. \right]^2 \\ &\left. - \gamma_{(2,m)} \frac{\partial \xi_{(i,m)}^*}{\partial \hat{\theta}_m} \hat{\theta}_m - \frac{\partial \xi_{(i,m)}^*}{\partial \hat{\theta}_m} \gamma_{(1,m)} \tau_{(i-1,m)} \right\} \quad (108)\end{aligned}$$

$$\psi_{(i,m)} = \psi_{(i-1,m)} - \eta_{(1,m)} \frac{\partial \xi_{(i,m)}^*}{\partial \hat{\theta}_m} \quad (109)$$

$$\begin{aligned}\tau_{(i,m)} &= \tau_{(i-1,m)} + \eta_{(i,m)}^2 \left(r_m^2 + \frac{1}{\phi_{(1,2,m)}^2(x_{(1,m)})} \right) \left[g_{(i,m)}(r_m, x_{(1,m)}) \right. \\ &+ \frac{\partial \xi_{(i,m)}^*}{\partial x_{(1,m)}} \phi_{(1,2,m)}(x_{(1,m)}) - \sum_{j=2}^{i-1} \frac{\partial \xi_{(i,m)}^*}{\partial \xi_{(j,m)}} g_{(j,m)}(r_m, x_{(1,m)}) \left. \right]^2 \quad (110)\end{aligned}$$

where $\zeta_{(i,m)}$ is any positive constant.

Step s_m: At this step, the control input u_m is designed as

$$\begin{aligned}u_m &= \xi_{(s_m+1,m)}^*(t, r_m, \hat{\theta}_m, x_{(1,m)}, \xi_{(2,m)}, \dots, \xi_{(s_m,m)}) \\ &- \phi_{(s_m,s_m+1,m)}(x_{(1,m)}) [\hat{x}_{(s_m+1,m)} + f_{(s_m+1,m)}(r_m, x_{(1,m)})]\end{aligned} \quad (111)$$

where $\xi_{(s_m+1,m)}^*$ is defined analogously to (108) with i substituted to be s_m and with $\mu_{(0,m)}(x_{(1,m)})$ in the denominator of the first term rather than $\phi_{(i,i+1,m)}(x_{(1,m)})$. $\psi_{(s_m,m)}$ and $\tau_{(s_m,m)}$ are defined as in (109) and (110), respectively, with i substituted to be s_m . The Lyapunov function

$$V_{(s_m,m)} = \frac{1}{2} \sum_{i=1}^{s_m} \eta_{(i,m)}^2 \quad (112)$$

satisfies

$$\begin{aligned} \dot{V}_{(s_m,m)} &\leq -\alpha_m(r_m, x_{(1,m)}) x_{(1,m)}^2 - \sum_{j=2}^{s_m} \zeta_{(j,m)} \eta_{(j,m)}^2 + \frac{s_m e_{(2,m)}^2}{4r_m^2 \theta_m} + \frac{s_m}{4\theta_m} \phi_{(1,m)}^2 \\ &\quad + (\bar{\theta}_m - \hat{\theta}_m - \gamma_{(1,m)} \psi_{(s_m,m)}) \tau_{(s_m,m)} + \psi_{(s_m,m)} [\dot{\hat{\theta}}_m + \gamma_{(2,m)} \hat{\theta}_m]. \end{aligned} \quad (113)$$

Designing the dynamics of the adaptation parameter $\hat{\theta}_m$ as

$$\dot{\hat{\theta}}_m = -\gamma_{(2,m)} \hat{\theta}_m + \gamma_{(1,m)} \tau_{(s_m,m)}, \quad (114)$$

and defining

$$\bar{V}_{(s_m,m)} = V_{(s_m,m)} + \frac{1}{2\gamma_{(1,m)}} (\hat{\theta}_m - \bar{\theta}_m)^2, \quad (115)$$

we have

$$\begin{aligned} \dot{\bar{V}}_{(s_m,m)} &\leq -\alpha_m(r_m, x_{(1,m)}) x_{(1,m)}^2 - \sum_{j=2}^{s_m} \zeta_{(j,m)} \eta_{(j,m)}^2 + \frac{s_m e_{(2,m)}^2}{4r_m^2 \theta_m} \\ &\quad + \frac{s_m \phi_{(1,m)}^2}{4\theta_m} - \frac{\gamma_{(2,m)}}{2\gamma_{(1,m)}} (\hat{\theta}_m - \bar{\theta}_m)^2 + \frac{\gamma_{(2,m)}}{2\gamma_{(1,m)}} \bar{\theta}_m^2. \end{aligned} \quad (116)$$

The design freedoms in the controller for the m th subsystem are the function $\alpha_m(r_m, x_{(1,m)})$, and the constants $\zeta_{(2,m)}, \dots, \zeta_{(s_m,m)}, \gamma_{(1,m)}$, and $\gamma_{(2,m)}$. The constants $\zeta_{(2,m)}, \dots, \zeta_{(s_m,m)}$, and $\gamma_{(1,m)}$ can be picked to be arbitrary positive constants, $\gamma_{(2,m)}$ can be picked to be an arbitrary nonnegative constant, and the function α_m must be chosen to satisfy a lower bound to be specified during the stability analysis in Sect. 3.4. Note that by picking the dynamics of r_m to be of the form $\dot{r}_m = w_m(r_m, x_{(1,m)})$, the functions $\xi_{(2,m)}^*, \dots, \xi_{(s_m+1,m)}^*$ are well-defined and continuous.

3.4 Stability Analysis

Define $M_m = [M_{(2,m)}, \dots, M_{(n_m,m)}]^T$ where

$$M_{(i,m)} = \phi_{(i,1,m)}(x_{(1,m)}) + |g_{(i,m)}(r_m, x_{(1,m)})| \frac{\phi_{(1,1,m)}(x_{(1,m)})}{|\phi_{(1,2,m)}(x_{(1,m)})|}$$

and define $\bar{\Phi}_m$ to be the $(n_m - 1) \times (n_m - 1)$ matrix with $(i, j)^{th}$ entry

$$\begin{aligned}\bar{\Phi}_{m(i,j)} &= \phi_{(i+1,j+1,m)}, i = 1, \dots, n_m - 1, j = 1, \dots, i \\ \bar{\Phi}_{m(i,j)} &= 0, i = 1, \dots, n_m - 2, j = i + 1, \dots, n_m - 1.\end{aligned}\quad (117)$$

The matrix $A_{(o,m)} + \bar{\Phi}_m$ satisfies the assumptions of Theorem 1 in [18]. Hence, given any positive constant ρ_m , nonnegative constants $q_{(1,m)}, \dots, q_{(n_m-1,m)}$, and a positive function $R_m(x_{(1,m)}) \geq 1$ exist such that $T_m(r_m)[A_{(o,m)} + \bar{\Phi}_m]T_m^{-1}(r_m)$ is w-CUDD(ρ_m) for all $r_m \geq R_m(x_{(1,m)})$ where

$$T_m(r_m) = [\text{diag}(r_m^{q_{(1,m)}}, \dots, r_m^{q_{(n_m-1,m)}})]^{-1}. \quad (118)$$

The w-CUDD (weak Cascading Upper Diagonal Dominance) property was defined in [17] and shown to be central in solvability of coupled Lyapunov inequalities [16]. From the construction in the proof of Theorem 2 in [18], $q_{(1,m)}$ can be taken to be 1 and⁷ $q_{(n_m-1,m)} - q_{(n_m-2,m)} = 1$. Using Theorem 1 from [18], a $(n_m - 1) \times 1$ vector $\tilde{G}_m(r_m, x_{(1,m)})$, a symmetric positive-definite matrix $P_{(o,m)}$, and positive constants $v_{(o,m)}, \underline{v}_{(o,m)}$, and $\bar{v}_{(o,m)}$ exist such that for all $r_m \geq R_m(x_{(1,m)})$ and all $x_{(1,m)} \in \mathcal{R}$,

$$\begin{aligned}P_{(o,m)} &\left\{ T_m(r_m)[A_{(o,m)} + Q_{(1,m)}\bar{\Phi}_m Q_{(2,m)}]T_m^{-1}(r_m) + \tilde{G}_m C_m \right\} \\ &+ \left\{ T_m(r_m)[A_{(o,m)} + Q_{(1,m)}\bar{\Phi}_m Q_{(2,m)}]T_m^{-1}(r_m) + \tilde{G}_m C_m \right\}^T P_{(o,m)} \\ &\leq -\frac{v_{(o,m)}}{r_m^{q_{(n_m-2,m)} - q_{(n_m-1,m)}}} |\phi_{(n_m-1,n_m,m)}| I \\ \underline{v}_{(o,m)} I &\leq P_{(o,m)} D_{(o,m)} + D_{(o,m)} P_{(o,m)} \leq \bar{v}_{(o,m)} I\end{aligned}\quad (119)$$

where $D_{(o,m)} = \text{diag}(q_{(1,m)}, \dots, q_{(n_m-1,m)})$ and $Q_{(1,m)}$ and $Q_{(2,m)}$ are arbitrary diagonal matrices of dimension $(n_m - 1) \times (n_m - 1)$ with each diagonal entry +1 or -1. By Theorem 1 in [18], the choice of \tilde{G}_m does not need to depend on $Q_{(1,m)}$ and $Q_{(2,m)}$. $G_m(r_m, x_{(1,m)}) = [g_{(2,m)}(r_m, x_{(1,m)}), \dots, g_{(n_m,m)}(r_m, x_{(1,m)})]^T$ is defined as

$$G_m(r_m, x_{(1,m)}) = r_m^{-q_{(1,m)}} T_m^{-1}(r_m) \tilde{G}_m(r_m, x_{(1,m)}) \quad (120)$$

so that $\tilde{G}_m C_m = T_m(r_m) G_m C_m T_m^{-1}(r_m) \cdot f_{(i,m)}$, $i = 2, \dots, n_m$, are obtained as

$$f_{(i,m)}(r_m, x_{(1,m)}) = - \int_0^{x_{(1,m)}} \frac{g_{(i,m)}(r_m, \pi)}{\phi_{(1,2,m)}(\pi)} d\pi. \quad (121)$$

The dynamics of $\varepsilon_m \stackrel{\triangle}{=} T_m(r_m) e_m$ are

⁷ Note that since the observer used is a reduced-order observer, $A_{(o,m)}$ is of dimension $(n_m - 1) \times (n_m - 1)$. Hence, by the construction in the proof of Theorem 2 in [18], $q_{(i,m)} = 1 + (n_m - 2)(n_m - 1)/2 - (n_m - i - 1)(n_m - i)/2$, $i = 1, \dots, n_m - 1$.

$$\begin{aligned}\dot{\varepsilon}_m &= T_m(r_m) \tilde{\Phi}_m + T_m(r_m) [A_{(o,m)}(x_{(1,m)}) + G_m(r_m, x_{(1,m)}) C_m] T_m^{-1}(r_m) \varepsilon_m \\ &\quad - \frac{\dot{r}_m}{r_m} D_{(o,m)} \varepsilon_m.\end{aligned}\tag{122}$$

The derivative of the Lyapunov function $V_{(o,m)} = \varepsilon_m^T P_{(o,m)} \varepsilon_m$ satisfies

$$\begin{aligned}\dot{V}_{(o,m)} &= 2\varepsilon_m^T P_{(o,m)} T_m(r_m) \tilde{\Phi}_m + \varepsilon_m^T \{P_{(o,m)} T_m(r_m) [A_{(o,m)} + G_m C_m] T_m^{-1}(r_m) \\ &\quad + T_m^{-1}(r_m) [A_{(o,m)} + G_m C_m]^T T_m(r_m) P_{(o,m)}\} \varepsilon_m \\ &\quad - \frac{\dot{r}_m}{r_m} \varepsilon_m^T [P_{(o,m)} D_{(o,m)} + D_{(o,m)} P_{(o,m)}] \varepsilon_m.\end{aligned}\tag{123}$$

The scaling $\varepsilon_m = T_m(r_m) e_m$ which comprises a scaling with not necessarily successive powers $q_{(1,m)}, \dots, q_{(n_m-1,m)}$ of the scaling parameter r_m essentially yields a multiple time scaling and is the key ingredient in allowing the removal of the cascading dominance assumption by using the Theorems 1 and 2 in [18]. In common with [11] and as in Sect. 2, the dynamics of the high-gain parameter are designed to be of the form

$$\dot{r}_m = \lambda_m(R_m - r_m)\Omega_m(r_m, x_{(1,m)})\tag{124}$$

with initial value $r_m(0) \geq 1$ and with Ω_m being an appropriately designed function. λ_m is chosen to be any non-negative $(s_m - 2)$ -times continuously differentiable function such that $\lambda_m(\pi) = 1$ for $\pi \geq 0$ and $\lambda_m(\pi) = 0$ for $\pi \leq -\varepsilon_{rm}$ with ε_{rm} being any positive constant. In contrast with the design in [18] where a single subsystem of form (83) was considered without appended dynamics, the function Ω_m should be chosen through a careful bounding of the term $2\varepsilon_m^T P_{(o,m)} T_m(r_m) \tilde{\Phi}_m$ since this term will generate cross-products of the form $g_{(i,m)}(r_m, x_{(1,m)}) x_{(1,k)}$ which cannot be handled in the composite Lyapunov function framework. To see the origin of such cross-products, note that⁸

$$|\tilde{\Phi}_m|_e \leq_e |\overline{\Phi}_m|_e |e_m|_e + \Omega_m\tag{125}$$

$$\begin{aligned}2\varepsilon_m^T P_{(o,m)} T_m(r_m) \tilde{\Phi}_m &\leq 2\varepsilon_m^T P_{(o,m)} T_m(r_m) Q_{(1,m)} \overline{\Phi}_m \Omega_{(2,m)} T_m^{-1}(r_m) \varepsilon_m \\ &\quad + 2|\varepsilon_m^T P_{(o,m)}|_e T_m(r_m) \Omega_m\end{aligned}\tag{126}$$

and observe that the bound on $\phi_{(1,m)}$ arising from Assumption A3' involves $x_{(1,k)}$, z_k , and ϖ . In (125)-(126),

$$\begin{aligned}\Omega_{(i,m)} &= \theta_m \sum_{k=1}^M [\Gamma_{(m,k)}(|x_{(1,k)}|) |x_{(1,k)}| + \Lambda_{(m,k)}(|z_k|)] \\ &\quad + \Gamma_{(m,\varpi)}(|\varpi|) + \frac{|g_{(i,m)}|}{|\phi_{(1,2,m)}|} |\phi_{(1,m)}|, \quad i = 2, \dots, n_m\end{aligned}\tag{127}$$

$$\Omega_m = [\Omega_{(2,m)}, \dots, \Omega_{(n_m,m)}]^T,\tag{128}$$

⁸ $|\beta|_e$ denotes a matrix of the same dimension as β with each element replaced by its absolute value. \leq_e denotes an element-wise inequality between two matrices of equal dimension.

and $Q_{(1,m)}$ and $Q_{(2,m)}$ are diagonal matrices with each diagonal entry $+1$ or -1 such that $|P_{(o,m)}\varepsilon_m|_e = Q_{(1,m)}P_{(o,m)}\varepsilon_m$ and $|\varepsilon_m|_e = Q_{(2,m)}\varepsilon_m$. To handle the bounding of the term $2\varepsilon_m^T P_{(o,m)} T_m(r_m) \tilde{\Phi}_m$, consider the two cases:

Case \mathcal{A} : $r_m > R_m(x_{(1,m)})$

Case \mathcal{B} : $r_m \leq R_m(x_{(1,m)})$.

For each subsystem $m \in \{1, \dots, M\}$ at each time t , one of these two cases must hold. Under Case \mathcal{A} , it is inferred from Theorem 1 in [18] and the construction in the proof of Theorem 3 in [16] that a positive constant \bar{G} exists such that $|g_{(i,m)}| \leq \bar{G} r^{q(i-1,m)} |\phi_{(1,2,m)}|$ for $i = 2, \dots, n_m$. Also, since (119) holds under Case \mathcal{A} , it follows that

$$\dot{V}_{(o,m)} \leq -\frac{v_{(o,m)}\sigma_m}{2} |\varepsilon_m|^2 + \frac{2}{v_{(o,m)}\sigma_m} \lambda_{max}^2(P_{(o,m)}) |\tilde{\mathcal{Q}}_m|^2 \quad (129)$$

where $\tilde{\mathcal{Q}}_m = [\tilde{\mathcal{Q}}_{(2,m)}, \dots, \tilde{\mathcal{Q}}_{(n_m,m)}]^T$ with

$$\begin{aligned} \tilde{\mathcal{Q}}_{(i,m)} &= \theta_m \sum_{k=1}^M [\Gamma_{(m,k)}(|x_{(1,k)}|)|x_{(1,k)}| + \Lambda_{(m,k)}(|z_k|)] \\ &\quad + \Gamma_{(m,\overline{\omega})}(|\overline{\omega}|) + \bar{G} |\phi_{(1,m)}| \quad i = 2, \dots, n_m. \end{aligned} \quad (130)$$

Under Case \mathcal{B} , it follows from (124) that $\dot{r}_m = \Omega_m(r_m, x_{(1,m)})$. It can be shown that the term $2\varepsilon_m^T P_{(o,m)} T_m(r_m) \tilde{\Phi}_m$ can be bounded as

$$\begin{aligned} 2\varepsilon_m^T P_{(o,m)} T_m(r_m) \tilde{\Phi}_m &\leq 2\varepsilon_m^T P_{(o,m)} T_m(r_m) Q_{(1,m)} \bar{\Phi}_m Q_{(2,m)} T_m^{-1}(r_m) \varepsilon_m \\ &\quad + \frac{v_{(o,m)}\sigma_m}{2} |\varepsilon_m|^2 \max \left(\frac{\max \{g_{(i,m)}^2 |i = 2, \dots, n_m\}}{\bar{G}^2 \phi_{(1,2,m)}^2}, 1 \right) \\ &\quad + \frac{2}{v_{(o,m)}\sigma_m} \lambda_{max}^2(P_{(o,m)}) |\tilde{\mathcal{Q}}_m|^2. \end{aligned} \quad (131)$$

Hence, designing $\Omega_m(r_m, x_{(1,m)})$ to be

$$\begin{aligned} \Omega_m(r_m, x_{(1,m)}) &\geq \frac{r_m}{v_{(o,m)}} \left\{ r_m^* + 2\lambda_{max}(P_{(o,m)}) \frac{r_m^{q_{nm}-1}}{r_m^{q_1}} \left[||A_{(o,m)} + G_m C_m|| + ||\bar{\Phi}_m|| \right] \right. \\ &\quad \left. + \frac{v_{(o,m)}\sigma_m}{2} \max \left(\frac{\max \{g_{(i,m)}^2 |i = 2, \dots, n_m\}}{\bar{G}^2 \phi_{(1,2,m)}^2}, 1 \right) \right\} \end{aligned} \quad (132)$$

with r_m^* being any positive constant, it follows that

$$\dot{V}_{(o,m)} \leq -r_m^* |\varepsilon_m|^2 + \frac{2}{v_{(o,m)}\sigma_m} \lambda_{max}^2(P_{(o,m)}) |\tilde{\mathcal{Q}}_m|^2. \quad (133)$$

Therefore, in either of the Cases \mathcal{A} and \mathcal{B} , the inequality

$$\dot{V}_{(o,m)} \leq -\min\left(\frac{v_{(o,m)}\sigma_m}{2}, r_m^*\right)|\epsilon_m|^2 + \frac{2}{v_{(o,m)}\sigma_m}\lambda_{max}^2(P_{(o,m)})|\tilde{Q}_m|^2 \quad (134)$$

holds.

By Assumption A4', $\sum_{k=1}^M \Lambda_{(k,m)}^2(\pi) = O[\alpha_{z_m}(\pi)]$ as $\pi \rightarrow 0^+$. Using a reasoning similar to that used in the proof of Theorem 2 in [26], it is seen that this local order estimate implies the existence of a new Lyapunov function \tilde{V}_{z_m} , class K_∞ functions $\tilde{\alpha}_{z_m}$ and $\tilde{\beta}_{(z_m,k)}$, and continuous non-negative functions $\alpha_{\theta m}$ and $\tilde{\beta}_{(z_m,\varpi)}$ such that

$$\dot{\tilde{V}}_{z_m} \leq -\tilde{\alpha}_{z_m}(|z_m|) + \alpha_{\theta m}(\theta_m) \sum_{k=1}^M \tilde{\beta}_{(z_m,k)}(|x_{(1,k)}|) + \tilde{\beta}_{(z_m,\varpi)}(|\varpi|) \quad (135)$$

with $\tilde{\alpha}_{z_m}(\pi) = O[\alpha_{z_m}(\pi)]$ as $\pi \rightarrow 0^+$, $\tilde{\alpha}_{z_m}(|z_m|) \geq \sum_{k=1}^M \Lambda_{(k,m)}^2(|z_m|) \forall z_m \in \mathcal{R}^{n_{z_m}}$, $\tilde{\beta}_{(z_m,k)}$ independent of θ_m , and $\tilde{\beta}_{(z_m,k)}(\pi) = O[\beta_{(z_m,k)}(\pi)]$ as $\pi \rightarrow 0^+$. Hence, a continuous non-negative function $\bar{\beta}_{(z_m,k)}$ exists such that

$$\tilde{\beta}_{(z_m,k)}(|x_{(1,k)}|) \leq x_{(1,k)}^2 \bar{\beta}_{(z_m,k)}(x_{(1,k)}). \quad (136)$$

Furthermore, it can be shown that a class K_∞ function \bar{V}_{z_m} exists such that $\tilde{V}_{z_m}(z_m) \leq \bar{V}_{z_m}(|z_m|)$ for all $z_m \in \mathcal{R}^{n_{z_m}}$.

Defining

$$V_{x_m} = \bar{V}_{(s_m,m)} + \frac{s_m + 1}{2\bar{\theta}_m \min(\frac{v_{(o,m)}\sigma_m}{2}, r_m^*)} V_{(o,m)}, \quad (137)$$

and using (134) and (116), we obtain

$$\begin{aligned} \dot{V}_{x_m} &\leq -\alpha_m(r_m, x_{(1,m)}) x_{(1,m)}^2 - \sum_{j=2}^{s_m} \zeta_{(j,m)} \eta_{(j,m)}^2 - \frac{1}{2} s_m^* |\epsilon_m|^2 \\ &\quad - \frac{\gamma_{(2,m)}}{2\gamma_{(1,m)}} (\hat{\theta}_m - \bar{\theta}_m)^2 + \frac{\gamma_{(2,m)}}{2\gamma_{(1,m)}} \bar{\theta}_m^2 + \frac{Q_{m0}}{\bar{\theta}_m} \left\{ M\theta_m^2 \sum_{k=1}^M \Gamma_{(m,k)}^2(|x_{(1,k)}|) x_{(1,k)}^2 \right. \\ &\quad \left. + M\theta_m^2 \sum_{k=1}^M \Lambda_{(m,k)}^2(|z_k|) + \Gamma_{(m,\varpi)}^2(|\varpi|) \right\} \end{aligned} \quad (138)$$

where Q_{m0} is a constant given by

$$Q_{m0} = \frac{s_m}{4} + \frac{3\lambda_{max}^2(P_{(o,m)})[s_m + 1]}{v_{(o,m)}\sigma_m \min(\frac{v_{(o,m)}\sigma_m}{2}, r_m^*)} [3M\bar{G}^2 + 2(n_m - 1)]$$

and s_m^* is an unknown positive constant defined to be $1/[2\bar{\theta}_m]$. Note that Q_{m0} does not depend on $\bar{\theta}_m$. At this point, we can choose the (unknown) constants $\bar{\theta}_1, \dots, \bar{\theta}_M$ to be $\bar{\theta}$ where

$$\begin{aligned}\bar{\theta} &= \max \left\{ \max \{Q_{k0}|k=1, \dots, M\}, \right. \\ &\quad \left. M \max \{Q_{k0}\theta_k^2|k=1, \dots, M\} \max \{\alpha_{\theta_k}(\theta_k), 1\} \right\}.\end{aligned}\quad (139)$$

Note that $\bar{\theta}_m$ is a constant used only in stability analysis and does not enter anywhere into the observer or controller equations. The overall composite Lyapunov function of the large-scale interconnected system is picked to be

$$V = \sum_{m=1}^M \left[V_{x_m} + 2 \left(\sum_{k=1}^M \frac{Q_{k0}M\theta_k^2}{\bar{\theta}_k} \right) \tilde{V}_{z_m} \right]. \quad (140)$$

We obtain

$$\begin{aligned}\dot{V} \leq & - \sum_{m=1}^M \left[\alpha_m(r_m, x_{(1,m)}) x_{(1,m)}^2 + \sum_{j=2}^{s_m} \zeta_{(j,m)} \eta_{(j,m)}^2 + \frac{1}{2} s_m^* |\varepsilon_m|^2 \right. \\ & + \frac{\gamma_{(2,m)}}{2\gamma_{(1,m)}} (\hat{\theta}_m - \bar{\theta}_m)^2 + \left(\sum_{k=1}^M \frac{Q_{k0}M\theta_k^2}{\bar{\theta}_k} \right) \tilde{\alpha}_{z_m}(|z_m|) \\ & \left. - \frac{\gamma_{(2,m)}}{2\gamma_{(1,m)}} \bar{\theta}_m^2 - \bar{\Gamma}_m(|x_{(1,m)}|) x_{(1,m)}^2 \right] + \bar{\Gamma}_{\varpi}(|\varpi|)\end{aligned}\quad (141)$$

where

$$\bar{\Gamma}_m(|x_{(1,m)}|) = \sum_{k=1}^M [\Gamma_{(k,m)}^2(|x_{(1,m)}|) + \bar{\beta}_{(z_k,m)}(|x_{(1,m)}|)] \quad (142)$$

$$\bar{\Gamma}_{\varpi}(|\varpi|) = \sum_{m=1}^M [\tilde{\beta}_{(z_m,\varpi)}(|\varpi|) + \Gamma_{(m,\varpi)}^2(|\varpi|)]. \quad (143)$$

Picking the design function α_m to be

$$\alpha_m(r_m, x_{(1,m)}) = \zeta_{(1,m)} + \bar{\Gamma}_m(|x_{(1,m)}|), \quad (144)$$

where $\zeta_{(1,m)}, m = 1, \dots, M$, are any positive constants, (141) reduces to

$$\begin{aligned}\dot{V} \leq & - \sum_{m=1}^M \left[\sum_{j=1}^{s_m} \zeta_{(j,m)} \eta_{(j,m)}^2 + \frac{1}{2} s_m^* |\varepsilon_m|^2 + \frac{\gamma_{(2,m)}}{2\gamma_{(1,m)}} (\hat{\theta}_m - \bar{\theta}_m)^2 \right. \\ & \left. + \left(\sum_{k=1}^M \frac{Q_{k0}M\theta_k^2}{\bar{\theta}_k} \right) \tilde{\alpha}_{z_m}(|z_m|) - \frac{\gamma_{(2,m)}}{2\gamma_{(1,m)}} \bar{\theta}_m^2 \right] + \bar{\Gamma}_{\varpi}(|\varpi|)\end{aligned}\quad (145)$$

By straightforward signal chasing and using the BIBS property in Assumption A2', it can be shown that all closed-loop signals remain bounded on the maximal interval of existence $[0, t_f]$ implying that $t_f = \infty$ and that solutions exist for all time. Furthermore, using standard Lyapunov arguments, Theorems 1–3 follow from (145). Also, note that $\bar{\Gamma}_{\varpi}$ is a linear combination of $\Gamma_{(m,\varpi)}^2$ and $\tilde{\beta}_{(z_m,\varpi)}$. This implies that if ϖ enters into each subsystem in a linear fashion, i.e., if $\Gamma_{(m,\varpi)}(|\varpi|)$ is linear in

$|\varpi|$ and $\beta_{(z_m, \varpi)}(|\varpi|)$ is quadratic in $|\varpi|$, then it can be inferred from (145) that \mathcal{L}_2 disturbance attenuation is achieved by picking controller parameters appropriately.

Theorem 1. Under Assumptions A1 and A2'–A4', the designed dynamic compensator achieves BIBS stability and IOpS of the closed-loop system with

$(x, z, r_1, \dots, r_M, \hat{\theta}_1, \dots, \hat{\theta}_M, \hat{x}_1, \dots, \hat{x}_M)$ considered to be the state of the closed-loop system (where \hat{x}_m denotes $[\hat{x}_{(2,m)}, \dots, \hat{x}_{(n_m,m)}]^T$), ϖ being the input of the closed-loop system, and $(x_{(1,1)}, \dots, x_{(1,M)}, z_1, \dots, z_M)$ being the output. Furthermore, practical regulation of $x_{(1,m)}$, $m = 1 \dots, M$, to zero is achieved in the presence of bounded disturbances, i.e., $\sum_{m=1}^M |x_{(1,m)}(t)|$ can be asymptotically regulated to within as small a value as desired by appropriately tuning the controller parameters.

Theorem 2. Under Assumptions A1 and A2'–A4', given any values of the initial conditions $(x(0), z(0))$ for the overall plant state and $(r_m(0), \hat{\theta}_m(0), \hat{x}_m(0))$, $m = 1, \dots, M$, for the controller states with $r_m(0) \geq 1$, $m = 1, \dots, M$, if the disturbance input terms go to zero asymptotically, i.e., if $\sum_{m=1}^M [\Gamma_{(m,\varpi)}(\varpi(t)) + \beta_{(z_m,\varpi)}(\varpi(t))] \rightarrow 0$ as $t \rightarrow \infty$, then the signals $\bar{x}(t), z(t), e_1(t), \dots, e_M(t)$ where $\bar{x} = [\bar{x}_1^T, \dots, \bar{x}_N^T]^T$ with $\bar{x}_m = [x_{(1,m)}, \dots, x_{(s_m,m)}]^T$ go to zero asymptotically as $t \rightarrow \infty$ if the controller parameters $\gamma_{(2,m)}$, $m = 1, \dots, M$, are picked to be zero. Furthermore, if the BIBS Assumption A2' is strengthened to a minimum phase assumption, then $x(t), z(t), \hat{x}_1(t), \dots, \hat{x}_M(t)$ go to zero asymptotically as $t \rightarrow \infty$.

Theorem 3. Under Assumptions A1 and A2'–A4' and the additional Assumption A5' below, given any values of the initial conditions $(x(0), z(0))$ for the overall plant state and $(r_m(0), \hat{\theta}_m(0), \hat{x}_m(0))$, $m = 1, \dots, M$, for the controller states with $r_m(0) \geq 1$, $m = 1, \dots, M$, the designed dynamic controller achieves boundedness of all closed-loop states.

Assumption A5': The values of $\int_0^\infty \Gamma_{(m,\varpi)}^2(|\varpi(t)|)dt$ and $\int_0^\infty \beta_{(z_m,\varpi)}(|\varpi(t)|)dt$ are finite for all $m = 1, \dots, M$.

Acknowledgement This work was supported in part by the NSF under grant ECS-0501539.

References

1. Šiljak DD (1978) Large-Scale Dynamic Systems: Stability and Structure. North-Holland, New York
2. Ilchmann A (1996) High-gain adaptive control: an overview. In: IEE Colloquium on Adaptive Control (Digest No: 1996/139), London, UK, pp 1–4
3. Jain S, Khorrami F (1997) Decentralized adaptive control of a class of large-scale interconnected nonlinear systems. IEEE Transactions on Automatic Control 42(2):136–154
4. Jain S, Khorrami F (1997) Decentralized adaptive output feedback design for large-scale nonlinear systems. IEEE Transactions on Automatic Control 42(5):729–735
5. Jamshidi M (1983) Large-Scale Systems: Modeling and Control. North-Holland, New York
6. Jiang ZP, Teel A, Praly L (1994) Small-gain theorem for ISS systems and applications. Mathematics of Control, Signals and Systems 7:95–120

7. Khalil HK (1996) Adaptive output feedback control of nonlinear systems represented by input-output models. *IEEE Transactions on Automatic Control* 41(2):177–188
8. Khalil HK, Saberi A (1982) Decentralized stabilization of nonlinear interconnected systems using high-gain feedback. *IEEE Transactions on Automatic Control* 27(1):265–268
9. Krishnamurthy P, Khorrami F (2003) Decentralized control and disturbance attenuation for large-scale nonlinear systems in generalized output-feedback canonical form. *Automatica* 39:1923–1933
10. Krishnamurthy P, Khorrami F (2004) Conditions for uniform solvability of parameter-dependent Lyapunov equations with applications. In: *Proceedings of the American Control Conference*, Boston, MA, pp 3896–3901
11. Krishnamurthy P, Khorrami F (2004) Dynamic high-gain scaling: state and output feedback with application to systems with ISS appended dynamics driven by all states. *IEEE Transactions on Automatic Control* 49(12):2219–2239
12. Krishnamurthy P, Khorrami F (2004) A high-gain scaling technique for adaptive output feedback control of feedforward systems. *IEEE Transactions on Automatic Control* 49(12):2286–2292
13. Krishnamurthy P, Khorrami F (2005) Adaptive output-feedback for nonlinear systems with no a priori bounds on parameters. In: *Proceedings of the American Control Conference*, Portland, OR, pp 3713–3718
14. Krishnamurthy P, Khorrami F (2005) Generalized state scaling-based robust control of nonlinear systems and applications to triangular systems. In: *Proceedings of the American Control Conference*, Portland, OR, pp 3427–3432
15. Krishnamurthy P, Khorrami F (2006) Application of the dual high-gain scaling technique to decentralized control and disturbance attenuation. In: *Proceedings of the IEEE Conference on Decision and Control*, San Diego, CA
16. Krishnamurthy P, Khorrami F (2006) On uniform solvability of parameter-dependent lyapunov inequalities and applications to various problems. *SIAM Journal on Control and Optimization* 45(4):1147–1164
17. Krishnamurthy P, Khorrami F (2007) Generalized state scaling and applications to feedback, feedforward, and non-triangular nonlinear systems. *IEEE Transactions on Automatic Control* 52(1):102–108
18. Krishnamurthy P, Khorrami F (2007) High-gain output-feedback control for nonlinear systems based on multiple time scaling. *Systems and Control Letters* 56(1):7–15
19. Krishnamurthy P, Khorrami F, Jiang ZP (2002) Global output feedback tracking for nonlinear systems in generalized output-feedback canonical form. *IEEE Transactions on Automatic Control* 47(5):814–819
20. Krishnamurthy P, Khorrami F, Chandra RS (2003) Global high-gain-based observer and backstepping controller for generalized output-feedback canonical form. *IEEE Transactions on Automatic Control* 48(12):2277–2284
21. Krstić M, Kanellakopoulos I, Kokotović PV (1995) *Nonlinear and Adaptive Control Design*. Wiley, New York
22. Özgüner Ü (1979) Near optimal control of composite systems: the multi-time-scale approach. *IEEE Transactions on Automatic Control* 24(4):652–655
23. Praly L (2003) Asymptotic stabilization via output feedback for lower triangular systems with output dependent incremental rate. *IEEE Transactions on Automatic Control* 48(6):1103–1108
24. Sepulchre R, Janković M, Kokotović PV (1997) Integrator forwarding: a new recursive nonlinear robust design. *Automatica* 33(5):979–984
25. Shi L, Singh SK (1992) Decentralized adaptive controller design for large-scale systems with higher order uncertainties. *IEEE Transactions on Automatic Control* 37(8):1106–1118
26. Sontag ED, Teel A (1995) Changing supply functions in input/state stable systems. *IEEE Transactions on Automatic Control* 40(8):1476–1478

Decentralized Output Feedback Guaranteed Cost Control of Uncertain Markovian Jump Large-Scale Systems: Local Mode Dependent Control Approach*

Junlin Xiong, Valery A. Ugrinovskii, and Ian R. Petersen

1 Introduction

Large-scale systems provide a mathematical model to represent dynamic systems that have a complex structure and high dimensions. Many physical systems, such as power systems [16], can be modeled as large-scale systems. In a large-scale system setting, one often lacks the system information that is necessary to implement a centralized controller design, or the complexity of a centralized design and cost of implementing centralized controllers is prohibitive. In such cases, decentralized controllers that only use locally available subsystem information provide a viable alternative to centralized solutions. In many practical problems involving large-scale systems, the decentralized control technique can effectively deal with such issues as information structure constraints, large dimensions and a broad range of uncertainties. However, the design of decentralized controllers is challenging for at least two reasons. Firstly, not all the system information is available to the controllers. Secondly, the dynamics of each subsystem in the large-scale system are affected by other subsystems [19] via the interconnections; in many situations the effect of those interconnections is essentially uncertain.

A large-scale system may experience abrupt changes in system parameters, due to, for instance, random failures of its components, sudden variations of the environment, and changes of the subsystem working conditions. In this case, the Markovian jump large-scale system model can serve as a suitable model to capture those abrupt system parameter changes. Specifically, a Markovian jump large-scale system can

*This work was supported by the Australian Research Council.

J. Xiong

Department of Automation, University of Science and Technology of China, Hefei 230026, China
e-mail: junlin.xiong@gmail.com

V.A. Ugrinovskii and I.R. Petersen

School of Engineering and Information Technology, University of New South Wales
at the Australian Defence Force Academy, Northcott Drive, Canberra, ACT 2600, Australia
e-mail: v.ugrinovskii@gmail.com; i.r.petersen@gmail.com

be seen as an interconnection of a finite number of subsystems, and each subsystem has a finite number of operation modes. The abrupt changes of system parameters cause the subsystems to change their operation modes. Furthermore, in such a large-scale system, these operation mode changes are governed by a Markov process that takes values in a finite set. Markovian jump large-scale systems may be regarded as a generalization of Markovian jump systems [10] and normal large-scale systems [16]. Such a system reduces to a Markovian jump system when the system is regarded as consisting of only one subsystem. Also, such a Markovian jump large-scale system turns into a regular large-scale system when each subsystem has only one operation mode.

In the recent control literature, much attention has been given to Markovian large-scale systems subject to uncertain perturbations; e.g., see [7, 9, 18, 22]. In particular, [7, 18] considered a class of uncertain Markovian jump parameter systems in which both local uncertain perturbations and subsystem interconnections were described in terms of integral quadratic constraints (IQC) [14, 15]. In these references, necessary and sufficient conditions were derived for absolute stabilization of the system using decentralized state and output feedback controllers, respectively. An important underlying assumption required to implement the controllers proposed in many available results including the above references is that the global operation mode of the large-scale system must be known to every controller; we refer to such controllers as *global mode dependent controllers*. In a global mode dependent decentralized control design, each subsystem is equipped with its own switching feedback controller which utilizes the subsystem dynamic state or output. Furthermore, the number of operation modes for each such controller is equal to the number of global operation modes of the system, and hence is greater than the number of local operation modes of the subsystem it controls. As a result, the controller has to change its operation mode even if the subsystem it controls does not change. Also, to implement such a control algorithm, one needs to ensure that the global mode information is available to every subsystem controller. Thus, to implement a global mode dependent control scheme in practice, this information must be collected and broadcast to each subsystem in real time. Collecting and broadcasting operation modes of each subsystem may be expensive or even impossible in some situations. Hence it is desirable to lessen the communication overheads by reducing the dependency of the decentralized controller on the knowledge of the global operation mode information. A local mode dependent approach has been recently developed where the mode of the decentralized controller only depends on the mode of the subsystem it controls [20]. These types of decentralized controllers are referred to as *local mode dependent controllers*. The result in [20] gives a sufficient condition and an algorithm for the design of local mode dependent stabilizing controllers under the assumption of full state feedback.

This chapter is a further development of the local mode dependent approach proposed in [20]. It extends the results of [20] in several directions. Firstly, in this chapter we address an output feedback stabilization problem via the approach of local mode dependent control. Our result leads to an algorithm for the design of dynamic output feedback controllers of full order. In [20], static state feedback controllers were considered. Secondly, unlike [20] which focused on the stabilization

problem, this chapter addresses performance of the closed-loop system. Here we consider an output feedback guaranteed cost control problem that is similar to that in [7]. Our control design method leads to a set of local mode dependent controllers which are suboptimal with respect to a given quadratic performance cost functional. Thirdly, in [20], the system model was somewhat limited in that the uncertainty outputs employed in the definition of the admissible uncertainty used in that paper did not allow for control input feed-through. This chapter overcomes this limitation and reintroduces the control input feed-through term in the definition of the uncertainty output. This however leads to several additional technical difficulties, which precludes us from following the approach undertaken in [20]; in particular the results developed in [7, 18, 20] cannot be used in the derivation of local mode dependent output feedback controllers. Hence, the control design technique used in this chapter is different from that used in the previous work. To develop the solution, a version of bounded real lemma [21] is adopted to tackle the control input feed-through terms, and the projection lemma [5] is used to derive the controller.

The decentralized guaranteed cost control problem for a class of Markovian jump large-scale systems studied in this chapter involves two classes of uncertainties described using IQCs [15]. Local uncertainties are the uncertainties that affect the dynamics of the subsystems. Also, we consider subsystem interconnection uncertainties which describe the interconnections between the subsystems. Our approach proceeds from the observation that the replacement of a global mode dependent control action with a local mode dependent control action can be interpreted as a result of a fictitious “mode mismatch uncertainty” in the global mode dependent controller; such an fictitious uncertainty acts to cancel out the information gained from the knowledge of the global operation mode information. This idea prompts us to propose the following two-step controller design algorithm to design the local mode dependent output feedback controllers. Firstly, an auxiliary class of *uncertain* global mode dependent controllers that stabilize the class of uncertain Markovian jump large-scale systems under consideration is designed; the uncertainty in the controller is used to represent the mode mismatch uncertainty. To design such controllers, the uncertain system is augmented to include the possibility of the controller uncertainty. The controller uncertainty is also described using IQCs. Next, the local mode dependent controller parameters are designed to be the limit (as time approaches infinity) of the conditional mean value of the global mode dependent controller parameters conditioned on the corresponding subsystem modes. A sufficient condition in terms of a set of rank constrained linear matrix inequalities is established to show that the proposed global and local mode dependent output controllers are guaranteed cost controllers. That is, we show that the worst case value of the performance cost functional computed along trajectories of the closed-loop system is bounded.

The organization of the chapter is as follows. Section 2 formulates the problem to be studied. The main results are developed in Sect. 3. An illustrative example and simulations are given in Sect. 4, and Sect. 5 concludes the chapter.

Notation. \mathbb{R}^+ denotes the set of positive real numbers. \mathbb{R}^n , $\mathbb{R}^{n \times m}$, and \mathbb{S}^+ denote, respectively, the n -dimensional Euclidean space, the set of $n \times m$ real matrices,

and the set of real symmetric positive definite matrices of compatible dimensions. Given a matrix $A \in \mathbb{R}^{m \times n}$ with $r = \text{rank}(A) < m$, $A^\perp \in \mathbb{R}^{(m-r) \times m}$ is an orthogonal complement matrix of the matrix A ; A^\perp satisfies $A^\perp A = 0$ and $\text{rank}(A^\perp) = m - r$. $\text{rank}(A) \leq n$ means that the matrix A is real symmetric positive semi-definite and $\text{rank}(A) \leq n$.

2 Problem Formulation

Consider an uncertain Markovian jump large-scale system consisting of N subsystems. The i -th subsystem is described by the equation

$$\mathcal{S}_i : \begin{cases} \dot{x}_i(t) = A_i(\eta_i(t))x_i(t) + B_i(\eta_i(t))u_i(t) + E_i(\eta_i(t))\xi_i(t) + L_i(\eta_i(t))r_i(t), \\ \zeta_i(t) = H_i(\eta_i(t))x_i(t) + G_i(\eta_i(t))u_i(t), \\ y_i(t) = C_i(\eta_i(t))x_i(t) + D_i(\eta_i(t))\xi_i(t), \end{cases} \quad (1)$$

where $i \in \mathcal{N} \triangleq \{1, 2, \dots, N\}$ indicates that \mathcal{S}_i is the i th subsystem of the large-scale system, $x_i(t) \in \mathbb{R}^{n_i}$ is the system state of subsystem \mathcal{S}_i , $u_i(t) \in \mathbb{R}^{m_i}$ is the control input, $y_i(t) \in \mathbb{R}^{t_i}$ is the measured output which will be used for feedback. $\zeta_i(t) \in \mathbb{R}^{q_i}$ is the uncertainty output, $\xi_i(t) \in \mathbb{R}^{p_i}$ is the local uncertainty input, $r_i(t) \in \mathbb{R}^{s_i}$ is the interconnection input, which describes the interconnection effect of subsystems \mathcal{S}_j ($j \in \mathcal{N}, j \neq i$) on \mathcal{S}_i due to the interconnection between subsystem \mathcal{S}_i and subsystems \mathcal{S}_j . The interconnection input $r_i(t)$ is treated as an uncertainty input in this chapter. The random process $\eta_i(t)$ denotes the operation mode of subsystem \mathcal{S}_i ; it takes values in the finite state space $\mathcal{M}_i \triangleq \{1, 2, \dots, M_i\}$. The initial condition of subsystem \mathcal{S}_i is given by $x_{i0} \in \mathbb{R}^{n_i}$ and $\eta_{i0} \in \mathcal{M}_i$.

For the large-scale system, a vector mode process $(\eta_1(t), \eta_2(t), \dots, \eta_N(t))$ can be defined to indicate the collection of operation modes of all the subsystems. Also, an operation mode pattern set \mathcal{M}_p is introduced as the set of all vector mode states that can be visited by the vector mode process $(\eta_1(t), \eta_2(t), \dots, \eta_N(t))$. The operation mode pattern set \mathcal{M}_p is a non-empty subset of the set $\mathcal{M}_1 \times \dots \times \mathcal{M}_N$. Suppose \mathcal{M}_p has M elements where $\max_{i \in \mathcal{N}} M_i \leq M \leq \prod_{i=1}^N M_i$, then we say that the large-scale system has M global operation modes in total. Let $\mathcal{M} \triangleq \{1, 2, \dots, M\}$, then there exists a bijective function $\Psi : \mathcal{M}_p \rightarrow \mathcal{M}$ with $v = \Psi(v_1, \dots, v_N)$. Also we can define functions $\Psi_i^{-1} : \mathcal{M} \rightarrow \mathcal{M}_i$ with $v_i = \Psi_i^{-1}(v)$ for all $i \in \mathcal{N}$. By virtue of the bijective function, we say the large-scale system is in global mode j at time t if $\Psi(\eta_1(t), \dots, \eta_N(t)) = j$. Formally, the global mode process $\eta(t)$ is defined as $\eta(t) \triangleq \Psi(\eta_1(t), \dots, \eta_N(t))$. Now, the mechanism of mode change for the large-scale system can be described by the global mode process $\eta(t)$. The random process $\eta(t)$ in this chapter is assumed to be a stationary ergodic continuous-time Markov process defined on a complete probability space $(\Omega, \mathcal{F}, \text{Pr})$. Moreover, the state transition rate matrix of $\eta(t)$ is assumed to be known and given by $\mathbf{Q} = (q_{\mu\nu}) \in \mathbb{R}^{M \times M}$, in which $q_{\mu\nu} \geq 0$ if $\nu \neq \mu$, and $q_{\mu\mu} \triangleq -\sum_{\nu=1, \nu \neq \mu}^M q_{\mu\nu}$.

Remark 7.1. Some points about the local operation modes of the subsystems in the large-scale system and the global operation mode of the large-scale system need to be clarified.

1. Because of the existence of the bijective function between \mathcal{M}_p and \mathcal{M} , the global operation mode process $\eta(t)$ carries the same information as in the local operation mode vector process $(\eta_1(t), \eta_2(t), \dots, \eta_N(t))$. Therefore, when the global mode dependent control techniques (for example, those in [7, 18]) are used in practical applications, all the local operation modes need to be collected and broadcast throughout the large-scale system. However, collecting and broadcasting all the local operation modes may be expensive or even impossible in some cases. On the other hand, if local mode dependent control techniques (such as, for example, the one developed in [20] and the one to be developed in this work) are used, then collecting and broadcasting the global mode information is not required. This observation motivates our research into local mode dependent control as one approach to reducing communication overheads of the existing decentralized control techniques for Markov jump parameter systems.
2. The assumption that the global operation mode process $\eta(t)$ is a Markov process is reasonable and is motivated by the fact that such a stochastic process model can be used to describe random events in many applications [10].
3. In this work, the global operation mode process $\eta(t)$ is not assumed to be decomposable or nearly decomposable into smaller Markov processes. Therefore, the large-scale system cannot be broken into several independent subsystems, governed by independent Markov chains. On the contrary, the local mode processes $\eta_i(t)$ are in general non-Markovian, and dependent on each other; an illustrative example of this will be given in Sect. 4. This observation in general prohibits the design of local mode dependent controllers by considering each subsystem as an isolated Markovian jump system.

The uncertainties and interconnections in the large-scale system (1) are described by equations of the form

$$\begin{aligned}\xi_i(t) &= \phi_i^\xi(t, \zeta_i(\cdot)|_0^t, \eta_i(\cdot)|_0^t), \\ r_i(t) &= \phi_i^r(t, \zeta_1(\cdot)|_0^t, \dots, \zeta_{i-1}(\cdot)|_0^t, \zeta_{i+1}(\cdot)|_0^t, \dots, \zeta_N(\cdot)|_0^t, \eta(\cdot)|_0^t).\end{aligned}$$

These uncertainties are assumed to satisfy the following IQCs [7, 15, 18].

Definition 7.1. Given a set of matrices $\bar{S}_i \in \mathbb{S}^+$, $i \in \mathcal{N}$. A collection of local uncertainty inputs $\xi_i(t)$, $i \in \mathcal{N}$, is an admissible local uncertainty for the large-scale system if there exists a sequence $\{t_l\}_{l=1}^\infty$ such that $t_l \rightarrow \infty$, $t_l \geq 0$ and

$$E \left(\int_0^{t_l} \left[\|\xi_i(t)\|^2 - \|\zeta_i(t)\|^2 \right] dt \mid x_0, \eta_0 \right) \leq x_{i0}^T \bar{S}_i x_{i0} \quad (2)$$

for all l and for all $i \in \mathcal{N}$, where $x_0 \triangleq [x_{10}^T \cdots x_{N0}^T]^T$, and $\eta_0 \triangleq \eta(0)$. The set of the admissible local uncertainties is denoted by Ξ^ξ .

Definition 7.2. Given a set of matrices $\tilde{S}_i \in \mathbb{S}^+$, $i \in \mathcal{N}$. The large-scale system is said to have admissible interconnections between subsystems if there exists a sequence $\{t_l\}_{l=1}^\infty$ such that $t_l \rightarrow \infty$, $t_l \geq 0$ and

$$\mathbb{E} \left(\int_0^{t_l} \left[\left(\|r_i(t)\|^2 - \sum_{j=1, j \neq i}^N \|\zeta_j(t)\|^2 \right) \right] dt \mid x_0, \eta_0 \right) \leq x_{i0}^T \tilde{S}_i x_{i0} \quad (3)$$

for all l and for all $i \in \mathcal{N}$. The set of the admissible interconnections is denoted by Ξ^r .

Without loss of generality, we assume that the same sequence $\{t_l\}_{l=1}^\infty$ is employed in both definitions.

Remark 7.2. The IQC-type descriptions in (2), (3) can capture a broad class of linear and nonlinear, time-invariant and time-varying, static and dynamic uncertainties and interconnections. Time delays may also be included provided the solution of (1) exists and the process $(x(t), \eta(t))$ is jointly Markov. For example, the norm-bounded uncertainty of the form $\xi_i(t) = \Delta_i(t)\zeta_i(t)$ satisfying $\|\Delta_i(t)\| \leq 1$ belongs to Ξ^ξ because it satisfies the IQC in (2) for any $\tilde{S}_i \in \mathbb{S}^+$ and any sequence $\{t_l\}_{l=1}^\infty$. Also, the dynamic interconnections represented in the operator form as $r_i(s) = \Delta_{r,i}(s)\zeta_{r,i}(s)$ belong to Ξ^r if the transfer function $\Delta_{r,i}(s)$ is from the Hardy space RH^∞ such that $\|\Delta_{r,i}(s)\|_\infty \leq 1$, where $\zeta_{r,i}(s) = [\zeta_1(s) \cdots \zeta_{i-1}(s) \zeta_{i+1}(s) \cdots \zeta_N(s)]$.

Consider a decentralized local mode dependent output feedback controller, whose i -th component has the form

$$\mathcal{K}_i : \begin{cases} \dot{x}_{K,i}(t) = A_{K,i}(\eta_i(t))x_{K,i}(t) + B_{K,i}(\eta_i(t))y_i(t), \\ u_i(t) = C_{K,i}(\eta_i(t))x_{K,i}(t) + D_{K,i}(\eta_i(t))y_i(t), \end{cases} \quad (4)$$

where $x_{K,i}(t) \in \mathbb{R}^{n_i}$ is the state of the controller. The initial controller state is set to zero, and the initial operation mode is set to that of the i -th subsystem \mathcal{S}_i ; see (1). The matrices $A_{K,i}(\nu_i)$, $B_{K,i}(\nu_i)$, $C_{K,i}(\nu_i)$, $D_{K,i}(\nu_i)$, $\nu_i \in \mathcal{M}_i$, are the parameters of the controller to be designed. It is worthwhile to emphasize that for controllers of the form (4) under consideration, the controller parameters are determined by the local operation mode of the subsystem while the controllers proposed in [7, 18] were dependent on the global operation mode defined by values of $\eta(t)$.

Definition 7.3. The closed-loop system corresponding to the uncertain system (1), (2), (3) with the controller (4) is said to be robustly stochastically stable if there exists a constant $c_1 \in \mathbb{R}^+$ such that $x_i(\cdot) \in \mathbb{L}_2[0, \infty)$ for all $i \in \mathcal{N}$ and

$$\sum_{i=1}^N \mathbb{E} \left(\int_0^\infty \|x_i(t)\|^2 dt \mid x_0, \eta_0 \right) \leq c_1 \sum_{i=1}^N \|x_{i0}\|^2 \quad (5)$$

for any initial conditions x_0 , η_0 , any admissible local uncertainty $\xi_i(t)$ and any admissible interconnection $r_i(t)$, $i \in \mathcal{N}$.

Associated with the large-scale system (1) is the quadratic cost functional of the form

$$J \triangleq \sum_{i=1}^N \mathbb{E} \left(\int_0^\infty [x_i^T(t) Q_i(\eta_i(t)) x_i(t) + u_i^T(t) R_i(\eta_i(t)) u_i(t)] dt \mid x_0, \eta_0 \right) \quad (6)$$

where $Q_i(v_i) \in \mathbb{S}^+$, $R_i(v_i) \in \mathbb{S}^+$, $v_i \in \mathcal{M}_i$, $i \in \mathcal{N}$, are given weighting matrices.

The objective of the chapter is to design a dynamic output feedback controller of form (4) for the uncertain system (1), (2), (3), such that the resulting closed-loop system is robustly stochastically stable and the corresponding worst case value of the cost functional (6) subject to the constraints (2) and (3) is upper bounded.

3 Guaranteed Cost Controller Design

This section presents the main results of the chapter. Our local mode dependent controller design technique is based on the decentralized global mode dependent control with controller uncertainties. The design methodology involves augmenting the uncertain system to include effects of mismatch between the global operation mode and the local operation mode controllers. This design methodology is described in Sect. 3.1, where we show how a local mode dependent controller can be derived from a given global mode dependent controller; the result of that section is a sufficient condition to ensure that such a derivation is possible. The design of a suitable auxiliary global mode dependent controller is described in Sect. 3.2. Here we present a sufficient condition for the existence of such an auxiliary output feedback controller. Note that the results of [7] cannot be used in the derivation since there are controller uncertainties in the global mode dependent controllers considered in this work, and also the controllers considered here are more general than those considered in [7]. As a result, the algorithm to design the auxiliary global output feedback controller is different from that proposed in [7]. In Sect. 3.3 we propose a local mode dependent controller design technique based on the auxiliary controller presented in Sect. 3.2. Section 3.4 summarizes the design procedure for local mode dependent controllers.

3.1 Design Methodology

As the global operation mode process $\eta(t)$ is Markovian, we can enlarge the mode state space of the subsystems in the large-scale system (1), and form a new class of uncertain large-scale systems with parameter constraints. The new class of uncertain large-scale systems is described by

$$\tilde{\mathcal{S}}_i : \begin{cases} \dot{\tilde{x}}_i(t) = \tilde{A}_i(\eta(t))\tilde{x}_i(t) + \tilde{B}_i(\eta(t))\tilde{u}_i(t) + \tilde{E}_i(\eta(t))\tilde{\xi}_i(t) + \tilde{L}_i(\eta(t))\tilde{r}_i(t), \\ \dot{\tilde{\xi}}_i(t) = \tilde{H}_i(\eta(t))\tilde{x}_i(t) + \tilde{G}_i(\eta(t))\tilde{u}_i(t), \\ \tilde{y}_i(t) = \tilde{C}_i(\eta(t))\tilde{x}_i(t) + \tilde{D}_i(\eta(t))\tilde{\xi}_i(t), \end{cases} \quad (7)$$

where $\tilde{A}_i(\mu) = A_i(\mu_i)$, $\tilde{B}_i(\mu) = B_i(\mu_i)$, $\tilde{E}_i(\mu) = E_i(\mu_i)$, $\tilde{L}_i(\mu) = L_i(\mu_i)$, $\tilde{H}_i(\mu) = H_i(\mu_i)$, $\tilde{G}_i(\mu) = G_i(\mu_i)$, $\tilde{C}_i(\mu) = C_i(\mu_i)$, $\tilde{D}_i(\mu) = D_i(\mu_i)$, $\mu \in \mathcal{M}$, $\mu_i = \Psi_i^{-1}(\mu)$ and $\Psi_i^{-1}(\cdot)$ is the function $\Psi_i^{-1} : \mathcal{M} \rightarrow \mathcal{M}_i$ introduced in Sect. 2. The uncertainty inputs $\tilde{\xi}_i(t)$ and $\tilde{r}_i(t)$ are described, respectively, by the same functions as $\xi_i(t)$ and $r_i(t)$ in (1). So $\tilde{\xi}_i(t) \in \Xi^\xi$ and $\tilde{r}_i(t) \in \Xi^r$. The initial condition is given by $\tilde{x}_{i0} = x_{i0}$, $i \in \mathcal{N}$, and η_0 . It can be seen that the system (7) and the system (1) are in fact the same. However, the operation mode of each subsystem in this new class of systems is now Markovian and hence the corresponding control problem is easier to deal with.

Associated with the uncertain system (7) is the following quadratic cost functional of the form

$$\tilde{J} \triangleq \sum_{i=1}^N \mathbb{E} \left(\int_0^\infty [\tilde{x}_i^T(t)\tilde{Q}_i(\eta(t))\tilde{x}_i(t) + \tilde{u}_i^T(t)\tilde{R}_i(\eta(t))\tilde{u}_i(t)] dt \mid \tilde{x}_0, \eta_0 \right), \quad (8)$$

where $\tilde{Q}_i(\mu) = Q_i(\mu_i)$, $\tilde{R}_i(\mu) = R_i(\mu_i)$, $\mu \in \mathcal{M}$, $\mu_i = \Psi_i^{-1}(\mu) \in \mathcal{M}_i$ for $i \in \mathcal{N}$, and $\tilde{x}_0 = x_0$.

We now consider the problem of guaranteed cost control of the system (7) by means of an uncertain decentralized global mode dependent output feedback controller of the form

$$\tilde{\mathcal{K}}_i : \begin{cases} \dot{\tilde{x}}_{K,i}(t) = \tilde{A}_{K,i}(\eta(t))\tilde{x}_{K,i}(t) + \tilde{B}_{K,i}(\eta(t))\tilde{y}_i(t) + \tilde{\xi}_{1i}(t) + \tilde{\xi}_{2i}(t), \\ \tilde{u}_i(t) = \tilde{C}_{K,i}(\eta(t))\tilde{x}_{K,i}(t) + \tilde{D}_{K,i}(\eta(t))\tilde{y}_i(t) + \tilde{\xi}_{3i}(t) + \tilde{\xi}_{4i}(t). \end{cases} \quad (9)$$

The initial state of the controller is zero, and the initial operation mode is the same as that of the system (7). The controller dynamics are assumed to be subject to controller uncertainties of the form

$$\begin{aligned} \tilde{\xi}_{1i}(t) &= \phi_{1i}(t, \tilde{x}_{K,i}(t), \eta(t)), & \tilde{\xi}_{2i}(t) &= \phi_{2i}(t, \tilde{y}_i(t), \eta(t)), \\ \tilde{\xi}_{3i}(t) &= \phi_{3i}(t, \tilde{x}_{K,i}(t), \eta(t)), & \tilde{\xi}_{4i}(t) &= \phi_{4i}(t, \tilde{y}_i(t), \eta(t)), \end{aligned}$$

which satisfy the following IQCs.

Definition 7.4. Given $\beta_{1i}(\mu), \beta_{2i}(\mu), \beta_{3i}(\mu), \beta_{4i}(\mu) \in \mathbb{R}^+$, $\mu \in \mathcal{M}$, $i \in \mathcal{N}$. A collection of controller uncertainties $\tilde{\xi}_{1i}(t), \tilde{\xi}_{2i}(t), \tilde{\xi}_{3i}(t), \tilde{\xi}_{4i}(t)$, $i \in \mathcal{N}$, is an admissible uncertainty input for the dynamic controller in (9) if there exists a sequence $\{t_l\}_{l=1}^\infty$ such that $t_l \rightarrow \infty$, $t_l \geq 0$ and

$$\mathbb{E} \left(\int_0^{t_l} \left[\left\| \tilde{\xi}_{1i}(t) \right\|^2 - \beta_{1i}^2(\eta(t)) \|\tilde{x}_{K,i}(t)\|^2 \right] dt \mid \tilde{x}_0, \eta_0 \right) \leq 0, \quad (10)$$

$$\mathbb{E} \left(\int_0^{t_l} \left[\left\| \tilde{\xi}_{2i}(t) \right\|^2 - \beta_{2i}^2(\eta(t)) \|\tilde{y}_i(t)\|^2 \right] dt \mid \tilde{x}_0, \eta_0 \right) \leq 0, \quad (11)$$

$$\mathbb{E} \left(\int_0^{t_l} \left[\left\| \tilde{\xi}_{3i}(t) \right\|^2 - \beta_{3i}^2(\eta(t)) \|\tilde{x}_{K,i}(t)\|^2 \right] dt \mid \tilde{x}_0, \eta_0 \right) \leq 0, \quad (12)$$

$$\mathbb{E} \left(\int_0^{t_l} \left[\left\| \tilde{\xi}_{4i}(t) \right\|^2 - \beta_{4i}^2(\eta(t)) \|\tilde{y}_i(t)\|^2 \right] dt \mid \tilde{x}_0, \eta_0 \right) \leq 0, \quad (13)$$

for all l and for all $i \in \mathcal{N}$. The set of the admissible controller uncertainties is denoted by Ξ^K .

Again, one can assume that the same sequence $\{t_l\}_{l=1}^\infty$ is selected in Definitions 7.1, 7.2 and 7.4.

Remark 7.3. In the uncertain controller (9), we use four (instead of two) controller uncertainty inputs. Also, we use unity matrices as the coefficients of the controller uncertainty inputs. One reason for this choice is to facilitate the derivation of the result in the following theorem. Another reason is that the design of controller (9) is not the main objective of the work and only serves as an intermediate step to obtain a local mode dependent controller of the form (4).

The following theorem demonstrates the auxiliary role of the global mode dependent controller (9) in the design of the local mode dependent controller (4). It establishes a sufficient condition for the controller in (4) to stabilize the uncertain system (1) provided the controller (9) stabilizes the uncertain system (7). The condition is expressed in terms of the differences between the controller parameter matrices.

Theorem 7.1. *Suppose an uncertain controller (9) stochastically stabilizes the uncertain large-scale system (7) subject to constraints (2), (3), (10)–(13) and leads to the cost bound $\sup_{\Xi^{\xi}, \Xi^r, \Xi^K} \tilde{J} < c$ for some $c \in \mathbb{R}^+$. If the controller parameter matrices in (4) are chosen so that*

$$\left\| \tilde{A}_{K,i}(\mu) - A_{K,i}(\mu_i) \right\| \leq \beta_{1i}(\mu), \quad (14)$$

$$\left\| \tilde{B}_{K,i}(\mu) - B_{K,i}(\mu_i) \right\| \leq \beta_{2i}(\mu), \quad (15)$$

$$\left\| \tilde{C}_{K,i}(\mu) - C_{K,i}(\mu_i) \right\| \leq \beta_{3i}(\mu), \quad (16)$$

$$\left\| \tilde{D}_{K,i}(\mu) - D_{K,i}(\mu_i) \right\| \leq \beta_{4i}(\mu), \quad (17)$$

for all $\mu \in \mathcal{M}$, $i \in \mathcal{N}$, where $\mu_i = \Psi_i^{-1}(\mu) \in \mathcal{M}_i$, then the controller in (4) stochastically stabilizes the uncertain large-scale system in (1) subject to constraints (2), (3) and also leads to the cost bound $\sup_{\Xi^{\xi}, \Xi^r} J < c$.

Proof. Define the matrices

$$\Delta_{1i}(\mu) \triangleq \tilde{A}_{K,i}(\mu) - A_{K,i}(\mu_i), \quad \Delta_{2i}(\mu) \triangleq \tilde{B}_{K,i}(\mu) - B_{K,i}(\mu_i),$$

$$\Delta_{3i}(\mu) \triangleq \tilde{C}_{K,i}(\mu) - C_{K,i}(\mu_i), \quad \Delta_{4i}(\mu) \triangleq \tilde{D}_{K,i}(\mu) - D_{K,i}(\mu_i),$$

for $\mu \in \mathcal{M}$, where $\mu_i = \Psi_i^{-1}(\mu) \in \mathcal{M}_i$. Then the inequalities in (14)–(17) imply that

$$\begin{aligned}\|\Delta_{1i}(\mu)\| &\leq \beta_{1i}(\mu), & \|\Delta_{2i}(\mu)\| &\leq \beta_{2i}(\mu), \\ \|\Delta_{3i}(\mu)\| &\leq \beta_{3i}(\mu), & \|\Delta_{4i}(\mu)\| &\leq \beta_{4i}(\mu).\end{aligned}$$

Now consider a set of particular controller uncertainties of the form

$$\tilde{\xi}_{1i}(t) = -\Delta_{1i}(\eta(t))\tilde{x}_{K,i}(t), \quad \tilde{\xi}_{2i}(t) = -\Delta_{2i}(\eta(t))\tilde{y}_i(t), \quad (18)$$

$$\tilde{\xi}_{3i}(t) = -\Delta_{3i}(\eta(t))\tilde{x}_{K,i}(t), \quad \tilde{\xi}_{4i}(t) = -\Delta_{4i}(\eta(t))\tilde{y}_i(t). \quad (19)$$

We have

$$\begin{aligned}\left\|\tilde{\xi}_{1i}(t)\right\|^2 &\leq \beta_{1i}^2(\eta(t))\|\tilde{x}_{K,i}(t)\|^2, & \left\|\tilde{\xi}_{2i}(t)\right\|^2 &\leq \beta_{2i}^2(\eta(t))\|\tilde{y}_i(t)\|^2, \\ \left\|\tilde{\xi}_{3i}(t)\right\|^2 &\leq \beta_{3i}^2(\eta(t))\|\tilde{x}_{K,i}(t)\|^2, & \left\|\tilde{\xi}_{4i}(t)\right\|^2 &\leq \beta_{4i}^2(\eta(t))\|\tilde{y}_i(t)\|^2.\end{aligned}$$

Hence the IQCs in (10)–(13) are satisfied. Thus the uncertainty inputs $\tilde{\xi}_{1i}(t)$, $\tilde{\xi}_{2i}(t)$, $\tilde{\xi}_{3i}(t)$ and $\tilde{\xi}_{4i}(t)$ defined in (18), (19), are an admissible uncertainty for controller (9). Also we have

$$\begin{aligned}\tilde{A}_{K,i}(\eta(t))\tilde{x}_{K,i}(t) + \tilde{\xi}_{1i}(t) &= A_{K,i}(\eta_i(t))\tilde{x}_{K,i}(t), \\ \tilde{B}_{K,i}(\eta(t))\tilde{y}_i(t) + \tilde{\xi}_{2i}(t) &= B_{K,i}(\eta_i(t))\tilde{y}_i(t), \\ \tilde{C}_{K,i}(\eta(t))\tilde{x}_{K,i}(t) + \tilde{\xi}_{3i}(t) &= C_{K,i}(\eta_i(t))\tilde{x}_{K,i}(t), \\ \tilde{D}_{K,i}(\eta(t))\tilde{y}_i(t) + \tilde{\xi}_{4i}(t) &= D_{K,i}(\eta_i(t))\tilde{y}_i(t).\end{aligned}$$

This means that the controller in (4) is a special case of the controller in (9) corresponding to the particular uncertainties given in (18), (19). Also note that the uncertain system (7) is in fact the same as the system in (1). It follows that the stability of the system (7) implies that of system (1).

Finally, the upper bound of the cost functional follows from the following observation:

$$\sup_{\Xi^\xi, \Xi^r} J = \sup_{\substack{\Xi^\xi, \Xi^r, \tilde{\xi}_{1i}(t), \tilde{\xi}_{2i}(t), \\ \tilde{\xi}_{3i}(t), \tilde{\xi}_{4i}(t), \text{given in (18) (19)}}} \tilde{J} \leq \sup_{\Xi^\xi, \Xi^r, \Xi^K} \tilde{J} < c.$$

Here the “ \leq ” inequality holds because the uncertainty inputs $\tilde{\xi}_{1i}(t)$, $\tilde{\xi}_{2i}(t)$, $\tilde{\xi}_{3i}(t)$ and $\tilde{\xi}_{4i}(t)$ defined in (18), (19), belong to Ξ^K . \square

3.2 Design of Global Mode Dependent Controllers

In this section, a sufficient condition is established for the design of the uncertain guaranteed cost controllers of the form (9). This condition, together with

Theorem 7.1, provides a basis for the design of local mode dependent guaranteed cost controllers of the form (4). The following notation is defined in advance to facilitate the presentation.

For every $\mu \in \mathcal{M}$ and $i \in \mathcal{N}$, let $\beta_{1i}(\mu), \beta_{2i}(\mu), \beta_{3i}(\mu), \beta_{4i}(\mu)$ be positive constants from Definition 7.4. Also, let us be given symmetric matrices $P_i(\mu) \in \mathbb{S}^+$, $X_i(\mu) \in \mathbb{S}^+$ and a collection of positive scalars $\tau_i, \theta_i, \tau_{1i}, \tau_{2i}, \tau_{3i}, \tau_{4i}, \bar{\tau}_i, \bar{\tau}_{1i}, \bar{\tau}_{2i}, \bar{\tau}_{3i}, \bar{\tau}_{4i}$. Using these matrices and constants define the following matrices:

$$\begin{aligned}
\hat{A}_i(\mu) &= \begin{bmatrix} \tilde{A}_i(\mu) & 0 \\ 0 & 0 \end{bmatrix} = N_i \tilde{A}_i(\mu) N_i^T, \quad N_i = \begin{bmatrix} I_{n_i \times n_i} \\ 0_{n_i \times n_i} \end{bmatrix}, \\
\hat{B}_i(\mu) &= \begin{bmatrix} \tilde{E}_i(\mu) & \tilde{L}_i(\mu) & 0 & 0 & \tilde{B}_i(\mu) & \tilde{B}_i(\mu) \\ 0 & 0 & I & I & 0 & 0 \end{bmatrix}, \\
\hat{C}_i(\mu) &= \begin{bmatrix} \tilde{Q}_i^{\frac{1}{2}}(\mu) & 0 \\ 0 & 0 \\ \tilde{H}_i(\mu) & 0 \\ 0 & \beta_{1i}(\mu)I \\ \beta_{2i}(\mu)\tilde{C}_i(\mu) & 0 \\ 0 & \beta_{3i}(\mu)I \\ \beta_{4i}(\mu)\tilde{C}_i(\mu) & 0 \end{bmatrix}, \\
\hat{D}_i(\mu) &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \tilde{R}_i^{\frac{1}{2}}(\mu) & \tilde{R}_i^{\frac{1}{2}}(\mu) \\ 0 & 0 & 0 & \tilde{G}_i(\mu) & \tilde{G}_i(\mu) \\ 0 & 0 & 0 & 0 & 0 \\ \beta_{2i}(\mu)\tilde{D}_i(\mu) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \beta_{4i}(\mu)\tilde{D}_i(\mu) & 0 & 0 & 0 & 0 \end{bmatrix}, \\
\underline{B}_i(\mu) &= \begin{bmatrix} 0 & \tilde{B}_i(\mu) \\ I & 0 \end{bmatrix}, \quad \underline{C}_i(\mu) = \begin{bmatrix} 0 & I \\ \tilde{C}_i(\mu) & 0 \end{bmatrix}, \quad \underline{D}_i(\mu) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \tilde{D}_i(\mu) & 0 & 0 & 0 & 0 \end{bmatrix}, \\
\underline{E}_i(\mu) &= \begin{bmatrix} 0 & 0 \\ 0 & \tilde{R}_i^{\frac{1}{2}}(\mu) \\ 0 & \tilde{G}_i(\mu) \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \Gamma_{2i}(\mu) = \begin{array}{c|ccccc|c} V_{1i}(\mu) & 0 & 0 & V_{2i}(\mu) & 0 & 0 & 0 & 0 \\ \hline 0 & I & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I \end{array}, \\
[V_{1i}(\mu) & V_{2i}(\mu)] = \begin{bmatrix} \tilde{B}_i(\mu) \\ \tilde{R}_i^{\frac{1}{2}}(\mu) \\ \tilde{G}_i(\mu) \end{bmatrix}^\perp, \quad \tilde{K}_i(\mu) = \begin{bmatrix} \tilde{A}_{K,i}(\mu) & \tilde{B}_{K,i}(\mu) \\ \tilde{C}_{K,i}(\mu) & \tilde{D}_{K,i}(\mu) \end{bmatrix}, \tag{20}
\end{aligned}$$

$$\Phi_i(\mu) = \begin{bmatrix} \hat{A}_i^T(\mu)P_i(\mu) + P_i(\mu)\hat{A}_i(\mu) + \sum_{v=1}^M q_{\mu v}P_i(v) & P_i(\mu)\hat{B}_i(\mu) & \hat{C}_i^T(\mu) \\ \hat{B}_i^T(\mu)P_i(\mu) & Q_{2i} & \hat{D}_i^T(\mu) \\ \hat{C}_i(\mu) & \hat{D}_i(\mu) & Q_{3i} \end{bmatrix}, \quad (21)$$

$$\Psi_{li}(\mu) = \begin{bmatrix} P_i(\mu)\underline{B}_i(\mu) \\ 0 \\ \underline{E}_i(\mu) \end{bmatrix}, \quad \Psi_{ri}(\mu) = [C_i(\mu) \ D_i(\mu) \ 0], \quad (22)$$

and

$$Q_{1i}(\mu) = N_i^T X_i(\mu) N_i \tilde{A}_i^T(\mu) + \tilde{A}_i(\mu) N_i^T X_i(\mu) N_i + q_{\mu\mu} N_i^T X_i(\mu) N_i,$$

$$Q_{2i} = -\text{diag}(\tau_i I, \theta_i I, \tau_{1i} I, \tau_{2i} I, \tau_{3i} I, \tau_{4i} I),$$

$$Q_{3i} = -\text{diag}(I, I, \bar{\tau}_i I, \bar{\tau}_{1i} I, \bar{\tau}_{2i} I, \bar{\tau}_{3i} I, \bar{\tau}_{4i} I),$$

$$\Gamma_{li}(\mu) = \left[\begin{array}{cccccc} \sqrt{q_{\mu,1}} N_i^T X_i(\mu) & \cdots & \sqrt{q_{\mu,\mu-1}} N_i^T X_i(\mu) & \sqrt{q_{\mu,\mu+1}} N_i^T X_i(\mu) & & \\ & \cdots & & \sqrt{q_{\mu,M}} N_i^T X_i(\mu) & & \end{array} \right],$$

$$Q_{4i}(\mu) = -\text{diag}(X_i(1), \dots, X_i(\mu-1), X_i(\mu+1), \dots, X_i(M)),$$

$$Q_{5i}(\mu) = \tilde{A}_i^T(\mu) N_i^T P_i(\mu) N_i + N_i^T P_i(\mu) N_i \tilde{A}_i(\mu) + \sum_{v=1}^M q_{\mu v} N_i^T P_i(v) N_i.$$

Now, we are in a position to state the main result of this section, which provides a sufficient condition for the design of an output feedback controller of the form (9) such that the resulting closed-loop system is robustly stochastically stable and the worst case value of the cost functional in (8) is bounded. The result involves the following coupled linear matrix inequalities with respect to the variables $P_i(\mu) \in \mathbb{S}^+$, $X_i(\mu) \in \mathbb{S}^+$, and $\tau_i, \theta_i, \tau_{1i}, \tau_{2i}, \tau_{3i}, \tau_{4i}, \bar{\tau}_i, \bar{\tau}_{1i}, \bar{\tau}_{2i}, \bar{\tau}_{3i}, \bar{\tau}_{4i} \in \mathbb{R}^+$, $\mu \in \mathcal{M}$, $i \in \mathcal{N}$:

$$\Gamma_{2i}(\mu) \begin{bmatrix} Q_{1i}(\mu) & N_i^T \hat{B}_i(\mu) & N_i^T X_i(\mu) \hat{C}_i^T(\mu) & \Gamma_{1i}(\mu) \\ \hat{B}_i^T(\mu) N_i & Q_{2i} & \hat{D}_i^T(\mu) & 0 \\ \hat{C}_i(\mu) X_i(\mu) N_i & \hat{D}_i(\mu) & Q_{3i} & 0 \\ \Gamma_{1i}^T(\mu) & 0 & 0 & Q_{4i}(\mu) \end{bmatrix} \Gamma_{2i}^T(\mu) < 0, \quad (23)$$

$$\begin{aligned} & \left[\begin{bmatrix} \tilde{C}_i^T(\mu) \\ \tilde{D}_i^T(\mu) \end{bmatrix}^\perp \ 0 \right] \left[\begin{array}{ccc} Q_{5i}(\mu) & N_i^T P_i(\mu) \hat{B}_i(\mu) & N_i^T \hat{C}_i^T(\mu) \\ \hat{B}_i^T(\mu) P_i(\mu) N_i & Q_{2i} & \hat{D}_i^T(\mu) \\ \hat{C}_i(\mu) N_i & \hat{D}_i(\mu) & Q_{3i} \end{array} \right] \\ & \times \left[\begin{bmatrix} \tilde{C}_i^T(\mu) \\ \tilde{D}_i^T(\mu) \end{bmatrix}^\perp \ 0 \\ 0 & I \end{bmatrix}^T < 0, \quad (24) \end{aligned}$$

with rank constraints

$$\overline{\text{rank}} \left(\begin{bmatrix} P_i(\mu) & I \\ I & X_i(\mu) \end{bmatrix} \right) \leq 2n, \quad (25)$$

$$\overline{\text{rank}} \left(\begin{bmatrix} \tau_i + \bar{\theta}_i & 1 \\ 1 & \bar{\tau}_i \end{bmatrix} \right) \leq 1, \quad \overline{\text{rank}} \left(\begin{bmatrix} \tau_{1i} & 1 \\ 1 & \bar{\tau}_{1i} \end{bmatrix} \right) \leq 1, \quad (26)$$

$$\overline{\text{rank}} \left(\begin{bmatrix} \tau_{2i} & 1 \\ 1 & \bar{\tau}_{2i} \end{bmatrix} \right) \leq 1, \quad \overline{\text{rank}} \left(\begin{bmatrix} \tau_{3i} & 1 \\ 1 & \bar{\tau}_{3i} \end{bmatrix} \right) \leq 1, \quad \overline{\text{rank}} \left(\begin{bmatrix} \tau_{4i} & 1 \\ 1 & \bar{\tau}_{4i} \end{bmatrix} \right) \leq 1, \quad (27)$$

where $\bar{\theta}_i \triangleq \sum_{j=1, j \neq i}^N \theta_j$.

Theorem 7.2. Given $\beta_{1i}(\mu), \beta_{2i}(\mu), \beta_{3i}(\mu), \beta_{4i}(\mu) \in \mathbb{R}^+$. Suppose there exist matrices $P_i(\mu) \in \mathbb{S}^+$, $X_i(\mu) \in \mathbb{S}^+$ and scalars $\tau_i, \theta_i, \tau_{1i}, \tau_{2i}, \tau_{3i}, \tau_{4i}, \bar{\tau}_i, \bar{\tau}_{1i}, \bar{\tau}_{2i}, \bar{\tau}_{3i}, \bar{\tau}_{4i} \in \mathbb{R}^+$, $\mu \in \mathcal{M}$, $i \in \mathcal{N}$, such that the coupled linear matrix inequalities (23)–(27) hold for all $\mu \in \mathcal{M}$, $i \in \mathcal{N}$. Consider an uncertain decentralized global mode dependent output feedback controller of the form (9) whose matrix parameters $\tilde{K}_i(\mu)$ defined in (20) are obtained by solving the coupled linear matrix inequalities

$$\Phi_i(\mu) + \Psi_{li}(\mu)\tilde{K}_i(\mu)\Psi_{ri}(\mu) + \Psi_{ri}^T(\mu)\tilde{K}_i^T(\mu)\Psi_{li}^T(\mu) < 0, \quad (28)$$

where the matrices $\Phi_i(\mu)$, $\Psi_{li}(\mu)$ and $\Psi_{ri}(\mu)$ are defined in (21), (22) using the solutions to (23)–(27). Then the closed-loop large-scale system, consisting of the uncertain subsystems (7) and this controller and subject to (2), (3) and (10)–(13), is robustly stochastically stable. Also, the worst-case cost functional (8) computed on the trajectories of this closed-loop system is bounded as follows:

$$\sup_{\Xi^\xi, \Xi^r, \Xi^K} \tilde{J} < \sum_{i=1}^N \tilde{x}_{i0}^T [N_i^T P_i(\eta_{i0}) N_i + \tau_i \bar{S}_i + \theta_i \tilde{S}_i] \tilde{x}_{i0}. \quad (29)$$

Proof. The proof is given in Appendix 1. \square

Remark 7.4. The design condition given in Theorem 7.2 takes into account the IQC structures of the uncertainties and the interconnections in the large-scale system. In fact, the IQC structure information is embedded into the design condition through the positive scalars $\tau_i, \theta_i, \tau_{1i}, \tau_{2i}, \tau_{3i}$ and τ_{4i} . Also, the following optimization problem

$$\inf_{\substack{P_i(\mu), X_i(\mu), \tau_i, \theta_i, \tau_{1i}, \tau_{2i}, \tau_{3i}, \tau_{4i}, \\ \text{subject to (23)–(27)}}} \sum_{i=1}^N \tilde{x}_{i0}^T [N_i^T P_i(\eta_{i0}) N_i + \tau_i \bar{S}_i + \theta_i \tilde{S}_i] \tilde{x}_{i0} \quad (30)$$

can be used to minimize the upper bound of cost functional (8). Solving this optimization problem would lead to a suboptimal guaranteed cost controller for the auxiliary control problem.

Remark 7.5. Due to the rank constraints (25)–(27), the solution set to (23)–(27) is not convex. Although in general it is difficult to solve such problems, several

numerical algorithms have been proposed for this purpose, such as the cone complementarity linearization algorithm in [4], the tangent and lift method in [12]. In this work, we used the algorithm from [11, 12] with good results.

In the next section, some additional conditions will be imposed on the matrices $\tilde{K}_i(\mu)$ from Theorem 7.2 to obtain a local mode dependent guaranteed cost controller from $\tilde{K}_i(\mu)$; see Theorem 7.3.

3.3 The Main Result: Design of Local Mode Dependent Controllers

In this section, we show that the local mode dependent controller (4) can be designed from a global mode dependent controller of the form (9) using the stochastic projection method proposed in [20]. Specifically, we show that a robust stabilizing dynamic local mode dependent controller (4) can be obtained by taking the expectation of a controller (9) conditioned on the subsystem operation modes as time approaches infinity. In combination with the results of Theorems 7.1 and 7.2, the result of this section yields a computational method for the design of a local mode dependent guaranteed cost controller (4). The method is presented in Theorem 7.3 given below.

Proposition 7.1 ([20]). *Given the matrices $\tilde{K}_i(\mu)$, $\mu \in \mathcal{M}$, $i \in \mathcal{N}$, defined in Theorem 7.2, let*

$$K_i(v_i) = \begin{bmatrix} A_{K,i}(v_i) & B_{K,i}(v_i) \\ C_{K,i}(v_i) & D_{K,i}(v_i) \end{bmatrix} \triangleq \frac{\sum_{\mu=1}^M \{\tilde{K}_i(\mu)\pi_{\infty\mu}\mathbb{I}_i(\mu, v_i)\}}{\sum_{\mu=1}^M \{\pi_{\infty\mu}\mathbb{I}_i(\mu, v_i)\}} \quad (31)$$

for all $v_i \in \mathcal{M}_i$, $i \in \mathcal{N}$, where

$$\mathbb{I}_i(\mu, v_i) = \begin{cases} 1 & \text{if } v_i = \Psi_i^{-1}(\mu), \\ 0 & \text{otherwise,} \end{cases}$$

$\pi_{\infty\mu}$ is the μ -th component of the vector $\pi_\infty = \mathbf{e}(\mathbf{Q} + \mathbf{E})^{-1}$ with $\mathbf{e} = [1 \ 1 \ \dots \ 1]^T \in \mathbb{R}^{1 \times M}$ and $\mathbf{E} = [\mathbf{e}^T \ \mathbf{e}^T \ \dots \ \mathbf{e}^T]^T \in \mathbb{R}^{M \times M}$. Then

$$K_i(v_i) = \lim_{t \rightarrow \infty} \mathbb{E} (\tilde{K}_i(\eta(t)) \mid \eta_i(t) = v_i).$$

Proof. First we observe that the vector π_∞ is the steady state distribution of the global mode process $\eta(t)$. The ergodic property of the Markov process $\eta(t)$ implies that $\lim_{t \rightarrow \infty} e^{\mathbf{Q}t} = [\pi_\infty^T \ \dots \ \pi_\infty^T]^T$. Then the probability distribution of the matrix-valued random process $\tilde{K}_i(\eta(t))$ has a limit as $t \rightarrow \infty$, which is $\lim_{t \rightarrow \infty} \Pr(\tilde{K}_i(\eta(t)) = \tilde{K}_i(\mu)) = \pi_{\infty\mu}$ for all $\mu \in \mathcal{M}$. So the expected value of $\tilde{K}_i(\eta(t))$ conditioned on the subsystem operation mode, as $t \rightarrow \infty$, is given by

$$\begin{aligned}
& \lim_{t \rightarrow \infty} \mathbb{E} (\tilde{K}_i(\eta(t)) \mid \eta_i(t) = v_i) \\
&= \sum_{\mu=1}^M \left\{ \tilde{K}_i(\mu) \lim_{t \rightarrow \infty} \Pr(\eta(t) = \mu \mid \eta_i(t) = v_i) \right\} \\
&= \sum_{\mu=1}^M \left\{ \tilde{K}_i(\mu) \lim_{t \rightarrow \infty} \frac{\Pr(\eta(t) = \mu, \eta_i(t) = v_i)}{\Pr(\eta_i(t) = v_i)} \right\} \\
&= \sum_{\mu=1}^M \left\{ \tilde{K}_i(\mu) \frac{\pi_{\infty \mu} \mathbb{I}_i(\mu, v_i)}{\sum_{\mu=1}^M \{\pi_{\infty \mu} \mathbb{I}_i(\mu, v_i)\}} \right\} \\
&= K_i(v_i).
\end{aligned}$$

This completes the proof. \square

Note that by choosing the matrix $K_i(v_i)$ to be the limit of the conditional expectation of $\tilde{K}_i(\eta(t))$ given the subsystem operation mode, we introduce the asymptotically most accurate local mode dependent approximation of $\tilde{K}_i(\eta(t))$. Indeed, the parameters of the local mode dependent controller (4) given in (31) are the minimum variance approximation of the corresponding controller parameters in (9) in the sense that $\lim_{t \rightarrow \infty} \mathbb{E}(\|\tilde{K}_i(\eta(t)) - K_i(\eta_i(t))\|_F^2 \mid \eta_i(t))$ is minimal [1, Theorem 3.1]; here $\|\cdot\|_F$ denotes the Frobenius norm.

Also, we have

$$\Delta_i(\mu) \triangleq \tilde{K}_i(\mu) - K_i(\mu_i) = \frac{\sum_{v=1, v \neq \mu}^M \{\mathbb{I}_i(v, \mu_i) \pi_{\infty v} [\tilde{K}_i(\mu) - \tilde{K}_i(v)]\}}{\sum_{v=1}^M \{\mathbb{I}_i(v, \mu_i) \pi_{\infty v}\}}, \quad (32)$$

where $\mu_i = \Psi_i^{-1}(\mu) \in \mathcal{M}_i$, $\mu \in \mathcal{M}$. Note that $\Delta_i(\mu)$ is a linear matrix function of $\tilde{K}_i(\mu)$, $\mu \in \mathcal{M}$.

A computational method for the design of guaranteed cost controller (4) is presented in the following result, which is based upon Theorem 7.1 and Theorem 7.2. The idea behind this result is to construct a global mode dependent controller of the form (9), using Theorem 7.2, such that the corresponding Δ 's given in (32) also satisfy the conditions of Theorem 7.1. Then, one will be able to construct the corresponding local mode dependent controller (4) with parameters given in (31), and concludes that such a controller stabilizes the system (1) and ensures the worst case quadratic performance cost which does not exceed that of the underlying auxiliary global mode dependent controller (9). Let

$$N_{1i} = \begin{bmatrix} I_{n_i \times n_i} \\ 0_{m_i \times n_i} \end{bmatrix}, \quad N_{2i} = \begin{bmatrix} I_{n_i \times n_i} \\ 0_{t_i \times n_i} \end{bmatrix}, \quad N_{3i} = \begin{bmatrix} 0_{n_i \times t_i} \\ I_{t_i \times t_i} \end{bmatrix}, \quad N_{4i} = \begin{bmatrix} 0_{n_i \times m_i} \\ I_{m_i \times m_i} \end{bmatrix}.$$

Theorem 7.3. Given a set of $\beta_{1i}(\mu)$, $\beta_{2i}(\mu)$, $\beta_{3i}(\mu)$, $\beta_{4i}(\mu) \in \mathbb{R}^+$. Suppose a set of solutions $P_i(\mu) \in \mathbb{S}^+$, $X_i(\mu) \in \mathbb{S}^+$, τ_i , θ_i , τ_{1i} , τ_{2i} , τ_{3i} , τ_{4i} , $\bar{\tau}_i$, $\bar{\tau}_{1i}$, $\bar{\tau}_{2i}$, $\bar{\tau}_{3i}$, $\bar{\tau}_{4i} \in \mathbb{R}^+$, $\mu \in \mathcal{M}$, $i \in \mathcal{N}$ is found for (23)–(27).

If there exist matrices $\tilde{K}(\mu)$ such that the following LMIs

$$\begin{bmatrix} \beta_{1i}(\mu)I & N_{2i}^T \Delta_i^T(\mu) N_{1i} \\ N_{1i}^T \Delta_i(\mu) N_{2i} & \beta_{1i}(\mu)I \end{bmatrix} \geq 0, \quad (33)$$

$$\begin{bmatrix} \beta_{2i}(\mu)I & N_{3i}^T \Delta_i^T(\mu) N_{1i} \\ N_{1i}^T \Delta_i(\mu) N_{3i} & \beta_{2i}(\mu)I \end{bmatrix} \geq 0, \quad (34)$$

$$\begin{bmatrix} \beta_{3i}(\mu)I & N_{2i}^T \Delta_i^T(\mu) N_{4i} \\ N_{4i}^T \Delta_i(\mu) N_{2i} & \beta_{3i}(\mu)I \end{bmatrix} \geq 0, \quad (35)$$

$$\begin{bmatrix} \beta_{4i}(\mu)I & N_{3i}^T \Delta_i^T(\mu) N_{4i} \\ N_{4i}^T \Delta_i(\mu) N_{3i} & \beta_{4i}(\mu)I \end{bmatrix} \geq 0, \quad (36)$$

and the linear matrix inequalities in (28) hold for all $\mu \in \mathcal{M}$, $i \in \mathcal{N}$, where $\Delta_i(\mu)$ is the linear function of $\tilde{K}_i(\mu)$ defined in (32), then the local mode dependent controller of the form (4) with parameters defined in (31) robustly stabilizes the uncertain system (1) subject to the constraints (2), (3) and leads to the cost bound

$$\sup_{\Xi^\xi, \Xi^r} J < \sum_{i=1}^N x_{i0}^T [N_i^T P_i(\eta_{i0}) N_i + \tau_i \bar{S}_i + \theta_i \tilde{S}_i] x_{i0}. \quad (37)$$

Proof. According to Theorem 7.2, the controller of the form (9) given by the solution of (28), (33)–(36), robustly stochastically stabilizes the uncertain system (7), and leads to the cost bound (29).

In view of (32), we have that

$$\begin{aligned} N_{1i}^T \Delta_i(\mu) N_{2i} &= \tilde{A}_{K,i}(\mu) - A_{K,i}(\mu_i), \\ N_{1i}^T \Delta_i(\mu) N_{3i} &= \tilde{B}_{K,i}(\mu) - B_{K,i}(\mu_i), \\ N_{4i}^T \Delta_i(\mu) N_{2i} &= \tilde{C}_{K,i}(\mu) - C_{K,i}(\mu_i), \\ N_{4i}^T \Delta_i(\mu) N_{3i} &= \tilde{D}_{K,i}(\mu) - D_{K,i}(\mu_i). \end{aligned}$$

Therefore, the LMIs in (33)–(36) are equivalent to the inequalities in (14)–(17), respectively. In view of Theorem 7.1, the local mode dependent controller given in (31) robustly stabilizes the uncertain system (1) and leads to the cost bound in (37). \square

3.4 Design Procedure

The proposed controller design procedure is summarized as follows.

- Given $\beta_{1i}(\mu), \beta_{2i}(\mu), \beta_{3i}(\mu), \beta_{4i}(\mu) \in \mathbb{R}^+$, $\mu \in \mathcal{M}$, $i \in \mathcal{N}$, find a feasible solution to the rank constrained LMI problem in (23)–(27). As noted above, although there does not exist an algorithm that is guaranteed to solve this problem, the algorithm in [12] can be applied with good results;

2. Using this feasible solution, find a global mode dependent controller (9) via solving the LMIs in (28), (33)–(36);
3. Construct the local mode dependent controller (4) using equation (31).

4 An Illustrative Example

In this section, we present a numerical example to illustrate the theory which has been developed. The uncertain large-scale system in the example has 3 subsystems and each subsystem can operate in 2 operation modes. The system data for the system (1) are as follows.

$$\begin{aligned}
A_1(1) &= \begin{bmatrix} 1 & 0 \\ -0.5 & -0.5 \end{bmatrix}, \quad B_1(1) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad C_1(1) = [0.6 \ 0], \\
A_1(2) &= \begin{bmatrix} 1 & 0 \\ 0.1 & -0.5 \end{bmatrix}, \quad B_1(2) = \begin{bmatrix} 1 \\ 0.1 \end{bmatrix}, \quad C_1(2) = [1 \ 0], \\
A_2(1) &= \begin{bmatrix} -0.6 & 0.5 \\ 0 & 0.5 \end{bmatrix}, \quad B_2(1) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C_2(1) = [0.1 \ 1], \\
A_2(2) &= \begin{bmatrix} -1 & 0 \\ -0.5 & 0.5 \end{bmatrix}, \quad B_2(2) = \begin{bmatrix} 0.1 \\ 1 \end{bmatrix}, \quad C_2(2) = [0 \ 1], \\
A_3(1) &= \begin{bmatrix} -1 & 0 \\ -0.1 & 0.1 \end{bmatrix}, \quad B_3(1) = \begin{bmatrix} 0.1 \\ 1 \end{bmatrix}, \quad C_3(1) = [0 \ 1], \\
A_3(2) &= \begin{bmatrix} -0.2 & 0 \\ 0.1 & -0.2 \end{bmatrix}, \quad B_3(2) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad C_3(2) = [1 \ 1], \\
E_i(1) &= \begin{bmatrix} 0 & 0 \\ 0.01 & 0 \end{bmatrix}, \quad L_i(1) = \begin{bmatrix} 0.01 \\ 0 \end{bmatrix}, \quad H_i(1) = \begin{bmatrix} 0.1 & 0 \\ 0 & 0 \end{bmatrix}, \\
E_i(2) &= \begin{bmatrix} 0.01 & 0 \\ 0 & 0 \end{bmatrix}, \quad L_i(2) = \begin{bmatrix} 0 \\ 0.01 \end{bmatrix}, \quad H_i(2) = \begin{bmatrix} 0 & -0.1 \\ 0 & 0.1 \end{bmatrix}, \\
G_i(1) &= \begin{bmatrix} 0 \\ 0.1 \end{bmatrix}, \quad G_i(2) = \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix}, \\
D_i(1) &= [0 \ 0.1], \quad D_i(2) = [0 \ 0.1], \quad i = 1, 2, 3.
\end{aligned}$$

Here the matrices $E_i(v_i)$, $H_i(v_i)$, $G_i(v_i)$ and $D_i(v_i)$ are chosen to satisfy the assumptions in [7]. The weighting matrices in the cost functional are given by

$$Q_i(v_i) = \begin{bmatrix} 0.001 & 0 \\ 0 & 0.001 \end{bmatrix}, \quad R_i(v_i) = 0.01, \quad v_i = 1, 2, \quad i = 1, 2, 3.$$

Table 1 Relationship between local operation modes and global operation mode

(v_1, v_2, v_3)	μ
(1, 1, 1)	1
(1, 2, 2)	2
(2, 1, 2)	3
(2, 2, 1)	4

The initial condition of the system is assumed to be

$$x_{10} = \begin{bmatrix} 5 \\ -5 \end{bmatrix}, \quad x_{20} = \begin{bmatrix} 3 \\ -3 \end{bmatrix}, \quad x_{30} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \eta_{10} = \eta_{20} = \eta_{30} = 1.$$

In this example, we assume that there are constraints on the operation modes of the subsystems: the operation mode of subsystem S_1 is assumed to be dependent on the operation modes of subsystems S_2 and S_3 . Specifically, $\eta_1(t) = 1$ if $\eta_2(t) = \eta_3(t)$, and $\eta_1(t) = 2$ otherwise. Therefore, the operation mode pattern set \mathcal{M}_p is given by $\{(1, 1, 1), (1, 2, 2), (2, 1, 2), (2, 2, 1)\}$. Hence, the large-scale system has four global operation modes. A relationship between the local operation modes and the global operation mode is described in Table 1. This table can also be used to define the bijective function $\Psi : \mathcal{M}_p \rightarrow \mathcal{M}$, and the functions $\Psi_i^{-1} : \mathcal{M} \rightarrow \mathcal{M}_i$, $i = 1, 2, 3$. The mode transition rate matrix of the Markov process $\eta(t)$ is assumed to be

$$\mathbf{Q} = \begin{bmatrix} -2 & 0.5 & 0.1 & 1.4 \\ 0.2 & -0.5 & 0.1 & 0.2 \\ 0.4 & 0.8 & -1.3 & 0.1 \\ 0.1 & 0.3 & 0.2 & -0.6 \end{bmatrix}.$$

It can be verified by direct calculations, for instance, that the local mode process $\eta_1(t)$ is not Markovian and is dependent on $\eta_2(t)$.

The software packages used to obtain a solution to the set of rank constrained LMI problem (23)–(27) are the Matlab LMIRank toolbox [11] with the YALMIP interface [8] and the underlying SeDuMi solver [17]. In Step 1 of the design procedure outlined in Sect. 3.4, we let $\beta_{ji}(\mu) = 1$ for $i = 1, \dots, 3$, $j, \mu = 1, \dots, 4$, and solve the rank constrained optimization problem

$$\min \gamma \quad \text{subject to} \quad \sum_{i=1}^N \tilde{x}_{i0}^T [N_i^T P_i(\eta_{i0}) N_i + \tau_i \bar{S}_i + \theta_i \tilde{S}_i] \tilde{x}_{i0} \leq \gamma$$

and the rank constrained LMIs in (23)–(27). A suboptimal worst-case controller of the form (4) can be found for $\gamma = 57$. Surprisingly, the controller we obtained may be approximated by a static output controller of the form

$$u_i(t) = D_{K,i}(\eta_i(t)) y_i(t)$$

since the values of $B_{K,i}(v_i)$ and $C_{K,i}(v_i)$ are small. Hence we only provide the values of $D_{K,i}(v_i)$:

$$\begin{aligned} D_{K,1}(1) &= -6.9269, & D_{K,1}(2) &= -3.7007, \\ D_{K,2}(1) &= -6.1663, & D_{K,2}(2) &= -3.8897, \\ D_{K,3}(1) &= -2.9246, & D_{K,3}(2) &= -1.7758. \end{aligned}$$

Using the resulting static output feedback controller, an upper bound of the cost functional in (6) was evaluated by solving the worst-case performance analysis problem for the corresponding closed-loop system; the bound was found to be $\sup_{\Xi^{\xi}, \Xi^r} J \leq 1.1594$. For comparison, the algorithm given in [7] was used to design a global mode dependent output feedback controller. Firstly, the system (1) with the cost functional (6) was regarded as a special class of the systems (7) with the cost functional (8), which were studied in [7].

Then, following the design procedure proposed in [7], it was found that the best global mode dependent controller that could be constructed yields the upper bound of the cost functional $\sup_{\Xi^{\xi}, \Xi^r} J \leq 1.0526$. It can be seen that the local mode dependent controller achieved almost the same level of performance as the global mode dependent controller (the difference is less than 10%). However, such a performance was achieved by a local mode dependent schedule between two operation modes for each subsystem. In contrast, the global mode dependent controller in [7] leads to a global mode dependent schedule between four operation modes for each subsystem. Moreover, using the local mode dependent controller eliminates collecting and broadcasting of the subsystem mode information throughout the entire large-scale system.

We now present some simulations to illustrate the properties of the static output feedback controller which was designed using our approach. The admissible uncertainties were chosen to be of the following form

$$\xi_i(t) = \alpha F_i \zeta_i(t), \quad \text{and} \quad \begin{cases} \dot{x}_{r,i}(t) = \beta A_{r,i} x_{r,i}(t) + \beta \sum_{j=1, j \neq i}^3 B_{r,ij} \zeta_j(t), \\ r_i(t) = C_{r,i} x_{r,i}(t), \end{cases}$$

for $i = 1, 2, 3$, where $\alpha \in [-1, 1]$ and $\beta > 0$ are unknown parameters. The parameters in the uncertainties are chosen to be

$$F_i = \begin{bmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{bmatrix}, \quad A_{r,i} = \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}, \quad B_{r,ij} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad C_{r,i} = [-0.3 \ 0.8], \quad i, j = 1, 2, 3.$$

The initial condition of the dynamic interconnections is set to zero. The reason for this particular linear uncertainty choice is as follows. With these linear uncertainties and the controller designed using our approach, the open-loop and the closed-loop large-scale systems are linear. As a result, we were able to use the existing results on stability of Markovian jump linear systems to check stability of the open-loop and closed-loop systems. It turned out that the open-loop system with $u_i(t) \equiv 0$

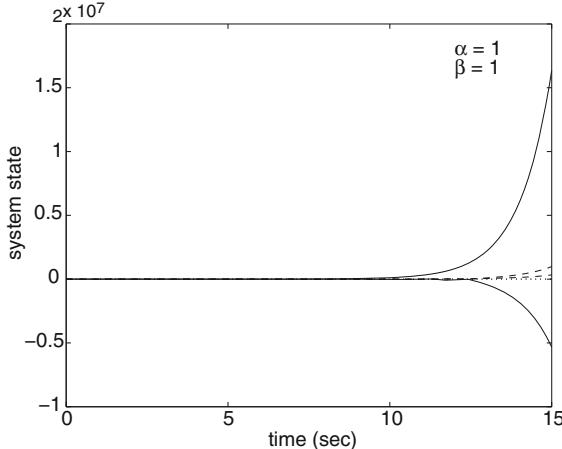


Fig. 1 Initial condition response of the open-loop system with uncertainty

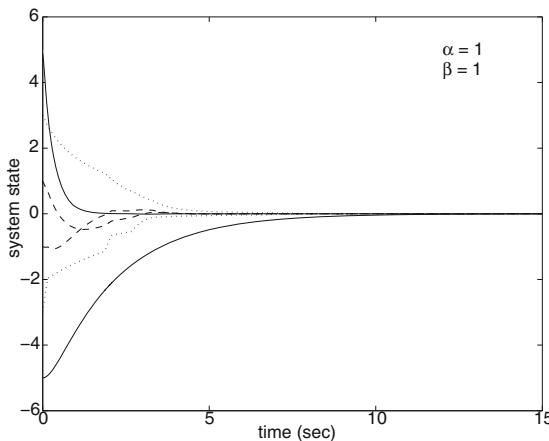


Fig. 2 Initial condition response of the closed-loop system with uncertainty

was not stochastically stable while the closed-loop system was found to be robustly stochastically stable. This confirms the result of Theorem 7.3. To further illustrate these findings, a random sample of $\eta(t)$ was generated and used in the simulations, and the sample state trajectories of the open-loop and the closed-loop large-scale systems are shown in Figs. 1 and 2, respectively.

5 Conclusions

This chapter has studied the decentralized output feedback guaranteed cost control problem for a class of uncertain Markovian jump large-scale systems via a local mode dependent approach. The controllers are entirely decentralized with respect to

the subsystems. They use the subsystem states and the subsystem operation modes to produce the subsystem control inputs. A sufficient condition in terms of rank constrained linear matrix inequalities has been developed to construct the controllers. Also, the theory has been illustrated by a numerical example and simulations.

Appendix 1: Proof of Theorem 7.2

The proof is divided into three steps. Firstly, we show that the feasibility of the rank constrained LMIs in (23)–(27), implies the solvability of the LMIs in (28). Secondly, we show that the solvability of (28) implies that a certain augmented system is stable and has certain H_∞ performance. Thirdly, we prove that the stability of the augmented system implies the robust stochastic stability of the closed-loop system composed of the uncertain system (7) and the controller (9) designed using the solution of (28). Also, we show that the H_∞ performance of the augmented system leads to the bound (29) on the cost functional (8) of the auxiliary system (7).

Step 1

Firstly, the rank constraint in (25), together with the positive definiteness of $P_i(\mu)$ and $X_i(\mu)$, implies that $X_i(\mu) = P_i^{-1}(\mu)$.

In view of the projection lemma (see, e.g. [5, Theorem 1]), the inequality in (28) has feasible solution $\tilde{K}_i(\mu)$ if and only if the following two inequalities hold:

$$(\Psi_{li}(\mu))^\perp \Phi_i(\mu) \left((\Psi_{li}(\mu))^\perp \right)^T < 0; \quad (38)$$

$$(\Psi_{ri}^T(\mu))^\perp \Phi_i(\mu) \left((\Psi_{ri}^T(\mu))^\perp \right)^T < 0. \quad (39)$$

Consider (38). We have

$$\Psi_{li}(\mu) = \begin{bmatrix} P_i(\mu) & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \underline{B}_i(\mu) \\ 0 \\ E_i(\mu) \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} \underline{B}_i(\mu) \\ 0 \\ E_i(\mu) \end{bmatrix} = \begin{bmatrix} 0 & \tilde{B}_i(\mu) \\ I & 0 \\ \hline 0 & 0 \\ \hline 0 & 0 \\ 0 & \tilde{R}_i^{\frac{1}{2}}(\mu) \\ 0 & \tilde{G}_i(\mu) \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

An orthogonal complement matrix of $\Psi_{li}(\mu)$ is found to be

$$\begin{aligned}
 (\Psi_{li}(\mu))^\perp &= \begin{bmatrix} B_i(\mu) \\ 0 \\ E_i(\mu) \end{bmatrix}^\perp \begin{bmatrix} P_i^{-1}(\mu) & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \\
 &= \begin{bmatrix} V_{1i}(\mu) & 0 & 0 & V_{2i}(\mu) & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} P_i^{-1}(\mu) & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \\
 &= \Gamma_{2i}(\mu) \begin{bmatrix} N_i^T & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} P_i^{-1}(\mu) & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \\
 &= \Gamma_{2i}(\mu) \begin{bmatrix} N_i^T P_i^{-1}(\mu) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},
 \end{aligned}$$

where $\Gamma_{2i}(\mu)$ is defined in (20). So, the inequality in (38) becomes

$$\Gamma_{2i}(\mu) \begin{bmatrix} \bar{Q}_{1i}(\mu) & N_i^T \hat{B}_i(\mu) & N_i^T X_i(\mu) \hat{C}_i^T(\mu) \\ \hat{B}_i^T(\mu) N_i & Q_{2i} & \hat{D}_i^T \\ \hat{C}_i(\mu) X_i(\mu) N_i & \hat{D}_i & Q_{3i} \end{bmatrix} \Gamma_{2i}^T(\mu) < 0$$

where

$$\bar{Q}_{1i}(\mu) = N_i^T X_i(\mu) N_i \tilde{A}_i^T(\mu) + \tilde{A}_i(\mu) N_i^T X_i(\mu) N_i + \sum_{v=1}^M q_{\mu v} N_i^T X_i(\mu) P_i(v) X_i(\mu) N_i.$$

The above inequality is further equivalent to (23) in view of Schur complement equivalence.

Consider (39). We have

$$\Psi_{ri}^T(\mu) = \begin{bmatrix} C_i^T(\mu) \\ D_i^T(\mu) \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & \tilde{C}_i^T(\mu) \\ I & 0 \\ 0 & \tilde{D}_i^T(\mu) \end{bmatrix},$$

so an orthogonal complement matrix of $\Psi_{ri}^T(\mu)$ is given by

$$\begin{aligned} (\Psi_{ri}^T(\mu))^\perp &= \left[\begin{array}{c|ccccc|c} V_{3i}(\mu) & 0 & V_{4i}(\mu) & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & I \end{array} \right] \\ &= \left[\begin{array}{c|ccccc|c} V_{3i}(\mu) & V_{4i}(\mu) & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & I \end{array} \right] \left[\begin{array}{c|ccccc|c} N_i^T & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & I & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & I \end{array} \right] \\ &= \begin{bmatrix} [\tilde{C}_i^T(\mu)]^\perp & 0 \\ [\tilde{D}_i^T(\mu)]^\perp & I \\ 0 & I \end{bmatrix} \begin{bmatrix} N_i^T & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}, \end{aligned}$$

where

$$[V_{3i}(\mu) \ V_{4i}(\mu)] = [\tilde{C}_i^T(\mu)]^\perp.$$

So the inequality in (39) becomes the inequality (24).

Therefore, if the linear matrix inequalities (23), (24) with rank constraints (25)–(27) have solutions, then the matrix inequalities (38), (39) hold. According to the projection lemma, the LMI in (28) is feasible after the solution of (23)–(27) is substituted into (28). As a result, the parameter matrices of the controller (9) can be obtained by solving (28).

Step 2

In this step, we will consider an augmented system and show that the solvability of (28) implies that the augmented system is stable and has a certain H_∞ performance.

Firstly, the rank constraints in (26), (27) imply that $\bar{\tau}_i = (\tau_i + \bar{\theta}_i)^{-1}$, $\bar{\tau}_{1i} = \tau_{1i}^{-1}$, $\bar{\tau}_{2i} = \tau_{2i}^{-1}$, $\bar{\tau}_{3i} = \tau_{3i}^{-1}$ and $\bar{\tau}_{4i} = \tau_{4i}^{-1}$.

Now, we consider an augmented system. Let

$$\tilde{z}_i(t) = \begin{bmatrix} \tilde{Q}_i^{\frac{1}{2}}(\eta(t)) \\ 0 \end{bmatrix} \tilde{x}_i(t) + \begin{bmatrix} 0 \\ \tilde{R}_i^{\frac{1}{2}}(\eta(t)) \end{bmatrix} \tilde{u}_i(t), \quad (40)$$

$$\tilde{\zeta}_{1i}(t) = \beta_{1i}(\eta(t))\tilde{x}_{K,i}(t), \quad (41)$$

$$\tilde{\zeta}_{2i}(t) = \beta_{2i}(\eta(t))\tilde{y}_i(t), \quad (42)$$

$$\tilde{\zeta}_{3i}(t) = \beta_{3i}(\eta(t))\tilde{x}_{K,i}(t), \quad (43)$$

$$\tilde{\zeta}_{4i}(t) = \beta_{4i}(\eta(t))\tilde{y}_i(t). \quad (44)$$

When a controller of form (9) is applied to the uncertain system (7), augmented with the additional outputs defined in (40)–(44), and subject to the IQCs in (2), (3), (10), (11), (12), (13), the resulting closed-loop uncertain system can be described by the state equations:

$$\bar{s}_i : \begin{cases} \dot{\bar{x}}_i(t) = \bar{A}_i(\eta(t))\bar{x}_i(t) + \bar{B}_i(\eta(t))U_i\bar{w}_i(t), \\ \bar{z}_i(t) = V_i\bar{C}_i(\eta(t))\bar{x}_i(t) + V_i\bar{D}_i(\eta(t))U_i\bar{w}_i(t), \end{cases} \quad (45)$$

where

$$\bar{x}_i(t) = \begin{bmatrix} \tilde{x}_i(t) \\ \tilde{x}_{K,i}(t) \end{bmatrix}, \quad \bar{w}_i(t) = U_i^{-1} \begin{bmatrix} \tilde{\xi}_i(t) \\ \tilde{r}_i(t) \\ \tilde{\zeta}_{1i}(t) \\ \tilde{\zeta}_{2i}(t) \\ \tilde{\zeta}_{3i}(t) \\ \tilde{\zeta}_{4i}(t) \end{bmatrix} = \begin{bmatrix} \sqrt{\tau_i}\tilde{\xi}_i(t) \\ \sqrt{\bar{\theta}_i}\tilde{r}_i(t) \\ \sqrt{\tau_{1i}}\tilde{\zeta}_{1i}(t) \\ \sqrt{\tau_{2i}}\tilde{\zeta}_{2i}(t) \\ \sqrt{\tau_{3i}}\tilde{\zeta}_{3i}(t) \\ \sqrt{\tau_{4i}}\tilde{\zeta}_{4i}(t) \end{bmatrix},$$

$$\bar{z}_i(t) = V_i \begin{bmatrix} \tilde{z}_i(t) \\ \tilde{\xi}_i(t) \\ \tilde{\zeta}_{1i}(t) \\ \tilde{\zeta}_{2i}(t) \\ \tilde{\zeta}_{3i}(t) \\ \tilde{\zeta}_{4i}(t) \end{bmatrix} = \begin{bmatrix} \tilde{z}_i(t) \\ \sqrt{\tau_i + \bar{\theta}_i}\tilde{\xi}_i(t) \\ \sqrt{\tau_{1i}}\tilde{\zeta}_{1i}(t) \\ \sqrt{\tau_{2i}}\tilde{\zeta}_{2i}(t) \\ \sqrt{\tau_{3i}}\tilde{\zeta}_{3i}(t) \\ \sqrt{\tau_{4i}}\tilde{\zeta}_{4i}(t) \end{bmatrix},$$

$$\bar{A}_i(\mu) = \begin{bmatrix} \tilde{A}_i(\mu) + \tilde{B}_i(\mu)\tilde{D}_{K,i}(\mu)\tilde{C}_i(\mu) & \tilde{B}_i(\mu)\tilde{C}_{K,i}(\mu) \\ \tilde{B}_{K,i}(\mu)\tilde{C}_i(\mu) & \tilde{A}_{K,i}(\mu) \end{bmatrix},$$

$$\bar{B}_i(\mu) = \begin{bmatrix} \tilde{E}_i(\mu) + \tilde{B}_i(\mu)\tilde{D}_{K,i}(\mu)\tilde{D}_i(\mu) & \tilde{L}_i(\mu) & 0 & 0 & \tilde{B}_i(\mu) & \tilde{B}_i(\mu) \\ \tilde{B}_{K,i}(\mu)\tilde{D}_i(\mu) & 0 & I & I & 0 & 0 \end{bmatrix},$$

$$\bar{C}_i(\mu) = \begin{bmatrix} \tilde{Q}_i^{\frac{1}{2}}(\mu) & & & 0 \\ \tilde{R}_i^{\frac{1}{2}}(\mu)\tilde{D}_{K,i}(\mu)\tilde{C}_i(\mu) & \tilde{R}_i^{\frac{1}{2}}(\mu)\tilde{C}_{K,i}(\mu) \\ \tilde{H}_i(\mu) + \tilde{G}_i(\mu)\tilde{D}_{K,i}(\mu)\tilde{C}_i(\mu) & \tilde{G}_i(\mu)\tilde{C}_{K,i}(\mu) \\ 0 & \beta_{1i}(\mu)I \\ \beta_{2i}(\mu)\tilde{C}_i(\mu) & 0 \\ 0 & \beta_{3i}(\mu)I \\ \beta_{4i}(\mu)\tilde{C}_i(\mu) & 0 \end{bmatrix},$$

$$\bar{D}_i(\mu) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \tilde{R}_i^{\frac{1}{2}}(\mu)\tilde{D}_{K,i}(\mu)\tilde{D}_i(\mu) & 0 & 0 & \tilde{R}_i^{\frac{1}{2}}(\mu) & \tilde{R}_i^{\frac{1}{2}}(\mu) \\ \tilde{G}_i(\mu)\tilde{D}_{K,i}(\mu)\tilde{D}_i(\mu) & 0 & 0 & \tilde{G}_i(\mu) & \tilde{G}_i(\mu) \\ 0 & 0 & 0 & 0 & 0 \\ \beta_{2i}(\mu)\tilde{D}_i(\mu) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \beta_{4i}(\mu)\tilde{D}_i(\mu) & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$U_i = \text{diag}\left(\frac{1}{\sqrt{\tau_i}}I, \frac{1}{\sqrt{\theta_i}}I, \frac{1}{\sqrt{\tau_{1i}}}I, \frac{1}{\sqrt{\tau_{2i}}}I, \frac{1}{\sqrt{\tau_{3i}}}I, \frac{1}{\sqrt{\tau_{4i}}}I\right),$$

$$V_i = \text{diag}(I, \frac{1}{\sqrt{\bar{\tau}_i}}I, \frac{1}{\sqrt{\bar{\tau}_{1i}}}I, \frac{1}{\sqrt{\bar{\tau}_{2i}}}I, \frac{1}{\sqrt{\bar{\tau}_{3i}}}I, \frac{1}{\sqrt{\bar{\tau}_{4i}}}I).$$

The introduction of these positive scalars τ_i , θ_i , τ_{1i} , τ_{2i} , τ_{3i} and τ_{4i} allows us to use the IQC uncertainty structure information in the controller design; see Step 3 of the proof and Remark 7.4 for more explanations about these parameters.

Since the global mode process $\eta(t)$ is a continuous-time Markov process, the closed-loop system in (45) is in fact a special class of Markovian jump linear systems [2, 10]. In order to use the theory which has been developed for Markovian jump linear systems, we re-write the system (45) as

$$\begin{cases} \dot{\check{x}}(t) = \check{A}(\eta(t))\check{x}(t) + \check{B}(\eta(t))\check{w}(t), \\ \dot{\check{z}}(t) = \check{C}(\eta(t))\check{x}(t) + \check{D}(\eta(t))\check{w}(t), \end{cases} \quad (46)$$

where

$$\check{x}(t) = \begin{bmatrix} \bar{x}_1(t) \\ \vdots \\ \bar{x}_N(t) \end{bmatrix}, \quad \check{w}(t) = \begin{bmatrix} \bar{w}_1(t) \\ \vdots \\ \bar{w}_N(t) \end{bmatrix}, \quad \check{z}(t) = \begin{bmatrix} \bar{z}_1(t) \\ \vdots \\ \bar{z}_N(t) \end{bmatrix},$$

$$\check{A}(\mu) = \text{diag}(\bar{A}_1(\mu), \dots, \bar{A}_N(\mu)),$$

$$\check{B}(\mu) = \text{diag}(\bar{B}_1(\mu)U_1, \dots, \bar{B}_N(\mu)U_N),$$

$$\check{C}(\mu) = \text{diag}(V_1\bar{C}_1(\mu), \dots, V_N\bar{C}_N(\mu)),$$

$$\check{D}(\mu) = \text{diag}(V_1\bar{D}_1(\mu)U_1, \dots, V_N\bar{D}_N(\mu)U_N).$$

According to the bounded real lemma (see, e.g. [21, Proposition 2]), the Markovian jump linear system in (46) is stochastically stable and has H_∞ performance $\|T_{\check{z}\check{w}}\|_\infty < 1$ if there exist matrices $\check{P}(\mu) = \text{diag}(P_1(\mu), \dots, P_N(\mu))$, $P_i(\mu) \in \mathbb{S}^+$, $\mu \in \mathcal{M}$, $i \in \mathcal{N}$, such that the matrix inequality

$$\begin{bmatrix} \check{A}^T(\mu)\check{P}(\mu) + \check{P}(\mu)\check{A}(\mu) + \sum_{v=1}^M q_{\mu v} \check{P}(v) \check{P}(\mu) \check{B}(\mu) \check{C}^T(\mu) \\ \check{B}^T(\mu)\check{P}(\mu) & -I & \check{D}^T(\mu) \\ \check{C}(\mu) & \check{D}(\mu) & -I \end{bmatrix} < 0 \quad (47)$$

holds for all $\mu \in \mathcal{M}$. Here $T_{\check{z}\check{w}}$ denotes the operator from \check{w} to \check{z} defined by system (46). For similar results, we refer to [2, 3, 6, 13].

Because all the matrices in (47) are diagonal, using a congruence transformation (see Appendix 2 for an illustrative example), the matrix inequality (47) is equivalent to

$$\begin{bmatrix} \bar{A}_i^T(\mu)P_i(\mu) + P_i(\mu)\bar{A}_i(\mu) + \sum_{v=1}^M q_{\mu v} P_i(v) P_i(\mu)\bar{B}_i(\mu)U_i & \bar{C}_i^T(\mu)V_i \\ U_i\bar{B}_i^T(\mu)P_i(\mu) & -I & U_i\bar{D}_i^T(\mu)V_i \\ V_i\bar{C}_i(\mu) & V_i\bar{D}_i(\mu)U_i & -I \end{bmatrix} < 0 \quad (48)$$

for all $\mu \in \mathcal{M}$ and $i \in \mathcal{N}$. Pre- and post-multiplying both sides of the above inequality by $\text{diag}(I, U_i^{-1}, V_i^{-1})$, we have the inequality

$$\begin{bmatrix} \bar{A}_i^T(\mu)P_i(\mu) + P_i(\mu)\bar{A}_i(\mu) + \sum_{v=1}^M q_{\mu v} P_i(v) P_i(\mu)\bar{B}_i(\mu) \bar{C}_i^T(\mu) \\ \bar{B}_i^T(\mu)P_i(\mu) & Q_{1i} & \bar{D}_i^T(\mu) \\ \bar{C}_i(\mu) & \bar{D}_i(\mu) & Q_{2i} \end{bmatrix} < 0. \quad (49)$$

After a direct computation, the above inequality is the same inequality as (28).

Therefore, if the inequality in (28) holds, then the system in (46) is stochastically stable and has the H_∞ performance $\|T_{\check{z}\check{w}}\|_\infty < 1$.

Step 3

Because the system in (46) is stochastically stable, the closed-loop system composed of the uncertain system (7) and the controller (9) designed via solving (28) is robustly stochastically stable as well. Next we prove that the corresponding closed-loop value of the cost functional satisfies the performance given in (29).

Define the Lyapunov functional for the system in (46) as

$$V(t, \check{x}(t), \eta(t)) \triangleq \sum_{i=1}^N V_i(t, \bar{x}_i(t), \eta(t))$$

where $V_i(t, \bar{x}_i(t), \eta(t)) = \bar{x}_i^T(t)P_i(\eta(t))\bar{x}_i(t)$. Then, the stability of system (46) implies that

$$\lim_{t \rightarrow \infty} \mathbb{E}(V(t, \tilde{x}(t), \eta(t)) \mid \tilde{x}_0, \eta_0) = 0.$$

Here, the reason that we take the expectation conditioned on (\tilde{x}_0, η_0) instead of (\check{x}_0, η_0) is that the controller state initial condition is fixed to zero.

On the other hand, it has been shown that the inequality in (28) is equivalent to the inequality in (48). Also, the inequality in (48) is equivalent to

$$\begin{aligned} Z_i(\mu) &\triangleq \begin{bmatrix} \bar{A}_i^T(\mu)P_i(\mu) + P_i(\mu)\bar{A}_i(\mu) + \sum_{v=1}^M q_{\mu v}P_i(v) & P_i(\mu)\bar{B}_i(\mu)U_i \\ U_i\bar{B}_i^T(\mu)P_i(\mu) & -I \end{bmatrix} \\ &\quad + \begin{bmatrix} \bar{C}_i^T(\mu)V_i \\ U_i\bar{D}_i^T(\mu)V_i \end{bmatrix} [V_i\bar{C}_i(\mu) V_i\bar{D}_i(\mu)U_i] < 0. \end{aligned}$$

Therefore, we have

$$\begin{aligned} \mathcal{L}V_i(t, \bar{x}_i(t), \eta(t)) &= \dot{\bar{x}}_i^T(t)P_i(\mu)\bar{x}_i(t) + \bar{x}_i^T(t)P_i(\mu)\dot{\bar{x}}_i(t) + \bar{x}_i^T(t) \left(\sum_{v=1}^M q_{\mu v}P_i(v) \right) \bar{x}_i(t) \\ &= -\bar{z}_i^T(t)\bar{z}_i(t) + \bar{w}_i^T(t)\bar{w}_i(t) + \begin{bmatrix} \bar{x}_i(t) \\ \bar{w}_i(t) \end{bmatrix}^T Z_i(\mu) \begin{bmatrix} \bar{x}_i(t) \\ \bar{w}_i(t) \end{bmatrix} \\ &< -\bar{z}_i^T(t)\bar{z}_i(t) + \bar{w}_i^T(t)\bar{w}_i(t) \end{aligned}$$

for any $\begin{bmatrix} \bar{x}_i(t) \\ \bar{w}_i(t) \end{bmatrix} \neq 0$. So,

$$\begin{aligned} &\sum_{i=1}^N \mathbb{E} \left(\int_0^\infty \|\bar{z}_i(t)\|^2 dt \mid \tilde{x}_0, \eta_0 \right) \\ &< \sum_{i=1}^N \mathbb{E} \left(\int_0^\infty \|\bar{w}_i(t)\|^2 dt \mid \tilde{x}_0, \eta_0 \right) - \sum_{i=1}^N \mathbb{E} \left(\int_0^\infty \mathcal{L}V_i(t)dt \mid \tilde{x}_0, \eta_0 \right). \end{aligned}$$

In view of the definitions of $\bar{z}_i(t)$ and $\bar{w}_i(t)$ in (45), the above inequality becomes

$$\begin{aligned} &\sum_{i=1}^N \mathbb{E} \left(\int_0^\infty \left[\|\tilde{z}_i(t)\|^2 + (\tau_i + \bar{\theta}_i) \|\tilde{\zeta}_i(t)\|^2 + \tau_{1i} \|\tilde{\zeta}_{1i}(t)\|^2 + \tau_{2i} \|\tilde{\zeta}_{2i}(t)\|^2 \right. \right. \\ &\quad \left. \left. + \tau_{3i} \|\tilde{\zeta}_{3i}(t)\|^2 + \tau_{4i} \|\tilde{\zeta}_{4i}(t)\|^2 \right] dt \mid \tilde{x}_0, \eta_0 \right) \\ &< \sum_{i=1}^N \mathbb{E} \left(\int_0^\infty \left[\tau_i \|\tilde{\xi}_i(t)\|^2 + \theta_i \|\tilde{r}_i(t)\|^2 + \tau_{1i} \|\tilde{\xi}_{1i}(t)\|^2 + \tau_{2i} \|\tilde{\xi}_{2i}(t)\|^2 \right. \right. \\ &\quad \left. \left. + \tau_{3i} \|\tilde{\xi}_{3i}(t)\|^2 + \tau_{4i} \|\tilde{\xi}_{4i}(t)\|^2 \right] dt \mid \tilde{x}_0, \eta_0 \right) \\ &\quad + \sum_{i=1}^N \tilde{x}_{i0}^T N_i^T P_i(\eta_{i0}) N_i \tilde{x}_{i0}. \end{aligned}$$

Because the uncertainties $\tilde{\xi}_i(t)$, $\tilde{r}_i(t)$, $\tilde{\xi}_{1i}(t)$, $\tilde{\xi}_{2i}(t)$, $\tilde{\xi}_{3i}(t)$ and $\tilde{\xi}_{4i}(t)$ satisfy the IQCs in (2), (3), (10)–(13), respectively, we conclude that

$$\begin{aligned} \sup_{\Xi^{\tilde{\xi}}, \Xi^r, \Xi^K} \tilde{J} &= \sup_{\Xi^{\tilde{\xi}}, \Xi^r, \Xi^K} \sum_{i=1}^N E \left(\int_0^\infty \|\tilde{z}_i(t)\|^2 \mid \tilde{x}_0, \eta_0 \right) \\ &< \sum_{i=1}^N \tilde{x}_{i0}^T (N_i^T P_i(\eta_{i0}) N_i + \tau_i \tilde{S}_i + \theta_i \tilde{S}_i) \tilde{x}_{i0}. \end{aligned}$$

That is, the inequality in (29) holds. This completes the proof. \square

Appendix 2: Equivalence Between (47) and (48)

The example in this appendix illustrates that the matrix inequality in (47) is equivalent to the matrix inequality in (48) via a congruence transformation. An extension to a general case is straightforward.

Example 7.1. Let the left hand side of (47) be

$$A = \left[\begin{array}{cc|cc|cc} a_1 & 0 & d_1 & 0 & f_1 & 0 \\ 0 & a_2 & 0 & d_2 & 0 & f_2 \\ \hline d_1 & 0 & b_1 & 0 & e_1 & 0 \\ 0 & d_2 & 0 & b_2 & 0 & e_2 \\ \hline f_1 & 0 & e_1 & 0 & c_1 & 0 \\ 0 & f_2 & 0 & e_2 & 0 & c_2 \end{array} \right].$$

If an elementary matrix is chosen as

$$T = \left[\begin{array}{cccccc} I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 \\ 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I \end{array} \right],$$

then we have

$$T A T^T = \left[\begin{array}{cccccc} a_1 & d_1 & f_1 & 0 & 0 & 0 \\ d_1 & b_1 & e_1 & 0 & 0 & 0 \\ f_1 & e_1 & c_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_2 & d_2 & f_2 \\ 0 & 0 & 0 & d_2 & b_2 & e_2 \\ 0 & 0 & 0 & f_2 & e_2 & c_2 \end{array} \right].$$

Acknowledgment This work was supported by the Australian Research Council.

References

- Anderson, B.D.O., Moore, J.B.: Optimal Filtering. Prentice-Hall, Englewood Cliffs (1979)
- de Farias, D.P., Geromel, J.C., do Val, J.B.R., Costa, O.L.V.: Output feedback control of Markov jump linear systems in continuous-time. *IEEE Transactions on Automatic Control* **45**(5), 944–949 (2000)
- de Souza, C.E., Fragoso, M.D.: H_∞ control for linear systems with Markovian jumping parameters. *Control Theory and Advanced Technology* **9**(2), 457–466 (1993)
- El Ghaoui, L., Oustry, F., Rami, M.A.: A cone complementarity linearization algorithm for static output-feedback and related problems. *IEEE Transactions on Automatic Control* **42**(8), 1171–1176 (1997)
- Iwasaki, T., Skelton, R.E.: All controllers for the general H_∞ control problem: LMI existence conditions and state space formulas. *Automatica* **30**(8), 1307–1317 (1994)
- Li, L., Ugrinovskii, V.A.: On necessary and sufficient conditions for H_∞ output feedback control of Markov jump linear systems. *IEEE Transactions on Automatic Control* **52**(7), 1287–1292 (2007)
- Li, L., Ugrinovskii, V.A., Orsi, R.: Decentralized robust control of uncertain Markov jump parameter systems via output feedback. *Automatica* **43**(11), 1932–1944 (2007)
- Löfberg, J.: YALMIP : A toolbox for modeling and optimization in MATLAB. In: Proceedings of the CACSD Conference (2004). URL <http://control.ee.ethz.ch/~joloef/yalmip.php>
- Mahmoud, M.S., Shi, P., Shi, Y.: H_∞ and robust control of interconnected systems with Markovian jump parameters. *Discrete and Continuous Dynamical Systems – Series B* **5**(2), 365–384 (2005)
- Mariton, M.: Jump Linear Systems in Automatic Control. Marcel Dekker, New York (1990)
- Orsi, R.: LMIRank: software for rank constrained LMI problems (2005). URL <http://users.rsise.anu.edu.au/~robert/lmirank/>
- Orsi, R., Helmke, U., Moore, J.B.: A Newton-like method for solving rank constrained linear matrix inequalities. *Automatica* **42**(11), 1875–1882 (2006)
- Pan, Z., Başar, T.: H_∞ -control of Markovian jump systems and solutions to associated piecewise-deterministic differential games. In: G.J. Olsder (ed.) *New Trends in Dynamic Games and Applications*, pp. 61–94. Birkhäuser, Basel (1995)
- Petersen, I.R.: Robust H^∞ control of an uncertain system via a stable decentralized output feedback controller. *Kybernetika* **45**(1), 101–120 (2009)
- Petersen, I.R., Ugrinovskii, V.A., Savkin, A.V.: Robust Control Design Using H_∞ Methods. Springer, Berlin (2000)
- Sandell, N.R., Varaiya, P., Athans, M., Safonov, M.G.: Survey of decentralized control methods for large scale systems. *IEEE Transactions on Automatic Control* **23**(2), 108–128 (1978)
- Sturm, J.F.: Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software* **11–12**, 625–653 (1999). URL <http://sedumi.mcmaster.ca/>
- Ugrinovskii, V.A., Pota, H.R.: Decentralized control of power systems via robust control of uncertain Markov jump parameter systems. *International Journal of Control* **78**(9), 662–677 (2005)
- Wang, W.J., Chen, Y.H.: Decentralized robust control design with insufficient number of controllers. *International Journal of Control* **65**(6), 1015–1030 (1996)
- Xiong, J., Ugrinovskii, V.A., Petersen, I.R.: Local mode dependent decentralized control of uncertain Markovian jump large-scale systems. In: Proceedings of the Conference on Decision and Control, pp. 4345–4350 (2008) (also in *IEEE-TAC* **54**(11))

21. Zhang, L., Huang, B., Lam, J.: H_∞ model reduction of Markovian jump linear systems. *Systems and Control Letters* **50**(2), 103–118 (2003)
22. Zhang, X., Zheng, Y., Lu, G.: An interconnected Markovian jump system and its decentralized observer-based control with sector nonlinearities. In: International Conference on Control and Automation, pp. 271–276 (2005)

Consensus Based Multi-Agent Control Algorithms

Miloš S. Stanković, Dušan M. Stipanović, and Srdjan S. Stanković

1 Introduction

Control of complex systems can be achieved via *hierarchical multilayered agent-based structures* benefiting from their inherent properties such as *modularity, scalability, adaptability, flexibility and robustness*. The agent-based structures consist of a number of simpler *subsystems* (or *agents*), each of which addresses in a coordinated manner a specific *sub-objective* or sub-task so as to attain the overall design objectives. The complexity of the behavior of such systems arises as a result of *interactions* between multiple agents and the environment in which they operate. More specifically, multi-agent control systems are fundamental parts of a wide range of safety-critical engineering systems, and are commonly found in aerospace, traffic control, chemical processes, power generation and distribution, flexible manufacturing, robotic system design and self-assembly structures. A multi-agent system can be considered as a loosely coupled network of problem-solver entities that work together to find answers to problems that are beyond the individual capabilities or knowledge of each entity, where local control law has to satisfy *decentralized information structure constraints* (see, e.g., [20]), and where no global system control (or supervision) is desired. Different aspects of multi-agent control systems are

M.S. Stanković

ACCESS Linnaeus Center, School of Electrical Engineering, Royal Institute of Technology,
100-44 Stockholm, Sweden

e-mail: milsta@kth.se

D.M. Stipanović

Department of Industrial and Enterprise Systems Engineering and the Coordinated Science
Laboratory, University of Illinois at Urbana-Champaign, Illinois, USA
e-mail: dusan@illinois.edu

S.S. Stanković

School of Electrical Engineering, University of Belgrade, Serbia
e-mail: stankovic@etf.rs

covered by a vast literature within the frameworks of computer science, artificial intelligence, network and system theory; for some aspects of multi-agent control systems and sensor networks see, e.g., [4, 7, 16, 31].

Considering methodologies for achieving *agreement* between the agents upon some decisions, important results were obtained in relation with *distributed iterations* in parallel computation and distributed optimization as early as in the 1980s, e.g., [1–3, 29, 30]. A very intensive research has been carried out recently in this direction, including numerous applications (see, e.g., [4, 6, 7, 12–14, 17, 18]). The majority of the cited references share a common general methodology: they all use some kind of *dynamic consensus strategy*.

In this chapter an attempt is made to approach the problem of *overlapping decentralized control of complex systems* by using a *multi-agent strategy*, where the agents (subsystems) communicate in order to achieve direct or indirect agreement upon a *control action* by using a dynamic consensus methodology. The aim of the chapter is to propose several different novel control structures derived from:

- (a) the choice of the variables upon which the agreement is made;
- (b) basic local controller structures derived from the decentralized control laws implemented by the agents.

The chapter is organized as follows. Section 2 deals with the problem definition, including the subsystem models and the distribution of control tasks among the agents. In Sect. 3 several new control structures are proposed based on the agreement between the agents upon the *control variables*. In the most general setting, it is assumed that each agent is able to formulate its local feedback control law starting from the local information structure constraints in the form of a general four-term dynamic output controller. The subsystem inputs generated by the agents by means of the local controllers enter the consensus process which generates the control signals to be applied to the system by some a priori specified agents. In the general case, the consensus scheme, determining, in fact, the control law for the whole system, is constructed on the basis of an *aggregation* of the local dynamic controllers. It is shown how the proposed scheme can be adapted to either static local output feedback controllers, or static local state feedback controllers. In Sec. 4 an alternative to the approach presented in Sect. 3 is proposed, based on the introduction of a dynamic consensus at the level of *state estimation* [24, 25]. Namely, it is assumed that the agents are able to generate local estimates of parts of the overall state vector using their own subsystem models. The dynamic consensus scheme is introduced to provide each agent with a reliable estimate of the whole system state. The control signal is obtained by applying the known global LQ optimal state feedback gain to the locally available estimates. A number of selected examples illustrate the applicability of all the proposed consensus based control schemes. Section 5 is devoted to the problem of decentralized overlapping control of a formation of unmanned aerial vehicles (UAVs). Starting from a specific formation model, the global LQ optimal state feedback is defined. Further, it is demonstrated that a decentralized consensus based estimator can be formulated on the basis of extracted subsystems attached to the vehicles. Efficiency of this approach is illustrated by a simulation example.

2 Problem Formulation

Let a complex system be represented by a linear model

$$\begin{aligned} \mathbf{S} : \quad & \dot{x} = Ax + Bu \\ & y = Cx, \end{aligned} \quad (1)$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$ and $y \in \mathbb{R}^v$ are the state, input and output vectors, respectively, while A , B and C are constant matrices of appropriate dimensions.

Assume that N agents have to control the system \mathbf{S} according to their own resources. The agents have their *local models* of the parts of \mathbf{S}

$$\begin{aligned} \mathbf{S}_i : \quad & \dot{\zeta}^{(i)} = A^{(i)}\zeta^{(i)} + B^{(i)}v^{(i)} \\ & y^{(i)} = C^{(i)}\zeta^{(i)}, \end{aligned} \quad (2)$$

where $\zeta^{(i)} \in \mathbb{R}^{n_i}$, $v^{(i)} \in \mathbb{R}^{m_i}$ and $y^{(i)} \in \mathbb{R}^{v_i}$ are the corresponding state, input and output vectors, and $A^{(i)}$, $B^{(i)}$ and $C^{(i)}$ constant matrices, $i = 1, \dots, N$. Components of the input vectors $v^{(i)} = (v_1^{(i)}, \dots, v_{m_i}^{(i)})^T$ represent subsets of the global input vector u of \mathbf{S} , so that $v_j^{(i)} = u_{p_j^i}$, $j = 1, \dots, m_i$, and $p_j^i \in \mathcal{V}^i$, where $\mathcal{V}^i = \{p_1^i, \dots, p_{m_i}^i\}$ is the *input index set* defining $v^{(i)}$. Similarly, for the outputs $y^{(i)}$ we have $y_j^{(i)} = y_{q_j^i}$, $j = 1, \dots, v_i$, and $q_j^i \in \mathcal{Y}^i$, where $\mathcal{Y}^i = \{q_1^i, \dots, q_{v_i}^i\}$ is the *output index set*; according to these sets, it is possible to find such constant $v_i \times n$ matrices C_i that $y^{(i)} = C_i x$, $i = 1, \dots, N$. The state vectors $\zeta^{(i)}$ do not necessarily represent parts of the global state vector x . They can be chosen, together with the matrices $A^{(i)}$, $B^{(i)}$ and $C^{(i)}$, according to the local criteria for modeling the input–output relations $v^{(i)} \rightarrow y^{(i)}$. In the particular case when $\zeta^{(i)} = x^{(i)}$, $x_j^{(i)} = x_{r_j^i}$, $j = 1, \dots, n_i$, $n_i \leq n$ and $r_j^i \in \mathcal{X}^i$, where $\mathcal{X}^i = \{r_1^i, \dots, r_{n_i}^i\}$ is the *state index set* defining $x^{(i)}$. In the last case, models \mathbf{S}_i , in general, represent overlapping subsystems of \mathbf{S} in a more strict sense; matrices $A^{(i)}$, $B^{(i)}$ and $C^{(i)}$ can represent in this case sub-matrices of A , B and C .

The task of the i th agent is to generate the control vector $v^{(i)}$ and to implement the control action $u^{(i)} \in \mathcal{U}^i$, satisfying $u_j^{(i)} = u_{s_j^i}$, $j = 1, \dots, \mu_i$, and $s_j^i \in \mathcal{U}^i$, where $\mathcal{U}^i = \{s_1^i, \dots, s_{\mu_i}^i\}$ is the *control index set* defining $u^{(i)}$. It is assumed that $\mathcal{U}^i \subseteq \mathcal{V}^i$ and $\mathcal{U}^i \cap \mathcal{U}^j = \emptyset$, so that $\sum_{i=1}^N \mu_i = m$, that is, the control vector $u^{(i)}$ of the i -th agent is a part of its input vector $v^{(i)}$, while one and only one agent is responsible for generation of each component of u within the considered control task. Consequently, all agents include the entire vectors $v^{(i)}$ of \mathbf{S}_i in the control design considerations, but they implement only those components of $v^{(i)}$ for which they are responsible.

In the case when the inputs $v^{(i)}$ do not overlap, the agents perform their tasks autonomously, without interactions with each other; that is, we have the case of decentralized control of \mathbf{S} , when the control design is based entirely on the local models \mathbf{S}_i . However, in the case when the model inputs $v^{(i)}$ overlap, more than one model \mathbf{S}_i can be used for calculation of a particular component of the input vector u .

Obviously, it would be beneficial for the agent responsible for implementation of that particular input component to use different suggestions about the control action and to calculate the numerical values of the control signal to be implemented on the basis of an agreement between the agents. The agents that do not implement any control action ($\mathcal{U}^i = \emptyset$) could, in this context, represent “advisors” to the agents responsible for control implementation. Our aim is to propose several overlapping decentralized feedback control structures for \mathbf{S} based on a dynamic consensus between multiple agents.

3 Consensus at the Control Input Level

In this section algorithms based on consensus at the control input level are presented. The first subsection deals with the algorithms derived from local dynamic feedback controllers, while the second subsection is related to the important special case of static local output feedback controllers.

3.1 Algorithms Derived from the Local Dynamic Output Feedback Control Laws

We assume that each agent generates its input vector $v^{(i)}$ in \mathbf{S}_i using the *local dynamic controller* in the form

$$\begin{aligned}\mathbf{C}_i : \quad \dot{w}^{(i)} &= F^{(i)}w^{(i)} + G^{(i)}y^{(i)} \\ v^{(i)} &= K^{(i)}w^{(i)} + H^{(i)}y^{(i)}\end{aligned}\tag{3}$$

where $w^{(i)} \in \mathbb{R}^{\rho_i}$ represents the controller state, and matrices $F^{(i)}$, $G^{(i)}$, $K^{(i)}$ and $H^{(i)}$, $i = 1, \dots, N$, are constant, with appropriate dimensions. Local controllers are designed according to the local models and local design criteria. Assuming that the agents can communicate between each other, the goal is to generate the control signal u for \mathbf{S} based on a mutual agreement, starting from the inputs $v^{(i)}$ generated by \mathbf{C}_i . The idea about reaching an agreement upon the components of u stems from the fact that the index sets $\mathcal{V}^{(i)}$ are, in general, overlapping, so that the agents responsible for control implementation according to the index sets $\mathcal{U}^{(i)}$ can improve their local control laws by getting “suggestions” from the other agents.

Algorithm 1. The second relation in (3) gives rise to $\dot{v}^{(i)} = K^{(i)}\dot{w}^{(i)} + H^{(i)}\dot{y}^{(i)}$, wherefrom we get

$$\begin{aligned}\dot{v}^{(i)} &= K^{(i)}[F^{(i)}w^{(i)} + G^{(i)}y^{(i)}] + H^{(i)}C^{(i)}[A^{(i)}\zeta^{(i)} + B^{(i)}v^{(i)}] \\ &= K^{(i)}F^{(i)}w^{(i)} + K^{(i)}G^{(i)}y^{(i)} + H^{(i)}C^{(i)}A^{(i)}\zeta^{(i)} + H^{(i)}C^{(i)}B^{(i)}v^{(i)}.\end{aligned}\tag{4}$$

Since $y^{(i)}$ are the available signals, and $v^{(i)}$ vectors to be locally generated for participation in the agreement process, we will use the following approximation

$$\dot{v}^{(i)} \approx [F_*^{(i)} + H^{(i)}C^{(i)}B^{(i)}]v^{(i)} + [K^{(i)}G^{(i)} + H^{(i)}A_*^{(i)} - F_*^{(i)}H^{(i)}]y^{(i)}, \quad (5)$$

where $F_*^{(i)} = K^{(i)}F^{(i)}K^{(i)+}$ and $A_*^{(i)} = C^{(i)}A^{(i)}C^{(i)+}$ are approximate solutions of the aggregation relations $K^{(i)}F^{(i)} = F_*^{(i)}K^{(i)}$ and $C^{(i)}A^{(i)} = A_*^{(i)}C^{(i)}$, respectively, where A^+ denotes the pseudo-inverse of a given matrix A [11, 20].

We will assume in the sequel, for the sake of presentation clarity, that all the agents can have their “suggestions” for all the components of u ; that is, we assume that the vector $v^{(i)} = U_i \in \mathbb{R}^m$ is a “local version” of u proposed by the i th agent to the other agents. Furthermore, we introduce $m \times \rho_i$ and $m \times v_i$ constant matrices K_i and H_i , obtained by taking the rows of $K^{(i)}$ and $H^{(i)}$ at the row indices defined by the index set $\mathcal{V}^{(i)}$ and leaving zeros elsewhere, and $n_i \times m$ matrix B_i obtained from $B^{(i)}$ by taking its columns at the indices defined by \mathcal{V}^i and leaving zeros elsewhere. Let $U = \text{col}\{U_1, \dots, U_N\}$, $Y = \text{col}\{y^{(1)}, \dots, y^{(N)}\}$, $\tilde{K} = \text{diag}\{K_1, \dots, K_N\}$ and $\tilde{H} = \text{diag}\{H_1, \dots, H_N\}$. Similarly, let $\tilde{A} = \text{diag}\{A^{(1)}, \dots, A^{(N)}\}$, $\tilde{B} = \text{diag}\{B_1, \dots, B_N\}$, $\tilde{C} = \text{diag}\{C^{(1)}, \dots, C^{(N)}\}$, $\tilde{F} = \text{diag}\{F^{(1)}, \dots, F^{(N)}\}$, and $\tilde{G} = \text{diag}\{G^{(1)}, \dots, G^{(N)}\}$. Assume that the agents communicate between each other in such a way that they send current values of U_i to each other according to a communication strategy determined by the *consensus matrix* $\tilde{\Gamma} = [\Gamma_{ij}]$, where Γ_{ij} , $i, j = 1, \dots, N$, $i \neq j$, are $m \times m$ diagonal matrices with positive entries and $\Gamma_{ii} = -\sum_{j=1, j \neq i}^N \Gamma_{ij}$, $i = 1, \dots, N$. Then, the algorithm for generating U , i.e., the vector containing all the agent input vectors U_i , $i = 1, \dots, N$, representing the result of the overall consensus process, is given by

$$\begin{aligned} \dot{U}_i &= \sum_{j=1, j \neq i}^N \Gamma_{ij}(U_j - U_i) + [K_i F^{(i)} K_i^+ + H_i C^{(i)} B^{(i)}] U_i \\ &\quad + [K_i G^{(i)} + H_i C^{(i)} A^{(i)} C^{(i)+} - K_i F^{(i)} K_i^+ H_i] y^{(i)}, \end{aligned} \quad (6)$$

$i = 1, \dots, N$, or, in a compact form,

$$\dot{U} = [\tilde{\Gamma} + \tilde{K}\tilde{F}\tilde{K}^+ + \tilde{H}\tilde{C}\tilde{B}]U + [\tilde{K}\tilde{G} + \tilde{H}\tilde{C}\tilde{A}\tilde{C}^+ - \tilde{K}\tilde{F}\tilde{K}^+\tilde{H}]Y. \quad (7)$$

The vector U generated by (7) is used for control implementation in such a way that the i th agent picks up the components of U_i selected by the index set $\mathcal{U}^{(i)}$ and applies them to the system \mathbf{S} . If Q is an $m \times mN$ matrix with zeros everywhere except one place in each row, where it contains 1; for the j th row with $j \in \mathcal{U}^{(i)}$, 1 is placed at the column index $(i-1)m + j$. Then, we have $u = QU$, and system (1) can be written as

$$\dot{x} = Ax + BQU. \quad (8)$$

Also, according to the adopted notation, $y^{(i)} = C_i x$, so that $Y = \tilde{C}x$, where $\tilde{C}^T = [C_1^T | \dots | C_N^T]$. Therefore, the whole closed-loop system is represented by

$$\begin{bmatrix} \dot{U} \\ \dot{x} \end{bmatrix} = \begin{bmatrix} \tilde{\Gamma} + \tilde{K}\tilde{F}\tilde{K}^+ + \tilde{H}\tilde{C}\tilde{B} & (\tilde{K}\tilde{G} + \tilde{H}\tilde{C}\tilde{A}\tilde{C}^+ - \\ \cdots & \cdots \\ BQ & A \end{bmatrix} \begin{bmatrix} U \\ x \end{bmatrix}. \quad (9)$$

Obviously, the system is stabilized by the controller (7) if the state matrix in (9) is asymptotically stable. In general, analysis of the stability of (9) is not an easy task.

Algorithm 2. One alternative to the above algorithm is the algorithm using explicitly the regulator state $w^{(i)}$. It has a disadvantage of being of higher order than Algorithm 1; however, it does not utilize any approximation of $w^{(i)}$ with $v^{(i)}$. Recalling (4), we obtain

$$\dot{v}^{(i)} \approx K^{(i)}F^{(i)}w^{(i)} + H^{(i)}C^{(i)}B^{(i)}v^{(i)} + [K^{(i)}G^{(i)} + H^{(i)}C^{(i)}A^{(i)}C^{(i)+}]y^{(i)},$$

since $w^{(i)}$ is generated by the first relation in (3). If $W = \text{col}\{w^{(1)}, \dots, w^{(N)}\}$, then we have, similarly as in the case of (7), that

$$\dot{U} = [\tilde{\Gamma} + \tilde{H}\tilde{C}\tilde{B}]U + \tilde{K}\tilde{F}W + [\tilde{K}\tilde{G} + \tilde{H}\tilde{C}\tilde{A}\tilde{C}^+]Y. \quad (10)$$

The whole closed-loop system can be represented as

$$\begin{bmatrix} \dot{U} \\ \dot{W} \\ \dot{x} \end{bmatrix} = \begin{bmatrix} \tilde{\Gamma} + \tilde{H}\tilde{C}\tilde{B} & \tilde{K}\tilde{F} & (\tilde{K}\tilde{G} + \tilde{H}\tilde{C}\tilde{A}\tilde{C}^+)\tilde{C} \\ 0 & \tilde{F} & \tilde{G}\tilde{C} \\ BQ & 0 & A \end{bmatrix} \begin{bmatrix} U \\ W \\ x \end{bmatrix}. \quad (11)$$

Both control Algorithms 1 and 2 have the structure which reduces to the local controllers when $\tilde{\Gamma} = 0$. In the case of Algorithm 1, the local controllers are derived from \mathbf{C}_i after aggregating (3) to one vector-matrix differential equation for $v^{(i)}$, while in the case of Algorithm 2 the differential equation for $v^{(i)}$ contains explicitly the term $w^{(i)}$, generated by the local observer in \mathbf{C}_i . The form of these controllers is motivated by the idea to introduce a first order dynamic consensus scheme. Namely, without the local controllers, relation $\dot{U} = \tilde{\Gamma}U$ provides asymptotically a weighted sum of the initial conditions $U_i(t_0)$, if the graphs corresponding to the particular components of U_i have center nodes (see, e.g., [16, 25]). Combination of the two terms provides a possibility to improve the overall performance by exploiting potential advantages of each local controller. However, the introduction of additional dynamics, required by the consensus scheme, can contribute to deterioration of the overall performance, and make the choice of the local controller parameters dependent upon the overall control scheme.

Example 1. An insight into the possibilities of the proposed algorithms can be obtained from a simple example in which the system \mathbf{S} is represented by (1), with

$$A = \begin{bmatrix} 0.8 & 2 & 0 \\ -2.5 & -5 & -0.3 \\ 0 & 10 & -2 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \text{ and } C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \text{ Assume that we have two}$$

agents characterized by \mathbf{S}_1 with $A^{(1)} = \begin{bmatrix} 0.8 & 2 \\ -2.5 & -5 \end{bmatrix}$, $B^{(1)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $C^{(1)} = [1 \ 0]$,

and \mathbf{S}_2 with $A^{(2)} = \begin{bmatrix} -5 & -0.3 \\ 10 & -2 \end{bmatrix}$, $B^{(2)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $C^{(2)} = [0 \ 1]$. Obviously, there is only one control signal u . Assume that the second agent is responsible for control implementation, so that $u = u^{(2)} = v^{(2)}$, according to the adopted notation. Assume that both agents have their own controllers \mathbf{C}_1 and \mathbf{C}_2 , obtained by the LQG methodology, assuming a low measurement noise level, so that we obtain $F^{(1)} = \begin{bmatrix} 1.6502 & 2.0000 \\ -2.4717 & -2.8223 \end{bmatrix}$, $G^{(1)} = \begin{bmatrix} -0.8502 \\ -0.26970 \end{bmatrix}$, $K^{(1)} = [0.7414 \ 0.82231]$ and $H^{(1)} = 0$, and $F^{(2)} = \begin{bmatrix} -2.2361 & -24.3071 \\ 0.1000 & 1.1200 \end{bmatrix}$, $G^{(2)} = \begin{bmatrix} 24.2068 \\ -3.1200 \end{bmatrix}$, $K^{(2)} = [0.2361 \ 0.0003]$ and $H^{(2)} = 0$. The system \mathbf{S} with the local controller \mathbf{C}_2 is unstable. Algorithm 1 has been applied according to (7), after introducing $Q = [0 \ 1]$ and $\Gamma_{12} = \Gamma_{21} = 100I_2$. Figure 1 presents the impulse response for all three components of the state vector x for \mathbf{S} . Algorithm 2 has then been applied according to (11); the corresponding responses are presented in Fig. 2.

It is to be emphasized that the consensus scheme puts together two local controllers, influencing in such a way both performance and robustness. Here, the role of the first controller is only to help the second controller in defining the control signal. The importance of the consensus effects can be seen from Fig. 3 in which the responses in the case when $\tilde{\Gamma} = 0$ are presented for the Algorithm 1. It is obvious that the response is worse than in Fig. 1. In the case of Algorithm 2, the system without consensus is even unstable (Fig. 4).

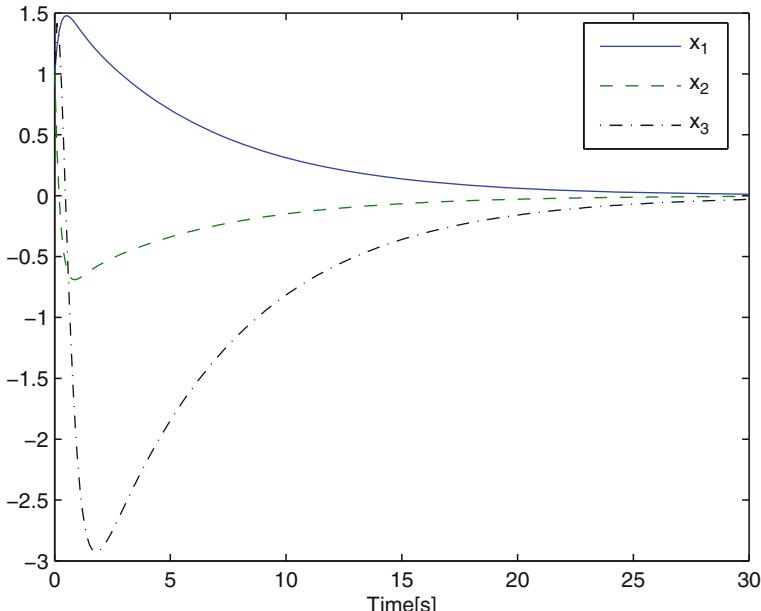


Fig. 1 Impulse response for Algorithm 1

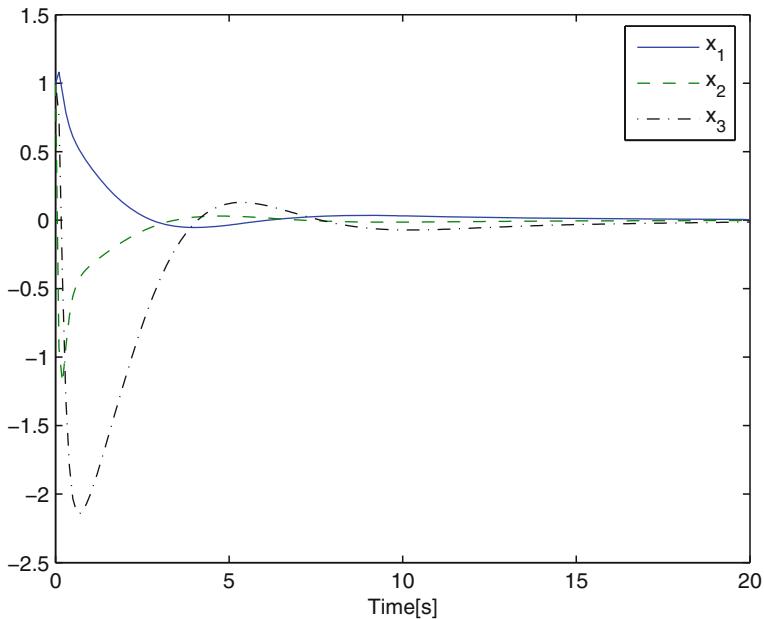


Fig. 2 Impulse response for Algorithm 2

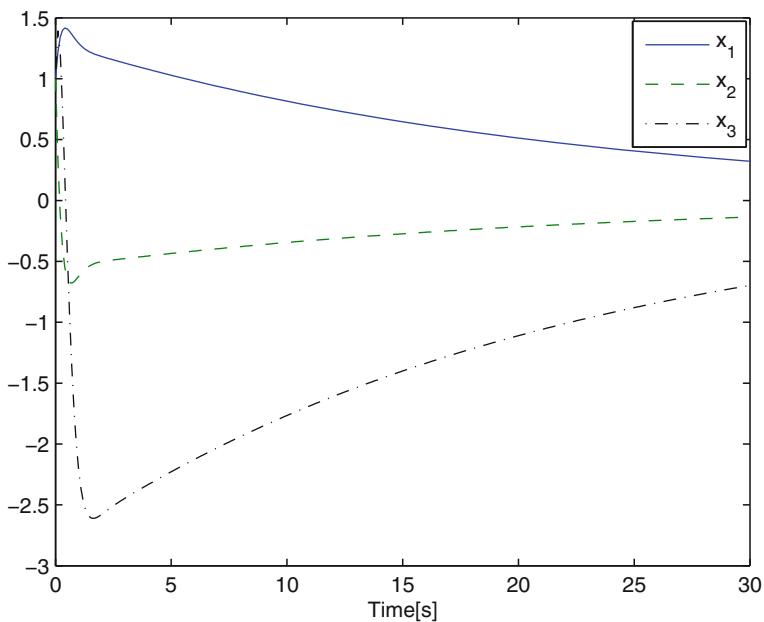


Fig. 3 Algorithm 1: local controllers without consensus

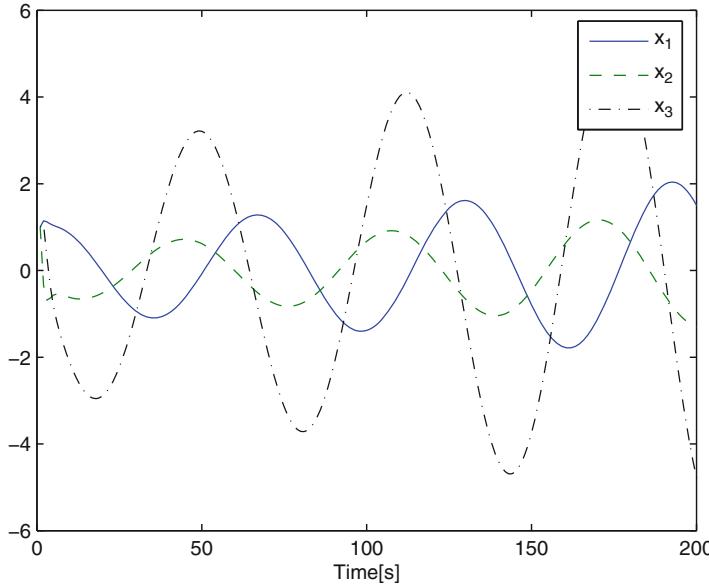


Fig. 4 Algorithm 2: local controllers without consensus

The control algorithms can be made more flexible by introducing some adjustable parameters, so that, for example, the terms $\tilde{K}\tilde{F}\tilde{K}^+$ in (7) and $\tilde{K}\tilde{F}$ in (10) are multiplied by a parameter α , and the term $\tilde{K}\tilde{G}$ in both algorithms by β ; it has been found to be beneficial to have $\alpha > 1$ and $\beta < 1$.

The problem of stabilizability of \mathbf{S} by the proposed algorithms is, in general, very difficult having in mind the supposed diversity of local models and dynamic controllers. Any analytic insight from this point of view into the system matrices in (9) and (11) seems to be very complicated. It is, however, logical to expect that the introduction of the consensus scheme can, in general, contribute to the stabilization of \mathbf{S} . Selection of the elements of $\tilde{\Gamma}$ can, obviously, be done in accordance with the expected performance of the local controllers and the confidence in their suggestions (see, for example, an analogous reasoning related to the estimation problem addressed in the next section). In this sense, connectedness of the agents network contributes, in general, to the overall control performance. The methodology of the vector Lyapunov functions offers good possibilities for at least qualitative conclusions [19, 20].

3.2 Algorithms Derived from the Local Static Feedback Control Laws

Algorithm 3. Assume now that we have static local output controllers, obtained from \mathbf{C}_i in (3) by introducing $F^{(i)} = 0$, $G^{(i)} = 0$ and $K^{(i)} = 0$, so that we have $v^{(i)} = H^{(i)}y^{(i)}$. Both Algorithms 1 and 2 give in this case

$$\dot{U} = \tilde{\Gamma}U + \tilde{H}\tilde{C}[\tilde{B}U + \tilde{A}\tilde{C}^+Y]. \quad (12)$$

The closed-loop system is now given by

$$\begin{bmatrix} \dot{U} \\ \dot{x} \end{bmatrix} = \begin{bmatrix} \tilde{\Gamma} + \tilde{H}\tilde{C}\tilde{B} & \tilde{H}\tilde{C}\tilde{A}\tilde{C}^+\tilde{C} \\ \tilde{B}Q & A \end{bmatrix} \begin{bmatrix} U \\ x \end{bmatrix}. \quad (13)$$

A special case of the above controller deserves particular attention. Assume in the Algorithm 3 that $C^{(i)} = I_{n_i}$ and that that $y^{(i)} = x^{(i)}$ and $\zeta^{(i)} = x^{(i)}$ represents a part of the vector x . In the special case when all the agents possess the knowledge about the entire model of \mathbf{S} , $y^{(i)} = x$, and the agents can differ by their control laws and responsibilities for control actions. Under these assumptions, Algorithm 3 becomes

$$\dot{U} = \tilde{\Gamma}U + \tilde{H}[\tilde{B}U + \tilde{A}\tilde{x}], \quad (14)$$

where $\tilde{x} = \text{col}\{x^{(1)}, \dots, x^{(N)}\}$, $\dim\{\tilde{x}\} = \sum_{i=1}^N n_i$ represents the expanded vector x , available through measurements. Notice that it is always possible to find a full rank $\sum_{i=1}^N n_i \times n$ matrix V that $\tilde{x} = Vx$ (for a general discussion about state expansion, see [20]). The closed-loop system is now

$$\begin{bmatrix} \dot{U} \\ \dot{x} \end{bmatrix} = \begin{bmatrix} \tilde{\Gamma} + \tilde{H}\tilde{B} & \tilde{H}\tilde{A}V \\ \tilde{B}Q & A \end{bmatrix} \begin{bmatrix} U \\ x \end{bmatrix}. \quad (15)$$

Example 2. Assume that we have the same system as in Example 1, and that we have the local stabilizing output feedback gains $H^{(1)} = 0.7414$ and $H^{(2)} = 30$, obtained from the state feedback part of the local controllers formulated in Example 1, and increasing the gain $H^{(2)}$ in order to have any significant effect of the feedback. Figure 5 depicts the responses in the case when the consensus with $\Gamma_{12} = \Gamma_{21} = 100I_2$ is applied. The system without consensus (with only one controller applied) is unstable.

Remark 1. Properties of (14), can be somewhat clarified by the following consideration using a simple example. Assume that $\dot{x} = ax + bu$, where x, u, a and b are scalars, and assume that $u = hx$ is a stabilizing feedback. Differentiating the last relation one obtains $\dot{u} = h\dot{x} = h(ax + bu)$. The obtained relation, together with the system model, defines a system with one pole of the closed-loop system at $a + hb$, and one at the origin. If one modifies the controller relation in such a way that $\dot{u} = cu + h(ax + bu)$, one obtains a root locus with respect to c with two branches, one of which goes from 0 to a , and the other from $a + bk$ to $-\infty$. Formally, in (14), the role of c is given to $\tilde{\Gamma}$; it comes out that the introduction of a structure modeling \dot{x} in (14) by $\tilde{B}U + \tilde{A}\tilde{x}$ generates a pole at the origin when $\tilde{B}U + \tilde{A}\tilde{x} = Bu + Ax$. When using local models, the introduced additional dynamics obviously deteriorates the overall performance, so that the consensus matrix (together with the other controller parameters), has as one of its duties to cope with this effect. Consequently, special parameters can be introduced in the algorithm, such as negative diagonal terms in $\tilde{\Gamma}$, or already mentioned parameters α and β , multiplying \tilde{B} and \tilde{A} , respectively (a more rigorous analysis can be based on vector Lyapunov functions).

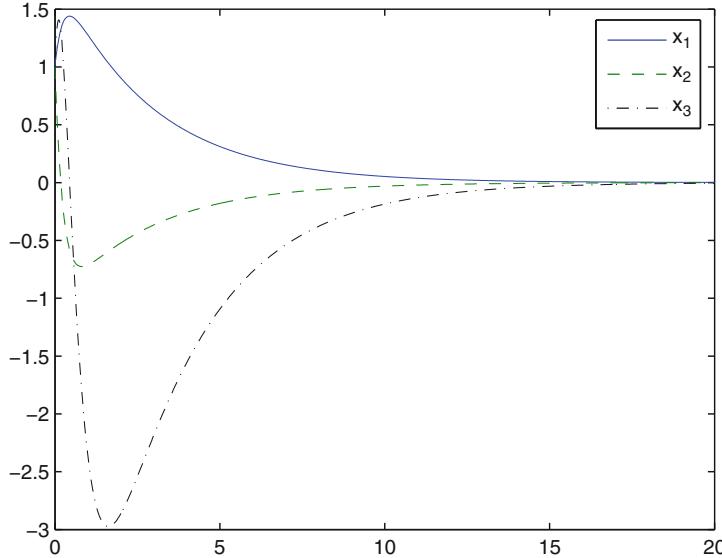


Fig. 5 Impulse response for Algorithm 3

Remark 2. The proposed multi-agent control schemes can be compared to those overlapping decentralized control schemes for complex systems that are derived by using the expansion/contraction paradigm and the inclusion principle (especially in the case of Algorithm 3), e.g., [8, 9, 11, 20, 26, 28], having in mind that both approaches follow similar lines of thought (the above presented approach is, however, much more general). From this point of view, formulation of the local controllers connected to the agents corresponds to the controller design in the expanded space in the case of the inclusion based design, and the application of the dynamic consensus strategy to the contraction to the original space for control action implementation, see, e.g., [8, 9, 20]. The proposed methodology offers, evidently, much more flexibility (local model structure, agreement strategy), at the expense of additional closed loop dynamics introduced by the consensus scheme itself. Moreover, it is interesting to notice that numerous numerical simulations show a pronounced advantage of the proposed scheme (smoother and even faster responses). The reason could be found in the advantage of the consensus strategy over the contraction transformation. In Sect. 5 an application of the mentioned expansion/contraction methodology to the control of formations of UAVs will be presented and compared to a consensus based methodology.

4 Consensus at the State Estimation Level

The previous section was devoted to general structures for introducing consensus at the input level in a multi agent system in which a number of agents with overlapping resources and different competencies participate in defining the global control law.

In this section we will approach the problem in a different way, in which the consensus strategy is introduced at the level of state estimation. A consensus based estimation scheme has been proposed in [25] for the continuous time case, and in [24] for the discrete time case.

Assume that the local models are such that $\zeta^{(i)} = x^{(i)}$, so that the dynamic systems \mathbf{S}_i are *overlapping subsystems* of \mathbf{S} [5, 20, 22, 23]. Starting from the model \mathbf{S}_i and the accessible measurements $y^{(i)}$, each agent is able to generate autonomously its own local estimate $\hat{x}^{(i)}$ of the vector $x^{(i)}$ using a *local estimator* which can be defined in the following Luenberger form:

$$\bar{\mathbf{E}}_i : \quad \dot{\hat{x}}^{(i)} = A^{(i)}\hat{x}^{(i)} + B^{(i)}v^{(i)} + L^{(i)}(y^{(i)} - C^{(i)}\hat{x}^{(i)}) \quad (16)$$

where $L^{(i)}$ is a constant matrix, which can be taken to be the steady state Kalman gain, and $v^{(i)}$ is the input which is supposed to be known.

The *overlapping decentralized estimators* (16) provide a set of *overlapping estimates* $\hat{x}^{(i)}$. However, if the final goal is to get an estimate \hat{x} of the whole state vector x of \mathbf{S} , a *consensus scheme* can be introduced which would enable all the agents to get reliable estimates of the whole state vector x on the basis of:

- (1) the local estimates $\hat{x}^{(i)}$, and
- (2) communications between the nodes based on a decentralized strategy uniform for all the nodes.

In [25] an algorithm providing a solution to this problem has been proposed. If X_i is an estimate of x generated by the i th agent, the following set of estimators is attached to the agents in the network:

$$\mathbf{E}_i : \quad \dot{X}_i = A_i X_i + B_i^* u + \sum_{j=1, j \neq i}^N \Gamma_{ij}(X_j - X_i) + L_i(y^{(i)} - C_i X_i), \quad (17)$$

$i = 1, \dots, N$, A_i is an $n \times n$ matrix with $n_i \times n_i$ nonzero elements being equal to those of $A^{(i)}$ but being placed at the indices defined by $\mathcal{X}^i \times \mathcal{X}^i$ while leaving zeros elsewhere. L_i is an $n \times v_i$ matrix obtained similarly as A_i in such a way that its nonzero elements are those of $L^{(i)}$ placed row by row at row-indices defined by \mathcal{X}^i leaving zeros elsewhere. B_i^* is an $n \times m$ matrix obtained from B_i by putting its rows at the indices defined by \mathcal{X}^i while leaving zeros elsewhere, and Γ_{ij} , $i \neq j$, are constant $n \times n$ diagonal matrices with positive entries. The algorithm is, in fact, based on a combination of decentralized overlapping estimators and a consensus scheme with matrix gains Γ_{ij} , tending to make the local estimates X_i as close as possible. If $X = \text{col}\{X_1, \dots, X_N\}$ is the vector composed of all the state estimates in the agents network, the following model describes its global behavior:

$$\mathbf{E} : \quad \dot{X} = (\tilde{\Gamma} + \tilde{A}^* - \tilde{L}^* \tilde{C}^*)X + \tilde{B}^* U^* + \tilde{L}^* Y, \quad (18)$$

where $\tilde{\Gamma}$ represents now an $nN \times nN$ matrix composed of the blocks Γ_{ij} , $i \neq j$, with $\Gamma_{ii} = -\sum_{i=1, i \neq j}^N \Gamma_{ij}$, $i = 1, \dots, N$, $\tilde{A}^* = \text{diag}\{A_1, \dots, A_N\}$, $\tilde{L}^* = \text{diag}\{L_1, \dots, L_N\}$, $\tilde{C}^* = \text{diag}\{C_1, \dots, C_N\}$, $\tilde{B}^* = \text{diag}\{B_1^*, \dots, B_N^*\}$ and $U^* = \text{col}\{u, \dots, u\}$.

Moreover, we shall assume that all the agents have the a priori knowledge about some convenient or optimal state feedback control law for \mathbf{S} , expressed as $u = K^o x$. Using this knowledge and the estimation scheme (17), the agents can calculate the corresponding inputs $U_i = K^o X_i$ on the basis of the separation principle. Accordingly, the implementation of the control signals is done according to the index sets \mathcal{U}^i .

Algorithm 4. The described decentralized overlapping estimation scheme with consensus, which provides state estimates of the whole state vector x to all the agents, used in conjunction with the globally optimal state feedback control law, represents a specific control algorithm which provides a solution to the general multi-agent control problem of \mathbf{S} .

Defining $\tilde{K}^o = \text{diag}\{K^o, \dots, K^o\}$, we have, according to the above given notation, that $u = Q\tilde{K}^o X$, so that the whole closed-loop system becomes

$$\begin{bmatrix} \dot{X} \\ \dot{x} \end{bmatrix} = \begin{bmatrix} \tilde{\Gamma} + \tilde{A}^* - \tilde{L}^* \tilde{C}^* + \tilde{B}^* \tilde{K} & \tilde{L} \\ BQ\tilde{K}^o & A \end{bmatrix} \begin{bmatrix} X \\ x \end{bmatrix}, \quad (19)$$

where $\tilde{K} = \text{col}\{QK^o, \dots, QK^o\}$ and $\tilde{L} = \text{col}\{L_1 C_1, \dots, L_N C_N\}$. A more realistic version of the above algorithm is obtained by replacing the actual input u in (17) by the local estimates of the input vector $U_i = K^o X_i$, having in mind the local availability of X_i . This imposes, obviously, additional problems related to stability of the closed loop system.

Example 3. In this example, performance of the above algorithm is demonstrated on the same system as in the Example 1. The local estimators are performing the local state estimation using the gains $L_1 = [-4 \ 9]^T$ and $L_2 = [2 \ -7]^T$. The consensus gains in the matrix $\tilde{\Gamma}$ are selected to be $\Gamma_{12} = \Gamma_{21} = 100I_2$. The global LQ optimal control matrix K^o is implemented by both agents. Since only the second agent implements the input u , we assume that the first one uses the estimate $U_1 = K^o X_1$ in the local state estimation algorithm. The impulse response of the proposed control algorithm, which is shown in Fig. 6, is comparable to the the impulse response of the globally LQ optimal controller shown in the same figure.

Stability analysis of Algorithm 4 represents in general a very complex task. It is possible to apply the methodology of [21] under very simplifying assumptions, and to show that the eigenvalues of (19) are composed of the eigenvalues of $\tilde{A}^* - \tilde{L}^* \tilde{C}^*$, $\tilde{A}^* + \tilde{B}^* \tilde{K}$ and $\tilde{A}^* - \tilde{L}^* \tilde{C}^* + \tilde{B}^* \tilde{K}$ modified by a term depending on the eigenvalues of the Laplacian of the network and the consensus gain matrices. However, the underlying assumptions in [21] include the one that all the agents have the exact system model, as well as that the control inputs are transmitted throughout the network; in the overlapping decentralized case, which is in the focus of this work, these assumptions are violated, making the stability analysis much more complex, dependent on the accuracy of the local models and the related estimators.

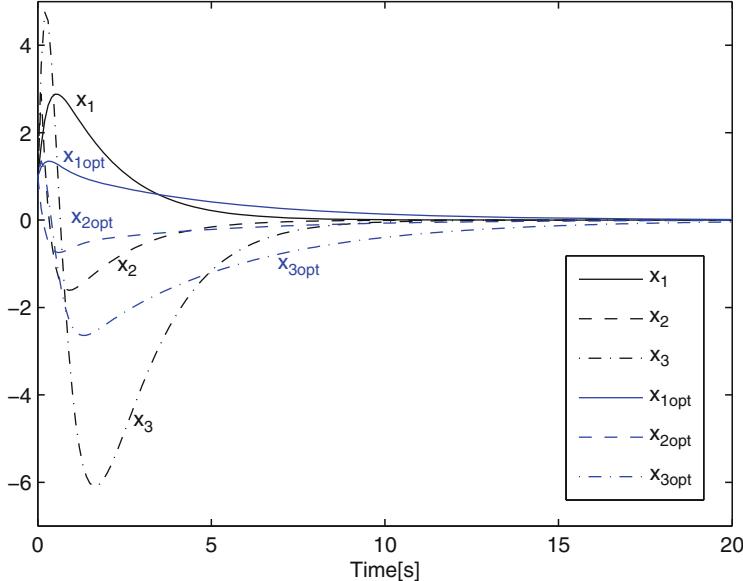


Fig. 6 Algorithm 4 and the globally LQ optimal controller

5 Consensus Based Decentralized Control of UAV Formations

In [27] decentralized control and estimation strategies are proposed for formations of UAVs starting from a specific formation model and the definition of specific subsystems attached to the vehicles. The design of the corresponding control and estimation algorithms has been based on the expansion/contraction paradigm and the inclusion principle [9, 20, 26–28]. As a result, each vehicle, according to the available measurements, is able to generate its own control signal. In this section, we shall show how the Algorithm 4 described in Sect. 4 can be applied to the control of formations of UAVs. It will be demonstrated that superior performance can be obtained when the state estimation is based on consensus between the agents and when the control signals are generated by using the globally LQ optimal state feedback gain matrix.

5.1 Formation Model

Consider a set of N vehicles moving in a plane, where the i th vehicle is represented by the linear double integrator model

$$\dot{z}_i = A_v z_i + B_v u_i = \begin{bmatrix} 0_{2 \times 2} & I_2 \\ 0_{2 \times 2} & 0_{2 \times 2} \end{bmatrix} z_i + \begin{bmatrix} 0_{2 \times 2} \\ I_2 \end{bmatrix} u_i, \quad (20)$$

($i = 1, \dots, N$), where $z_i \in \mathbb{R}^4$ and $u_i \in \mathbb{R}^2$ are the state and the control input vectors, respectively ($0_{m \times n}$ denotes the $m \times n$ zero matrix, and I_n the $n \times n$ identity matrix) (see [28] for possible physical interpretations). The vehicle models (20) are coupled through the control inputs. We shall assume that the i th vehicle is provided with the information about distances with respect to the vehicles whose indices belong to a set of indices of the *sensed vehicles* $S_i = \{s_1^i, \dots, s_{m_i}^i\}$. Accordingly, after decomposing z_i as $z_i = [z_i'^T \ z_i''^T]^T$, where $z_i' = [z_{i,1}' \ z_{i,2}']^T$ and $z_i'' = [z_{i,1}'' \ z_{i,2}'']^T$, we define

$$x_i' = \sum_{j \in S_i} \alpha_j^i z_j' - z_i', \quad x_i'' = z_i'', \quad (21)$$

where $\alpha_j^i \geq 0$ and $\sum_{j \in S_i} \alpha_j^i = 1$; $x_i' = [x_{i,1}' \ x_{i,2}']^T$ represents the distance between the i th vehicle and a “centroid” of the set of vehicles selected by S_i obtained by using a priori selected weights α_j^i . In the case of formation leaders, when $S_i = \emptyset$, we have $x_i' = -z_i'$. Therefore,

$$\dot{x}_i' = \sum_{j \in S_i} \alpha_j^i z_j'' - z_i'' = \sum_{j \in S_i} \alpha_j^i x_j'' - x_i'' = \dot{x}_i'' = u_i, \quad (22)$$

$i = 1, \dots, N$, using the fact that $z_i'' = z_i'$, so that $x_i'' = [x_{i,1}'' \ x_{i,2}'']^T = z_i'' = [z_{i,1}'' \ z_{i,2}'']^T = \dot{z}_i' = [z_{i,1}' \ z_{i,2}']^T$.

The above described set of N vehicles with their sensing indices and the corresponding weights can be considered as a directed weighted graph \mathcal{G} in which each vertex represents a vehicle, and an arc with the weight α_j^i leads from vertex j to vertex i if $j \in S_i$. Consequently, the *weighted adjacency matrix* $G = [G_{ij}]$ is an $N \times N$ square matrix defined by $G_{ij} = \alpha_j^i$ for $j \in S_i$, and $G_{ij} = 0$ otherwise. We shall define the *weighted Laplacian* of the graph as $L = [L_{ij}]$, $L_{ij} = G_{ij}$, $i \neq j$, $L_{ii} = -\sum_j \alpha_{ij}$ (e.g., see [6]).

Defining the vehicle state and control input vectors as $x_i = [x_i'^T \ x_i''^T]^T$ and u_i , $i = 1, \dots, N$, respectively, we obtain from (22) the following *formation state model*

$$\mathbf{S} : \dot{x} = Ax + Bu = [(G - I) \otimes A_v]x + [I \otimes B_v]u, \quad (23)$$

where x and u are the formation state and control vectors defined as concatenations of the vehicle state and control vectors, while \otimes denotes the Kronecker's product. We shall assume that each vehicle has the information about the reference state trajectories $r_i = [r_i'^T \ r_i''^T]^T$, so that the control task to be considered is the task of tracking the desired, possibly time varying, references.

5.2 Global LQ Optimal State Feedback

We shall attach the following quadratic criterion to (23)

$$J = \int_0^\infty (x^T Q x + u^T R u) dt, \quad (24)$$

where $Q \geq 0$ and $R > 0$ are appropriately defined matrices. The design of the state feedback gain minimizing J is faced with the problem that the model (23) is in general not completely controllable. Namely, one can directly observe that the part of the state vector of (23) which corresponds, for example, to the relative positions with respect to the first axis $x'_1 = [x'_{i,1} \dots x''_{N,1}]^T$, satisfies the relation $x'_1 = (I - G)p_1$, where p_1 is the vector of absolute vehicle positions with respect to a reference frame. If the graph \mathcal{G} has a spanning tree, the Laplacian L has one eigenvalue at the origin, and the rest in the open left-half plane, e.g., [15, 18]. This means that in the case when $I - G = L$, we have $r^T x = 0$, where r^T is the left eigenvector of L corresponding to the zero eigenvalue [27]. The controllability matrix does not have full rank since $\text{rank}[B|AB] = 2(N-1)$ (having in mind that $A^2 = 0$). However, it is possible to observe that the system is in this case controllable for the admissible initial conditions for (23) which have to satisfy $r^T x_0 = 0$ for a real formation. Notice that in the case of no formation leader, i.e., when $I - G \neq L$, the matrix $I - G$ is nonsingular provided \mathcal{G} has a spanning tree. A way of solving the problem can be seen after applying to x a nonsingular transformation $T = \begin{bmatrix} r^T \\ \vdots \\ W \end{bmatrix}$ in which W is a full rank matrix such that r^T is linearly independent of the rows of W . It can be seen that \mathbf{S} is controllable for all the admissible initial conditions provided the formation model is controllable with respect to $v = Wx$. Models for v in the form

$$\mathbf{S}^a : \dot{v} = A^a v + B^a u. \quad (25)$$

represent *aggregations* of \mathbf{S} having in mind that W is a full rank matrix [20, 22, 26]. The system matrices satisfy then the aggregation conditions $WA = A^a W$ and $B^a = WB$ [20]. In order to take care of optimality, we shall attach to (25) the following criterion

$$J^a = \int_0^\infty (v^T Q^a v + u^T R^a u) dt. \quad (26)$$

Obviously, the criterion J includes the criterion J^a , i.e., $J = J^a$, if $W^T Q^a W = Q$ and $R = R^a$ (see [10, 20] for a general discussion about the inclusion of performance indices). If one starts from J , an approximate solution to the posed optimization problem can be found by formulating J^a using the approximate relation $Q^a = W^{+T} Q W^+$ (where W^+ denotes the pseudo-inverse of W) and solving the minimization problem of J^a taking \mathbf{S}^a as a constraint. If K^a is obtained as the corresponding optimal feedback gain matrix, a feedback gain matrix K for \mathbf{S} can be found simply by applying the relation $K = K^a W$ [22], since in this case

the closed-loop system (\mathbf{S}^a, K^a) is an aggregation of the closed-loop system (\mathbf{S}, K) and $J = J^a$. Notice that W can be chosen in many different ways: a simple choice

is, for example, $W = \begin{bmatrix} 0.5 & 0 & 0.5 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \cdots & & & & 1 \end{bmatrix}$, which does not substantially change

the structure of Q^a with respect to Q . Also, for a given W , different choices of A^a are possible; having in mind the sparsity of A , A^a can be found for adequate choices of W by using simple linear combinations of the rows of A and deleting its columns.

5.3 Decentralized State Estimation

We shall consider formation state estimation starting from a convenient formation model involving all the inter-vehicle distances measured by the agents

$$\bar{\mathbf{S}} : \dot{\bar{x}} = \bar{A}\bar{x} + \bar{B}u, \quad (27)$$

where $\bar{x} = [\bar{x}_1^{cT} \dots \bar{x}_N^{cT}]^T$, $\bar{x}_i^c = [(z'_{s_i} - z'_i)^T \dots (z'_{s_{m_i}} - z'_i)^T (z''_i)^T]^T$, $\bar{A} = \text{blockcol}\{\bar{A}_1^c \dots \bar{A}_N^c\}$, $\bar{A}_i^c = \begin{bmatrix} \bar{A}_i^{c*} \\ \vdots \\ 0_{2 \times 2 \sum_{k=1}^N (m_k+1)} \end{bmatrix}$ in which \bar{A}_i^{c*} is composed of N blocks in a block-row having dimensions $2m_i \times 2(m_k+1)$, $k = 1, \dots, N$; these blocks are nonzero

only for: 1) $k = i$, when it has the form $\begin{bmatrix} \vdots & -I_2 \\ 0 & \vdots \\ \vdots & -I_2 \end{bmatrix}_{2m_i \times 2(m_i+1)}$ and for 2) all $k \in S_i$

when they have the form $\begin{bmatrix} \vdots & I_2 \\ 0 & \vdots \\ \vdots & I_2 \end{bmatrix}_{2m_i \times 2(m_k+1)}$ in which I_2 is placed at the row index

j satisfying $s_j^i = k$, and $\bar{B} = \text{blockcol}\{\bar{B}_1^c \dots \bar{B}_N^c\}$, $\bar{B}_i^c = \begin{bmatrix} 0_{2m_i \times 2} \\ \vdots \\ I_2 \end{bmatrix}$.

It is easy to verify that $x = T\bar{x}$, where $T = \text{blockdiag}\{T_1 \dots T_N\}$ and $T_i = \begin{bmatrix} \alpha_{s_1^i}^i I_2 & \cdots & \alpha_{s_{m_i}^i}^i I_2 & \cdots \\ \vdots & \ddots & \vdots & \ddots \\ \vdots & & \vdots & \\ \vdots & & & I_2 \end{bmatrix}$, so that \mathbf{S} represents an aggregation of $\bar{\mathbf{S}}$; the system matrices are, consequently, related by $T\bar{A} = AT$ and $\bar{B} = TB$ [9, 20].

Following the basic idea exposed in [27], we shall extract the following *overlapping subsystems* from $\bar{\mathbf{S}}$

$$\bar{\mathbf{S}}^{(i)} : \dot{\bar{x}}^{(i)} = \bar{A}^{(i)}\bar{x}^{(i)} + \bar{B}^{(i)}\bar{u}^{(i)}, \quad (28)$$

$$i = 1, \dots, N, \text{ where } \bar{x}^{(i)} = \begin{bmatrix} (z''_{s_1^i})^T & \cdots & (z''_{s_{m_i}^i})^T & (z'_{s_1^i} - z'_i)^T & \cdots & (z'_{s_{m_i}^i} - z'_i)^T & (z''_i)^T \end{bmatrix}^T,$$

$$\bar{u}^{(i)} = \begin{bmatrix} u_{s_1^i}^T & \cdots & u_{s_{m_i}^i}^T & u_i^T \end{bmatrix}^T, \bar{A}^{(i)} = \begin{bmatrix} 0_{2m_i \times 2(m_i+1)} \\ \vdots \\ \bar{A}_i^* \\ \vdots \\ 0_{2 \times 2(m_i+1)} \end{bmatrix}, \text{ in which } \bar{A}_i^* \text{ has the form}$$

$$\begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ I_{2m_i} & -I_2 \\ 0 & \vdots \\ I_{2m_i} & -I_2 \end{bmatrix} \text{ and } \bar{B}^{(i)} = \begin{bmatrix} I_{2m_i} & 0_{2m_i \times 2} \\ \vdots & \vdots \\ 0_{2m_i \times 2m_i} & 0_{2m_i \times 2} \\ 0_{2 \times 2m_i} & I_2 \end{bmatrix}.$$

Following the idea exposed in [27], models $\tilde{\mathbf{S}}^{(i)}$ can be used for constructing local observers of Luenberger type providing overlapping estimates $\hat{x}^{(i)}$ of the subsystem states:

$$\mathbf{E}^{(i)} : \quad \dot{\hat{x}}^{(i)} = \bar{A}^{(i)}\hat{x}^{(i)} + \bar{B}^{(i)}\bar{u}^{(i)} + L^{(i)}[y^{(i)} - C^{(i)}\hat{x}^{(i)}], \quad (29)$$

where $y^{(i)}$ is the local measurement vector available in the i th vehicle, $C^{(i)}$ a matrix defining mapping from the local state to the local output, and $L^{(i)}$ is the estimator gain (compare with (17)). It should be noticed that the formulated local observers (29) can be implemented provided the control vectors $\bar{u}^{(i)}$ are appropriately defined. When we have decentralized control like in the case treated in [27], the local control law for the i th vehicle gives u_i , while the remaining elements of $\bar{u}^{(i)}$ are then obtained by considering the sensed vehicles as leading vehicles, with their own velocity feedbacks adapted to the given velocity reference (see, for example, the methodology based on LQ optimization applied to the platooning problem [26]).

In order to obtain the estimates of the whole state vector \bar{x} of $\tilde{\mathbf{S}}$ by all the agents, a *consensus scheme* can be introduced as in Sect. 4. If \hat{x}_i is the estimate of the whole formation state vector \bar{x} generated by the i th agent, the following consensus based estimator results now directly from (17)

$$\mathbf{E}_i : \quad \dot{\hat{x}}_i = A_i\hat{x}_i + B_iU_i + \sum_{j=1, j \neq i}^N \Gamma_{ij}(\hat{x}_j - \hat{x}_i) + L_i(y^{(i)} - C_i\hat{x}_i), \quad (30)$$

$i = 1, \dots, N$, where matrices A_i , B_i and L_i are obtained from $\bar{A}^{(i)}$, $\bar{B}^{(i)}$ and $L^{(i)}$ similarly as in the case of (17) using the model (27) for $\tilde{\mathbf{S}}$, while matrices Γ_{ij} contain the consensus parameters. The main point here is generation of the global input vector which should be introduced in (30) and which is not available to all the agents. Vector U_i introduced in (30) represents an approximation of u achievable by the i th agent based on the assumption that all the agents are supposed to know the global state feedback gain K in the optimal mapping $u = Kx = K^aWx$, where K^a is obtained by minimizing J^a in (26). Therefore, we have, according to Sect. 4, that $U_i = KT\hat{x}_i$; this means that the implemented control signal u_i is the i -th component of the N -dimensional vector U_i . The resulting closed loop system model can be obtained directly using (19).

5.4 Experiments

In this section we shall illustrate the proposed method for control of formations of UAVs. A formation of four vehicles without a formation leader has been simulated. It has been assumed that the second and the third vehicle observe the first, the fourth observes the second and the third, while the first vehicle observes the fourth one. The globally LQ optimal feedback gain has been found on the basis of Sect. 5.2. The consensus based estimator, proposed in Sect. 5.3 has been implemented by each agent, assuming the adopted information flow between the agents. The consensus gains are all set to be the same, equal to 100. In Figs. 7 and 8 x-components of the distances and velocities of all four vehicles in the formation are depicted, assuming step distance reference change. On the other hand, Figs. 9 and 10 represent the responses of the same formation with the controllers designed using the expansion/contraction paradigm and inclusion principle with local estimators [8, 20, 26, 28]. It is obvious that better performance is obtained using the consensus based control structure, at the expense of additional communications between the vehicles in the formation.

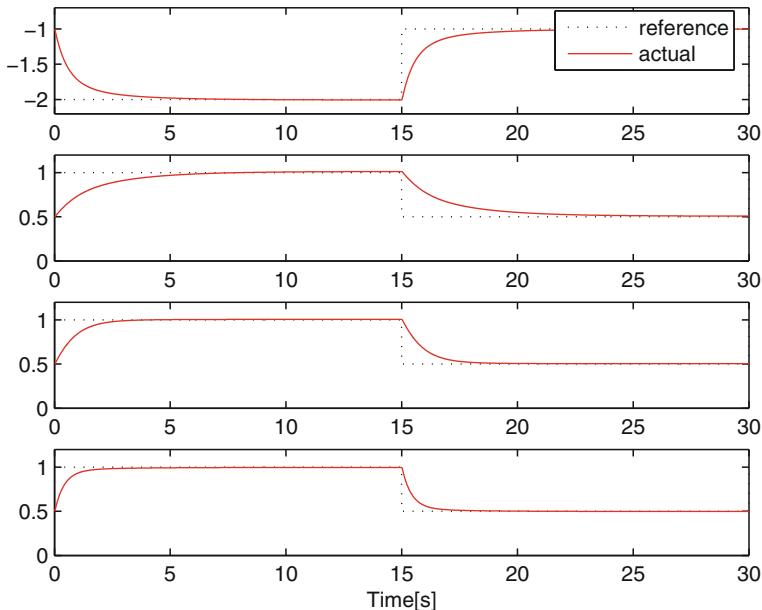


Fig. 7 Distance plots: consensus based controllers

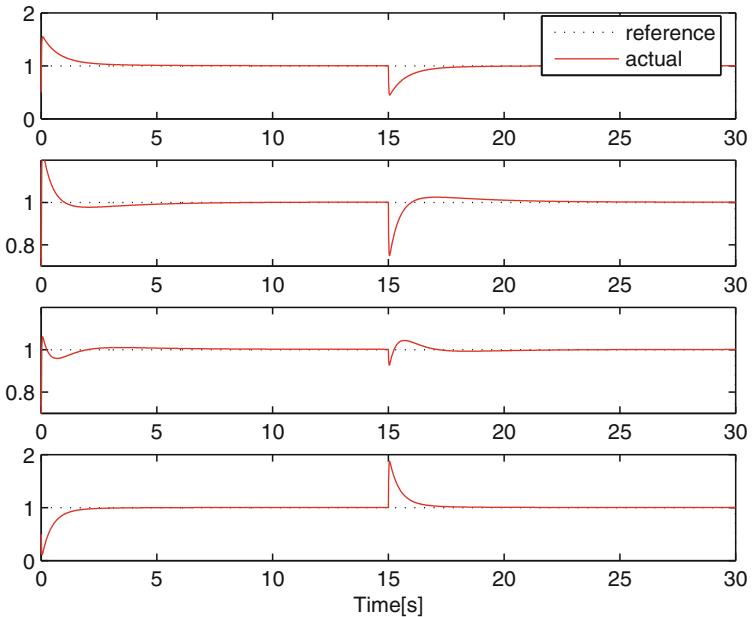


Fig. 8 Velocity plots: consensus based controllers

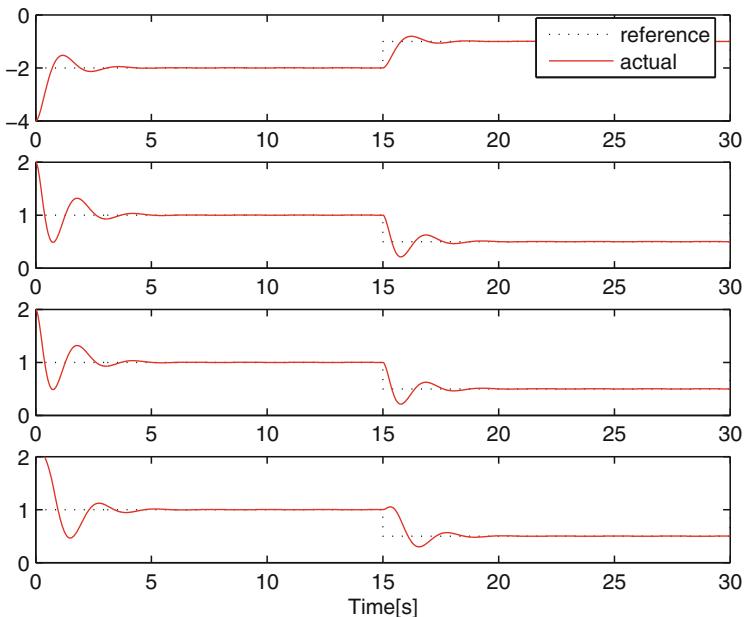


Fig. 9 Distance plots: expansion/contraction based controllers

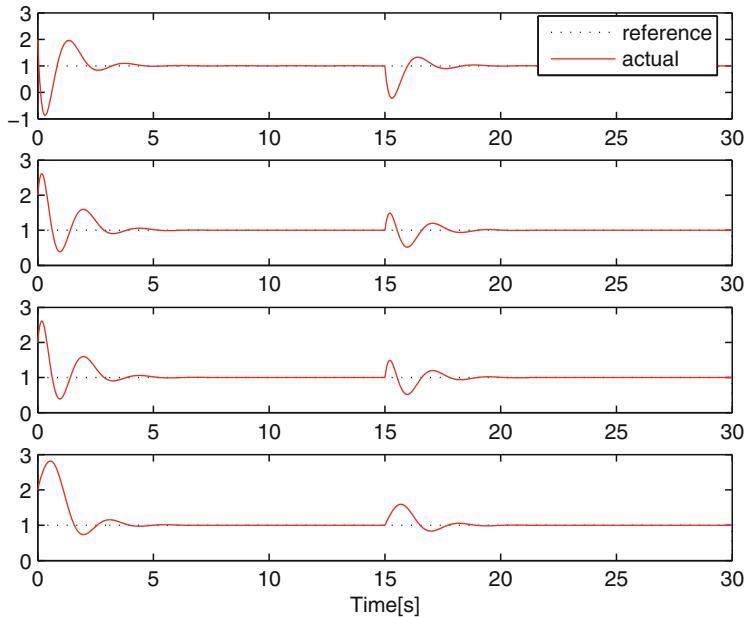


Fig. 10 Velocity plots: expansion/contraction based controllers

References

1. B. Baran, E. Kaszkurewicz, and A. Bhaya, *Parallel asynchronous team algorithms: convergence and performance analysis*, IEEE Trans. Parallel Distrib. Syst. **7** (1996), 677–688
2. D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: Numerical methods*, Prentice-Hall, Englewood Cliffs, 1989
3. V. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis, *Convergence in multiagent coordination, consensus and flocking*, Proc. IEEE Conf. Decision Contr., 2005
4. C. G. Cassandras and W. Li, *Sensor networks and cooperative control*, Eur. J. Contr. **11** (2005), 436–463
5. X. B. Chen and S. S. Stanković, *Decomposition and decentralized control of systems with multi-overlapping structure*, Automatica **41** (2005), 1765–1772
6. A. Fax and R.M Murray, *Information flow and cooperative control of vehicle formations*, IEEE Trans. Automat. Contr. **49** (2004), 1465–1476
7. H. Gharavi and S. Kumar (eds.), *Proceedings of the IEEE: Special Issue on Sensor Networks and Applications*, vol. 91, August 2003
8. M. Ikeda and D. D. Šiljak, *Decentralized control with overlapping information sets*, J. Optim. Theory Applic. **34** (1981), 279–310
9. _____, *Overlapping decentralized control with input, state and output inclusion*, Contr. Theory Adv. Technol. **2** (1986), 155–172
10. M. Ikeda, D. D. Šiljak, and D.E. White, *Decentralized control with overlapping information sets*, J. Optim. Theory Applic. **34** (1981), 279–310
11. _____, *An inclusion principle for dynamic systems*, IEEE Trans. Autom. Contr. **29** (1984), 244–249
12. A. Jadbabaie, J. Lin, and A. Morse, *Coordination of groups of mobile autonomous agents using nearest neighbor rules*, IEEE Trans. Automat. Contr. **48** (2003), 988–1001

13. Z. Lin, B. Francis, and M. Maggiore, *Necessary and sufficient conditions for formation control of unicycles*, IEEE Trans. Automat. Contr. **50** (2005), 121–127
14. R. Olfati-Saber and R. Murray, *Consensus problems in networks of agents with switching topology and time-delays*, IEEE Trans. Automat. Contr. **49** (2004), 1520–1533
15. W. Ren and E. Atkins, *Distributed multi-vehicle coordinated control via local information exchange*, Int. J. Robust Nonlinear Contr. **17** (2007), 1002–1033
16. W. Ren, R. W. Beard, and E. M. Atkins, *A survey of consensus problems in multi-agent coordination*, Proceedings of American Control Conference, 2005, pp. 1859–1864
17. W. Ren, R. W. Beard, and D. B. Kingston, *Multi-agent Kalman consensus with relative uncertainty*, Proceedings of American Control Conference, 2005
18. W. Ren and R.W. Beard, *Consensus seeking in multi-agent systems using dynamically changing interaction topologies*, IEEE Trans. Autom. Contr. **50** (2005), 655–661
19. D. D. Šiljak, *Large scale dynamic systems: Stability and structure*, North-Holland, New York 1978
20. D. D. Šiljak, *Decentralized control of complex systems*, Academic, New York, 1991
21. R. S. Smith and F. Y. Hadaegh, *Closed-loop dynamics of cooperative vehicle formations with parallel estimators and communication*, IEEE Trans. Autom. Contr. **52** (2007), 1404–1414
22. S. S. Stanković and D. D. Šiljak, *Contractibility of overlapping decentralized control*, Syst. Contr. Lett. **44** (2001), 189–199
23. S. S. Stanković and D. D. Šiljak, *Stabilization of fixed modes in expansions of LTI systems*, Syst. Contr. Lett. **57** (2008), 365–370
24. S. S. Stanković, M. S. Stanković, and D. M. Stipanović, *Consensus based overlapping decentralized estimation with missing observations and communication faults*, Automatica **45** (2009), 1397–1406
25. S. S. Stanković, M. S. Stanković, and D. M. Stipanović, *Consensus based overlapping decentralized estimator*, IEEE Trans. Autom. Contr. **54** (2009), 410–415
26. S. S. Stanković, M. J. Stanojević, and D. D. Šiljak, *Decentralized overlapping control of a platoon of vehicles*, IEEE Trans. Contr. Syst. Technol. **8** (2000), 816–832
27. S. S. Stanković, D. M. Stipanović, and M. S. Stanković, *Decentralized overlapping tracking control of a formation of autonomous unmanned vehicles*, Proceedings of American Control Conference, 2009
28. D. M. Stipanović, G. İnalan, R. Teo, and C. Tomlin, *Decentralized overlapping control of a formation of unmanned aerial vehicles*, Automatica **40** (2004), 1285–1296
29. J. N. Tsitsiklis, *Problems in decentralized decision making and computation*, Ph.D. thesis, Dep. Electrical Eng. Comput. Sci., M.I.T., Cambridge, MA, 1984
30. J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, *Distributed asynchronous deterministic and stochastic gradient optimization algorithms*, IEEE Trans. Autom. Contr. **31** (1986), 803–812
31. P. Yang, R.A. Freeman, and K.M. Lynch, *Multi-agent coordination by decentralized estimation and control*, IEEE Trans. Autom. Contr. **53** (2008), 2480–2496

Graph-Theoretic Methods for Networked Dynamic Systems: Heterogeneity and \mathcal{H}_2 Performance

Daniel Zelazo and Mehran Mesbahi

1 Introduction

The analysis and synthesis of large-scale systems pose a range of challenges due to the number of their subsystems and the complexity of their interactions. In the meantime, the importance of these systems has become increasingly prevalent in science and engineering, especially in the realm of multi-agent systems such as coordination of autonomous vehicles on ground, in sea, air, and space, as well as localization and sensor fusion, energy networks, and distributed computation [1, 2, 4–7, 9, 19, 21]. One aspect of the complexity of large-scale systems is that their subsystems, which we occasionally refer to as ‘agents,’ may not be described by the same set of input-output dynamics. The difference between the dynamics of distinct subsystems may be the result of manufacturing inconsistencies or intentional differences due to varying requirements for each subsystem in the ensemble. An important component to the analysis of these systems, therefore, is to understand how heterogeneity in the subsystem dynamics affects the behavior and performance of the entire system. Another facet of this complexity relates to the interconnection of different subsystems. The underlying interconnection topology of a large-scale system may be determined by the governing dynamic equations of each subsystem, e.g., when the interconnection is a function of some state of each subsystem, or it may be designed as part of the engineering process. In both cases, the interconnection topology has a profound impact on the overall system in terms of its stability, controllability, observability, and performance. Hence, it becomes crucial to understand and explicitly

D. Zelazo

Institute for Systems Theory and Automatic Control, University of Stuttgart,
Pfaffenwaldring 9, 70550 Stuttgart, Germany
e-mail: Daniel.Zelazo@ist.uni-stuttgart.de

M. Mesbahi

Department of Aeronautics and Astronautics, University of Washington,
Box 352400, Seattle, Washington, USA
e-mail: mesbahi@aa.washington.edu

parameterize the role of the interconnection topology for both analysis and synthesis of large-scale systems.

This chapter aims to address these issues via the development of four canonical models of large-scale systems, those that will be collectively referred to as *networked dynamic systems* (NDS). A crucial tool that we employ to derive these models are algebraic representations of graphs and networks [12].

The four systems are NDS coupled at the output, NDS coupled at the input, NDS coupled at the state, and NDS coupled at a combination of input, output, and state. We further subdivide each class by distinguishing between NDS with homogeneous agent dynamics and NDS with heterogeneous agent dynamics. For simplicity of presentation, we will focus on continuous linear time-invariant systems; an analogous framework can be developed for the discrete-time case.

Graph-centric analysis of networked dynamic systems has been extensively treated in the literature. Consensus-type problems, which fall under the category of NDS coupled at the state, are prime examples of how notions from graph theory can be applied in a dynamic systems setting [27]. Examples of such studies include Nyquist-based stability analysis for consensus-based feedback systems [9], graph-centric notion of controllability in consensus problems [25], and consensus algorithms with guaranteed \mathcal{H}_∞ performance [18]. Works related to formation flying applications which rely on relative sensing, falling under the category of NDS coupled at the output, also use results from algebraic and spectral graph theory [19, 24, 31, 39].

The outline of this chapter is as follows. First, we describe the models for various classes of NDS in Sect. 2. In Sect. 3 we proceed to present a graph-theoretic analysis for four classes of NDS models. The goal of this section is to highlight the role of heterogeneity and the interconnection topology on the system-theoretic properties of NDS. In many instances, the differences between agent dynamics and the interconnection topology can be embedded into the system matrices resulting in a single state-space representation of the system. The discussion in Sect. 3 aims to emphasize the importance of keeping the interconnection topology *explicit* in the formulation of the model. That is, for analysis purposes, we prefer to consider a quintuple representation of the system, $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathcal{G})$, where \mathcal{G} denotes the underlying connection topology. Using this approach we will then examine the controllability and observability properties of different NDS models as well as a graph-theoretic characterization of their \mathcal{H}_2 performance. The results obtained from the *analysis* of NDS will then motivate techniques for the *design* of the underlying interconnection topology in §4. In this section, we will rely on results from combinatorial optimization and semi-definite programming that lead to numerically tractable solutions for topology design. For NDS coupled at the output model, we show that with an appropriate representation of the corresponding network synthesis problem, the celebrated Kruskal's algorithm can be used to find an optimal topology in the \mathcal{H}_2 setting. For NDS coupled at the state, we propose a convex relaxation for the minimum cost sensor placement problem that leads to a semi-definite program. Finally in Sect. 5, we present some concluding remarks regarding the implications of the framework discussed in this chapter.

1.1 Preliminaries and Notations

We provide some mathematical preliminaries and notations here. Matrices are denoted by capital letters, e.g., A , and vectors by lower case letters, e.g., x . Diagonal matrices will be written as $D = \text{diag}\{d_1, \dots, d_n\}$; this notation will also be employed for block-diagonal matrices and linear operators. A matrix and/or a vector that consists of all zero entries will be denoted by $\mathbf{0}$; whereas, ‘0’ will simply denote the scalar zero. Similarly, the vector $\mathbf{1}$ denotes the vector of all ones, and $\mathbf{J} = \mathbf{1}\mathbf{1}^T$. The set of real numbers will be denoted as \mathbb{R} , and $\|\cdot\|_p$ denotes the p -norm of its argument, e.g., $p = 2, \infty$, which is used for vector, matrix, and system norms. The adjoint of a linear operator f is denoted by f^* . The notation $A \circ B$ denotes the Hadamard product of the two matrices [15]. The Kronecker product of two matrices A and B on the other hand is written as $A \otimes B$. The following result for Kronecker products will be used subsequently.

Theorem 9.1 ([14]). *Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times q}$ each have a singular value decomposition of $A = U_A \Sigma_A V_A^T$ and $B = U_B \Sigma_B V_B^T$. The singular value decomposition of the Kronecker product of A and B is then*

$$A \otimes B = (U_A \otimes U_B)(\Sigma_A \otimes \Sigma_B)(V_A^T \otimes V_B^T). \quad (1)$$

An immediate consequence of Theorem 9.1 is the identity

$$\|A \otimes B\|_2 = \|A\|_2 \|B\|_2, \quad (2)$$

for the matrix 2-norm. We also make use of the multiplication property

$$(A \otimes B)(C \otimes D) = (AC \otimes BD), \quad (3)$$

for Kronecker products, where the matrices are all of commensurate dimension.

Graphs and the matrices associated with them form a convenient construct for much of the work presented in this chapter. The reader is referred to [12] for a detailed treatment of the subject and we present here only a minimal summary of relevant notions and results. An undirected (simple) graph \mathcal{G} is specified by a vertex set \mathcal{V} and an edge set \mathcal{E} whose elements characterize the incidence relation between distinct pairs of \mathcal{V} . Two vertices i and j are called *adjacent* (or neighbors) when $\{i, j\} \in \mathcal{E}$; we denote this by writing $i \sim j$. The cardinalities of the vertex and edge sets of \mathcal{G} will be denoted by $|\mathcal{V}|$ and $|\mathcal{E}|$, respectively. An *orientation* of an undirected graph \mathcal{G} is the assignment of directions to its edges, i.e., an edge e_k is an ordered pair (i, j) such that i and j are, respectively, the initial and the terminal nodes of e_k . In our discussion, we make extensive use of the $|\mathcal{V}| \times |\mathcal{E}|$ incidence matrix, $E(\mathcal{G})$, for a graph with arbitrary orientation. The incidence matrix is a $\{0, \pm 1\}$ -matrix with rows and columns indexed by the vertices and edges of \mathcal{G} such that $[E(\mathcal{G})]_{ik}$ has the value ‘1’ if node i is the initial node of edge e_k , ‘−1’ if it is the terminal node, and ‘0’ otherwise. The degree of vertex i , d_i , is the cardinality of the set of vertices adjacent to it; we define the degree matrix as $\Delta(\mathcal{G}) = \text{diag}\{d_1, \dots, d_{|\mathcal{V}|}\}$. The adjacency matrix

of an undirected graph, $A(\mathcal{G})$, is the symmetric $|\mathcal{V}| \times |\mathcal{V}|$ matrix such that $[A(\mathcal{G})]_{ij}$ takes the value ‘1’ if node i is connected to node j , and ‘0’ otherwise.

A connected graph \mathcal{G} can be written as the union of two edge-disjoint subgraphs on the same vertex set as $\mathcal{G} = \mathcal{G}_\tau \cup \mathcal{G}_c$, where \mathcal{G}_τ is a spanning tree subgraph and \mathcal{G}_c contains the remaining edges that necessarily complete the cycles in \mathcal{G} . Similarly, the columns of the incidence matrix for the graph \mathcal{G} can always be permuted such that $E(\mathcal{G})$ is written as

$$E(\mathcal{G}) = [E(\mathcal{G}_\tau) \ E(\mathcal{G}_c)]. \quad (4)$$

The cycle edges can be constructed from linear combinations of the tree edges via a linear transformation [28], as

$$E(\mathcal{G}_\tau)T_\tau^c = E(\mathcal{G}_c), \quad (5)$$

where

$$T_\tau^c = (E(\mathcal{G}_\tau)^T E(\mathcal{G}_\tau))^{-1} E(\mathcal{G}_\tau)^T E(\mathcal{G}_c). \quad (6)$$

Using (5) we obtain the following alternative representation of the incidence matrix of the graph

$$E(\mathcal{G}) = E(\mathcal{G}_\tau) [I \ T_\tau^c] = E(\mathcal{G}_\tau)R(\mathcal{G}); \quad (7)$$

the rows of the matrix

$$R(\mathcal{G}) = [I \ T_\tau^c] \quad (8)$$

are viewed as the basis for the *cut space* of \mathcal{G} [12]. The matrix $[-T_\tau^c \ I]^T$, on the other hand, forms a basis for the *flow space*.

The matrix $R(\mathcal{G})$ has a close connection with a number of structural properties of the underlying network. For example, the number of spanning trees in a graph, $\tau(\mathcal{G})$, can be determined from the cut space basis [12], as

$$\tau(\mathcal{G}) = \det [R(\mathcal{G})R(\mathcal{G})^T]. \quad (9)$$

The (graph) Laplacian of \mathcal{G} ,

$$L(\mathcal{G}) := E(\mathcal{G})E(\mathcal{G})^T = \Delta(\mathcal{G}) - A(\mathcal{G}), \quad (10)$$

is a rank deficient positive semi-definite matrix. The eigenvalues of the graph Laplacian are real and will be ordered and denoted as

$$0 = \lambda_1(\mathcal{G}) \leq \lambda_2(\mathcal{G}) \leq \dots \leq \lambda_{|\mathcal{V}|}(\mathcal{G}).$$

The *edge Laplacian* is defined as [40]

$$L_e(\mathcal{G}) := E(\mathcal{G})^T E(\mathcal{G}). \quad (11)$$

The edge Laplacian is intimately related to the graph Laplacian, as shown through the following similarity transformation.

Theorem 9.2. The graph Laplacian for a connected graph $L(\mathcal{G})$ containing cycles is similar to

$$\begin{bmatrix} L_e(\mathcal{G}_\tau)R(\mathcal{G})R(\mathcal{G})^T & 0 \\ 0 & 0 \end{bmatrix},$$

where \mathcal{G}_τ is a spanning tree subgraph of \mathcal{G} and the matrix $R(\mathcal{G})$ is defined via (8).

Proof. We define the transformation

$$S_v(\mathcal{G}) = \left[E(\mathcal{G}_\tau) (E(\mathcal{G}_\tau)^T E(\mathcal{G}_\tau))^{-1} \mathbf{1} \right], \quad S_v(\mathcal{G})^{-1} = \begin{bmatrix} E(\mathcal{G}_\tau)^T \\ (1/\|\mathcal{V}\|)\mathbf{1}^T \end{bmatrix}, \quad (12)$$

where $E(\mathcal{G}_\tau)$ is the incidence matrix of $\mathcal{G}(\mathcal{G}_\tau)$. Applying the transformation as

$$\begin{aligned} S_v(\mathcal{G})^{-1}L(\mathcal{G})S_v(\mathcal{G}) &= \begin{bmatrix} E(\mathcal{G}_\tau)^T E(\mathcal{G}_\tau) \\ 0 \end{bmatrix} R(\mathcal{G})R(\mathcal{G})^T \begin{bmatrix} I & 0 \end{bmatrix} \\ &= \begin{bmatrix} L_e(\mathcal{G}_\tau)R(\mathcal{G})R(\mathcal{G})^T & 0 \\ 0 & 0 \end{bmatrix}, \end{aligned} \quad (13)$$

leads to the desired result.

The transformation (13) provides a transparent way to separate the zero eigenvalue of the Laplacian for a connected graph while preserving algebraic properties of the graph via the edge Laplacian.

In order to apply the framework developed here to specific graphs, we will work with the complete graph and its generalization in terms of k -regular graphs, which are defined as follows. The *complete graph* on n nodes, K_n , is the graph where all possible pairs of vertices are adjacent, or equivalently, if the degree of all vertices is $|V| - 1$. Figure 1(a) depicts K_{10} , the complete graph on ten nodes. When every node

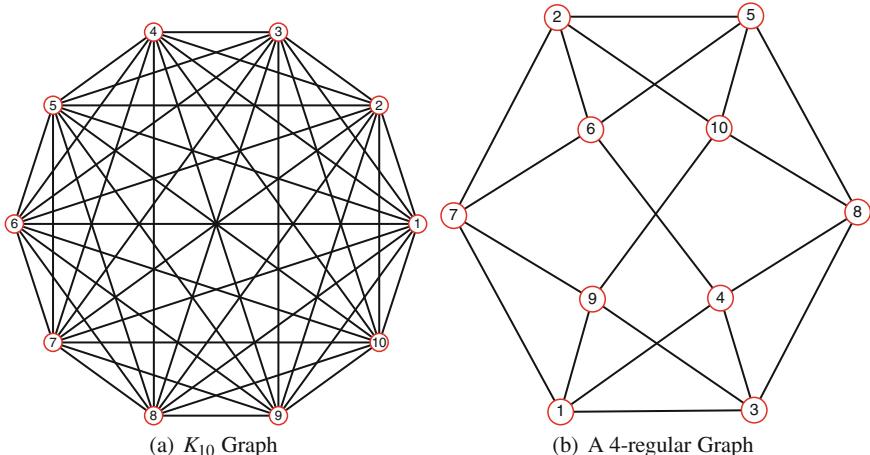


Fig. 1 Example of regular graphs on ten vertices

in a graph with n nodes has the same degree $k \leq n - 1$, it is called a k -regular graph. The k -regular graph on n nodes for $k = 2$ is called the *cycle graph*, C_n . Figure 1(b) shows a 4-regular graph.

2 Canonical Models of Networked Dynamic Systems

In this section we develop a general linear time-invariant model for networked dynamic systems with an emphasis on the means by which the underlying connection topology enters into the system. As alluded to in the introduction, we develop a model which explicitly highlights how the underlying connection topology interacts with each agent in the ensemble.

Fundamental to all NDS is the notion of a “local” and “global” dynamic system layer. The local layer corresponds to the dynamics of each individual agent in the ensemble. This layer captures both the dynamic behavior of each agent in addition to local performance criteria that may or may not be related to certain global or team objectives. For example, the formation control for a team of unmanned vehicles may require each agent to perform a local control and estimation in order to accept higher level navigation commands relating to the team objective. In this direction, we identify two broad classes of NDS: (1) homogeneous, and (2) heterogeneous. For both cases, we will work with a group of n dynamic systems, referred to as agents, each modeled as a linear and time-invariant system of the form

$$\Sigma_i : \begin{cases} \dot{x}_i(t) = A_i x_i(t) + B_i u_i(t) + \Gamma_i w_i(t) \\ z_i(t) = C_i^z x_i(t) + D_i^{zu} u_i(t) + D_i^{zw} w_i(t) \\ y_i(t) = C_i^y x_i(t) + D_i^{yw} w_i(t), \end{cases} \quad (14)$$

where each agent is indexed by the sub-script i . Here, $x_i(t) \in \mathbb{R}^{n_i}$ represents the state, $u_i(t) \in \mathbb{R}^{m_i}$ the control, $w_i(t) \in \mathbb{R}^{r_i}$ an exogenous input (e.g., disturbances and noises), $z_i(t) \in \mathbb{R}^{p_i}$ the controlled variable, and $y_i(t) \in \mathbb{R}^{b_i}$ the locally measured output.

When working with homogeneous NDS, the subscript is dropped, as each agent is described by the same set of linear state-space dynamics (e.g., $\Sigma_i = \Sigma_j$ for all i, j). It should be noted that in a heterogeneous system, the dimension of each agent need not be the same; however, without loss of generality, we assume each agent to have the same dimension.

The parallel interconnection of all the agents has a state-space description

$$\Sigma : \begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \Gamma\mathbf{w}(t), \\ \mathbf{z}(t) = \mathbf{C}^z\mathbf{x}(t) + \mathbf{D}^{zu}\mathbf{u}(t) + \mathbf{D}^{zw}\mathbf{w}(t), \\ \mathbf{y}(t) = \mathbf{C}^y\mathbf{x}(t) + \mathbf{D}^{yw}\mathbf{w}(t), \end{cases} \quad (15)$$

with $\mathbf{x}(t)$, $\mathbf{u}(t)$, $\mathbf{w}(t)$, $\mathbf{z}(t)$, and $\mathbf{y}(t)$ denoting, respectively, the concatenated state vector, control vector, exogenous input vector, controlled vector, and output vector

of all the agents in the NDS. The bold faced matrices represent the block diagonal aggregation of each agent's state-space matrices, e.g., $\mathbf{A} = \text{diag}\{A_1, \dots, A_n\}$.

Given the above model for each agent and motivated by the diverse applications of multi-agent systems, we can begin to incorporate the role of the interconnection topology \mathcal{G} . To begin, we first define four canonical classes of NDS models. Such a classification is useful for analysis purposes; we will also show in the sequel that under certain conditions they are, in a sense, equivalent.

NDS Coupled at the Output

In this class of NDS, the underlying network topology couples each agent through their outputs. Systems relying on relative sensing to achieve global objectives such as formation flying fall under this classification [10, 17, 31]. The block diagram in Fig. 2 shows how the connection topology interacts with each agent. Here we have shown disturbances entering each agent and the global output of the entire system. An important feature of these types of systems is the underlying connection topology does not affect, in the open-loop, the dynamic behavior of each agent.

Motivated by applications that rely on relative sensing, we now derive a mathematical model to capture the global layer of this type of NDS. The sensed output of the system is the vector $\mathbf{y}_{\mathcal{G}}(t)$ containing relative state information of each agent and its neighbors. The incidence matrix of a graph naturally captures differences and will be the algebraic construct used to define the relative outputs. For example, the output sensed between agent i and agent j would be of the form $y_i(t) - y_j(t)$. This can be compactly written using the incidence matrix for the entire system as

$$\mathbf{y}_{\mathcal{G}}(t) = (E(\mathcal{G})^T \otimes I)\mathbf{y}(t). \quad (16)$$

Here, \mathcal{G} is the graph that describes the connection topology; the node set is given as $\mathcal{V} = \{1, \dots, n\}$.

When considering the analysis of the global layer, we are interested in studying the map from the agent's exogenous inputs, $\mathbf{w}(t)$, to the sensed output of the NDS, $\mathbf{y}_{\mathcal{G}}(t)$. Therefore, for the homogeneous NDS, the system Σ in (15) is augmented to include the sensed output

$$\mathbf{y}_{\mathcal{G}}(t) = (E(\mathcal{G})^T \otimes C^y)\mathbf{x}(t). \quad (17)$$

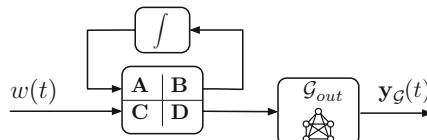


Fig. 2 NDS coupled at the output; the feedback connection represents an upper fractional transformation [8]

For the heterogeneous case, the sensed output is

$$\mathbf{y}_{\mathcal{G}}(t) = (E(\mathcal{G})^T \otimes I)\mathbf{C}^y \mathbf{x}(t). \quad (18)$$

Remark 9.1. For relative sensing, the observation matrix \mathbf{C}^y used in (17) and (18) may in fact be different from the local observation of each agent, as described in (14).

In the context of NDS coupled at the output, we denote by $\Sigma_{hom}(\mathcal{G})$ the homogeneous system (15) with the additional sensed output (17). The heterogeneous system will be denoted by $\Sigma_{het}(\mathcal{G})$ and corresponds to the system (15) with the additional sensed output (18).

For notational simplicity, we denote $T_{hom}^{w_i \rightarrow \mathcal{G}}$ and $T_{het}^{w_i \rightarrow \mathcal{G}}$ as the map from the exogenous inputs to the NDS sensed output for homogeneous and heterogeneous systems respectively.

NDS Coupled at the Input

In this class of NDS, the underlying network topology enters at the system input. The block diagram in Fig. 3 shows a networked input being distributed to each agent via an interconnection topology. Large physically coupled systems where actuation affects multiple components might be modeled in this way. In fact, this class of NDS may even be considered the “dual” of the output coupled NDS presented above. The agents are therefore coupled via the inputs. To maintain a close connection with the NDS coupled at the output model, we will assume the network input, $\mathbf{u}_{\mathcal{G}}(t)$, is distributed to each agent via the incidence matrix. The control applied to each agent, therefore, is the net contribution of the control applied to all the edges incident to that agent. When the underlying graph is directed and connected, each agent’s control can be written as

$$u_i(t) = \sum_{(i,j) \in \mathcal{E}} [\mathbf{u}_{\mathcal{G}}(t)]_{(i,j)} - \sum_{(j,i) \in \mathcal{E}} [\mathbf{u}_{\mathcal{G}}(t)]_{(j,i)}, \quad (19)$$

where $[\mathbf{u}_{\mathcal{G}}(t)]_{(i,j)}$ denotes the component of the input vector corresponding to the directed edge (i,j) . This can be compactly written using the incidence matrix and Kronecker products to obtain a complete model for NDS coupled at the input.

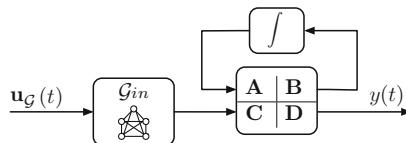


Fig. 3 NDS coupled at the input; the feedback connection represents an upper fractional transformation [8]

For the homogeneous case, we have

$$\mathbf{u}(t) = (E(\mathcal{G}) \otimes B)\mathbf{u}_{\mathcal{G}}(t), \quad (20)$$

and for the heterogeneous case

$$\mathbf{u}(t) = \mathbf{B}(E(\mathcal{G}) \otimes I)\mathbf{u}_{\mathcal{G}}(t). \quad (21)$$

The parallel interconnected system (15) can be modified to include the network input using (20) and (21) as

$$\Sigma_{hom}(\mathcal{G}) : \begin{cases} \dot{\mathbf{x}}(t) = (I_n \otimes A)\mathbf{x}(t) + (E(\mathcal{G}) \otimes B)\mathbf{u}_{\mathcal{G}}(t) + (I_n \otimes \Gamma)\mathbf{w}(t), \\ \mathbf{z}(t) = (I_n \otimes C^z)\mathbf{x}(t) + (I_n \otimes D^{zu})\mathbf{u}(t) + (I_n \otimes D^{zw})\mathbf{w}(t), \\ \mathbf{y}(t) = (I_n \otimes C^y)\mathbf{x}(t) + (I_n \otimes D^{yw})\mathbf{w}(t), \end{cases} \quad (22)$$

and

$$\Sigma_{het}(\mathcal{G}) : \begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}(E(\mathcal{G}) \otimes I)\mathbf{u}_{\mathcal{G}}(t) + \Gamma\mathbf{w}(t), \\ \mathbf{z}(t) = \mathbf{C}^z\mathbf{x}(t) + \mathbf{D}^{zu}\mathbf{u}(t) + \mathbf{D}^{zw}\mathbf{w}(t), \\ \mathbf{y}(t) = \mathbf{C}^y\mathbf{x}(t) + \mathbf{D}^{yw}\mathbf{w}(t). \end{cases} \quad (23)$$

NDS Coupled at the State

This class of NDS is perhaps one of the most studied in the systems and control community. In this type of NDS, the underlying topology couples each agent at the state level, resulting in an important connection between the dynamic evolution of each agent and the underlying topology. The block diagram in Fig. 4 shows the connection topology entering a dynamic system at the state level.

The most general way to model such systems is to simply denote the dependence of the state-matrix on the network with the notation $\mathbf{A}(\mathcal{G})$. For our purposes, however, we will focus on a special instance of this system, known as the agreement or consensus protocol [20, 23, 26]. Consequently, we will only focus on homogeneous systems for this case.

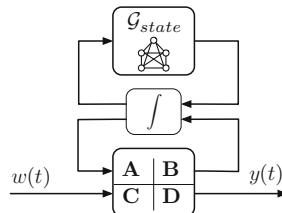


Fig. 4 NDS coupled at the state; the feedback connection between the plant matrices and the integrator represents an upper fractional transformation [8], whereas the feedback connection between the integrator and the graph represents a relation such as (26)

The consensus model is built upon a general setup consisting of a group of n identical single integrator units,

$$\dot{x}_i(t) = u_i(t), \quad i = 1, \dots, n, \quad (24)$$

each connected to a fixed number of other units in the ensemble, determined by the interconnection topology \mathcal{G} . The interaction or coupling between units' dynamics is realized through the control input $u_i(t)$ in (24), assumed to be the sum of the differences between states of an agent and its neighbors, i.e.,

$$u_i(t) = \sum_{j \sim i} (x_j(t) - x_i(t)). \quad (25)$$

Expressing the dynamic evolution of the resulting system in a compact matrix form one has

$$\dot{\mathbf{x}}(t) = -L(\mathcal{G}) \mathbf{x}(t), \quad (26)$$

where $L(\mathcal{G})$ is the graph Laplacian. This model can be extended to include exogenous inputs and controlled variables, which will be discussed in the sequel. Extensions of this model have been extensively treated, including random networks [13, 32], switching topologies [22], and noisy networks [36].

NDS Coupled by Combinations of State, Input, and Output

A natural extension of the above models is to consider systems that have a network coupling the agents at all component levels. Figure 5 shows a dynamic system where there are different connection topologies at the input, output, and state level. Clearly, this type of model represents the most complex and intricate connection between the dynamic properties and interconnection topology of a system. It is worth noting that although this type of system can be exhaustively studied on its own, we only present it here for completeness and as a vehicle to illustrate how each of the previous types can be interrelated.

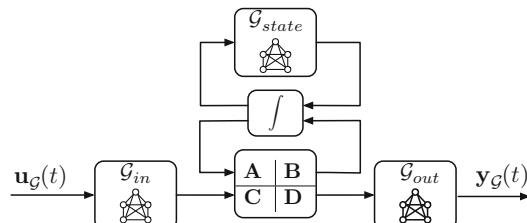


Fig. 5 NDS coupled at the state, input, and output; the feedback connection between the plant matrices and the integrator represents an upper fractional transformation [8], whereas the feedback connection between the integrator and the graph represents a relation such as (26)

3 Analysis and Graph-Theoretic Performance Bounds

The development of the NDS models in Sect. 2 allows us to examine systems-theoretic properties and performance bounds for networked system from a graph-theoretic perspective. The objective of this section, therefore, is to develop explicit connections between control-theoretic concepts such as observability, controllability, and performance in terms of the underlying interconnection graph. In this section we will first discuss the observability properties of NDS coupled at the output, and then proceed to highlight how duality streamlines the controllability analysis of NDS structure coupled at the input. We then focus on characterizing the \mathcal{H}_2 performance of NDS coupled at the output and at the state via constructs from algebraic graph theory.

3.1 Observability and Controllability of NDS

Studying the observability and controllability properties of a linear system can provide qualitative as well as quantitative insights into the design of the corresponding controllers and estimators. In the context of NDS, we also consider how the underlying topology affects these properties in addition to examining the effects of homogeneity and heterogeneity of the agent dynamics comprising the NDS. In this direction, we consider simplified versions of the models presented in Sect. 2. For example, for observability analysis, we consider the following simplified model of an NDS coupled at the output,

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) \\ \mathbf{y}_{\mathcal{G}}(t) = (\mathbf{E}(\mathcal{G})^T \otimes \mathbf{I})\mathbf{C}^y\mathbf{x}(t) \end{cases}. \quad (27)$$

Analogously, the following simplified model for an NDS coupled at the input will be used to study the controllability properties,

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}(\mathbf{E}(\mathcal{G}) \otimes \mathbf{I})\mathbf{u}_{\mathcal{G}}(t); \quad (28)$$

for both systems, we will examine the homogeneous and heterogeneous realizations.

Our observability and controllability analysis relies on the observability and controllability gramians for networked systems. Recall that the observability gramian of a stable linear system with state matrix A and observation matrix C can be written as

$$Y_o = \int_0^\infty e^{A^T t} C^T C e^{At} dt. \quad (29)$$

Similarly, the controllability gramian of a stable linear system with state matrix A and input matrix B can be written as

$$X_c = \int_0^\infty e^{At} B B^T e^{A^T t} dt. \quad (30)$$

For this analysis, we will assume that each agent is stable, e.g., A_i is Hurwitz, the pair (A_i, C_i^y) is observable, and (A_i, B_i) is a controllable pair.

Observability of NDS Coupled at the Output

A natural question for this analysis is whether *the initial condition of each agent in an NDS coupled at the output can be inferred from their relative states*. The answer to this question can have profound implications for the design of estimators for such systems.

For homogeneous NDS, the observability gramian can be written as

$$\mathbf{Y}_o = L(\mathcal{G}) \otimes \int_0^\infty e^{A^T t} (C^y)^T C^y e^{At} dt = L(\mathcal{G}) \otimes Y_o, \quad (31)$$

where Y_o is the gramian for an individual agent in the NDS.

Theorem 9.3. *The homogeneous NDS coupled at the output in (27) is unobservable.*

Proof. Using the gramian expression in (31) and Theorem 9.1 we conclude that \mathbf{Y}_o has precisely n eigenvalues at the origin, leading to an unobservable system. \square

The unobservable modes of (27), in fact, correspond to the inertial position of the entire formation; these modes lie in the subspace $\text{span}\{\mathbf{1} \otimes I\}$. The importance of this result is that when each agent has identical dynamics, relative measurements alone are insufficient to reconstruct their inertial states. If in addition to the relative output, an additional inertial measurement is available, say one that corresponds to the inertial position of a single agent, then the observability of the system can be recovered.

An interesting consequence of this result highlights how the underlying connection topology influences the relative degree of observability of the observable modes. We denote and index each singular value of Y_o as σ_i , and using the results of Theorem 9.1 we can express the non-zero singular values of \mathbf{Y}_o as $\lambda_j(\mathcal{G})\sigma_i$ for $j = 2, \dots, n$ and all i . The eigenvalues of the graph Laplacian, therefore, can amplify or attenuate the relative degree of observability of the system. For example, the complete graph as in Fig. 1(a), has $\lambda_i(\mathcal{G}) = \lambda_j(\mathcal{G}) = n$ for $i, j \geq 2$. In this case, the connection topology does not favor any particular modes of the system as each is scaled by the same amount. Conversely, when the graph is disconnected with two connected components, then $\lambda_2(\mathcal{G}) = 0$ and n additional unobservable modes are introduced into the system.

In the heterogeneous case, the observability gramian of (27) has a non-trivial form. We define the observability operator for an individual agent as $\Psi_i(x) = C_i^y e^{A_i^T t} x$, and its adjoint as $\Psi_i^*(y(t)) = \int_0^\infty e^{A_i^T t} (C_i^y)^T y(t) dt$ [8]. The observability gramian of (27) can be written as

$$\mathbf{Y}_o = \mathbf{diag}\{\Psi^*\}(L(\mathcal{G}) \otimes I)\mathbf{diag}\{\Psi\} = (L(\mathcal{G}) \otimes I) \circ \Psi^* \Psi, \quad (32)$$

where $\Psi = [\Psi_1 \dots \Psi_n]$. The derivation of (32) can be found in [37].

Theorem 9.4. *The heterogeneous NDS coupled at the output in (27) is unobservable if and only if the following conditions are met:*

1. *There exists an eigenvalue, μ^* , of \mathbf{A} that is common to each A_i , and*
2. *One has $C_i^y q_i = C_j^y q_j$ for all i, j with $A_i q_i = \mu^* q_i$ for all i .*

Proof. The necessary condition is verified when all agents have identical dynamics, as shown in Theorem 9.3. For the sufficient condition, assume that there exists μ^* that is an eigenvalue for each A_i . We can then construct an eigenvector for \mathbf{A} as $q = [q_1^T \cdots q_n^T]^T$, with $A_i q_i = \mu^* q_i$. By condition 2, we have that $\mathbf{C}q = \mathbf{1} \otimes r$, where $r = C_i q_i \neq 0$ for all i . Using properties of the Kronecker product we then have

$$(E(\mathcal{G})^T \otimes I)\mathbf{C}^y q = (E(\mathcal{G})^T \otimes I)(\mathbf{1} \otimes r) = (E(\mathcal{G})^T \mathbf{1} \otimes r) = 0. \quad (33)$$

This shows the system is unobservable with q the corresponding unobservable mode. \square

Theorem 9.4 shows that a heterogeneous NDS becomes unobservable only when the outputs of each agent associated with a certain initial condition direction becomes indistinguishable. For general heterogeneous NDS – therefore – the system is ‘expected’ to be observable. This is a rather non-trivial, as it suggests that the inertial position of each agent can be reconstructed solely from relative measurements.

As in the homogeneous case, the underlying connection topology can have a profound affect on the relative degree of observability of the system. The form of (32) is appealing in how it separates the role of the network from each agent. Although the precise characterization of the eigenvalues of (32) is non-trivial, bounds on those values can be derived, as presented in [15]. In particular, since both terms in the Hadamard product are positive semi-definite matrices, we can apply Schur’s Theorem to obtain the bound,

$$\underline{\sigma}(\Psi^* \Psi) \leq \underline{\sigma}(\mathbf{Y}_o) \leq \overline{\sigma}(\mathbf{Y}_o) \leq \bar{d} \overline{\sigma}(\Psi^* \Psi), \quad (34)$$

where $\underline{\sigma}(\mathbf{Y}_o)$ and $\overline{\sigma}(\mathbf{Y}_o)$ correspond, respectively, to the smallest and largest singular values of \mathbf{Y}_o , and

$$\underline{d} = \min_i [L(\mathcal{G}) \otimes J_n]_{ii}, \quad \bar{d} = \max_i [L(\mathcal{G}) \otimes J_n]_{ii};$$

the quantities \underline{d} and \bar{d} correspond, respectively, to the minimum and maximum degree vertices of the underlying graph. We note that the bounds (34) become tight when agent have homogeneous dynamics. Such observations point to interesting connections between the degree of each agent in the ensemble and the relative observability of the modes of the system. This theme will be revisited when we study the \mathcal{H}_2 performance of heterogeneous NDS coupled at the output.

Controllability of NDS Coupled at the Input

In this section, we consider whether *from any initial condition, each agent can be driven to an arbitrary final state when the input is distributed to each agent through a network*. For homogeneous NDS, the controllability gramian can be written by inspection as

$$\mathbf{X}_c = L(\mathcal{G}) \otimes \int_0^{\infty} e^{At} BB^T e^{A^T t} dt = L(\mathcal{G}) \otimes X_c, \quad (35)$$

where X_c is the controllability gramian for an individual agent in the NDS.

Theorem 9.5. *The homogeneous NDS coupled at the input (28) is uncontrollable.*

Proof. The gramian (35) has precisely n eigenvalues at the origin, leading to an uncontrollable system. \square

Here, the dual nature of the NDS coupled at the output and input becomes immediately apparent. As in the former case, the uncontrollable mode corresponds to the inertial position of the entire ensemble, lying in the subspace $\text{span}\{\mathbf{1} \otimes I\}$. We also note that the relative degree of controllability can be inferred from the gramian, but this analysis is omitted as it mirrors that of the observability analysis for NDS coupled at the output.

For the heterogeneous case, we arrive at the following result.

Theorem 9.6. *The heterogeneous NDS coupled at the input (28) is uncontrollable if and only if the following conditions are met:*

1. *There exists an eigenvalue, μ^* , of \mathbf{A} that is common to each A_i , and*
2. *One has $q_i^T B_i = q_j^T B_j$ for all i, j with $q_i^T A_i = \mu^* q_i^T$ for all i .*

The proof of Theorem 9.6 follows the same procedure as for Theorem 9.5, and is omitted. The conclusion, as expected, shows that heterogeneity in the dynamics of each agent can lead to a fully controllable system.

The dual structure between the NDS coupled at the input and coupled at the output should now be clear. A simple exercise will show that the controllability gramian for (28) has a similar form to the observability gramian (32), with the role of Ψ^* replaced with the controllability operator. It becomes apparent that the degree of each agent in the ensemble can have a profound affect on the overall controllability properties of the system.

3.2 Graph-Theoretic Bounds on NDS Performance

In this section we explore a graph-theoretic characterization of the \mathcal{H}_2 performance of different NDS models. The main goal is to again make explicit the role of the underlying connection topology on the system performance norms. We will assume throughout this section that the underlying connection graph \mathcal{G} is connected. For analysis, we also assume each agent is stable.

\mathcal{H}_2 Performance of NDS Coupled at the Output

The sensed output of an NDS coupled at the output can be used to achieve a variety of objectives, including localization- for which the observability results of §3.1 apply- and formation control. It is important therefore to examine how noise entering the dynamics of each agent in the NDS propagates through the network to the sensed output. A natural measure for quantifying this property is the \mathcal{H}_2 system norm. This section, therefore, aims to explicitly characterize the affect of the network on the \mathcal{H}_2 norm of the system. For this analysis, we assume that each agent is driven by a Gaussian white noise. A simplified version of the agent dynamics is given as

$$\Sigma_i : \begin{cases} \dot{x}_i(t) = A_i x_i(t) + \Gamma_i w_i(t) \\ y_i(t) = C_i^y x_i(t). \end{cases} \quad (36)$$

The corresponding model for the NDS coupled at the output takes the form

$$\Sigma_{het}(\mathcal{G}) : \begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \Gamma\mathbf{w}(t) \\ \mathbf{y}(t) = \mathbf{C}^y\mathbf{x}(t) \\ \mathbf{y}_{\mathcal{G}}(t) = (E(\mathcal{G})^T \otimes I)\mathbf{C}^y\mathbf{x}(t). \end{cases} \quad (37)$$

The \mathcal{H}_2 -norm of a system can be calculated using the controllability and observability gramians of the system, as discussed in §3.1. The \mathcal{H}_2 norm of each agent from the exogenous input channel to the measured output can be expressed in terms of the gramians as

$$\|\Sigma_i\|_2 = \sqrt{\text{trace}(\Gamma_i^T Y_o^i \Gamma_i)} \quad (38)$$

$$= \sqrt{\text{trace}(C_i^y X_c^i (C_i^y)^T)}, \quad (39)$$

where Y_o^i and X_c^i denote, respectively, the observability gramian and controllability gramian of agent i .

Theorem 9.7. *The \mathcal{H}_2 norm of the homogeneous NDS coupled at the output (37) is given by*

$$\left\| T_{hom}^{w \rightarrow \mathcal{G}} \right\|_2 = \|E(\mathcal{G})\|_F \|\Sigma\|_2. \quad (40)$$

Proof. The \mathcal{H}_2 norm can be written directly from (31) as

$$\left\| T_{hom}^{w \rightarrow \mathcal{G}} \right\|_2 = \sqrt{\text{trace}((I_n \otimes \Gamma)^T (L(\mathcal{G}) \otimes Y_o)(I_n \otimes \Gamma)).}$$

Using the properties of the Kronecker product defined in §1.1 and the definition of the Frobenius norm of a matrix, $\|M\|_F = \sqrt{\text{trace}(M^T M)}$, leads to the expression in (40). \square

The expression in (40) gives an explicit characterization of how the network affects the the system performance. For homogeneous systems, we find that the \mathcal{H}_2

performance changes with the addition or removal of an edge. Recall that the Frobenius norm of a matrix can be expressed in terms of the 2-norm of each column, as

$$\|M\|_F = \left(\sum_{i=1}^n \|m_i\|_2^2 \right)^{1/2},$$

where m_i is the i th column of the matrix M . As each column of $E(\mathcal{G})$ represents an edge in \mathcal{G} , the Frobenius norm can be expressed in terms of the number of edges in the graph, $|\mathcal{E}|$, as

$$\|E(\mathcal{G})\|_F = (2|\mathcal{E}|)^{1/2}. \quad (41)$$

This highlights the importance of *the number of edges* as opposed to the actual structure of the graph (e.g., a star graph or k -regular graph). This makes intuitive sense, as more edges would correspond to additional amplification of the disturbances entering the system.

If we consider only connected graphs, we arrive at the following corollaries providing lower and upper bounds on the \mathcal{H}_2 norm of the system.

Corollary 9.1. *The \mathcal{H}_2 norm of the homogeneous NDS coupled at the output (37) for an arbitrary connected graph \mathcal{G} is lower bounded by an NDS where \mathcal{G} is a spanning tree, as*

$$\left\| T_{hom}^{w \mapsto \mathcal{G}} \right\|_2^2 \geq 2 \|\Sigma\|_2^2 (n-1); \quad (42)$$

the lower bound is attained with equality whenever the underlying graph is a spanning tree.

It is clear from the definition of the Frobenius norm that the choice of tree is irrelevant (e.g., a star or a path).

Corollary 9.2. *The \mathcal{H}_2 norm of the homogeneous NDS coupled at the output (37) for an arbitrary connected graph \mathcal{G} is upper bounded by an NDS where $\mathcal{G} = K_n$, the complete graph, as*

$$\left\| T_{hom}^{w \mapsto \mathcal{G}} \right\|_2^2 \leq 2 \|\Sigma\|_2^2 n (n-1); \quad (43)$$

the upper bound is attained with equality whenever the underlying graph is complete.

For the heterogeneous case we rely on (39) to derive the \mathcal{H}_2 norm. The connection topology only couples agents at the output leading to a block diagonal description for the controllability gramian, with each block corresponding to each agent's controllability gramian.

Theorem 9.8. *The \mathcal{H}_2 norm of the heterogeneous NDS coupled at the output (37) is given as*

$$\left\| T_{het}^{w \mapsto \mathcal{G}} \right\|_2 = \left(\sum_i d_i \|\Sigma_i\|_2^2 \right)^{1/2}, \quad (44)$$

where d_i is the degree of the i th agent in the graph.

Proof. The norm expression in (44) can be derived using (39) as,

$$\left\| T_{het}^{w \rightarrow \mathcal{G}} \right\|_2 = (\text{trace}\{(E(\mathcal{G})^T \otimes I) \mathbf{C}^y \mathbf{X}_c (\mathbf{C}^y)^T (E(\mathcal{G}) \otimes I)\})^{1/2}, \quad (45)$$

where \mathbf{X}_c denotes the block diagonal aggregation of each agent's controllability gramian. First, we make the following observation,

$$\text{trace}\{\mathbf{C}^y \mathbf{X}_c (\mathbf{C}^y)^T\} = \sum_{i=1}^n \|\Sigma_i\|_2^2.$$

Using the cyclic property of the trace operator [41] and exploiting the block diagonal structure of the argument leads to the following identity simplification,

$$\begin{aligned} \text{trace}\{\mathbf{C}^y \mathbf{X}_c (\mathbf{C}^y)^T ((\Delta(\mathcal{G}) - A(\mathcal{G})) \otimes I)\} &= \sum_i \text{trace}\{C_i^y X_c^i (C_i^y)^T (d_i \otimes I)\} \\ &= \sum_i d_i \|\Sigma_i\|_2^2. \end{aligned} \quad (46)$$

This leads to the desired result. \square

A further examination of (44) reveals that it can be written as the Frobenius norm of a node-weighted incidence matrix,

$$\left\| T_{het}^{w \rightarrow \mathcal{G}} \right\|_2 = \left\| \begin{bmatrix} \|\Sigma_1\|_2 & & \\ & \ddots & \\ & & \|\Sigma_n\|_2 \end{bmatrix} E(\mathcal{G}) \right\|_F. \quad (47)$$

When each agent has the same dynamics, (47) reduces to the expression in (40). This characterization paints a clear picture of how the placement of an agent within a certain topology affects the overall system gain. In order to minimize the gain, it is beneficial to keep systems with high norm in locations with minimum degree.

For certain graph structures, a more explicit characterization of the \mathcal{H}_2 performance can be derived, leading to the following corollaries.

Corollary 9.3. *The \mathcal{H}_2 norm of the heterogeneous NDS coupled at the output (37) when the underlying connection graph is k -regular is*

$$\left\| T_{het}^{w \rightarrow \mathcal{G}} \right\|_2 = \left(k \sum_i \|\Sigma_i\|_2^2 \right)^{1/2}, \quad (48)$$

where every node has degree k .

Note that having regularity in the connection topology introduces homogeneity into the heterogeneous NDS. As in the homogeneous case, the placement of an agent in the network will not affect the overall performance. The system norm for this topology becomes a scaled version of the parallel connection of the n sub-systems.

\mathcal{H}_2 Performance of NDS Coupled at the State

The consensus model derived in Sect. 2 is an important model for an increasingly wide range of applications. In the literature there are many variations of the model (26) including its nonlinear extensions and those evolving over random and switching graph topologies. The common theme in all these scenarios is the focus on the convergence of the system's state to the agreement subspace when arbitrarily initialized.

In this section, we like to consider a more system-theoretic approach to the analysis of consensus seeking systems. To this end, we will work with a consensus model that is corrupted by noise at the process and the corresponding measurements. We will then characterize the \mathcal{H}_2 performance of the model, which can be used to *reason about how noise in the network result in the asymptotic deviation of each node's state from the consensus state.*

For the derivation of the consensus model with noise we highlight again the relationship of the different type of NDS models derived in Sect. 2. The consensus model with noise can be considered as an NDS with coupling at both the input and output with a unity feedback applied to close the loop, as shown in Fig. 6.

For this model, we will combine the NDS coupled at the input and output for homogeneous agent dynamics, each described by a single integrator dynamics. The process noise entering each agent, $w_i(t)$, is assumed to be a Gaussian white noise with covariance

$$\mathbf{E}[w_i(t)w_i(t)^T] = \sigma_w^2 I \quad \text{and} \quad \mathbf{E}[w_i(t)w_j(t)^T] = 0 \text{ when } i \neq j.$$

The NDS measurement, which is the sensed relative measurement between neighboring agents, is also corrupted by a Gaussian white noise $v(t)$ with covariance $\mathbf{E}[v(t)v(t)^T] = \sigma_v^2 I$. The open-loop model, therefore, can be written as

$$\begin{cases} \dot{\mathbf{x}}(t) = E(\mathcal{G})\mathbf{u}_{\mathcal{G}}(t) + \mathbf{w}(t) \\ \mathbf{y}_{\mathcal{G}}(t) = E(\mathcal{G})^T \mathbf{x}(t) + v(t) \end{cases}. \quad (49)$$

When the output-feedback control $\mathbf{u}_{\mathcal{G}}(t) = -\mathbf{y}_{\mathcal{G}}(t)$ is applied, we obtain a closed-loop consensus system driven by noise. To complete the input-output description of the system, we recall that in consensus seeking systems, the objective is for each

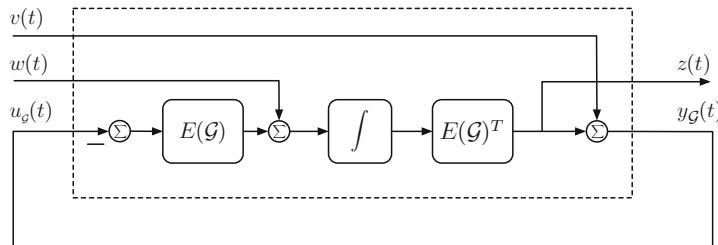


Fig. 6 Consensus as an NDS with coupling at the input and output

state to agree on a common value. Therefore, we include the performance variable $z(t)$ that captures the difference between states.

$$\Sigma(\mathcal{G}) : \begin{cases} \dot{\mathbf{x}}(t) = -L(\mathcal{G})\mathbf{x}(t) + [I - E(\mathcal{G})] \begin{bmatrix} \mathbf{w}(t) \\ v(t) \end{bmatrix}, \\ z(t) = E(\mathcal{G})^T \mathbf{x}(t) \end{cases}. \quad (50)$$

Note that in the absence of noise, the system is driven by its initial conditions and the original consensus model derived in §2 is recovered. More interestingly, an input-output description of the consensus model can be considered as an NDS with coupling at the input, output, and state.

The final hurdle to overcome in an \mathcal{H}_2 performance analysis of (50) is that the system state matrix, $-L(\mathcal{G})$, has an eigenvalue at the origin, which will lead to an unbounded \mathcal{H}_2 norm. The observability and controllability results of Sect. 3.1, however, can immediately be applied to conclude that (50) is neither controllable nor observable. The uncontrollable and unobservable modes are precisely the modes that lie in the agreement subspace, $\text{span}\{\mathbf{1}\}$. Therefore, we can consider our analysis on a minimal representation of (50) instead. Theorem 9.2 can be applied as a coordinate transformation on (50) that has the desired feature of separating the uncontrollable and unobservable modes while maintaining an algebraic representation of the underlying connection topology via the edge Laplacian (see Sect. 1.1). In this direction, we introduce the coordinate transformation $S_v x_e(t) = x(t)$, where S_v is defined in (12). The minimal system representation can therefore be expressed as

$$\Sigma_\tau : \begin{cases} \dot{x}_\tau(t) = -L_e(\mathcal{G}_\tau)R(\mathcal{G})R(\mathcal{G})^T x_\tau(t) + \sigma_w E(\mathcal{G}_\tau)^T \hat{w}(t) \\ \quad -\sigma_v L_e(\mathcal{G}_\tau)R(\mathcal{G})\hat{v}(t) \\ z(t) = R(\mathcal{G})^T x_\tau(t); \end{cases} \quad (51)$$

the signals $\hat{w}(t)$ and $\hat{v}(t)$ are the normalized process and measurement noise signals. We also note that $x_\tau(t)$ corresponds to the states on the edges of the spanning tree sub-graph \mathcal{G}_τ . The performance variable, $z(t)$, contains information on the tree states in addition to the cycle states. Here we recall that the cycle states are a linear combination of the tree states and we note that $z(t)$ actually contains redundant information. This is highlighted by recognizing that the tree states converging to the origin forces the cycle states to do the same. Consequently, we will consider the system with cycles as well as a system containing only the tree states at the output, which we denote as $\hat{\Sigma}_\tau$,

$$\hat{\Sigma}_\tau : \begin{cases} \dot{x}_\tau(t) = -L_e(\mathcal{G}_\tau)R(\mathcal{G})R(\mathcal{G})^T x_\tau(t) + \sigma_w E(\mathcal{G}_\tau)^T \hat{w}(t) \\ \quad -\sigma_v L_e(\mathcal{G}_\tau)R(\mathcal{G})\hat{v}(t) \\ z(t) = x_\tau(t). \end{cases} \quad (52)$$

This distinction will subsequently be employed to quantify the effect of cycles on the system performance.

The \mathcal{H}_2 norm of Σ_τ and $\hat{\Sigma}_\tau$ can be calculated using the controllability gramian as,

$$\|\Sigma_\tau\|_2^2 = \text{trace}[R^T X^* R], \quad \text{and} \quad \|\hat{\Sigma}_\tau\|_2^2 = \text{trace}[X^*], \quad (53)$$

where R is defined in (8) and X^* is the positive-definite solution to the Lyapunov equation

$$-L_e(\mathcal{G}_\tau)RR^TX - XRR^T L_e(\mathcal{G}_\tau) + \sigma_w^2 L_e(\mathcal{G}_\tau) + \sigma_v^2 L_e(\mathcal{G}_\tau)RR^T L_e(\mathcal{G}_\tau) = 0. \quad (54)$$

The structure of (54) suggests that the solution will be dependent on certain properties of the graph. In fact, the solution can be found by inspection by first noting that

$$\sigma_w^2 L_e(\mathcal{G}_\tau) + \sigma_v^2 L_e(\mathcal{G}_\tau) RR^T L_e(\mathcal{G}_\tau) = L_e(\mathcal{G}_\tau) (\sigma_w^2 (L_e(\mathcal{G}_\tau))^{-1} + \sigma_v^2 RR^T) L_e(\mathcal{G}_\tau).$$

The solution to (54) is therefore

$$X^* = \frac{1}{2} (\sigma_w^2 (R R^T)^{-1} + \sigma_v^2 L_e(\mathcal{G}_\tau)), \quad (55)$$

and we arrive at the following result.

Theorem 9.9. *The \mathcal{H}_2 norm of the system Σ_τ (51) is*

$$\|\Sigma_\tau\|_2^2 = \frac{\sigma_w^2}{2} (n-1) + \sigma_v^2 |\mathcal{E}|. \quad (56)$$

On the other hand, the \mathcal{H}_2 norm of the $\hat{\Sigma}_\tau$ system (52) is

$$\|\hat{\Sigma}_\tau\|_2^2 = \frac{\sigma_w^2}{2} \text{trace}[(R R^T)^{-1}] + \sigma_v^2 (n-1). \quad (57)$$

Proof. The proof follows from (55) and noting that $\text{trace}[L_e(\mathcal{G}_\tau)] = 2(n-1)$, or twice the number of edges in a spanning tree. \square

We observe that $\|\Sigma_\tau\|_2^2$ is a linear function of the number of edges in the graph. This has a clear practical relevance, as it indicates that the addition of each edge corresponds to an amplification of the noise in the consensus-type network. Let us consider the implications of the graph-theoretic characterization of the \mathcal{H}_2 norm for two classes of graphs.

(a) **Spanning Trees:** The first case resulting in a simplification of (56) arises when \mathcal{G} is a spanning tree. In this case $R = I$ and (57) simplifies to

$$\|\hat{\Sigma}_\tau\|_2^2 = (n-1) \left(\frac{\sigma_w^2}{2} + \sigma_v^2 \right). \quad (58)$$

A direct consequence of this result is that *all* spanning trees result in the same \mathcal{H}_2 system performance. That is, the choice of spanning tree (e.g., a path or

a star) does not affect this performance metric. As expected, in this scenario $\|\Sigma_\tau\|_2^2 = \|\hat{\Sigma}_\tau\|_2^2$.

- (b) **k -Regular Graphs:** Regular graphs also lead to a simplification of (57). In general, any connected k -regular graph will contain cycles resulting in a non-trivial expression for matrix product RR^T . The \mathcal{H}_2 norm is therefore intimately related to the cut space of the graph.

Denote the eigenvalues of RR^T by μ_i and note that

$$\text{trace}[(RR^T)^{-1}] = \sum_{i=1}^{n-1} \frac{1}{\mu_i} = \frac{1}{\tau(\mathcal{G})} \sum_{i=1}^{n-1} \prod_{j \neq i}^{n-1} \mu_j, \quad (59)$$

where $\tau(\mathcal{G})$ is the number of spanning trees in \mathcal{G} . The quantity $\prod_{j \neq i}^{n-1} \mu_j$ is recognized as a first minor of the matrix RR^T .

Corollary 9.4. *The cycle graph C_n has n spanning trees and hence*

$$\text{trace}[(R(C_n)R(C_n)^T)^{-1}] = \frac{(n-1)^2}{n}. \quad (60)$$

Thereby, the \mathcal{H}_2 norm of the $\hat{\Sigma}_\tau$ system when the underlying graph is the cycle graph C_n is given as

$$\|\hat{\Sigma}_\tau\|_2^2 = (n-1) \left(\frac{\sigma_w^2(n-1)}{n} + \sigma_v^2 \right). \quad (61)$$

Proof. Without loss of generality, we consider a directed path graph on n nodes, with initial node v_1 and terminal node v_n as the spanning tree subgraph \mathcal{G}_τ . Index the edges as $e_i = (v_i, v_{i+1})$. The cycle graph is formed by adding the edge $e_n = (v_n, v_1)$. For this graph, we have $T_\tau^c = \mathbf{1}_{n-1}$ and $R(C_n)R(C_n)^T = I + \mathbf{J}$. From this it follows that $\det[R(C_n)R(C_n)^T] = n$ and all its first minors have value $n-1$. Combined with (56) yields the desired result. \square

Corollary 9.5. *The complete graph K_n has n^{n-2} spanning trees, and therefore*

$$\text{trace}[(R(K_n)R(K_n)^T)^{-1}] = \frac{2(n-1)n^{n-3}}{n^{n-2}} = \frac{2(n-1)}{n}. \quad (62)$$

Thereby, the \mathcal{H}_2 norm of the $\hat{\Sigma}_\tau$ system when the underlying graph is the complete graph K_n is given as

$$\|\hat{\Sigma}_\tau\|_2^2 = (n-1) \left(\frac{\sigma_w^2}{n} + \sigma_v^2 \right). \quad (63)$$

Proof. Without loss of generality, we consider a star graph with center at node v_1 and all edges are of the form $e_k = (v_1, v_{k+1})$. Then the cycles in the graph are created by adding the edges $e = (v_i, v_j)$, $i, j \neq 1$ and $R(K_n)R(K_n)^T = nI - \mathbf{J}$. It then follows that

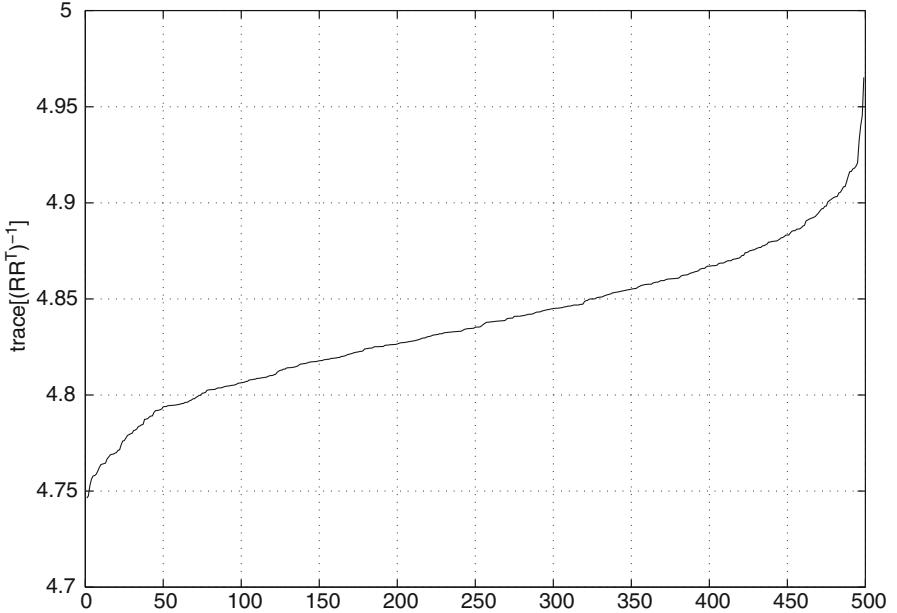


Fig. 7 $\text{trace}[(R(\mathcal{G})R(\mathcal{G})^T)^{-1}]$ for random 5-regular graphs

$$\det [R(K_n)R(K_n)^T] = n^{n-2}$$

and all the first minors have value $2n^{n-3}$. Combined with (56) yields the desired result. \square

Figure 7 depicts the sorted values of $\text{trace}[(RR^T)^{-1}]$ for 500 randomly generated regular graphs of degree five. As this figure shows, although the degree of each node remains constant, the actual cycle structure of each graph instance varies, effecting the resulting \mathcal{H}_2 norm of the corresponding consensus-type input-output system.

Using the above analysis, we now proceed to characterize how the cycle structure of the graph effects the \mathcal{H}_2 performance for the corresponding consensus-type system. In fact, examining the ratio

$$\frac{\|\Sigma_\tau(\mathcal{G})\|_2^2}{\|\Sigma_{\tau}(\mathcal{G}_\tau)\|_2^2}$$

provides an indication of how the cycles increase the \mathcal{H}_2 norm; recall that \mathcal{G} is in general a graph containing cycles and $\mathcal{G}_\tau \subseteq \mathcal{G}$ is the spanning tree subgraph. For example, consider the cycle graph C_n and assume unit covariance for both the process and measurement noise. Then, as the number of nodes increase, the ratio of the two \mathcal{H}_2 norms behaves as

$$\lim_{n \rightarrow \infty} \frac{\|\Sigma_\tau(C_n)\|_2^2}{\|\Sigma_\tau(P_n)\|_2^2} = \lim_{n \rightarrow \infty} \frac{4n-2}{3n} = \frac{4}{3} \quad (64)$$

indicating that for large cycles, the \mathcal{H}_2 performance is a constant multiple of the \mathcal{H}_2 performance for the path graph P_n .

In the meantime, for the complete graph K_n we have

$$\frac{\|\Sigma_\tau(K_n)\|_2^2}{\|\Sigma_\tau(\mathcal{G}_\tau)\|_2^2} = \frac{n+1}{3}; \quad (65)$$

in this case, we see that the norm is amplified linearly as a function of the number of vertices in the graph. It is worth mentioning here that typical performance measures for consensus problems, such as $\lambda_2(\mathcal{G})$, would favor the complete graph over the cycle graph. However, in terms of the \mathcal{H}_2 performance, we see that there is a penalty to be paid for faster convergence offered by the complete graph due to its cycle structure.

Alternatively, insight is also gained by considering the ratio

$$\frac{\|\Sigma_\tau(\mathcal{G})\|_2^2}{\|\hat{\Sigma}_\tau(\mathcal{G})\|_2^2},$$

which highlights the effects of including cycles in the performance variable $z(t)$. For the cycle graph we have

$$\lim_{n \rightarrow \infty} \frac{\|\Sigma_\tau(C_n)\|_2^2}{\|\hat{\Sigma}_\tau(C_n)\|_2^2} = \lim_{n \rightarrow \infty} \frac{n(3n-1)}{2(n-1)(2n-1)} = \frac{3}{4}, \quad (66)$$

suggesting that the effect of including the cycle for performance does not vary significantly with the size of the graph.

For the complete graph, on the other hand, one has

$$\frac{\|\Sigma_\tau(K_n)\|_2^2}{\|\hat{\Sigma}_\tau(K_n)\|_2^2} = \frac{n}{2}, \quad (67)$$

suggesting that the inclusion of cycles results in \mathcal{H}_2 performance that increases linearly as a function of vertices in the graph.

4 Topology Design for NDS

The analysis results of §3 points to the importance of the underlying interconnection topology on the overall performance of NDS. In fact, these results can be used to motivate *network synthesis* problems for NDS. For multi-agent systems, in addition to designing control and estimation algorithms for each agent in the ensemble, it should be a primary objective of the designer to also consider what the underlying

connection topology should look like. In its most general form, therefore, we would like to consider the problem of finding the underlying connection topology \mathcal{G} , such that the resulting NDS, $\Sigma(\mathcal{G})$, achieves a specified performance. This problem can be stated formally as

$$\begin{aligned} \min_{\mathcal{G}} \quad & \|\Sigma(\mathcal{G})\|_p \\ \text{s.t. } & \mathcal{G} \text{ is connected.} \end{aligned} \tag{68}$$

The challenges associated with this problem is to find numerically tractable algorithms for (68). The difficulty arises from the combinatorial nature of (68); the decision to allow agent i and j to be connected is binary.

In this section we present a solution to (68) for NDS coupled at the output for the \mathcal{H}_2 norm. Our results point to an intriguing connection between results in combinatorial optimization and systems theory. We also present a variation of (68) related to sensor placement for NDS coupled at the state.

4.1 \mathcal{H}_2 Topology Design for NDS Coupled at the Output

We now present a polynomial time algorithm for the design of the interconnection topology for heterogeneous agent dynamics in an NDS that is coupled at the output. Recall from Sect. 3.2 that in terms of the \mathcal{H}_2 norm objective, an optimal topology should always correspond to a spanning tree. Hence, the design problem is to determine which spanning tree achieves the smallest \mathcal{H}_2 norm for the NDS. Note that the design of the topology reduces to the design of the incidence matrix, $E(\mathcal{G})$. This problem is combinatorial in nature, as there are only a finite number of graphs that can be constructed from a set of n vertices. As the number of agents in the NDS grows, solving this problem becomes prohibitively hard [16]. However, as we will shortly show, with an appropriate modification of the problem statement, a celebrated algorithm in combinatorial optimization can be used for solving the topology design problem in polynomial time. Specifically, we will show that the *minimum spanning tree* (MST) problem captures the essential features of this problem. The MST can be solved using Kruskal's algorithm in $\mathcal{O}(|\mathcal{E}| \log(|\mathcal{V}|))$ time. The algorithm is given below and a proof of its correctness can be found, for example, in [16].

In order to apply the MST to the \mathcal{H}_2 synthesis problem we must reformulate the original problem statement. To begin, we first write the expression for the \mathcal{H}_2 norm of the system in (37) as

$$\begin{aligned} \|T_{het}^{w \mapsto \mathcal{G}}\|_2^2 &= \sum_i^n d_i \mathbf{trace}\{C^y X_i (C^y)^T\} \\ &= \sum_i^n d_i \|T_i^{w \mapsto y}\|_2^2, \end{aligned} \tag{69}$$

Data: A connected undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ and weights $w : \mathcal{E} \mapsto \mathbb{R}$.

Result: A spanning tree \mathcal{G}_t of minimum weight.

begin

 Sort the edges such that $w(e_1) \leq w(e_2) \leq \dots \leq w(e_{|\mathcal{E}|})$, where $e_i \in \mathcal{E}$

 Set $\mathcal{G}_t := \mathcal{G}_t(\mathcal{V}, \emptyset)$

for $i := 1$ **to** $|\mathcal{E}|$ **do**

if $\mathcal{G}_t + e_i$ *contains no cycle* **then**

 set $\mathcal{G}_t := \mathcal{G}_t + e_i$

end

end

end

Algorithm 1: Kruskal's Algorithm

where $T_i^{w_i \rightarrow y}$ is the map from the exogenous input entering agent i to its position, $C^y x_i(t)$. We reiterate here that the NDS norm description is related to the degree of each node in the network. Using the weighted incidence graph interpretation of the norm, as in (47), we see that the gain of each agent, $\|T_i^{w_i \rightarrow y}\|_2^2$, acts as a weight on the nodes. As each agent is assumed to have fixed dynamics, the problem of minimizing the NDS \mathcal{H}_2 norm reduces to finding the degree of each agent while ensuring the resulting topology is a spanning tree. This objective is related to properties of the nodes of the graph. In order to use the MST framework, we must convert the objective from weights on the nodes to weights on the edges.

To develop this transformation, consider the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with fixed weights w_i on each node $i = 1, \dots, n$. The node-weighted Frobenius norm of the incidence matrix is then

$$\|WE(\mathcal{G})\|_F^2 = \sum_i d_i w_i^2, \quad (70)$$

where $W = \text{diag}(w_1, \dots, w_n)$.

Next, consider the effect of adding an edge $\hat{e} = (i, j)$ to \mathcal{E} in terms of the Frobenius norm of the augmented incidence matrix,

$$\|W [E(\mathcal{G}) \hat{e}] \|_F^2 = \left(\sum_k d_k w_k^2 \right) + w_i^2 + w_j^2, \quad (71)$$

where d_k represents the degree of node k *before* adding the new edge \hat{e} . This shows that each edge $\hat{e} = (i, j)$ contributes $(w_i^2 + w_j^2)$ to the overall norm. Therefore, weights on the edges can be constructed by adding the node weights corresponding to the nodes adjacent to each edge as

$$\mathbf{w}_e = |E(\mathcal{G})^T| \mathbf{w}_n^2. \quad (72)$$

Using the above transformation from node weights to edge weights, we arrive at the following result.

Theorem 9.10. *The connection topology that minimizes the \mathcal{H}_2 norm of (37), can be found using Kruskal's MST algorithm with input data \mathcal{G} , and weights*

$$w = |E(\mathcal{G})|^T \begin{bmatrix} \|T_1^{w \mapsto \mathcal{G}}\|_2^2 \\ \vdots \\ \|T_n^{w \mapsto \mathcal{G}}\|_2^2 \end{bmatrix}. \quad (73)$$

Proof. The proof follows from (69) and the transformation from node weights to edge weights described in (70)–(72). \square

Remark 9.2. The choice of the input graph \mathcal{G} may be application specific, and can capture certain communication or sensing constraints between agents. For example, one may consider a scenario where agents are initially randomly distributed (a geometric random graph) upon deployment and can only sense neighboring agents within a specified range. The results of Theorem 9.10 can be used to determine the optimal spanning tree for that initial configuration.

Remark 9.3. There are a number of distributed algorithms that solve the MST problem [3, 11]. These could be used in place of the centralized version when the optimal spanning tree topology needs to be reconfigured. This scenario can arise due to the initialization problem discussed in Remark 9.2, or in situations when certain agents are disabled, lost, or reallocated for different mission purposes.

If there are no initial constraints on the input graph for Theorem 9.10, then we arrive at the following result.

Corollary 9.6. *When the input graph in Theorem 9.10 is the complete graph, then the star graph with center node corresponding to the agent with minimum norm is the (non-unique) optimal topology.*

Proof. The degree of the center node in a star graph is $n - 1$, and all other nodes have degree one. Assume the node weights are sorted as $w_1 \leq \dots \leq w_n$, then the \mathcal{H}_2 norm of the RSN is $\|T_{\text{het}}^{w \mapsto \mathcal{G}}\|_2^2 = (n - 1)w_1 + \sum_{i=2}^n w_i$. Any other tree can be obtained by removing and adding a single edge, while ensuring connectivity. With each such operation, the cost is non-decreasing, as any new edge will increase the degree of node $i > 1$ and by assumption $w_1 \leq w_i$. \square

Corollary 9.6 shows that if there are no restrictions on the initial configuration, the optimal topology can be obtained without the MST algorithm. The computational effort required is only to determine the agent with smallest norm. The non-uniqueness of the star graph can occur if certain agents have identical norm, resulting in other possible configuration with an equivalent overall cost.

4.2 Sensor Placement with \mathcal{H}_2 Performance for NDS Coupled at the State

The input–output description of consensus-type systems derived in §3.2 highlight how noise entering the system can affect the performance. A common challenge in

the design of engineered systems is to minimize the overall cost while achieving the best performance. In this direction, we can formulate a variation of the topology design (68) problem that focuses on choosing sensors for a consensus-seeking system that aims to minimize both the cost of each sensor and the \mathcal{H}_2 performance of the system.

In this direction, consider a modification of the system in (51) in the form

$$\Sigma_\tau : \begin{cases} \dot{x}(t) = -L_e(\mathcal{G}_\tau)RR^T x_\tau(t) + \sigma_w E(\mathcal{G}_\tau)^T \hat{w}(t) - L_e(\mathcal{G}_\tau)^T R\Gamma \hat{v}(t) \\ z(t) = R^T x_\tau(t), \end{cases} \quad (74)$$

where $\hat{w}(t)$ and $\hat{v}(t)$ are the normalized noise signals, and the matrix Γ is a diagonal matrix with elements σ_i corresponding to the variance of the sensor on edge i . We note that the most general version of this problem considers a finite set of p sensors each with an associated variance,

$$P = \{\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2\}, \quad (75)$$

where for each element $\sigma_i^2 \in P$ there is an associated cost $c(\sigma_i^2)$. The cost function has the property that $c(\sigma_i^2) > c(\sigma_j^2)$ if $\sigma_i^2 < \sigma_j^2$. Using (53)–(54), in order to find the optimal placement of these sensors, one can consider the mixed-integer program [16],

$$\begin{aligned} \mathcal{P}_1 : \quad & \min_{X, W} \lambda \text{trace}[R^T X R] + \sum_{i=1}^{|\mathcal{E}|} c(w_i) \\ \text{s.t. } & W = \text{diag}\{w_1, \dots, w_{|\mathcal{E}|}\}, w_i \in P, \sum_i w_i \leq \mu, \\ & -L_e(\mathcal{G}_\tau)^T R R^T X - X R R^T L_e(\mathcal{G}_\tau)^T + \sigma_w^2 L_e(\mathcal{G}_\tau) + L_e(\mathcal{G}_\tau)^T R W R^T L_e(\mathcal{G}_\tau) = 0, \end{aligned}$$

where λ represents a weighting on the \mathcal{H}_2 performance of the solution, and μ represents the maximum aggregated noise covariance. Note that in general $|\mathcal{E}| \min_i \sigma_i^2 \leq \mu \leq |\mathcal{E}| \max_i \sigma_i^2$.

The problem \mathcal{P}_1 is combinatorial in nature, as a binary decision needs to be made as to which sensor to use and place in the network. Although \mathcal{P}_1 can certainly be solved by using a mixed-integer programming solvers [16], certain relaxations can be made to convexify the resulting problem. Most notably, one approach involves relaxing the discrete nature of the set P (75) into a box-type constraint as

$$\hat{P} = [\underline{\sigma}^2, \bar{\sigma}^2]. \quad (76)$$

The cost function can now be written as a continuous map $c : \hat{P} \mapsto \mathbb{R}$ which is convex and a strictly decreasing function. The simplest version of such a function would be the linear map

$$c(\sigma_i^2) = -\beta \sigma_i^2$$

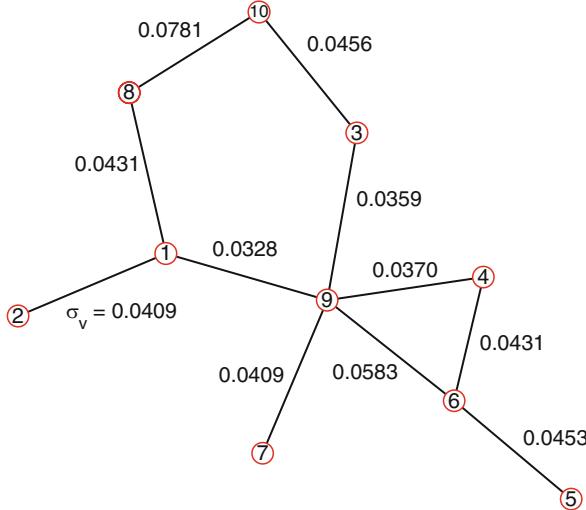


Fig. 8 A graph on ten nodes with optimal sensor selection; σ_v denotes the sensor variance

for some $\beta > 0$. This relaxation leads to the following modified program,

$$\begin{aligned} \mathcal{P}_2 : \quad & \min_{X, W} \lambda \text{trace}[R^T X R] - \beta \text{trace}[W] \\ \text{s.t. } & W = \mathbf{diag}\{w_1, \dots, w_{|\mathcal{E}|}\}, \underline{\sigma}^2 \leq w_i \leq \bar{\sigma}^2, \sum_i w_i \leq \mu, \\ & -L_e(\mathcal{G}_\tau)^T R R^T X - X R R^T L_e(\mathcal{G}_\tau)^T + \sigma_w^2 L_e(\mathcal{G}_\tau) + L_e(\mathcal{G}_\tau)^T R W R^T L_e(\mathcal{G}_\tau) = 0. \end{aligned}$$

As an example of the applicability of \mathcal{P}_2 , we considered the sensor selection for the graph in Fig. 8. A random graph on ten nodes with an edge probability of 0.15 was generated. The resulting graph is connected and contains two independent cycles, resulting in a more general problem instance. The sensor constraints were $\hat{P} = [0.001 \ 0.1]$ and $\mu^2 = 0.501$. Finally, the cost function weights were chosen as $\beta = 5$ and $\lambda = 1$.

Solving \mathcal{P}_2 then resulted in a non-trivial selection of sensors for each edge. The sensor covariance for each edge is labeled in Figure 8; we observe that the highest fidelity sensors tend to be concentrated around the node of highest degree. Also, the edge with the lowest fidelity sensor is placed in “low traffic” areas.

5 Concluding Remarks

The complexity of large-scale systems requires a systematic approach for their analysis and synthesis that blends constructs from system theory on one hand, and graph theory on the other. This chapter presents a viable framework for studying these systems that highlights their structural properties.

By defining four canonical models of networked dynamic systems, we proceeded to explicitly describe how interconnections effect some of the system-theoretic properties of the overall system. Furthermore, the generality of the models developed allows for a further specification of system complexity in the realm of the notions of homogeneity and heterogeneity.

One of the main themes of this work revolved around a characterization of the \mathcal{H}_2 performance of NDS for both analysis and synthesis. The performance of these systems was shown to be intimately related to the number of edges in the network, and the degree of each agent in the ensemble. A natural extension of this work is to consider other system norms for analysis and synthesis, such as \mathcal{H}_{∞} norm of NDS, that is the subject of forthcoming work by the authors. Perhaps a more subtle point of this chapter is the relationship between the different NDS models. Specifically, we noted that via appropriate transformations and the inclusion of structured decentralized control laws, distinct types of NDS models can be transformed to one another. This suggests, for example, that certain NDS models might be more advantageous to use than others when considering analysis and synthesis for a particular networked system. This last comment also relates to the necessity of finding tractable algorithms for the synthesis of NDS. The results of §4 highlighted a convenient connection between results in combinatorial optimization and systems theory. While the combinatorial structure of NDS may seem prohibitive at first glance, a complete understanding of how certain properties of the graph relate to the overall performance can lead to useful relaxations and efficient algorithms for synthesis of networks that support the operation of distributed dynamic systems.

Acknowledgement This work was supported by NSF grant ECS-0501606.

References

1. “Autonomous Nanotechnology Swarm,” <http://ants.gsfc.nasa.gov>
2. I. Akyildiz, W. Su, Y. Sankarasubramanian, and E. Cayirci, A survey on sensor networks, *IEEE Communications Magazine*, 102–114, August 2002
3. B. Awerbuch, “Optimal distributed algorithms for minimum weight spanning tree, counting, leader election, and related problems,” *Proceedings of the 19th Annual ACM Symposium on Theory of Computing*, New York, New York, 1987
4. P. Barooah, N. Machado da Silva, and J. P. Hespanha, “Distributed optimal estimation from relative measurements for localization and time synchronization,” *Proceedings of Distributed Computing in Sensor Systems*, 2006
5. A. Behar, J. Matthews, F. Carsey, and J. Jones, “NASA/JPL Tumbleweed Polar Rover,” *IEEE Aerospace Conference*, Big Sky, MT, March 2004
6. D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation*, Prentice-Hall, Englewood Cliffs, NJ, 1989
7. K. Danzmann and LISA Study Team, “LISA: Laser Interferometer Space Antenna for the Detection and Observation of Gravitational Waves,” Max-Planck Institute fur Quantenoptik, Garching, Germany, 1998
8. G. E. Dullerud and F. Paganini, *A Course in Robust Control Theory: A Convex Approach*, Springer, New York, 2000

9. J. A. Fax and R. M. Murray, Information flow and cooperative control of vehicle formations, *IEEE Transactions on Automatic Control*, (49) 9:1465–1476, 2004
10. C. V. M. Fridlund, Darwin-the infrared space interferometry mission, *ESA Bulletin*, (103):20–25, August 2000
11. R. G. Gallager, P. A. Humblet, and P. M. Spira, A distributed algorithm for minimum-weight spanning trees, *ACM TOPLAS*, (5) 1:66–77, 1983
12. C. Godsil and G. Royle, *Algebraic Graph Theory*, Springer, Berlin, 2001
13. Y. Hatano and M. Mesbahi, Agreement over random networks, *IEEE Transactions on Automatic Control*, (50) 11:1867–1872, 2005
14. R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991
15. C. R. Johnson, *Matrix Theory and Applications*, *Proceedings of Symposia in Applied Mathematics* (40), 1990
16. B. Korte and J. Vygen, *Combinatorial Optimization: Theory and Algorithms*, Springer, Berlin, 2000
17. P. R. Lawarence, The Terrestrial planet finder, *Proceedings of the IEEE Aerospace Conference*, March 2001
18. P. Lin, Y. Jia, and L. Li, Distributed robust \mathcal{H}_{∞} consensus control in directed networks of agents with time-delays, *Systems and Control Letters*, (57):643–653, 2007
19. M. Mesbahi and F. Y. Hadaegh, Formation flying control of multiple spacecraft via graphs, matrix inequalities, and switching, *AIAA Journal of Guidance and Control*, (24) 2:369–377, 2001
20. M. Mesbahi and M. Egerstedt, *Graph-theoretic Methods in Multi-agent Networks*, Princeton University Press, 2010
21. R. Olfati-Saber, “Distributed Kalman filtering for sensor networks,” *IEEE Conference on Decision and Control*, December 2007
22. R. Olfati-Saber and R. M. Murray, Consensus problems in networks of agents with switching topology and time-delays, *IEEE Transactions on Automatic Control*, (49) 9:1520–1533, 2004
23. R. Olfati-Saber, J. A. Fax, and R. M. Murray, Consensus and cooperation in networked multi-agent systems, *Proceedings of the IEEE*, (95) 1:215–233, 2007
24. G. Purcell, D. Kuang, S. Lichten, S. C. Wu, and L. Young, “Autonomous formation flyer (AFF) sensor technology development,” *American Astronautical Society Guidance and Control Conference*, February 1998
25. A. Rahmani, M. Ji, M. Mesbahi, and M. Egerstedt, Controllability of multi-agent systems from a graph theoretic perspective, *SIAM Journal of Control and Optimization*, 48 (1):162–186, 2009
26. W. Ren and R. W. Beard, *Distributed Consensus in Multi-vehicle Cooperative Control*, Springer, London, 2008
27. W. Ren, R. W. Beard, and E. M. Atkins, A survey of consensus problems in multi-agent coordination, *American Control Conference*, June 2005
28. J. Sandhu, M. Mesbahi and T. Tsukamaki, Relative Sensing Networks: Observability, Estimation, and the Control Structure, *IEEE Conference on Decision and Control*, 2005
29. C. Scherer, P. Gahinet, and M. Chilali, Multiobjective output-feedback control via LMI optimization, *IEEE Transactions on Automatic Control*, (42) 7:896–911, 1997
30. G. Sholomitsky, O. Prilutsky, and V. Rodin, Infra-red space interferometer, *International Astronautical Federation*, Paper IAF-77-68, 1977
31. R. S. Smith and F. Y. Hadaegh, Control of deep space formation flying spacecraft; relative sensing and switched information, *AIAA Journal of Guidance and Control*, (28) 1:106–114, 2005
32. A. Tahbaz-Salehi and A. Jadbabaie. Necessary and sufficient conditions for consensus over random networks, *IEEE Transactions on Automatic Control*, (53) 3:791–796, 2008
33. G. Walsh, H. Ye, and L. Bushnell, Stability analysis of networked control systems, *American Control Conference*, June 1999
34. P. K. C. Wang and F. Y. Hadaegh, Coordination and control of multiple microspacecraft moving in formation, *Journal of Astronautical Sciences*, 44:315–355, 1996

35. B. Wie, *Space Vehicle Dynamics and Control*, American Institute of Aeronautics and Astronautics, 1998
36. L. Xiao, S. Boyd, and S. Lall, A scheme for robust distributed sensor fusion based on average consensus, *Fourth International Symposium on Information Processing in Sensor Networks*, April 2005
37. D. Zelazo and M. Mesbahi, On the observability properties of homogeneous and heterogeneous networked dynamic systems, *IEEE Conference on Decision and Control*, December 2008
38. D. Zelazo and M. Mesbahi, Edge agreement: Graph-theoretic performance bounds and passivity analysis, *IEEE Transactions on Automatic Control* 2010 (to appear)
39. D. Zelazo and M. Mesbahi, “ \mathcal{H}_2 Performance of relative sensing networks: Analysis and synthesis,” *AIAA@Infotech Conference*, April 2009
40. D. Zelazo, A. Rahmani, and M. Mesbahi, “Agreement via the edge Laplacian,” *IEEE Conference on Decision and Control*, December 2007
41. F. Zhang, *Matrix Theory*. Springer, Berlin, 1999

A Novel Coordination Strategy for Multi-Agent Control Using Overlapping Subnetworks with Application to Power Systems

R.R. Negenborn, G. Hug-Glazmann, B. De Schutter, and G. Andersson

1 Introduction

Power networks [15, 16, 24] are one of the corner stones of our modern society. The dynamics of a power network as a whole are the result of the interactions between the millions of individual components. Conventionally, the power in power networks is generated using several large power generators. This power is then transported through the transmission and distribution network to the location where it is consumed, e.g., households and industry. Power flows are then relatively predictable, and the number of control agents is relatively low. Due to the ongoing deregulation in the power generation and distribution sector in the US and Europe, the number of players involved in the generation and distribution of power has increased

R.R. Negenborn

Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2,
2628 CD Delft, The Netherlands
e-mail: r.r.negenborn@tudelft.nl

G. Hug-Glazmann

Department of Electrical and Computer Engineering, Carnegie Mellon University,
5000 Forbes Avenue, Pittsburgh, Pennsylvania
e-mail: ghug@andrew.cmu.edu

B. De Schutter

Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2,
2628 CD Delft, The Netherlands
and
Department of Marine and Transport Technology, Delft University of Technology,
Mekelweg 2, 2628 CD Delft, The Netherlands
e-mail: b@deschutter.info

G. Andersson

Power Systems Laboratory, ETH Zürich, Physikstrasse 3, 8092 Zürich, Switzerland
e-mail: andersson@eeh.ee.ethz.ch

significantly. The number of source nodes of the power distribution network is increasing even further as also large-scale industrial suppliers and small-scale individual households start to feed electricity into the network [13].

As a consequence, the structure of the power network is changing from a hierarchical top-down structure into a much more decentralized system with many generating sources and distributing agencies. This causes that power flows become less predictable and may actually change their conventional directions. To still guarantee basic requirements and service levels, such as voltage magnitude and frequency levels, bounds on deviations, stability, elimination of transients, etc., and to meet the demands and requirements of the users, new infrastructure in the shape of transmission lines and so-called Flexible Alternating Current Transmission Systems (FACTS) [10] is installed. Transmission lines increase directly the capacity of the network on the one hand. FACTS devices can be used to actively change the way in which power flows over the network on the other hand. FACTS devices can change voltage magnitudes, line impedances, and phase angles, and therefore have the potential to improve the security of the network, to increase the dynamic and transient stability, to increase the quality of supply for sensitive industries, and to enable environmental benefits [10]. Two particular types of FACTS devices that frequently appear in practice and that also will be used later on in this chapter are Static Var Compensators (SVCs) and Thyristor Controlled Series Compensators (TCSCs) [5].

To optimally use and control such devices and to optimally use the existing infrastructure, new control techniques have to be developed and implemented [18]. A major challenge in this context is that the devices in the network, such as the various FACTS devices, are usually owned and operated by different authorities. Despite this, the operators of the various devices have as objective to determine their actions in such a way that the best overall network performance is obtained. Hence, multi-agent control, in which communication and cooperation between various control authorities is explicitly taken into account, has to be employed.

1.1 Multi-Agent Control of Power Networks

The control structure of power networks can be represented as a multi-agent system [18, 25–27], in which the control agents are organized in several layers as illustrated in Fig. 1. A control agent hereby is an entity, e.g., a human, a computer, or a hardware device, that on the one hand observes the state or situation of the network and on the other hand chooses actions to be taken in the network by changing settings of actuators, such as the reference for the power output of generators or the reference for settings of FACTS devices. A control agent has to choose its actions in such a way that the performance of the network in terms of safety, security, and stability, is the best possible, while respecting operational constraints and minimizing costs. High costs hereby indicate a bad performance of the network, whereas low costs indicate a good performance. In the control hierarchy that power networks are controlled by, at the lower layers control agents consider faster dynamics, more local information, smaller subnetworks, and shorter time spans. At the higher layers

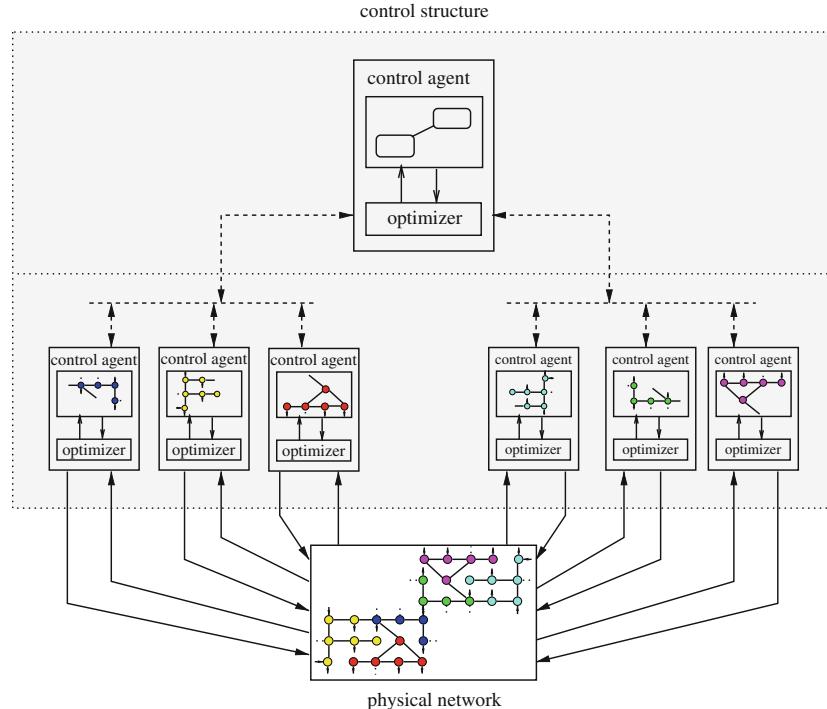


Fig. 1 Illustration of multi-layer control of large-scale networks (inspired by [18]). The control structure consists of several layers of control agents. The control agents make measurements of the state of the network and determine which actions to take

control agents consider slower dynamics, more global information, larger subnetworks, and longer time spans [20].

The control problem that an individual control agent in a control hierarchy faces can be cast as an optimization problem, based on a local objective function that encodes the control goals of the agent, subject to a model of the part of the network that the control agent controls, and additional constraints, e.g., on the range of the inputs. The model of the part of the network that the control agent controls is referred to as its prediction model. This prediction model describes how the values of variables of interest (such as voltage magnitudes, power flows, etc.) react to changes in inputs and can therefore be used to predict what the effect of certain input choices is going to be.

1.2 Control of Subnetworks

In a multi-agent system, control is distributed over several control agents. Each of the control agents controls only its own part of the network, i.e., its own subnetwork. Let for now a network be modeled at an abstract level using a number of nodes with arcs interconnecting the nodes. The nodes represent characteristics of the

components of the physical network, whereas the arcs model the direct interaction between the nodes. E.g., one node κ could model the characteristics of a power generator together with a bus and a transmission line, and another node ω could model the characteristics of a load and a bus. If the bus of this load is physically connected to the transmission line, then an arc is defined between the nodes κ and ω . The subnetwork of a control agent then constitutes a number of nodes together with the arcs connected to these nodes.¹

Usually subnetworks are defined through geographical or institutional borders, such as borders of cities, provinces, countries, the European Union, etc. Subnetworks can however also be defined differently, e.g., based on a fixed “radius” around input nodes. Nodes that are reachable within a certain number of arcs from a particular node with an actuator are then included in a particular subnetwork [9]. Or, subnetworks can be defined using an influence-based approach [8]. The idea of influence-based subnetworks is that the subnetworks are defined based on the nodes that a certain input and, hence, a control agent controlling that input, can influence. Sensitivities are then used to determine which variables an input can influence, and hence, which nodes should be considered part of a subnetwork. The fixed-radius and the influence-based approaches have as advantage that the subnetworks are defined taking a more actuator-centered perspective. Using the fixed-radius approach, this definition is somewhat ad hoc and heuristic. On the contrary, the influence-based approach is more flexible and allows for a structured determination of the subnetwork that a control agent has to consider.

When using the mentioned approaches for defining subnetworks, any pair of two resulting subnetworks can be categorized as *non-overlapping*, *touching*, or *overlapping*, as illustrated in Fig. 2. If for two subnetworks, the nodes belonging to one of

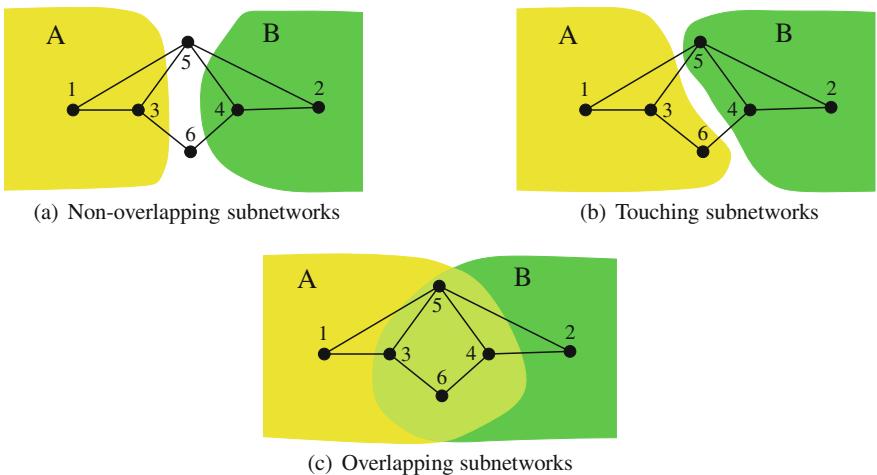


Fig. 2 Illustration of different types of subnetworks

¹ In Sect. 2 we define networks, nodes, arcs, and subnetworks more formally.

these do not coincide with the nodes belonging to the other subnetwork, and if there are no arcs going from nodes in the one subnetwork into nodes of the other subnetwork, then the subnetworks are non-overlapping. If for two subnetworks, the nodes belonging to one of these do not coincide with the nodes of the other subnetwork, but if there are arcs between nodes of the one subnetwork and nodes of the other subnetwork, then the subnetworks are touching. If for two subnetworks, the nodes belonging to one of these partially coincide with the nodes belonging to the other subnetwork, then the subnetworks are overlapping. In that case, a *common* sub-subnetwork is defined consisting of those nodes and arcs that belong to both subnetworks.

If the subnetworks are non-overlapping, then the values of the variables of the nodes that control agents can influence significantly do not overlap, so no coordination among control agents is necessary. In that case, adequate control performance can be obtained, as illustrated in [8]. If the subnetworks are touching, coordination can be obtained by adapting the technique of [3], as will be discussed in Sect. 3. For subnetworks that are overlapping, no techniques have been proposed so far for obtaining coordination. For overlapping subnetworks, the control agents will have to find agreement on the values of variables involved in the characteristics of the common sub-subnetworks. This topic is addressed in this chapter.

1.3 Optimal Power Flow Control

Optimal power flow control is a well known-method to optimize the operation of a power network at higher control layers [15]. Optimal power flow control is typically used to improve steady-state network security by improving the voltage profile, preventing lines from overloading, and minimizing active power losses. The optimal power flow control problem is usually stated as an optimization problem in which variables to be optimized consist of inputs or settings for generators, the objective function encodes the control goals (such as maintaining voltage magnitudes within desired bounds, preventing transmission lines from overloading, minimizing power losses, etc.), and the prediction model consists of the steady-state characteristics of the network.

To optimally make use of the FACTS devices installed in the power network we employ optimal power flow control to determine the settings for these devices. As mentioned, the devices in the network can be owned and controlled by different authorities. Traditional approaches for optimal power flow control in power networks using multiple control agents assume that control agents consider at most touching, and thus not overlapping, subnetworks [14, 22]. In these cases subnetworks are typically defined based on existing geographical borders of countries, states, provinces, cities, etc. However, when the subnetworks are overlapping, the traditional approaches may not be suitable. Therefore, a new coordination approach for control of overlapping subnetworks has to be developed. We have already made a first step in this with the proposal of the approach described in [12], of which the approach proposed in this chapter is a further elaboration and generalization.

1.4 Goal and Outline of This Chapter

In this chapter we propose a coordination scheme for control agents controlling overlapping subnetworks with the aim of obtaining the best overall network performance. This chapter is organized as follows. In Sect. 2, we formalize the modeling of networks, subnetworks, and control objectives used in this chapter. In Sect. 3, we first discuss a recently proposed approach that can be used for the multi-agent control of subnetworks that are *not* overlapping (i.e., non-overlapping or touching). We then propose an extension of this approach to multi-agent control of subnetworks that are overlapping in Sect. 4. In Sect. 5, we apply the proposed approach to an optimal power flow control problem from the domain of power networks. In particular, we employ the approach to control FACTS devices in an adjusted IEEE 57-bus power network, in which each FACTS is controlled by a different control agent. Section 6 contains conclusions and directions for future research.

2 Modeling of Network Characteristics and Control Objectives

In this section we formalize the way in which we describe the network characteristics, subnetworks, and control objectives in this chapter. An example of the application of this formalization is given in Sect. 5.

2.1 Network Characteristics

We consider the control of power networks by multiple control agents that operate in a higher control layer. At this layer, we are interested in controlling the very slow dynamics or the long-term behavior of the network, and therefore we can assume that dynamics of the lower control layers and physical network can be represented or approximated by instantaneous, steady-state characteristics.

Let a network be represented by a network model. Let the model consist of v nodes, and let κ , for $\kappa \in \{1, \dots, v\}$ denote a particular node. Each of the nodes in the network model is labeled with a set of variables (e.g., voltage magnitudes and angles) and constraints (e.g., power flow equations) used to compute the steady-state values for these variables, given values for inputs (e.g., amount of power to be generated) and disturbances (e.g., amount of power consumed). The constraints of a particular node κ involve variables of that particular node and variables of other nodes, referred to as the neighboring nodes $\mathcal{N}^\kappa = \{\omega_{\kappa,1}, \dots, \omega_{\kappa,n_\kappa}\}$. To indicate the interaction between node κ and its neighboring nodes in \mathcal{N}^κ , we define an arc between κ and each node $\omega \in \mathcal{N}^\kappa$.

Let for node $\kappa \in \{1, \dots, v\}$, the variables $\mathbf{z}^\kappa \in \mathbb{R}^{n_{z^\kappa}}$, $\mathbf{u}^\kappa \in \mathbb{R}^{n_{u^\kappa}}$, and $\mathbf{d}^\kappa \in \mathbb{R}^{n_{d^\kappa}}$, denote the (static) states,² the input variables, and the disturbance variables

² Sometimes the static states are also referred to as algebraic variables.

associated with node κ , respectively, and let the constraints of node κ be given by

$$\mathbf{g}^\kappa(\mathbf{z}^\kappa, \mathbf{u}^\kappa, \mathbf{d}^\kappa, \mathbf{z}^{\omega_{\kappa,1}}, \dots, \mathbf{z}^{\omega_{\kappa,n_\kappa}}) = 0, \quad (1)$$

where \mathbf{z}^ω are the variables of neighboring node $\omega \in \mathcal{N}^\kappa$, and \mathbf{g}^κ are the constraint functions of node κ . These constraint function are assumed to be smooth. A steady-state model for the overall network is obtained by aggregating the constraints (1) for all nodes $\kappa \in \{1, \dots, v\}$, and is compactly represented as

$$\mathbf{g}(\mathbf{z}, \mathbf{u}, \mathbf{d}) = 0, \quad (2)$$

where \mathbf{z} , \mathbf{u} , and \mathbf{d} are the state, input, and disturbance variables of the overall network, and \mathbf{g} defines the steady-state characteristics of the network. Given the inputs \mathbf{u} and the disturbance variables \mathbf{d} , the steady state in which the network settles is determined by solving the system of equations (2).

2.2 Control Objectives

With each node objective terms can be associated. These objective terms specify which behavior is desired by assigning costs to the values of the variables \mathbf{z}^κ and \mathbf{u}^κ of that node. The objective terms involve the variables of node κ and may in addition also involve the variables of the neighboring nodes $\omega \in \mathcal{N}^\kappa$. The summation of the objectives terms of all nodes in the network model gives the objective for the control of the overall network. E.g., if a node has assigned to it constraints representing the characteristics of a transmission line, then as objective term the costs on power losses of that transmission line may be associated to the node. In addition, if a node represents the characteristics of a bus, then an objective term representing costs on a voltage magnitude violation of that bus may be associated to this node.

2.3 Definition of Subnetworks

The values of the inputs \mathbf{u} should be adjusted in such a way that the objectives associated with the nodes are achieved as well as possible. Let for a control agent i the nodes that it controls define its subnetwork. The prediction model that control agent i then uses consists of the union of the constraints of each node that is part of its subnetwork. Let the subnetwork and the control goals of a control agent be defined using one of the approaches mentioned in Sect. 1.2.

In the following we first discuss an approach that can be used in the case that the subnetworks are touching. Then we extend this approach to be able to deal with overlapping subnetworks. For the sake of simplicity we assume below that there are no nodes that do not belong to any subnetwork.

3 Multi-Agent Control of Touching Subnetworks

In this section we discuss a technique for coordinating control agents that use touching subnetworks. This technique is based on an adaptation of the ideas of the modified Lagrange technique proposed in [3] to our network and objective formalization. The technique requires that subnetworks of any two control agents are touching, i.e., the nodes in the subnetwork of one control agent are only taken into account (i.e., modeled and controlled) by that control agent and not by any other. In short, when the control agents have to determine actions, they perform a series of iterations, in each of which the control agents perform a local optimization step and communicate information. The local optimization problems are formulated using local objective functions, local prediction models of the subnetworks, and local constraints. After each local optimization the control agents exchange information, reformulate their local optimization, and perform a new optimization. This continues until a stopping condition is satisfied. Below we first introduce some terminology, then formulate the control problem as considered by an individual control agent, and then we discuss the scheme used by multiple control agents for coordination and communication.

3.1 Internal and External Nodes

We define the following concepts that will be frequently used in the remainder of this chapter:

- We categorize the nodes that control agent i considers based on their location from the point of view of control agent i . For touching subnetworks, the nodes that control agent i considers can be *internal* nodes or *external* nodes. The internal nodes of control agent i are those nodes that belong exclusively to its subnetwork. The external nodes of control agent i are those nodes that do not belong to its subnetwork.
- Based on the distinction between internal and external nodes of control agent i , we make a distinction between internal and external variables of control agent i . The internal variables are those variables associated with the internal nodes of control agent i . The external variables are those variables associated with the external nodes of control agent i .
- For control agent i , the *localized constraint type* of a particular constraint associated with a node κ that control agent i considers is formed by the combination of the location and the types of variables involved in that constraint. The localized constraint type of a constraint associated with a node κ considered by control agent i is denoted by $C_{i,\text{Loc}}^{\text{Vars}}$, where $\text{Loc} \in \{\text{int}, \text{ext}\}$ indicates the location of the node to which the constraint is associated, and $\text{Vars} \in \{\text{int}, \text{int+ext}\}$ indicates the variables involved in the constraint. Recall that a constraint associated with a particular node κ involves variables of that particular node and possibly variables

Table 1 Overview of the localized constraint types of constraints associated with nodes in a sub-network that touches other subnetworks. The location indicates the location of the node from the point of view of control agent i . The variables involved in the constraint indicate which variables are involved in the constraint, from the point of view of control agent i

Type	Location	Variables involved in constraint
$\mathcal{C}_{i,int}^{int}$	Internal	Internal
$\mathcal{C}_{i,int}^{int+ext}$	Internal	Internal+external
$\mathcal{C}_{i,ext}^{ext}$	External	External
$\mathcal{C}_{i,ext}^{int+ext}$	External	Internal+external

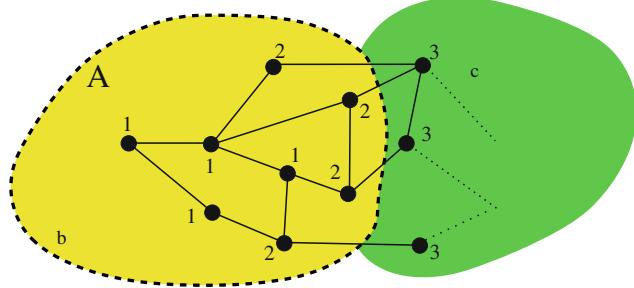


Fig. 3 Illustration of different localized constraint types that can be found at nodes considered by control agent i . The number next to a node in the figure corresponds as follows to the localized constraint types of the constraints that can be associated to that node: 1: $\mathcal{C}_{i,int}^{int}$; 2: $\mathcal{C}_{i,int}^{int+ext}$; 3: $\mathcal{C}_{i,ext}^{int+ext}$, $\mathcal{C}_{i,ext}^{ext}$

of neighboring nodes. The constraints associated with the nodes considered by control agent i can therefore have the localized constraint types listed in Table 1. Figure 3 illustrates for some nodes the localized constraint types that can be found at these nodes.

- In a similar way as we have defined localized constraint types $\mathcal{C}_{i,Loc}^{Vars}$, we also define localized objective term types $\mathcal{J}_{i,Loc}^{Vars}$, referring to the location of the node to which an objective term is associated and the variables that are involved in the objective function term.

3.2 Control Problem Formulation for One Agent

The local optimization problem of control agent i consists of minimizing the local objective function J_i , subject to the prediction model of subnetwork i and additional constraints on inputs and outputs. Below we focus on the issues arising due to the presence of subnetworks that touch the subnetwork of control agent i . We discuss the issues arising with respect to the prediction model and the objective function of

control agent i . For the sake of simplicity of explanation we consider two control agents, control agent i with neighboring agent j , that together control subnetworks that cover all nodes of the network model. The generalization to more than 2 control agents and not fully-covered networks is straightforward.

3.2.1 Prediction Model

The prediction model of control agent i consists of the constraints associated with all its internal nodes. In order to make predictions, control agent i has to know accurate values for all variables involved in the constraints of these nodes. The internal nodes that do not have external neighboring nodes do not require special attention, since the variables involved in the constraints of these internal nodes are of localized constraint type $\mathcal{C}_{i,int}^{int}$ and thus only involve variables of the subnetwork of control agent i . However, the internal nodes that are connected to external nodes do require special attention, since the constraints associated with these internal nodes can be of localized constraint type $\mathcal{C}_{i,int}^{int+ext}$, and thus involve not only variables of the subnetwork of control agent i , but also variables of the subnetwork of neighboring agent j . For the external variables, control agent i has to coordinate with the neighboring agents which values these variables should have. To obtain coordination on the values of the external variables, we apply an idea that was first proposed in [3] as follows.

Below a distinction is made between constraints that are considered as *hard*, and constraints that are considered as *soft*. The hard constraints are constraints that have to be satisfied at all costs. The soft constraints are constraints for which it is desirable that they are satisfied, but for which this should not be done at any price. The hard constraints are included in the formulation of optimization problems as explicit equality constraints; the soft constraints are included in the objective function of optimization problems through a penalty term, weighted by a parameter specifying the costs for violation of the soft constraint.

Recall that the control agents perform a series of iterations and that in each iteration the control agents solve a local optimization problem followed by an exchange of information. Note that internal and external nodes of control agent i correspond to external and internal nodes, respectively, of control agent j . Control agent i considers in its local optimization problem the constraints that are associated with its internal nodes and that are of localized constraint type $\mathcal{C}_{i,int}^{int+ext}$ as hard constraints, using fixed values for the external variables. The values for these external variables are obtained from the neighboring agent j . Control agent i solves its local optimization problem using these values for the external variables. The optimization yields values for the internal variables of control agent i , and for the Lagrange multipliers that are associated with the constraints of localized constraint type $\mathcal{C}_{i,int}^{int+ext}$. The Lagrange multipliers of these constraints and the values of the internal variables involved in these constraints are sent to neighboring agent j .

Neighboring agent j considers the constraints of the internal nodes of control agent i that involve external variables of control agent i in its decision making by including the associated constraints as soft constraints in its local objective function.

In the soft constraints of control agent j , the external variables, which correspond to internal variables of control agent i , are fixed to the values that control agent i has sent to control agent j . Also, the soft constraints are weighted by the Lagrange multipliers as given by control agent i . Neighboring agent j solves its optimization problem, yielding values for its internal variables. It sends the values of the internal variables that appear in the soft constraints to control agent i , such that control agent i can update its information about the corresponding external variables.

Based on this idea, Table 2 shows how control agent i deals with the different constraints when formulating its optimization problem.

3.2.2 Objectives

The local objective function for control agent i consists of objective function terms that are associated with the nodes in its subnetwork. Objective terms associated with internal nodes that are only connected to internal nodes are simply included in the local objective function. However, objective terms associated with internal nodes that are also connected to external nodes cause problems for the same reason as constraints associated with such nodes. Coordination on the values of these variables is achieved by obtaining the desired values for the external variables from neighboring agents.

Table 3 summarizes how the different localized objective term types that control agent i are considered, and how the agent deals with these types, when formulating its optimization problem.

Table 2 Overview of the constraints that control agent i can have and how it deals with these constraints. For the hard and soft constraints, the external variables are fixed to values obtained from neighboring agents. For the hard constraints with external variables Lagrange multipliers are determined. The soft constraints are weighted using the Lagrange multipliers received from neighboring agents

Localized constraint type	Constraint
$C_{i,int}^{int}$	Hard
$C_{i,int}^{int+ext}$	Hard
$C_{i,ext}^{int+ext}$	Soft

Table 3 Overview of the localized objective term types that control agent i considers and how it deals with these terms. External variables are fixed to values obtained from neighboring agents

Localized objective term type	How deal with the objective term
$J_{i,int}^{int}$	Include as is
$J_{i,int}^{int+ext}$	Include as is

3.3 Control Scheme for Multiple Agents

The outline of the scheme for coordination of control agents controlling touching subnetworks, based on the scheme proposed in [3], is as follows:

1. Each control agent i measures the current values for the state variables \mathbf{z}_i and the input variables \mathbf{u}_i that are associated with the nodes in its subnetwork. In addition, it obtains predictions of known disturbance variables \mathbf{d}_i . Furthermore, it obtains through communication from its neighbors values for the external variables and Lagrange multipliers associated with the external nodes that control agent i considers.
2. The iteration counter s is set to 1.
3. Let $\mathbf{w}_{\text{in},i}^{(s-1)}$ and $\lambda_{\text{soft},i}^{(s-1)}$ denote the external variables and Lagrange multipliers, respectively, of which control agent i has received the values from neighboring agents. Given $\mathbf{w}_{\text{in},i}^{(s-1)}$ and $\lambda_{\text{soft},i}^{(s-1)}$, each control agent $i \in \{1, \dots, n\}$ performs concurrently with the other control agents the following steps:

- a. Control agent i solves the local optimization problem:

$$\min_{\mathbf{z}_i, \mathbf{u}_i} J_i \left(\mathbf{z}_i, \mathbf{u}_i, \mathbf{w}_{\text{in},i}^{(s-1)} \right) + \left(\lambda_{\text{soft},i}^{(s-1)} \right)^T \tilde{\mathbf{g}}_{\text{soft},i} \left(\mathbf{z}_i, \mathbf{u}_i, \mathbf{w}_{\text{in},i}^{(s-1)} \right) \quad (3)$$

subject to

$$\tilde{\mathbf{g}}_{\text{hard},i} \left(\mathbf{z}_i, \mathbf{u}_i, \mathbf{d}_i \right) = 0 \quad (4)$$

$$\tilde{\mathbf{g}}_{\text{hard,ext},i} \left(\mathbf{z}_i, \mathbf{u}_i, \mathbf{d}_i, \mathbf{w}_{\text{in},i}^{(s-1)} \right) = 0 \quad (5)$$

$$\mathbf{z}_{i,\text{min}} \leq \mathbf{z}_i \leq \mathbf{z}_{i,\text{max}} \quad (6)$$

$$\mathbf{u}_{i,\text{min}} \leq \mathbf{u}_i \leq \mathbf{u}_{i,\text{max}}, \quad (7)$$

where $\mathbf{z}_{i,\text{min}}$ and $\mathbf{z}_{i,\text{max}}$ are upper and lower bounds on \mathbf{z}_i , $\mathbf{u}_{i,\text{min}}$ and $\mathbf{u}_{i,\text{max}}$ are upper and lower bounds on \mathbf{u}_i , $\tilde{\mathbf{g}}_{\text{soft},i}$ are the constraints of localized constraint type $\mathcal{C}_{i,\text{ext}}^{\text{int+ext}}$, $\tilde{\mathbf{g}}_{\text{hard},i}$ are the constraints of localized constraint type $\mathcal{C}_{i,\text{int}}^{\text{int}}$, $\tilde{\mathbf{g}}_{\text{hard,ext},i}$ are the constraints of localized constraint type $\mathcal{C}_{i,\text{int}}^{\text{int+ext}}$. Solving this local optimization results in values for the variables $\mathbf{z}_i^{(s)}$ and $\mathbf{u}_i^{(s)}$, as well as Lagrange multipliers $\lambda_{\text{hard,ext},i}^{(s)}$ associated with the constraints (5) for current iteration s . After solving this optimization problem the variables $\mathbf{w}_{\text{out},i}^{(s)}$ can be determined as:

$$\mathbf{w}_{\text{out},i}^{(s)} = \tilde{\mathbf{K}}_i \left[\left(\mathbf{z}_i^{(s)} \right)^T \left(\mathbf{u}_i^{(s)} \right)^T \left(\mathbf{d}_i \right)^T \right]^T, \quad (8)$$

where $\mathbf{w}_{\text{out},i}$ are the so-called interconnecting output variables, selected using a selection matrix $\tilde{\mathbf{K}}_i$. These variables represent the variables that control agent i uses in its communication to neighboring agents. Selection matrix $\tilde{\mathbf{K}}_i$ has in each row only zeros, except for a single 1 in the column

corresponding to the position of an element of $\left[\left(\mathbf{z}_i^{(s)} \right)^T, \left(\mathbf{u}_i^{(s)} \right)^T, (\mathbf{d}_i)^T \right]^T$ that is an interconnecting output variable.

- b. Control agent i sends the values of the Lagrange multipliers $\lambda_{\text{hard},\text{ext},i}^{(s)}$ of the hard constraints of localized constraint type $\mathcal{C}_{i,\text{int}}^{\text{int+ext}}$ and the values of $\mathbf{w}_{\text{out},i}$ corresponding to internal variables of these nodes to the neighboring agents that consider the involved external variables.
 - c. Control agent i receives from the neighboring agent j those Lagrange multipliers related to the localized constraint type $\mathcal{C}_{i,\text{ext}}^{\text{int+ext}}$ and those values of the internal variables of the neighboring agents that control agent i requires in order to fix its external variables. Control agent i uses this received information at the next iteration as $\lambda_{\text{soft},i}^{(s)}$ and $\mathbf{w}_{\text{in},i}^{(s)}$.
4. The next iteration is started by incrementing s and going back to step 3, unless a local stopping condition is satisfied for all control agents. The stopping condition is defined as the condition that the absolute changes in the Lagrange multipliers from iteration $s - 1$ to s are smaller than a pre-defined small positive constant $\gamma_{\epsilon,\text{term}}$.

A shortcoming of this method is that it requires that the subnetworks are touching, since it assumes that each node in the network model is assigned to only one of the subnetworks. However, in the case of control of overlapping subnetworks, some of the nodes are included in more than one subnetwork and the identification of internal and external nodes of a control agent is not straightforward any more. Therefore, the method is not directly applicable to overlapping subnetworks. In the following section we extend the method discussed above to control of overlapping subnetworks.

4 Multi-Agent Control for Overlapping Subnetworks

We first propose some new definitions, next we consider the issues appearing due to the overlap, and then we propose a way to deal with these issues. Again, for simplicity of explanation we consider two control agents, control agent i with neighboring control agent j , that together control the subnetworks, which are assumed to cover the full network model.

4.1 Common Nodes

In addition to internal and external nodes as defined before, for control of overlapping subnetworks we make the following definitions:

- *Common nodes* are nodes that belong to the subnetwork of control agent i and that also belong to the subnetwork of the control agent j . A sub-subnetwork

Table 4 Overview of the localized constraint types for overlapping subnetworks

Type	Location	Variables involved in constraint
$C_{i,int}^{int}$	Internal	Internal
$C_{i,int}^{int+com}$	Internal	Internal+common
$C_{i,int}^{int+ext}$	Internal	Internal+external
$C_{i,int}^{int+com+ext}$	Internal	Internal+common+external
$C_{i,com}^{int+com}$	Common	Internal+common
$C_{i,com}^{int+com+ext}$	Common	Internal+common+external
$C_{i,com}^{com}$	Common	Common
$C_{i,com}^{com+ext}$	Common	Common+external
$C_{i,ext}^{ext}$	External	External
$C_{i,ext}^{int+ext}$	External	Internal+external
$C_{i,ext}^{com+ext}$	External	Common+external
$C_{i,ext}^{int+com+ext}$	External	Internal+common+external

defined by the nodes common to several subnetworks is referred to as a common sub-subnetwork.

- The variables associated with the common nodes are referred to as the common variables.
- Given the definition of a common node, the number of possibilities for localized constraint types increases. Table 4 lists the localized constraint types that can be considered by a control agent when subnetworks can be overlapping. In total there are 12 different localized constraint types. Figure 4 illustrates some of the possible localized constraint types.
- In addition to the extension of the localized constraint types, the localized objective term types are extended as well, by also defining localized objective term types that are based on variables of common nodes.

4.2 Control Problem Formulation for One Agent

For multi-agent control of overlapping subnetworks an approach has to be found to deal with the common nodes. Since the common nodes are considered by several control agents, the constraints associated with these common nodes appear in the subnetwork models of multiple control agents. Even though the control agents have the same objective with respect to these nodes, combined with the objective for their internal nodes, conflicting values for the variables of the common nodes can be the result. Below we discuss how to extend the scheme of Sect. 3 for control of overlapping subnetworks. Again, for the sake of simplicity of explanation we focus on two control agents: control agent i with neighboring agent j .

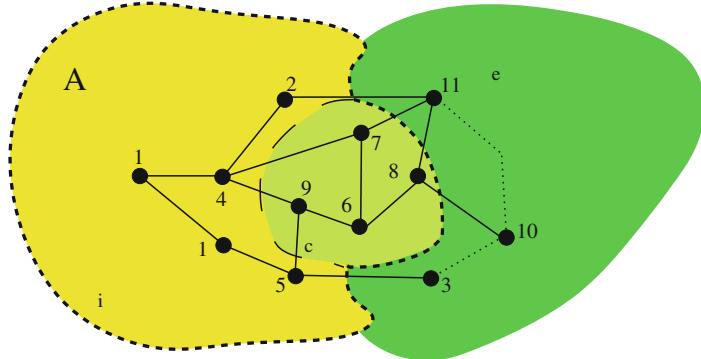


Fig. 4 Illustration of different localized constraint types that can be found at particular nodes considered by control agent i . The number next to a node in the figure corresponds as follows to the localized constraint types of the constraints that can be associated to that node: 1: $C_{i,int}^int$; 2: $C_{i,int}^{int+ext}$; 3: $C_{i,ext}^{int+ext}$, $C_{i,ext}^ext$; 4: $C_{i,int}^int$, $C_{i,int}^{int+com}$; 5: $C_{i,int}^int$, $C_{i,int}^{int+com}$, $C_{i,int}^{int+ext}$, $C_{i,int}^{int+com+ext}$; 6: $C_{i,com}^com$; 7: $C_{i,com}^{int+com}$, $C_{i,com}^{com+ext}$, $C_{i,com}^com$, $C_{i,com}^{int+com+ext}$; 8: $C_{i,com}^com$, $C_{i,com}^{com+ext}$; 9: $C_{i,com}^com$, $C_{i,com}^{int+com}$; 10: $C_{i,ext}^ext$, $C_{i,ext}^{ext+com}$, 11: $C_{i,ext}^{int+ext}$, $C_{i,ext}^{com+ext}$, $C_{i,ext}^ext$, $C_{i,ext}^{int+com+ext}$

4.2.1 Prediction Model

Similarly as for control of touching subnetworks, for control of overlapping subnetworks, internal nodes of control agent i that are connected to external nodes require special attention, since the constraints associated to these nodes may involve external variables. In addition to this, common nodes of control agent i that are connected to external nodes also require special attention. The extension of the approach for control of touching subnetworks to the control of overlapping subnetworks involves the following extension of the prediction model.

Control agent i considers as prediction model the constraints of all internal *and* common nodes. For the constraints of localized constraint types $C_{i,int}^{int+ext}$, $C_{i,int}^{int+ext+com}$, $C_{i,com}^{com+ext}$, and $C_{i,com}^{int+com+ext}$ the control agent takes for the external variables values that it has received from neighboring agent j . When control agent i has solved its optimization problem, it sends the values of the internal *and* the common variables of the constraints of these specialized constraint types to neighboring agents.

Neighboring agent j considers in its optimization problem the constraints of the internal and common nodes of control agent i that involve external variables of control agent i as soft constraints by including them in the objective function through a penalty term, weighted by the Lagrange multipliers provided by control agent i , and with fixed values for the external *and* common values in the soft constraints as received from control agent i . Note that although control agent j considers fixed values for the common variable in the soft constraints, it will not fix the values for the common variables in the hard constraints (similarly as control agent i). Hence, control agents i and j share the responsibility for the common variables. The result

Table 5 Overview of the way in which control agent i considers the constraints of particular localized constraint types in its optimization problem. For the hard constraints all external variables are fixed to values obtained from neighboring agents. For the soft constraints all external and common variables are fixed. For the hard constraints with external variables Lagrange multipliers are determined. The soft constraints are weighted with Lagrange multipliers obtained from neighboring agents. Note that the soft constraint part of the inclusion of constraints of type $\mathcal{C}_{i,\text{com}}^{\text{int+com+ext}}$ involves fixed external and common variables and a Lagrange multiplier as obtained from neighboring agents, whereas the hard constraint part of the inclusion of constraints of type $\mathcal{C}_{i,\text{com}}^{\text{int+com+ext}}$ involves only fixed external variables

Localized constraint type	Constraint
$\mathcal{C}_{i,\text{int}}^{\text{int}}$	Hard
$\mathcal{C}_{i,\text{int}}^{\text{int+ext}}, \mathcal{C}_{i,\text{int}}^{\text{int+com}}$	Hard
$\mathcal{C}_{i,\text{int}}^{\text{int+com+ext}}$	Hard
$\mathcal{C}_{i,\text{com}}^{\text{int+com}}$	Hard and soft
$\mathcal{C}_{i,\text{com}}^{\text{int+com+ext}}$	Hard and soft
$\mathcal{C}_{i,\text{com}}^{\text{com}}$	Hard
$\mathcal{C}_{i,\text{com}}^{\text{com+ext}}$	Hard
$\mathcal{C}_{i,\text{ext}}^{\text{int+ext}}$	Soft
$\mathcal{C}_{i,\text{ext}}^{\text{int+ext+com}}$	Soft

of solving the optimization problem of neighboring agent j therefore yields values for the internal, common, and external variables of control agent j . The internal variables of control agent j related to the soft constraints are sent to control agent i .

Table 5 summarizes how control agent i deals with the different localized constraint types.

4.2.2 Objectives

With the nodes that control agent i has in its subnetwork objective terms are associated. The objective function terms associated with each node can depend on the variables associated with that node and its neighboring nodes. As before, the objective terms involving only internal variables require no special attention. The objective terms involving both internal and external variables can be dealt with by fixing the external variables, as is also done for control of touching subnetworks. However, the common variables appearing in control of overlapping subnetworks do require special attention.

For control of overlapping subnetworks, multiple control agents will try to control the values of the common variables. To allow control agents to jointly achieve performance comparable to the performance that an overall centralized control agent can achieve, the responsibility for the objective terms involving only common variables, i.e., of localized objective term type $\mathcal{J}_{i,\text{com}}^{\text{com}}$, is shared equally by the control agents. Hence, each control agent i that considers a particular common node κ ,

Table 6 Overview of the localized objective term types that control agent i considers and how it deals with the associated objective terms. External variables are fixed. Variable N_κ is the number of control agents considering node κ as common node

Localized objective term type	How deal with the objective term
$\mathcal{J}_{i,int}^{int}$	Include as is
$\mathcal{J}_{i,int}^{int+ext}$	Include as is
$\mathcal{J}_{i,int}^{int+com}$	Include as is
$\mathcal{J}_{i,int}^{com}$	Include as is
$\mathcal{J}_{i,com}^{com}$	Include partially by weighting it with a factor $1/N_\kappa$
$\mathcal{J}_{i,com}^{int+com}$	Include as is

includes in its objective function $1/N_\kappa$ times the objective function terms of such nodes of localized objective term type $\mathcal{J}_{i,com}^{com}$, where N_κ is the number of control agents considering node κ as common node. Control agent i in addition includes into its objective function the objective terms of all its internal nodes, and the objective terms of these common nodes that involve only internal and common variables, i.e., the objective terms of localized objective term types $\mathcal{J}_{i,int}^{int}$, $\mathcal{J}_{i,int}^{int+ext}$, $\mathcal{J}_{i,int}^{int+com}$ and $\mathcal{J}_{i,com}^{int+com}$.

Table 6 summarizes how control agent i deals with the different localized objective term types.

4.3 Control Scheme for Multiple Agents

We have discussed how each control agent formulates its prediction model and objective function. The scheme that we propose for multi-agent control for overlapping subnetworks consists of the scheme proposed in Sect. 3 for touching subnetworks, with the following changes:

- Control agent i receives from the neighboring agents the following information at initialization and after each iteration:
 - Lagrange multipliers with respect to the constraints of localized constraint type $\mathcal{C}_{i,com}^{int+com}$, $\mathcal{C}_{i,com}^{int+com+ext}$, $\mathcal{C}_{i,ext}^{int+ext}$, and $\mathcal{C}_{i,ext}^{int+ext+com}$.
 - Values for the external variables and the common variables involved in these constraints.
- The optimization problem that each agent solves is changed accordingly to reflect the extensions discussed in this section, i.e., to take into account the constraints as given in Table 5 and the objective terms as given in Table 6.

The result is a control scheme that can be used by higher-layer control agents that control subnetworks that are overlapping. The control agents hereby share the responsibility for the common variables. In the next section we apply this scheme on an optimal flow control problem in power networks.

5 Application: Optimal Flow Control in Power Networks

In this section apply the scheme for multi-agent control of overlapping subnetworks, as discussed in Sect. 4, to the problem of optimal power flow control in power networks. A case study is carried out on the IEEE 57-bus power network [2], comprising as components generators, loads, transmission lines, and buses, with in addition FACTS devices installed at various locations, as illustrated in Fig. 5. Two configurations are considered: in the first configuration only SVCs are included; in the second configuration only TCSCs are present. Each of the FACTS devices is controlled by an individual control agent.

5.1 Parameters of the Power Network

The parameters of the IEEE 57-bus base network can be obtained from the Power Systems Test Case Archive [2]. Line limits on the apparent power flows have been assigned to all transmission lines in such a way that no lines are overloaded. In order to find an interesting and meaningful situation for FACTS control, the grid was adapted by placing an additional generator at bus 30 leading to increased power flows in the center of the grid. The parameters of this generator are as follows: $dV_{gen,m} = 1.03$ p.u., $dP_{gen,m} = 0.5$ p.u., for $m = 30$. The parameters of the base IEEE 57 network used in this chapter can be found in [2]. The limits on the apparent power flows are set as listed in Table 7.

In the following we make a distinction between a *bus* and a *node*. A bus refers to an element of the physical power network, whereas a node refers to an element of the model of the physical power network. Since for each physical bus a corresponding node is included in the model, references to a bus or its corresponding node can be interchanged, except for when assigning constraints relating to two buses, such as constraints imposed due to transmission lines, to a single node, as we will see next.

Below we formulate the steady-state models used to describe the network behavior, we assign the constraints to nodes, we set up the objective terms associated with the nodes, we discuss the way in which the subnetworks can be determined using the influence-based approach, and we illustrate the workings of the proposed approach.

5.2 Steady-state Characteristics of Power Networks

As the focus lies on improving the steady-state network security, the power network is modeled using equations describing the steady-state characteristics of the power network. As we will see, the aspects of the steady-state security that we are interested in can be determined from the voltage magnitude and voltage angle at each of the 57 (physical) buses in the network. We therefore define 57 nodes to model

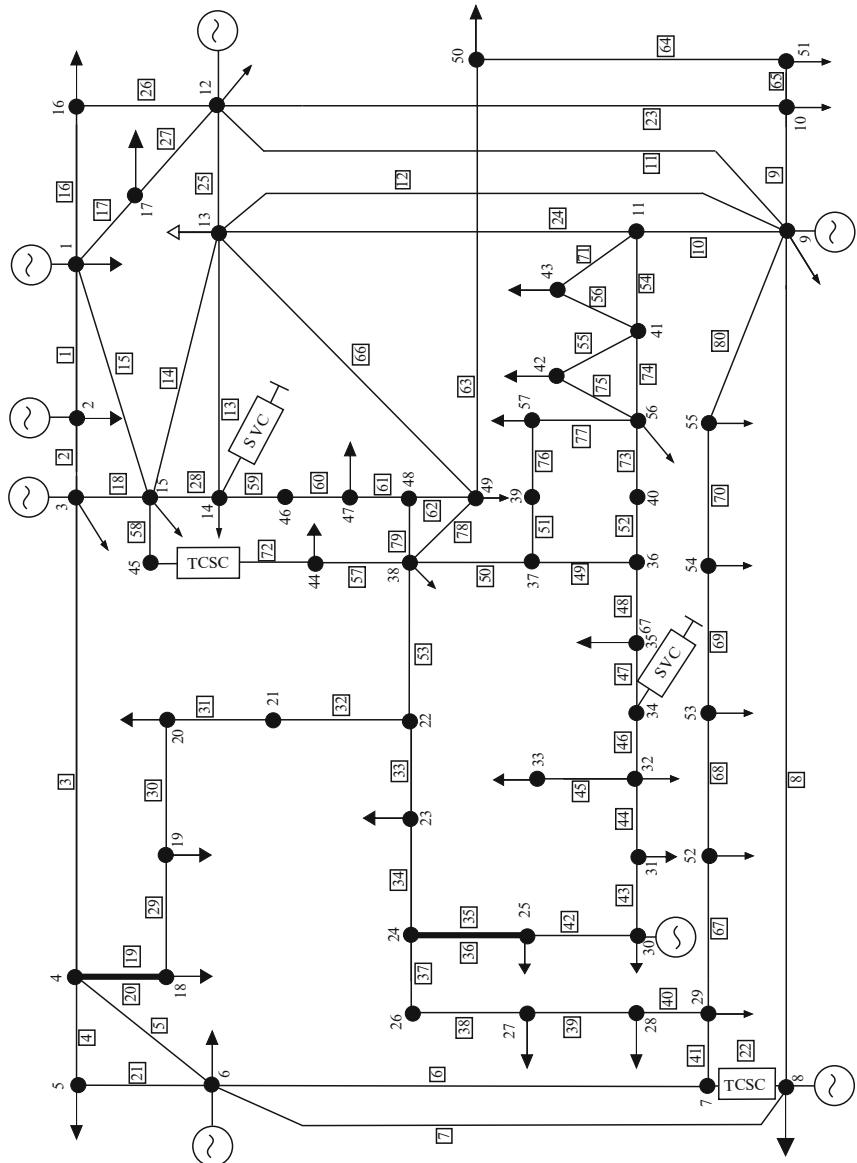


Fig. 5 IEEE 57-bus network extended with either SVCs installed at buses 14 and 34, or with TCSCs in lines 22 and 72

the network, and assign to each node m the voltage magnitude $z_{V,m}$ per unit (p.u.) and the voltage angle $z_{\theta,m}$ (degrees) as variables. In order to determine the values for these variables under different disturbance variables and actuator values, models for the components and their influence on the voltage magnitude and angle are defined. We model the transmission lines, the generators, the loads, and the FACTS devices.

Table 7 Line limits on the apparent power flows

Line nr.	Limit (p.u.)	Line nr.	Limit (p.u.)	Line nr.	Limit (p.u.)	Line nr.	Limit (p.u.)
1	1.800	21	0.550	41	1.100	61	0.350
2	1.650	22	2.160	42	0.500	62	0.300
3	0.721	23	0.700	43	0.500	63	0.500
4	0.412	24	0.700	44	0.400	64	0.300
5	0.800	25	0.900	45	0.300	65	0.450
6	0.750	26	0.600	46	0.400	66	0.600
7	1.200	27	0.750	47	0.400	67	0.550
8	2.200	28	1.000	48	0.400	68	0.500
9	0.600	29	0.300	49	0.700	69	0.350
10	0.450	30	0.300	50	0.800	70	0.400
11	0.300	31	0.300	51	0.300	71	0.300
12	0.500	32	0.300	52	0.300	72	0.700
13	0.700	33	0.400	53	0.400	73	0.300
14	0.936	34	0.450	54	0.300	74	0.300
15	1.900	35	0.300	55	0.300	75	0.300
16	1.050	36	0.300	56	0.300	76	0.300
17	1.200	37	0.350	57	0.500	77	0.300
18	1.200	38	0.350	58	0.600	78	0.550
19	0.300	39	0.350	59	0.636	79	0.550
20	0.300	40	0.400	60	0.650	80	0.500

5.2.1 Transmission Lines

For the transmission lines the well-known π -model is used [15]. The active power $z_{P,mn}$ (p.u.) and the reactive power $z_{Q,mn}$ (p.u.) flowing from bus m over the transmission line to bus n are then given by:

$$\begin{aligned} z_{P,mn} = & (z_{V,m})^2 \left(\frac{\eta_{R,mn}}{(\eta_{R,mn})^2 + (\eta_{X,mn})^2} \right) \\ & - z_{V,m} z_{V,n} \left(\frac{\eta_{R,mn}}{(\eta_{R,mn})^2 + (\eta_{X,mn})^2} \cos(z_{\theta,m} - z_{\theta,n}) \right) \\ & + z_{V,m} z_{V,n} \left(\frac{\eta_{X,mn}}{(\eta_{R,mn})^2 + (\eta_{X,mn})^2} \sin(z_{\theta,m} - z_{\theta,n}) \right) \end{aligned} \quad (9)$$

$$\begin{aligned} z_{Q,mn} = & (z_{V,m})^2 \left(\frac{\eta_{X,mn}}{(\eta_{R,mn})^2 + (\eta_{X,mn})^2} \right) \\ & - z_{V,m} z_{V,n} \left(\frac{\eta_{R,mn}}{(\eta_{R,mn})^2 + (\eta_{X,mn})^2} \sin(z_{\theta,m} - z_{\theta,n}) \right) \\ & - (z_{V,m})^2 \left(\frac{\eta_{B,mn}}{2} \right) - z_{V,m} z_{V,n} \left(\frac{\eta_{X,mn}}{(\eta_{R,mn})^2 + (\eta_{X,mn})^2} \cos(z_{\theta,m} - z_{\theta,n}) \right), \end{aligned} \quad (10)$$

where $\eta_{B,mn}$ (p.u.) is the shunt susceptance, $\eta_{R,mn}$ (p.u.) is the resistance, and $\eta_{X,mn}$ (p.u.) is the reactance of the line between buses m and n .

The constraints for each transmission line going from bus m to bus n , for $n \in N^m$ (where N^m is the set of neighboring buses of bus m , i.e., the buses that are physically connected to bus m through a transmission line), are assigned to node m , if $m < n$, and to node n otherwise.

5.2.2 Generators

Generators are assumed to have constant active power injection and constant voltage magnitude, and therefore

$$z_{P,gen,m} = d_{P,gen,m} \quad (11)$$

$$z_{V,m} = d_{V,gen,m}, \quad (12)$$

where $d_{P,gen,m}$ is the given active power that the generator produces, and $d_{V,gen,m}$ is the given voltage magnitude that the generator maintains. At most one generator can be connected to a bus, since a generator directly controls the voltage magnitude of that bus.

The generator connected to bus 1 is considered as a slack generator, i.e., a generator with infinite active and reactive power capacity, with fixed voltage magnitude and angle [15]. So, for this generator we have with $m = 1$

$$z_{V,m} = d_{V,gen,m} \quad (13)$$

$$z_{\theta,m} = d_{\theta,gen,m}, \quad (14)$$

where $d_{\theta,gen,m}$ is the given voltage angle ensured by the generator.

The constraints of a generator at bus m are assigned to node m .

5.2.3 Loads

The loads are constant active and constant reactive power injections, i.e.,

$$z_{P,load,m} = d_{P,load,m} \quad (15)$$

$$z_{Q,load,m} = d_{Q,load,m}, \quad (16)$$

where $d_{P,load,m}$ and $d_{Q,load,m}$ are the given active and reactive power consumption, respectively, of the load connected to bus m . For simplicity, only one load can be connected to a bus. Multiple loads can easily be aggregated to obtain a single load.

The constraints of the loads at bus m are assigned to node m .

5.2.4 FACTS Devices

SVC

An SVC is a FACTS device that is shunt-connected to a bus m and that injects or absorbs reactive power $z_{Q,SVC,m}$ to control the voltage $z_{V,m}$ at that bus [10]. The SVC connected to bus m accepts as control input the effective susceptance $u_{B,SVC,m}$. The injected reactive power $z_{Q,SVC,m}$ of the SVC is:

$$z_{Q,SVC,m} = (z_{V,m})^2 u_{B,SVC,m}. \quad (17)$$

The control input $u_{B,SVC,m}$ is limited to the domain:

$$u_{B,SVC,min,m} \leq u_{B,SVC,m} \leq u_{B,SVC,max,m}, \quad (18)$$

where the values of $u_{B,SVC,min,m}$ and $u_{B,SVC,max,m}$ are determined by the size of the device [7].

The constraints of an SVC at bus m are assigned to the node m .

TCSC

A TCSC is a FACTS device that can control the active power flowing over a line [10]. It can change the line reactance $z_{X,line,mn}$. The TCSC is therefore considered as a variable reactance $u_{X,TCSC,mn}$ connected in series with the line. If a TCSC is connected in series with a transmission line between buses m and n , the total reactance $z_{X,line,mn}$ of the line including the TCSC is given by:

$$z_{X,line,mn} = \eta_{X,mn} + u_{X,TCSC,mn}, \quad (19)$$

where $\eta_{X,mn}$ is the reactance of the line without the TCSC installed. The reactance $u_{X,TCSC,mn}$ is limited to the domain:

$$u_{X,TCSC,min,mn} \leq u_{X,TCSC,mn} \leq u_{X,TCSC,max,mn}, \quad (20)$$

where the values of $u_{X,TCSC,min,mn}$ and $u_{X,TCSC,max,mn}$ are determined by the size of the TCSC and the characteristics of the line in which it is placed, since due to the physics the allowed compensation rate of the line $u_{X,TCSC,mn}/\eta_{X,mn}$ is limited [7].

The constraints of the TCSC at the line between bus m and n are assigned to node m , if $m < n$, and to node n otherwise.

5.2.5 Power Balance

By Kirchhoff's laws, at each bus the total incoming power and the total outgoing power has to be equal. This yields the following additional constraints for bus m :

$$z_{P,\text{load},m} - z_{P,\text{gen},m} + \sum_{n \in \mathcal{N}^m} z_{P,mn} = 0 \quad (21)$$

$$z_{Q,\text{load},m} - z_{Q,\text{gen},m} - z_{Q,\text{SVC},m} + \sum_{n \in \mathcal{N}^m} z_{Q,mn} = 0. \quad (22)$$

If no generator is connected to bus m , then $z_{P,\text{gen},m}$ and $z_{Q,\text{gen},m}$ are zero. If no load is connected to bus m , then $z_{P,\text{load},m}$ and $z_{Q,\text{load},m}$ are zero. If no SVC is connected to bus m , then $z_{Q,\text{SVC},m}$ is zero.

The constraints resulting from Kirchhoff's laws for bus m are assigned to node m .

5.3 Control Objectives

The objectives of the control are to improve the system security through minimization of the deviations of the bus voltages from given references to improve the voltage profile, minimization of active power losses, and preventing lines from overloading, by choosing appropriate settings for the FACTS devices. These objectives are translated into objective terms associated with the buses as follows:

- To minimize the deviations of the bus voltage magnitude $z_{V,m}$ of bus m from a given reference $d_{V,\text{ref},m}$, an objective term $p_V (z_{V,m} - d_{V,\text{ref},m})^2$ is associated with node m , where p_V is a weighting coefficient.
- To minimize the active power losses over a line between bus m and bus n , an objective term $p_{\text{loss}} (z_{P,mn} + z_{P,nm})$, where p_{loss} is a weighting coefficient, is associated to node m , if $m < n$, and to node n otherwise. Note that the term $z_{P,mn} + z_{P,nm}$, which represents the power losses, is always nonnegative.
- To minimize the loading of the line between buses m and n , an objective term is associated to node m , if $m < n$, and to node n otherwise, as $p_{\text{load}} \left(\frac{z_{S,mn}}{z_{S,\text{max},mn}} \right)^2$, where p_{load} is a weighting coefficient, and where $z_{S,mn}$ is the apparent power flowing over the line from bus m to bus n , defined as $z_{S,mn} = \sqrt{(z_{P,mn})^2 + (z_{Q,mn})^2}$. The relative line loading is penalized in a quadratic way such that an overloaded line is penalized more severely than a line that is not overloaded.

The weighting coefficients p_V , p_{loss} , and p_{load} allow to change the weight given to each objective. In the following we take $p_V = 1,000$, $p_{\text{loss}} = 100$, and $p_{\text{load}} = 1$.

5.4 Setting Up the Control Problems

Each FACTS device is controlled by a different control agent. The influence-based subnetworks of the control agents controlling the FACTS devices can be overlapping, and therefore the control problems of the control agents are set up using the approach discussed in Sect. 4. To solve their subproblems at each iteration the control agents use the nonlinear problem solver SNOPT v5.8 [6], as implemented in Tomlab v5.7 [11], and accessed from Matlab v7.3 [17].

In the following we illustrate how the approach works for a particular assignment of nodes to subnetworks in two representative scenarios.

5.5 Simulations

Various test scenarios with different FACTS devices and subnetworks have been examined. Here we present two representative scenarios. The subnetworks used in these scenarios are shown in Fig. 6. It can be seen that these subnetworks are overlapping, since there are several nodes that are included in both subnetworks.

5.5.1 Scenario 1: Control of SVCs

In the first scenario, SVCs are placed at buses 14 and 34. As the SVCs are mainly used to influence the voltage profile, the line limits are chosen such that no line is at the risk of being overloaded.

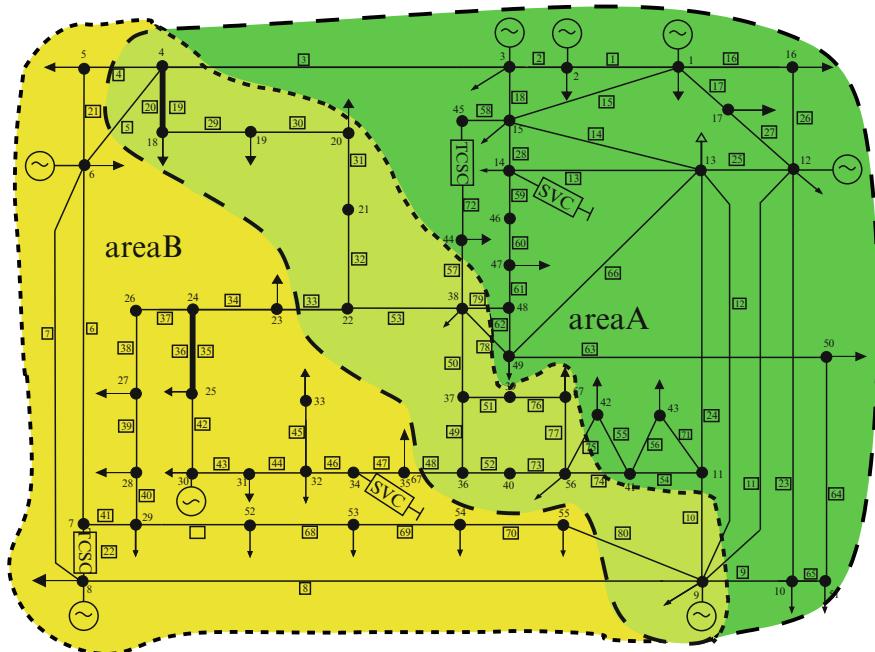


Fig. 6 IEEE 57-bus system with decomposition into two subnetworks. *Scenario 1:* SVCs at buses 14 and 34, *scenario 2:* TCSCs in lines 22 and 72. The dotted line indicates the borders of subnetwork 1 (light shaded); the dashed line indicates the borders of subnetwork 2 (dark shaded). The region encapsulated both by subnetwork 1 and subnetwork 2 is the common region (medium shaded)

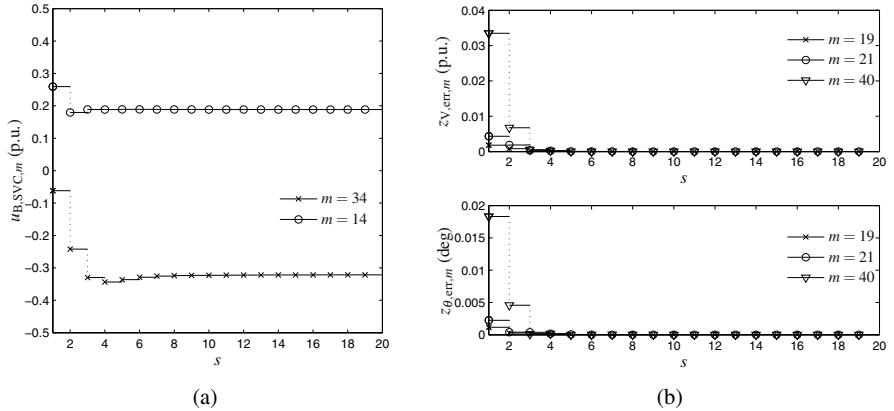


Fig. 7 (a) Convergence of the settings of the SVCs at buses 14 and 34, as a function of the iteration, for scenario 1. (b) Convergence of the difference between the values of the voltage magnitudes (top) and the voltage angles (bottom) as considered by both control agents for buses 19, 21, 40, as a function of the iteration, for scenario 1

Figure 7(a) shows the convergence of the SVC settings over the iterations. As can be seen, the settings of the SVCs converge within only a few iterations to the final values, which in this case are equal to the values obtained from a centralized optimization. Figure 7(b) shows the evolution of the deviations between the values determined by both subnetworks for the voltage magnitudes and angles at some common buses. In the figure the error $z_{V,err,m}$ is defined as the absolute difference between the values that control agents 1 and 2 want to give to the voltage magnitude $z_{V,m}$. Similarly, the error $z_{\theta,err,m}$ is defined as the absolute difference between the values that control agents 1 and 2 want to give to the voltage angles. As can be seen fast convergence is obtained.

5.5.2 Scenario 2: Control of TCSCs

In the second scenario, TCSCs are installed on lines 72 and 22. Since TCSCs are mainly used to influence active power flows and to resolve congestion, the line limits are chosen such that lines 7 and 60 are overloaded if the FACTS devices are not being used.

The results for the TCSC settings and the difference between the voltage magnitudes and angles for some common buses over the iterations are given in Figs. 8(a) and 8(b), respectively. The control agent of subnetwork 1 sets the TCSC to its upper limit at the first few iterations. But after some additional iterations, the values that the control agents choose converge to their final values, which are again equal to the values obtained by a centralized control agent.

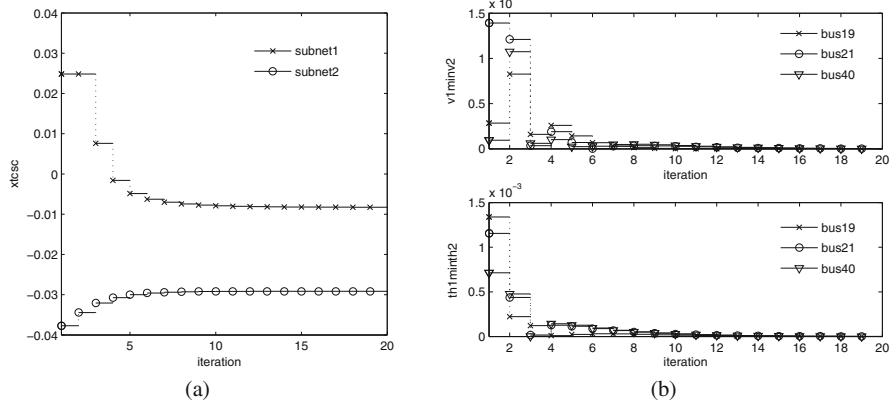


Fig. 8 (a) Convergence of the settings of the TCSCs in lines 22 and 72 (i.e., the lines between buses 7 and 8, and buses 44 and 45, respectively), as a function of the iteration number, for scenario 2. (b) Convergence of the difference between the values of the voltage magnitudes (top) and the voltage angles (bottom) as considered by both control agents for buses 19, 21, 40, as a function of the iteration number, for scenario 2

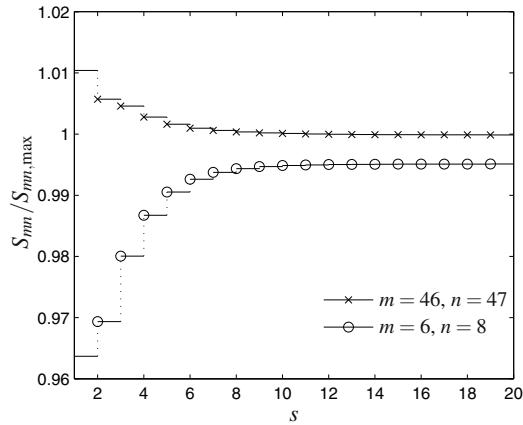


Fig. 9 Convergence of the relative line loadings of lines 7 and 60 (i.e., the lines between buses 6 and 8, and 46 and 47, respectively), as a function of the iteration number, for scenario 2

In Fig. 9 the line loadings of lines 7 and 60, i.e., the lines which are overloaded without FACTS devices in operation, are shown. Line 7 is immediately brought below its limit whereas for line 60, the loading approaches 100% in the course of the optimization process.

6 Conclusions and Future Research

In this chapter we have focused on an alternative way to define subnetworks for higher-layer multi-agent control. The higher control layer uses steady-state characteristics only. We have discussed how subnetworks can be defined based on the influence of inputs on the variables of nodes. When such an approach is used to define subnetworks, some subnetworks could be overlapping, resulting in constraints and objectives in common sub-subnetworks. We have proposed a method for higher-layer multi-agent control that can be used by control agents that control such overlapping subnetworks.

To illustrate the topics discussed and the proposed approach, we have defined overlapping subnetworks for Flexible Alternating Current Transmission Systems (FACTS) in an adjusted version of the IEEE 57-bus power network. Using the proposed control approach, we have then solved an optimal power flow control problem. The simulations illustrate that in the considered cases the proposed approach can achieve fast convergence to actuator values that are globally optimal.

Further research will address the following issues and topics. It will be determined formally when the approach converges and what the quality of the obtained solutions is, in particular when compared to an overall single-agent, centralized, control scheme. This will provide more insight into the quality of the solutions and the time required to obtain these solutions. Also, power networks are just a particular network from the general class of transportation networks. Other examples from the class of transportation networks to which the approach discussed in this chapter could be successfully applied in future work are traffic and transportation systems [4], natural gas networks [23], combined electricity and gas networks [1], water networks [21], etc. The approach will also be extended to deal with dynamics using ideas from [19].

Acknowledgement This research is supported by the BSIK project “Next Generation Infrastructures (NGI)”, the Delft Research Center Next Generation Infrastructures, the project “Coordinated control of FACTS devices”, supported by ABB Switzerland, the European STREP project “Hierarchical and distributed model predictive control (HD-MPC)”, and the project “Multi-Agent Control of Large-Scale Hybrid Systems” (DWV.6188) of the Dutch Technology Foundation STW. The authors thank Swissgrid for their technical support.

References

1. M. Arnold, R. R. Negenborn, G. Andersson, and B. De Schutter. Multi-area predictive control for combined electricity and natural gas systems. In *Proceedings of the European Control Conference 2009*, Budapest, Hungary, August 2009
2. R. D. Christie. Power Systems Test Case Archive. URL: <http://www.ee.washington.edu/research/pstca/>, 2008. Last accessed at May 23, 2008
3. A. J. Conejo, F. J. Nogales, and F. J. Prieto. A decomposition procedure based on approximate Newton directions. *Mathematical Programming, Series A*, 93(3):495–515, December 2002
4. C. F. Daganzo. *Fundamentals of Transportation and Traffic Operations*. Pergamon Press, New York, 1997

5. A. Edris, R. Adapa, M. H. Baker, L. Bohmann, K. Clark, K. Habashi, L. Gyugyi, J. Lemay, A. S. Mehraban, A. K. Meyers, J. Reeve, F. Sener, D. R. Torgerson, and R. R. Wood. Proposed terms and definitions for flexible AC transmission system (FACTS). *IEEE Transactions on Power Delivery*, 12(4):1848–1853, October 1997
6. P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Journal on Optimisation*, 12(4):979–1006, 2002
7. G. Glanzmann and G. Andersson. Using FACTS devices to resolve congestions in transmission grids. In *Proceedings of the CIGRE/IEEE PES International Symposium*, San Antonio, TX, October 2005
8. G. Glanzmann and G. Andersson. FACTS control for large power systems incorporating security aspects. In *Proceedings of X SEPOPE*, Florianopolis, Brazil, May 2006
9. P. Hines, L. Huawei, D. Jia, and S. Talukdar. Autonomous agents and cooperation for the control of cascading failures in electric grids. In *Proceedings of the 2005 IEEE International Conference on Networking, Sensing and Control*, pages 273–278, Tucson, AZ, March 2005
10. N. G. Hingorani and L. Gyugyi. *Understanding FACTS Concepts and Technology of Flexible AC Transmission Systems*. IEEE Press, New York, New York, 2000
11. K. Holmström, A. O. Göran, and M. M. Edvall. User's guide for Tomlab /SNOPT, December 2006
12. G. Hug-Glanzmann, R. R. Negenborn, G. Andersson, B. De Schutter, and J. Hellendoorn. Multi-area control of overlapping areas in power systems for FACTS control. In *Proceedings of Power Tech 2007*, Lausanne, Switzerland, July 2007. Paper 277
13. N. Jenkins, R. Allan, P. Crossley, D. Kirschen, and G. Strbac. *Embedded Generation*. TJ International, Padstow, UK, 2000
14. B. H. Kim and R. Baldick. A comparison of distributed optimal power flow algorithms. *IEEE Transactions on Power Systems*, 15(2):599–604, May 2000
15. P. Kundur. *Power System Stability and Control*. McGraw-Hill, New York, New York, 1994
16. J. Machowski, J. Bialek, and J. R. Bumby. *Power System Dynamics and Stability*. Wiley, New York, New York, 1997
17. Mathworks. Matlab. URL: <http://www.mathworks.com/>, 2007
18. R. R. Negenborn. *Multi-Agent Model Predictive Control with Applications to Power Networks*. PhD thesis, Delft University of Technology, Delft, The Netherlands, December 2007
19. R. R. Negenborn, B. De Schutter, and J. Hellendoorn. Multi-agent model predictive control for transportation networks: Serial versus parallel schemes. *Engineering Applications of Artificial Intelligence*, 21(3):353–366, April 2008
20. R. R. Negenborn, S. Leirens, B. De Schutter, and J. Hellendoorn. Supervisory nonlinear MPC for emergency voltage control using pattern search. *Control Engineering Practice*, 17(7):841–848, July 2009
21. R. R. Negenborn, P. J. van Overloop, T. Keviczky, and B. De Schutter. Distributed model predictive control for irrigation canals. *Networks and Heterogeneous Media*, 4(2):359–380, June 2009
22. F. J. Nogales, F. J. Prieto, and A. J. Conejo. Multi-area AC optimal power flow: A new decomposition approach. In *Proceedings of the 13th Power Systems Control Conference (PSCC)*, pages 1201–1206, Trondheim, Germany, 1999
23. A. J. Osiadacz. *Simulation and Analysis of Gas Networks*. Gulf Publishing Company, Houston, TX, 1987
24. P. W. Sauer and M. A. Pai. *Power System Dynamics and Stability*. Prentice-Hall, London, UK, 1998
25. K. P. Sycara. Multiagent systems. *AI Magazine*, 2(19):79–92, 1998
26. D. D. Šiljak. *Decentralized Control of Complex Systems*. Academic, Boston, Massachusetts, 1991
27. G. Weiss. *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. MIT Press, Cambridge, Massachusetts, 2000

Distributed Control Methods for Structured Large-Scale Systems

Justin Rice, Paolo Massioni, Tamás Keviczky, and Michel Verhaegen

1 Introduction

The development of efficient and tractable distributed control design procedures for large-scale systems has been an active area of research in the past three decades [1, 2]. The growing number of applications where such solutions offer increased functionality, flexibility or efficiency has spawned a renewed interest in this topic. The main challenges lie in the computational cost and complexity of efficient controller design and implementation. For a thorough overview of distributed and decentralized control research, see [1], and the introduction to [2].

In this work we describe three recently developed approaches and their extensions, and apply them to a benchmark problem involving an infinite platoon of vehicles.

For infinite arrays of spatially invariant interconnected systems, a spatial Fourier transform may be used for solving optimal control problems. The optimal controller is then often approximated using spatial truncation. Although there are significant differences between finite and infinite-dimensional interconnected systems, we will perform our numerical studies for the infinite-dimensional case, noting that all of the discussed approaches can be extended to the finite-dimensional problem as well, e.g. [16, 25, 30].

Our first approach to be considered is to exploit the special operator structure induced by such interconnections. Structure preserving iterative algorithms can be employed on these Laurent operators with rational symbols to solve a myriad of control problems. While the exact solutions to optimal control problems involving shift invariant systems are often Laurent operators with *irrational* symbols, using this

J. Rice, P. Massioni, T. Keviczky, and M. Verhaegen
Delft Center for Systems and Control, Delft University of Technology, 2628 CD, Delft,
The Netherlands
e-mail: j.k.rice@tudelft.nl; p.massioni@tudelft.nl; t.keviczky@tudelft.nl;
m.verhaegen@moesp.org

approach we can find arbitrarily close rational approximations to these solutions, which also have the favorable property of admitting spatially discrete distributed controller realizations.

Special structures in the problem formulation itself have also often offered possibilities to render the design problem tractable. Several approaches have been applying various types of decoupling transformations that allow significant reduction of the problem complexity in the transformed domain [3–5]. The second approach discussed in this work belongs to this class of methods and is applicable to large-scale systems composed of identical linear systems interconnected with each other according to a certain pattern. The proposed method exploits a decomposition of the global system that converts the global synthesis problem into a set of computationally simple problems, and the use of Linear Matrix Inequalities (LMIs) allows distributed controller synthesis under performance constraints.

If the large-scale system is composed of independent identical linear system dynamics, the system structure allows significant simplification of the control design procedure: the synthesis of stabilizing distributed control laws can be obtained by using a simple local LQR design. Such a local LQR controller based design procedure will be described as the third approach under consideration in this work. We will also illustrate how stability of the overall large-scale system is related to the robustness of local controllers and the spectrum of a Laurent matrix representing the desired sparsity pattern. Although this design method is extremely simple to calculate and implement, it relies heavily on the identical, dynamically decoupled subsystem assumption. Furthermore, although the stability result does not depend on the weighting matrices chosen in the cost function, optimality of the distributed controller is not guaranteed.

In Sect. 2, the benchmark problem is described, while Sects. 3–5 provide a summary and extensions of the three discussed control design methods. Each section contains an explanation of how the particular approach can be used to address the benchmark problem. Section 6 summarizes our numerical results, while Sect. 7 collects our observations and possible avenues of future research.

2 Problem Statement

As a benchmark problem used for comparison, we consider the problem of controlling the absolute and relative distances in an infinite-dimensional car platoon and assume a second order model for each vehicle [6]. The dynamics of the states is described by the following:

$$\begin{bmatrix} \dot{x}_i^1 \\ \dot{x}_i^2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_i^1 \\ x_i^2 \end{bmatrix} + \begin{bmatrix} 0 \\ q \end{bmatrix} w_i + \begin{bmatrix} 0 \\ g \end{bmatrix} u_i \quad (1)$$

for $i = -\infty \dots +\infty$; x_i^1 is the position, x_i^2 is the velocity, u_i is the control input, w_i is the disturbance input, and q and g are constants which we assume to be both 1 without loss of generality. As we are going to consider the state feedback \mathcal{H}_2

problem, the measured output will be the state itself, while as performance output z_i we choose the following:

$$z_i = \begin{bmatrix} u_i \\ f_1 x_i^1 \\ f_2 \left(\frac{1}{2} x_{i-1}^1 - x_i^1 + \frac{1}{2} x_{i+1}^1 \right) \end{bmatrix} \quad (2)$$

The performance outputs include the control effort, a symmetric measure of the relative position and absolute position (needed for the well-posedness, as explained in [6]), weighted by f_2 and f_1 .

The full, infinite order system, is then described by the following equations:

$$\begin{cases} \bar{x}(t) = \bar{A}\bar{x}(t) + \bar{B}_1\bar{w}(t) + \bar{B}_2\bar{u}(t) \\ \bar{z}(t) = \bar{C}\bar{x}(t) + \bar{D}\bar{u}(t) \end{cases} \quad (3)$$

where $\bar{x}, \bar{z}, \bar{u}, \bar{w}$ are in the Hilbert space $l_2(-\infty, \infty)$ with the usual inner product and represent the infinite dimensional vectors containing the stack of all x_i, z_i, u_i and w_i , and the letters with a bar over them represent bounded Laurent operators, of which $\bar{A}, \bar{B}_1, \bar{B}_2$, and \bar{D} are block diagonal, whereas \bar{C} is banded, with one non-zero band over the diagonal and one below. We will look for controllers of the form:

$$\bar{u}(t) = \bar{K}\bar{x}(t) \quad (4)$$

which minimize the \mathcal{H}_2 norm from the disturbance w to the output z . For some system, $G = \begin{bmatrix} \bar{A} & \bar{B} \\ \bar{C} & 0 \end{bmatrix}$, the continuous time, discrete-space, spatiotemporal \mathcal{H}_2 norm [7] is defined as:

$$\|G\|_2^2 = \left(\frac{1}{2\pi}\right)^2 \int_0^{2\pi} \int_{-\infty}^{\infty} \text{Tr}[G(e^{j\theta}, j\omega)G(e^{j\theta}, j\omega)^*] d\omega d\theta \quad (5)$$

For the benchmark problem under consideration in this work, solving the problem in a distributed fashion will imply restricting the search to operators \bar{K} with only a limited number of off-diagonal bands.

2.1 \mathcal{H}_2 Problem and Exact Solution

The Riccati equation for computing the \mathcal{H}_2 optimal controller is the following:

$$\bar{A}^T \bar{P} + \bar{P} \bar{A} - \bar{P} \bar{B}_2 \bar{B}_2^T \bar{P} + \bar{C}^T \bar{C} = 0 \quad (6)$$

Notice that the only non block-diagonal term is $\bar{C}^T \bar{C}$. Let us define $\bar{Q} = \bar{C}^T \bar{C}$. By using a Fourier transform as in [7], we can turn this infinite-size Riccati equation into an infinite set of parameter dependent equations:

$$\hat{A}_\lambda^T \hat{P}_\lambda + \hat{P}_\lambda \hat{A}_\lambda - \hat{P}_\lambda \hat{B}_{2,\lambda} \hat{B}_{2,\lambda}^T \hat{P}_\lambda + \hat{Q}_\lambda = 0 \quad (7)$$

with $\lambda = e^{j\theta}$, $\theta \in [0, 2\pi]$. Actually, only \hat{Q}_λ depends on λ (it derives from the only non-diagonal operator). We can easily see that the equation is equivalent to:

$$\begin{aligned} & \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \hat{P}_\lambda + \hat{P}_\lambda \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} - \hat{P}_\lambda \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \hat{P}_\lambda + \\ & + (f_1^2 + f_2^2 (\frac{3}{2} - \lambda^{-1} - \lambda + \frac{1}{4}\lambda^{-2} + \frac{1}{4}\lambda^2)) \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = 0 \end{aligned} \quad (8)$$

Partitioning \hat{P}_λ as:

$$\hat{P}_\lambda = \begin{bmatrix} p_1 & p_2 \\ p_2 & p_3 \end{bmatrix} \quad (9)$$

and doing the computations, we have:

$$\begin{cases} p_2^2 = k \\ 2p_2 - p_3^2 = 0 \\ p_1 = p_2 p_3 \end{cases} \quad \text{with } k = \left(f_1^2 + f_2^2 \left(\frac{3}{2} - \lambda^{-1} - \lambda + \frac{1}{4}\lambda^{-2} + \frac{1}{4}\lambda^2 \right) \right) \quad (10)$$

from which, the real positive definite stabilizing (and irrational) solution is:

$$\hat{P}_\lambda = \begin{bmatrix} \sqrt{2k\sqrt{k}} & \sqrt{k} \\ \sqrt{k} & \sqrt{2\sqrt{k}} \end{bmatrix} \quad (11)$$

In the following three sections, we will propose different methods for the case when the explicit solution cannot be calculated as simply as described above.

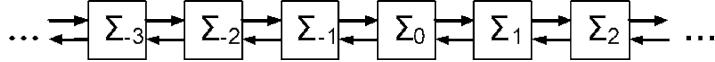
3 A Rational Laurent Operator Structure Preserving Iterative Approach to Distributed Control

The first distributed control method that we will discuss exploits the rationally symbolled Laurent matrix structure in distributed systems composed of identical subsystems connected on a string. We will first briefly introduce the type of interconnected systems that we will consider; see [2, 8–10] for more detailed discussions.

Consider the set of subsystems:

$$\begin{bmatrix} \dot{x}_s \\ v_{s-1}^p \\ v_{s+1}^m \\ z_s \\ y_s \end{bmatrix} = \begin{bmatrix} A & B^p & B^m & B^1 & B^2 \\ C^p & W^p & Z^m & L^p & V^p \\ C^m & Z^p & W^m & L^m & V^m \\ C^1 & J^p & J^m & D^{11} & D^{12} \\ C^2 & H^p & H^m & D^{21} & D^{22} \end{bmatrix} \begin{bmatrix} x_s \\ v_s^p \\ v_s^m \\ w_s \\ u_s \end{bmatrix} \quad (12)$$

where x_s is the local state, u_s and y_s are the local inputs and outputs, z_s and w_s are performance channel and disturbance input, and v_s^m and v_s^p are the coupling terms to the neighboring subsystems. In this work we will investigate doubly infinite one-dimensional strings of such systems ($s \in \mathbb{Z}$), as in Fig. 1.

**Fig. 1** String interconnection

Note that for ‘well-posed’ interconnections [2], Σ can always be transformed such that $Z_m = 0, Z_p = 0$ [11], and from now on we will make this assumption w.l.o.g. In this case the coupling terms, v_s^m and v_s^p , are superfluous, and if we resolve them we get a ‘lifted system’:

$$\bar{\Sigma} : \begin{bmatrix} \bar{x} \\ \bar{z} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} \bar{A} & \bar{B}_1 & \bar{B}_2 \\ \bar{C}_1 & \bar{D}_{11} & \bar{D}_{12} \\ \bar{C}_2 & \bar{D}_{12} & \bar{D}_{22} \end{bmatrix} \begin{bmatrix} \bar{x} \\ \bar{w} \\ \bar{u} \end{bmatrix} \quad (13)$$

with states $\bar{x} = [\dots x_{-1}^T x_0^T x_1^T x_2^T \dots]^T$, and similarly structured inputs \bar{u}, \bar{w} and outputs \bar{y}, \bar{z} in Hilbert space l_2 . \bar{A}, \bar{B}_1 , etc., are block Laurent operators (see e.g., [12, 13]) with rational symbols: doubly infinite Toeplitz matrices (see (14) below for the matrix realization of \bar{A}) that we may ‘diagonalize’ using the Fourier transform to obtain their operator symbols on the unit circle, $z \in \mathbb{T}$:

$$\mathcal{F}\bar{A}\mathcal{F}^{-1} := A(z) = D + B^m(zI - W^m)^{-1}C^m + B^p(z^{-1}I - W^p)^{-1}C^p$$

assuming that $\rho(W^p), \rho(W^m) < 1$, so that the series converge. $A(z)$, the symbol of \bar{A} , is a rational transfer function on $z \in \mathbb{T}$, but not all Laurent operators have rational symbols, and so we will distinguish them using the notation $\bar{A} = L_r\{B^m, W^m, C^m, D, B^p, W^p, C^p\}$.

The Fourier transform here is unitary ($\mathcal{F}\mathcal{F}^* = I$) and an isomorphism, so there is an equivalence in norm and spectrum between the block L_r -operator and its symbol [13]:

$$\|\bar{A}\| = \|A(z)\|_\infty = \sup_{z \in \mathbb{T}} \|A(z)\|_2$$

$$\lambda(\bar{A}) = \lambda(A(\mathbb{T}))$$

$$\underbrace{\begin{bmatrix} \vdots \\ \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \\ \vdots \end{bmatrix}}_{\bar{x}} = \underbrace{\begin{bmatrix} \ddots & & & & & & & \\ & \ddots & & & & & & \\ & & A & & & & & \\ & & & B^m C^m & & & & \\ & & & & A & & & \\ & & & & & B^p C^p & & \\ & & & & & & B^p W^p C^p & \\ & & & & & & & B^p W^p W^p C^p \\ & & & & & & & & B^p W^p W^p W^p C^p \\ & & & & & & & & \dots \end{bmatrix}}_{\bar{A}} \underbrace{\begin{bmatrix} \vdots \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ \vdots \end{bmatrix}}_{\bar{x}}$$

$$+ \bar{B}_1 \bar{w} + \bar{B}_2 \bar{u}$$

$$(14)$$

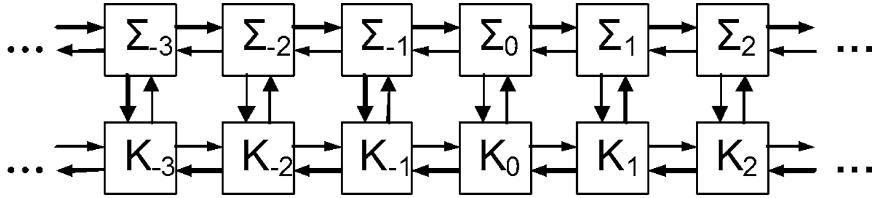


Fig. 2 String interconnection with similarly connected controller

and also, from a system theoretic point of view, an equivalence of exponential stability and performance, controllability, etc., [8, 9, 14] between the lifted system of Laurent operators $\begin{bmatrix} \bar{A} & \bar{B} \\ \bar{C} & \bar{D} \end{bmatrix}$ and the Fourier transformed family of systems: $\begin{bmatrix} A(z) & B(z) \\ C(z) & D(z) \end{bmatrix}$ parameterized over the unit circle $z \in \mathbb{T}$. Calculations for the lifted system can thus also be reduced to finite dimensional eigenvalue problems and Lyapunov and Riccati equations, rationally parametric on the unit circle [8, 9]. However, it is impossible to compute independently at every $z \in \mathbb{T}$, and it is not clear how to go about sampling \mathbb{T} , performing local computations, and then interpolating to all \mathbb{T} for guaranteed stability and performance (the eigenvalues of transfer matrices may not even have Lipschitz smoothness properties [15]), so we will use these properties solely for proofs, and develop another strategy for computations.

Also, in the same way that the interconnected subsystems Σ_s induce a system of L_r operators, $\bar{\Sigma}$, a controller of L_r operators, \bar{K} , can be directly distributed into interconnected subcontrollers K_s of the same structure as the plant, as we see in Fig. 2 (see [16] for the finite dimensional analog with explicit controller formulas). Hence in the following we will be careful to preserve the L_r structure.

3.1 L-Operator Sign Function

As one would expect given the mixed-causal LTI system input-output mapping interpretation of our rational Laurent operators, it is possible to add, multiply, and invert L_r operators while preserving their structure, and to calculate the operator norm and perform model order reductions. Simple formulas for these calculations are available in [17].

However, system analysis and controller synthesis calculations will require more than an arithmetic, we must be able to solve eigenvalue problems and Lyapunov and Riccati equations, which is possible using the matrix sign function, as we will now discuss.

The matrix sign function [18] has been shown to be a very powerful tool for finite dimensional linear systems analysis and control synthesis (see [19] for an overview). The Newton's method for calculations also converges extremely fast (locally quadratically), making it one of the most efficient computational techniques for solving Riccati equations and other common control problems. In the following

we will extend the sign function definition, some convergence bounds, and numerical robustness calculations from the finite dimensional case to the rational Laurent operator case.

3.2 Definition

We can define the L_r operator sign iteration and sign function as in Algorithm 2 below. As we see, the sign iteration only requires L_r -structure preserving arithmetic

$$\begin{aligned}\bar{Z}_0 &= \bar{X} \\ \bar{Z}_{k+1} &= \frac{1}{2}(\bar{Z}_k + \bar{Z}_k^{-1}) \quad \text{for } k = 0, 1, 2, \dots \\ \text{sign}(\bar{X}) &= \lim_{k \rightarrow \infty} \bar{Z}_k\end{aligned}$$

Algorithm 2: Sign Iteration

for its computation, and just as in the finite matrix case [18], if \bar{X} has no spectrum on the imaginary axis, then every \bar{Z}_k is regular. We should emphasize that for some L_r operator, \bar{X} , $\bar{Z}_\infty = \text{sign}(\bar{X})$ will be a Laurent operator, but **not** always with a rational symbol, since the space of rational functions is not complete. However, \bar{Z}_k will have a rational symbol $\forall k < \infty$ and thus we can approximate $\text{sign}(\bar{X})$ arbitrarily close using rational Laurent operators, and the approximation generated by the halted sign iteration will converge very fast to $\text{sign}(\bar{X})$ in operator norm, as we will next show.

3.3 Convergence

As we discussed in the introduction, the Fourier operator ‘block-diagonalizes’ an L_r operator to its symbol. Hence the sign iteration for L_r operators can be considered just as a sign iteration of the rational symbols, $Z_k(z)$, $z \in \mathbb{T}$, or equivalently of complex matrices $Z_k(z_0)$ at each $z_0 \in \mathbb{T}$.

For a complex finite dimensional matrix, X , the matrix sign can be considered basically as a sign iteration on each of the eigenvalues individually. Using this property, it can be shown ([17], using results in [20]) that the distance from the spectrum of some iterate \bar{Z}_k to that of $\text{sign}(\bar{X})$ can be upper bounded based on the locations of the spectrum of X :

$$\|\bar{Z}_k - \text{sign}(\bar{X})\| < \varepsilon$$

after $k = \sup_{z \in \mathbb{T}} \mathcal{O}(\log_2(\eta(z))^2 + \log_2 \log_2(\varepsilon^{-1} \kappa_2(V(z))))$ iterations, where

$$\eta(z) = \max_{\lambda_i \in \lambda(X(z))} \{1 + |\mathbb{R}(\lambda_i)| + |\mathbb{R}(\lambda_i)|^{-1} + \frac{|\text{Im}(\lambda_i)|}{|\mathbb{R}(\lambda_i)|}\}$$

$\kappa_2()$ is the condition number in the 2-norm, and $X(z) = V(z)\Lambda(z)V(z)^{-1}$ is an eigenvalue decomposition. So basically, assuming some bounds on the non-normality of $X(z)$ and on its spectral radius and the distance from its spectrum to the imaginary axis (bound on $\eta(z)$), we can bound the number of sign iterations necessary to get arbitrarily close to $\text{sign}(\bar{X})$ in the operator norm using the sign iteration. Note that the contribution from the eigenvectors is only a ‘local’ one, at each $V(z)$, since \bar{X} is block diagonalized by the unitary Fourier transform, and a large bound on non-normality is not much of a limitation anyway, since \log_2^2 is a very severe function (e.g. $\log_2 \log_2(10^{100}) < 9$).

Thus due to the ‘diagonalizability’ of L_r operators, the sign iteration on them will work essentially the same as it does on finite dimensional matrices, and we can compute an arbitrarily close approximation $\bar{Z}_k \approx \text{sign}(\bar{X})$, with an L_r structured \bar{Z}_k , in a reasonable number of iterations.

3.4 Applications

As discussed in [16] and references therein, in finite dimensions, the matrix sign function is useful for many things in control analysis and design, such as checking matrix stability ($\text{sign}(X) = -I \Leftrightarrow \lambda(X) \in \mathbb{C}_-$) and solving Lyapunov and Riccati equations. As it turns out, most of these results extend directly to the operator sign function on L_r operators in a straightforward way. In the following we will just show for solving Riccati equations, but the technique applies to the other applications as well.

It is well known that one can solve a continuous time finite dimensional Riccati equation: $XA + A^T X + Q - RX = 0$ where $Q = Q^T$, $R = R^T$, by applying sign iterations to the Hamiltonian matrix $H = \begin{bmatrix} A & -R \\ -Q & -A^T \end{bmatrix}$, to calculate $\text{sign}(H) = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$ and solving the linear system of equations:

$$\begin{bmatrix} S_{12} \\ S_{22} + I \end{bmatrix} X = - \begin{bmatrix} S_{11} + I \\ S_{21} \end{bmatrix}. \quad (15)$$

The Riccati equation has a unique stabilizing solution X if and only if (15) has a unique solution and H has no eigenvalues on the imaginary axis [21].

We now consider the L_r operator case. We can calculate:

$$\begin{bmatrix} \bar{S}_{11} & \bar{S}_{12} \\ \bar{S}_{21} & \bar{S}_{22} \end{bmatrix} \approx \text{sign}(\begin{bmatrix} \bar{A} & -\bar{R} \\ -\bar{Q} & -\bar{A}^T \end{bmatrix}) \quad (16)$$

to arbitrary accuracy using Algorithm 1, with L_r structured $\bar{S}_{11}, \bar{S}_{12}, \bar{S}_{21}, \bar{S}_{22}$. We then solve

$$\begin{bmatrix} \bar{S}_{12} \\ \bar{S}_{22} + I \end{bmatrix} \tilde{\bar{X}} = - \begin{bmatrix} \bar{S}_{11} + I \\ \bar{S}_{21} \end{bmatrix} \quad (17)$$

where we are solving for some approximate $\tilde{\bar{X}}$, since our sign calculation is only approximate. The Fourier transform can be used to show here that for ‘good enough’ approximations of (16), assuming exponential stabilizability and detectability of the operator system, then $\tilde{\bar{X}}$ is (exponentially) stabilizing and solves

$$\tilde{\bar{X}}\bar{A} + \bar{A}^T\tilde{\bar{X}} + \bar{Q} - \tilde{\bar{X}}\bar{R}\tilde{\bar{X}} \approx 0$$

The discrete time Riccati, and continuous and discrete time Sylvester and Lyapunov equations and stability checks will work in essentially the same way, and with the same results. Hence we can extend many finite dimensional computational techniques using the sign iteration, such as stability, and spatiotemporal \mathcal{H}_2 performance analysis, and controller synthesis to L_r operators.

3.5 Numerical Difficulties

Of course, a key assumption in the above discussion is that $\tilde{\bar{X}}$ is a close enough approximation to \bar{X} so as to be exponentially stabilizing and have good performance. There will always be some error $\|\tilde{\bar{X}} - \bar{X}\|$ since our calculation of the matrix sign in Algorithm 1 will never be taken to $k = \infty$, but will always be halted after some threshold of convergence $\|\bar{Z}_k - \bar{Z}_{k-1}\| < \varepsilon$. Also, repeated order-reducing approximations, must be used to prevent the complexity from blowing up [22]. Such iterative approximations if too aggressive, can cause numerical instability in the sign iteration.

However, in linear systems analysis and control applications, *a posteriori* closed loop stability and performance (e.g. \mathcal{H}_2 , \mathcal{H}_∞), can be relaxed to ε -sub-optimal problems involving Lyapunov and Riccati inequalities. The verification of solutions to such problems can thus be reduced to checking the positive definiteness of a Hermitian matrix [16, 22].

Similar arguments can be made to reduce the problem of the *a posteriori* checking of the closed loop exponential stability and spatiotemporal \mathcal{H}_2 norm of an L_r system to the checking of the positive definiteness of a symmetric L_r operator, which in turn can be checked by solving a Riccati equation, as follows:

Lemma 11.1. *Let A be a stable matrix and $D = D^T$. The following statements are equivalent:*

1. $\bar{X} = L_r\{C, A, B, D, B^*, A^*, C^*\} \succ 0$
2. $D + C(zI - A)^{-1}B + B^T(z^*I - A^T)^{-1}C^T \succ 0, \quad \forall z \in \mathbb{T}$
3. $\exists P \succeq 0$, unique, such that: $P = APA^* + (B - APC^*)(D - CPC^*)^{-1}(B - APC^*)^*$ and $D - CPC^* \succ 0$ and $\rho(A - KC) < 1$ where $K = (B - APC^*)(D - CPC^*)^{-1}$

Proof. $1 \Leftrightarrow 2$ is due to the isomorphism between an L_r operator and its symbol, and $2 \Leftrightarrow 3$ is an application of the Positive Real Lemma ([23], Lemma 8.C.2)

In summary, while the sign iteration for solving Laurent operator Riccati and Lyapunov equations might have numerical difficulties, checking to see if a solution satisfies a Laurent operator Riccati or Lyapunov inequality can be reduced to checking the positive realness of an operator, which is a simple calculation (a finite dimensional Riccati equation) as in Lemma 11.1 statement 3. Hence a posteriori stability and performance analysis is numerically feasible.

3.6 Application to the Example Problem

In order to apply these methods to the platoon example, we thus simply need to write it as a set of interconnected subsystems as in (12):

$$\left[\begin{array}{ccccc} A & B^p & B^m & B^1 & B^2 \\ C^p & W^p & Z^m & L^p & V^p \\ C^m & Z^p & W^m & L^m & V^m \\ C^1 & J^p & J^m & D^{11} & D^{12} \\ C^2 & H^p & H^m & D^{21} & D^{22} \end{array} \right] = \left[\begin{array}{c|c|c|c|c|c} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_s^1 & g_s \\ \hline 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \\ -f_s^2 & 0 & \frac{f_s^2}{2} & \frac{f_s^2}{2} & 0 & 0 \\ f_s^1 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{array} \right] \quad (18)$$

then resolve the interconnection variables to get the L_r operator system, in which we can apply our structure preserving iterative techniques. We consider the \mathcal{H}_2 optimal state feedback problem ([24] 14.8.1) and hence only have to solve one Riccati equation, which, for our problem, reduces to:

$$\bar{A}^* \bar{X} + \bar{X} \bar{A} + \bar{C}_1^* \bar{C} - \bar{X} \bar{B}_2^* \bar{B}_2 \bar{X} = 0 \quad (19)$$

with controller $\bar{K} = -\bar{B}_2^* \bar{X}$. Where we will instead solve for some L_r operator $\tilde{\bar{X}}$ such that $\|\bar{X} - \tilde{\bar{X}}\| < \varepsilon$, such that $\tilde{K} = -\bar{B}_2^* \tilde{\bar{X}}$ will have a discrete distributed implementation as in Fig. 2.

4 Distributed Control Design for Decomposable Systems

The second distributed control method to be discussed relies on the concept of decomposable systems. A distributed controller synthesis method has been introduced in [25] that applies to systems of finite order with a special structure in the state-space matrices. We call this class of systems “decomposable systems”, as the

synthesis method exploits a kind of modal decomposition that is made possible by this structure. LMIs are used in order to impose the same structure on the controller as well.

In this section we are first going to explain the idea at the base of the method, and then we will show how it can be extended to infinite dimensional systems and applied to the infinite platoon problem.

4.1 General Description

We start by defining the class of systems to which the method applies.

Definition 11.1 (Decomposable systems). Let us consider the N nth order linear dynamical system described by:

$$\begin{cases} \dot{x}(t) = Ax(t) + B_1w(t) + B_2u(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \quad (20)$$

where x is the state, $u \in \mathbb{R}^{Nm_u}$ is the control input, $w \in \mathbb{R}^{Nm_w}$ the disturbance input and $z \in \mathbb{R}^{Nr}$ is the output. We call such systems “decomposable systems” iff there exists a diagonalizable “pattern matrix” $P \in \mathbb{R}^{N \times N}$ (with $P = Z^{-1}\Lambda Z$, where Λ is the diagonal matrix containing the eigenvalues) such that the state-space matrices can be parameterized as:

$$\begin{aligned} A &= I_N \otimes A_a + P \otimes A_b \\ B_\bullet &= I_N \otimes B_{\bullet,a} + P \otimes B_{\bullet,b} \\ C &= I_N \otimes C_a + P \otimes C_b \\ D &= I_N \otimes D_a + P \otimes D_b \end{aligned} \quad (21)$$

If the pattern matrix P is symmetric, then we call the system a “symmetric decomposable system”.

We then introduce the theorem that is at the base of the synthesis method for this class of systems, and which explains the name. We focus on symmetric decomposable systems as the pattern matrix will have only real eigenvalues, making the reasoning simpler.

Theorem 11.1. *A symmetric decomposable system of order Nn as described in Definition 11.1 is equivalent to N independent “modal” subsystems of order n . Each of these subsystems has only m_u control inputs, m_w disturbance inputs and r outputs:*

$$\begin{cases} \dot{\hat{x}}_i(t) = \mathbf{A}_i \hat{x}_i(t) + \mathbf{B}_{1,i} \hat{w}_i(t) + \mathbf{B}_{2,i} \hat{u}_i(t) \\ \hat{z}_i(t) = \mathbf{C}_i \hat{x}_i(t) + \mathbf{D}_i \hat{u}_i(t) \end{cases} \quad \text{for } i = 1, \dots, N \quad (22)$$

where for all the matrices in bold font it holds that:

$$\mathbf{M}_i = M_a + \lambda_i M_b \quad (23)$$

where the λ_i are the eigenvalues of P and the matrices M_a, M_b are defined as in (21). Conversely, it is true that all the sets of systems as in (22) for which the parameterization (23) holds are equivalent to a decomposable system.

Proof. The system is decomposed with the following change of variables:

$$\begin{aligned} x &= (Z \otimes I_n)\hat{x} \\ u &= (Z \otimes I_{m_u})\hat{u} \\ w &= (Z \otimes I_{m_w})\hat{w} \\ z &= (Z \otimes I_r)\hat{z} \end{aligned} \quad (24)$$

Every state-space matrix M in (20) is equivalent to a matrix in (22) through the following relation:

$$M = I_N \otimes M_a + P \otimes M_b = (Z \otimes I)\mathbf{M}(Z \otimes I)^{-1} \quad (25)$$

where \mathbf{M} is a block diagonal matrix containing all the matrices \mathbf{M}_i in the correct order in its diagonal. The idea that is exploited is that if Z diagonalizes P , then $(Z \otimes I)$ block-diagonalizes any matrix parameterized as in (21) thanks to the properties of the Kronecker product. For the details of the proof see [25]. \square

This last theorem basically says that a decomposable system (that is, a system for which all the state-space matrices M can be parameterized as $M = I_N \otimes M_a + P \otimes M_b$) is equivalent to a set of N independent systems with state-space matrices $\mathbf{M}_i = M_a + \lambda_i M_b$. Conversely, if a set of N independent systems has matrices which are parameterized as $\mathbf{M}_i = M_a + \lambda_i M_b$, then they are equivalent to a decomposable system. This is of fundamental importance, because:

1. For a decomposable system, it is possible to simplify control problems by evaluating them in the framework of the N independent systems, which are of much smaller order.
2. If the state-space matrices of the controllers of the N independent systems are parameterized as in (23) (with the same P), then the controller in its untransformed form will be a decomposable system with the same sparsity as the plant.

From now on, we will always use the bold font to identify matrices which can be parameterized according to (23).

As just seen, as consequence of Theorem 11.1, for decomposable systems control problems can be approached in the domain of the transformed variables, where the system is equivalent to a set of smaller independent modal subsystems. Once the solution has been obtained independently for each subsystem, one can retrieve the solution to the original problems through the inverse of the transformation shown in the proof of Theorem 11.1. Notice that the fact of working with a decomposed system does not imply that the final controller will be distributed or sparse, or decomposable as well; for this purpose, additional care will be needed, which we will discuss next.

For example, let us now consider the problem of finding a stabilizing static state feedback for the system in (20). The basic LMI approach for solving the problem is to find a feasible solution to the following inequalities [26]:

$$\begin{cases} X \succ 0 \\ AX + XA^T + B_2L + L^T B_2^T \prec 0 \end{cases} \quad (26)$$

where $X = X^T$ and L are decision variables; the static state feedback is given by the expression $K = LX^{-1}$. In the transformed domain, the LMI above is equivalent to the following set of smaller independent LMIs:

$$\begin{cases} X_i \succ 0 \\ (A_a + \lambda_i A_b)X + X(A_a + \lambda_i A_b)^T + (B_{2,a} + \lambda_i B_{2,b})L_i + L_i^T(B_{2,a} + \lambda_i B_{2,b})^T \prec 0 \\ \text{for } i = 1, \dots, N \end{cases} \quad (27)$$

where now $X_i = X_i^T$ and L_i are decision variables. If we just solve each of the N LMIs independently, then there will be a gain $K_i = L_i X_i^{-1}$ for each subsystem; but if we stack all these gains in a block diagonal matrix \hat{K} and perform the inverse transformation shown in (25), that is:

$$K = (Z \otimes I)\hat{K}(Z \otimes I)^{-1} \quad (28)$$

then this K will not be the matrix of a decomposable system as in (21), because \hat{K} is not parameterized according to (23) (it is not “bold”).

This problem can be solved by introducing a constraint in the LMI optimization that will force the K_i to have the structure that we want. This can be obtained by introducing the following coupling constraints to (27):

$$\begin{aligned} X_i &= X \\ L_i &= L_a + \lambda_i L_b \quad \text{for } i = 1, \dots, N \end{aligned}$$

and thus, the gains K_i will be parameterized according to (23):

$$K_i = (L_a + \lambda_i L_b)X^{-1} = K_a + \lambda_i K_b = \mathbf{K}_i$$

thus yielding a K that is the matrix of a decomposable system, which means that the final controller will have a sparse structure described by the pattern matrix; in fact, $K = I_n \otimes K_a + P \otimes K_b$. This approach is similar to the so-called *multiobjective optimization* [27]; some conservatism is introduced because we have set the same X matrix for all the LMIs. Since X is associated to the Lyapunov function of the closed loop system, this method is also called *Lyapunov shaping*. The approach that has been used here for finding a stabilizing feedback can be extended to a wider range of problems. For example, it can be used to approach the \mathcal{H}_2 problem. The optimal state feedback law of the kind $u = Kx$ for the system:

$$\begin{cases} \dot{x}(t) = Ax(t) + B_1w(t) + B_2u(t) \\ z(t) = Cx(t) + Du(t) \end{cases} \quad (29)$$

minimizing the \mathcal{H}_2 norm from w to z can be obtained by solving the following problem:

$$\begin{array}{ll} \text{minimize}_{\gamma(\lambda)} & \left\{ \begin{array}{l} XA^T + AX + L^T B_2^T + B_2 L + B_1 B_1^T \prec 0 \\ \left[\begin{array}{cc} X & XC^T + L^T D^T \\ * & W \end{array} \right] \succ 0 \\ \text{trace}(W) < \gamma^2 \end{array} \right. \\ \text{under:} & \end{array} \quad (30)$$

The application of this formula to a decomposable system and the introduction of the necessary constraints lead to the following theorem.

Theorem 11.2. *Consider a symmetric decomposable system (Definition 11.1). There exists a “decomposable” state feedback law of the kind: $u = (I_N \otimes K_a + P \otimes K_b)x$ that yields an \mathcal{H}_2 norm from w to z smaller than γ if the following LMI set (for $i = 1 \dots N$) is feasible:*

$$\begin{array}{ll} \left\{ \begin{array}{l} XA_i^T + A_i X + L_i^T B_{2,i}^T + B_{2,i} L_i + B_{1,i} B_{1,i}^T \prec 0 \\ \left[\begin{array}{cc} X & XC_i^T + L_i^T D_i^T \\ * & W_i \end{array} \right] \succ 0 \\ \sum_i \text{trace}(W_i) < \gamma^2 \end{array} \right. \\ \end{array} \quad (31)$$

where x_0 is the initial state at time $t = 0$, $A_i = A_a + \lambda_i A_b$, $W_i = W_a + \lambda_i W_b$ etc.; $X = X^T$, $W_a = W_a^T$, $W_b = W_b^T$, L_a and L_b are the decision variables. The controller is retrieved by the relations: $K_a = L_a X^{-1}$, $K_b = L_b X^{-1}$. Notice that the theorem expresses a suboptimality result (it is only a sufficient condition) due to the presence of constraints in the LMI.

4.2 Application to the Example Problem

4.2.1 Generalization to Infinite Dimensional Systems

The synthesis method in Theorem 11.2 applies to systems of finite order, but it can be easily generalized to systems of infinite dimension like the infinite platoon example. This is possible by replacing the pattern matrix with a bounded Laurent operator (for which we use the notation \bar{P}), and by replacing the standard \mathcal{H}_2 norm with the spatiotemporal one introduced in [7]. In this way the same results still hold formally, with the difference that the eigenvalues λ_i of the pattern matrix are replaced by the real spectrum $\lambda(\bar{P})$ of the pattern operator, which we consider as a function of $z = e^{j\theta}$ (with $\theta \in [0, 2\pi]$). With this reasoning, the \mathcal{H}_2 suboptimal problem would be solved by the following set of LMIs:

$$\left\{ \begin{array}{l} X\mathbf{A}_\theta^T + \mathbf{A}_\theta X + \mathbf{L}_\theta^T \mathbf{B}_{2,\theta}^T + \mathbf{B}_{2,\theta} \mathbf{L}_\theta + \mathbf{B}_{1,\theta} \mathbf{B}_{1,\theta}^T \prec 0 \\ \left[\begin{array}{c c} X & X\mathbf{C}_\theta^T + \mathbf{L}_\theta^T \mathbf{D}_\theta^T \\ * & \mathbf{W}_\theta \end{array} \right] \succ 0 \\ \text{for } \theta \in [0, 2\pi] \\ \int_0^{2\pi} \text{trace}(W_a + \lambda(e^{j\theta})W_b) d\theta < \gamma 2 \end{array} \right. \quad (32)$$

where now $\mathbf{A}_\theta = A_a + \lambda(e^{j\theta})A_b$, $\mathbf{W}_\theta = W_a + \lambda(e^{j\theta})W_b$ etc.

The problem consists of an infinite number of LMIs, or better, a set of parameter-varying LMIs depending either on the parameter θ or on the real spectrum $\lambda(e^{j\theta})$. In the case of $B_b = 0$ and $D_b = 0$, the LMIs are all affine in the parameter λ , so the convexity property of the LMIs allows the replacement of the parameter varying inequalities with only those for the extreme (maximum and minimum) values of the spectrum, reducing the problem from infinite to finite complexity. The result is summarized in the following theorem.

Theorem 11.3. Consider an infinite dimensional decomposable system (Definition 11.1), with a symmetric Laurent operator \bar{P} as pattern, and with $B_b = 0$ and $D_b = 0$. There exists a “decomposable” state feedback law of the kind: $u = (\bar{I} \otimes K_a + \bar{P} \otimes K_b)x$ that yields a spatiotemporal \mathcal{H}_2 norm from w to z smaller than γ if the following LMI set is feasible:

$$\left\{ \begin{array}{l} X(A_a + \bar{\lambda}A_b)^T + (A_a + \bar{\lambda}A_b)X + (L_a + \bar{\lambda}L_b)^T B_{2,a}^T + \\ + B_{2,a}(L_a + \bar{\lambda}L_b) + (B_{1,a} + \bar{\lambda}B_{1,b})(B_{1,a} + \bar{\lambda}B_{1,b})^T \prec 0 \\ \left[\begin{array}{c c} X & X(C_a + \bar{\lambda}C_b)^T + (L_a + \bar{\lambda}L_b)^T D_a^T \\ * & W_a + \bar{\lambda}W_b \end{array} \right] \succ 0 \\ X(A_a + \underline{\lambda}A_b)^T + (A_a + \underline{\lambda}A_b)X + (L_a + \underline{\lambda}L_b)^T B_{2,a}^T + \\ + B_{2,a}(L_a + \underline{\lambda}L_b) + (B_{1,a} + \underline{\lambda}B_{1,b})(B_{1,a} + \underline{\lambda}B_{1,b})^T \prec 0 \\ \left[\begin{array}{c c} X & X(C_a + \underline{\lambda}C_b)^T + (L_a + \underline{\lambda}L_b)^T D_a^T \\ * & W_a + \underline{\lambda}W_b \end{array} \right] \succ 0 \\ \text{trace}(W_a + \lambda_{av}W_b) < \frac{\gamma^2}{2\pi} \end{array} \right. \quad (33)$$

where $\bar{\lambda}$, $\underline{\lambda}$ and λ_{av} are respectively the maximum, minimum and average value of λ ($\lambda_{av} = \int_0^{2\pi} \lambda(e^{j\theta}) d\theta / 2\pi$). The decision variables are $X = X^T$, $W_a = W_a^T$, $W_b = W_b^T$, L_a and L_b , and the controller is retrieved by the relations: $K_a = L_a X^{-1}$, $K_b = L_b X^{-1}$.

4.2.2 The Platoon

The tools of Theorem 11.3 can be applied directly to the platoon example, for which we will have:

$$A_a = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B_{1,a} = \begin{bmatrix} 0 \\ q \end{bmatrix}, \quad B_{2,a} = \begin{bmatrix} 0 \\ g \end{bmatrix},$$

$$C_a = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ f_1 & 0 \end{bmatrix}, \quad C_b = \begin{bmatrix} 0 & 0 \\ f_2 & 0 \\ 0 & 0 \end{bmatrix}, \quad D_a = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad (34)$$

and the other matrices are zero. The pattern is given by:

$$\bar{P} = \frac{1}{2} \begin{bmatrix} \ddots & \ddots & \ddots & \ddots \\ & 1 & -2 & 1 \\ & & 1 & -2 & 1 \\ & & & 1 & -2 & 1 \\ & & & & \ddots & \ddots & \ddots \end{bmatrix} \quad (35)$$

It is easy to show that the real spectrum of this pattern operator is in the interval $[-2, 0]$, with -1 as average value. The operator that diagonalizes \bar{P} is again simply the Fourier operator \mathcal{F} , which we can use to compute the symbol of the operator as a function of $z \in \mathbb{T}$:

$$\mathcal{F}\bar{P}\mathcal{F}^{-1} = \frac{1}{2}z^{-1} - 1 + \frac{1}{2}z = \cos \theta - 1 \in [-2, 0] \quad (36)$$

and as a consequence, $\lambda_{\text{av}} = \int_0^{2\pi} (\cos \theta - 1) d\theta / 2\pi = -1$.

We remark that the extension of the results of [25] to infinite dimensional, spatially invariant systems, brings to results that are quite close to the ones shown in the earlier paper [28]. In this reference, the reduction of the number of LMIs from infinite to finite is done by means of application of the discrete-time Kalman–Yakubovich–Popov lemma [29]. In this article instead we have shown that the reduction can be done just by exploitation of the convexity properties of the LMIs, provided that the pattern operator is symmetric (and thus it has a real spectrum).

5 Distributed LQR of Identical Systems

If stability and design simplicity of the distributed control system are more important than optimality, then the complexity of the control design task can be further reduced. In this section we provide a conceptually simple way of generating stabilizing structured controllers for the special class of systems under consideration, using the solution of small-sized local LQR problems. These results rely on robustness and other special properties of LQR solutions for identical dynamically decoupled systems. They enable the construction of structured stabilizing, suboptimal controllers for finite-dimensional systems as presented in [30]. In Sect. 5.1 we provide an overview of these properties and then extend their applicability to the infinite-dimensional string system under study in Sect. 5.2.

5.1 Special Properties of LQR for Dynamically Decoupled Systems

Reference [30] studies distributed LQR design for identical dynamically decoupled systems and proposes a structured controller design procedure based on local stabilizing LQR controllers. A similar problem setup with a special class of symmetrically interconnected systems is treated in [31]. In the following, we summarize the main properties of the local LQR solutions in [30].

Consider a set of N_L identical, decoupled linear time-invariant dynamical systems, the i th system being described by the continuous-time state equation:

$$\begin{aligned}\dot{x}_i &= Ax_i + Bu_i, \\ x_i(0) &= x_{i0}.\end{aligned}\tag{37}$$

where $x_i(t) \in \mathbb{R}^n$, $u_i(t) \in \mathbb{R}^m$ are states and inputs of the i th system at time t , respectively. Let $\tilde{x}(t) \in \mathbb{R}^{nN_L}$ and $\tilde{u}(t) \in \mathbb{R}^{mN_L}$ be the vectors which collect the states and inputs of the N_L systems at time t :

$$\begin{aligned}\dot{\tilde{x}} &= \tilde{A}\tilde{x} + \tilde{B}\tilde{u}, \\ \tilde{x}(0) &= \tilde{x}_0 \triangleq [x_{10}, \dots, x_{N_L 0}]^T,\end{aligned}\tag{38}$$

with

$$\begin{aligned}\tilde{A} &= I_{N_L} \otimes A, \\ \tilde{B} &= I_{N_L} \otimes B.\end{aligned}\tag{39}$$

We consider an LQR control problem for the set of N_L systems where the cost function couples the dynamic behavior of individual systems. This cost function contains terms which weigh the i th system states and inputs, as well as the difference between the i th and the j th system states and can be written using the following compact notation:

$$J(\tilde{u}(t), \tilde{x}_0) = \int_0^\infty (\tilde{x}(\tau)^T \tilde{Q} \tilde{x}(\tau) + \tilde{u}(\tau)^T \tilde{R} \tilde{u}(\tau)) d\tau,\tag{40}$$

where the matrices \tilde{Q} and \tilde{R} have a special structure defined next. \tilde{Q} and \tilde{R} can be decomposed into N_L^2 blocks of dimension $n \times n$ and $m \times m$ respectively:

$$\tilde{Q} = \begin{bmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} & \cdots & \tilde{Q}_{1N_L} \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{Q}_{N_L 1} & \cdots & \cdots & \tilde{Q}_{N_L N_L} \end{bmatrix}, \tilde{R} = \begin{bmatrix} R & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \cdots & R \end{bmatrix}\tag{41}$$

with

$$\begin{aligned}\tilde{Q}_{ii} &= Q_{ii} + \sum_{k=1, k \neq i}^{N_L} Q_{ik}, \quad \forall i \in \{1, \dots, N_L\}. \\ \tilde{Q}_{ij} &= -Q_{ij}, \quad \forall i, j \in \{1, \dots, N_L\}, \quad i \neq j. \\ \tilde{R} &= I_{N_L} \otimes R.\end{aligned}\tag{42}$$

where

$$R_{ii} = R_{ii}^T = R > 0, \quad Q_{ii} = Q_{ii}^T \geq 0 \quad \forall i,\tag{43a}$$

$$Q_{ij} = Q_{ij}^T = Q_{ji} \geq 0 \quad \forall i \neq j.\tag{43b}$$

Using the cost function defined above, let \tilde{K} and $\tilde{x}_0^T \tilde{P} \tilde{x}_0$ be the optimal controller and the value function corresponding to the following LQR problem:

$$\begin{aligned}\min_{\tilde{u}} \quad & J(\tilde{u}, \tilde{x}_0) \\ \text{subj. to} \quad & \dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}\tilde{u} \\ & \tilde{x}(0) = \tilde{x}_0\end{aligned}\tag{44}$$

We will assume that a stabilizing solution to the LQR problem (44) with finite performance index exists and is unique (see [32], p. 52 and references therein).

It is well known that

$$\tilde{K} = -\tilde{R}^{-1} \tilde{B}^T \tilde{P},\tag{45}$$

where \tilde{P} is the symmetric positive definite solution to the following ARE:

$$\tilde{A}^T \tilde{P} + \tilde{P} \tilde{A} - \tilde{P} \tilde{B} \tilde{R}^{-1} \tilde{B}^T \tilde{P} + \tilde{Q} = 0\tag{46}$$

Theorem 11.4. [30] Assume the weighting matrices (42) of the LQR problem (44) are chosen as

$$\begin{aligned}Q_{ii} &= Q_1 \quad \forall i \in \{1, \dots, N_L\} \\ Q_{ij} &= Q_2 \quad \forall i, j \in \{1, \dots, N_L\}, \quad i \neq j.\end{aligned}\tag{47}$$

Let $\tilde{x}_0^T \tilde{P} \tilde{x}_0$ be the value function of the LQR problem (44) with weights (47), and the blocks of the matrix \tilde{P} be denoted by $\tilde{P}_{ij} = \tilde{P}[(i-1)n : in, (j-1)n : jn]$ with $i, j \in \{1, \dots, N_L\}$.

Then,

(I) $\sum_{j=1}^{N_L} \tilde{P}_{ij} = P$ for all $i \in \{1, \dots, N_L\}$, where P is the symmetric positive definite solution of the ARE associated with a single node local problem:

$$A^T P + PA - PBR^{-1}B^T P + Q_1 = 0.\tag{48}$$

- (II) $\sum_{j=1}^{N_L} \tilde{K}_{ij} = K$ for all $i \in \{1, \dots, N_L\}$, where $K = -R^{-1}B^TP$.
(III) $\tilde{P}_{ij} = \tilde{P}_{lm} \triangleq \tilde{P}_2 \forall i \neq j, \forall l \neq m$ is a symmetric negative semidefinite matrix.

Proof. See in [30]. \square

The following corollary of Theorem 11.4 follows from the stability and the robustness of the LQR controller $-R^{-1}B^T(-N_L\tilde{P}_2)$ for system $A - XP$ as shown in [30].

Corollary 11.1. $A - XP + \alpha N_L X \tilde{P}_2$ is a Hurwitz matrix for all $\alpha > \frac{1}{2}$, and $\alpha = 0$.

This corollary can be relaxed to any $\alpha \geq 0$ for a wide class of systems as discussed in [30].

In Sect. 5.2 we show how to construct a distributed stabilizing suboptimal controller for the infinite-dimensional application example at hand using the properties described above.

5.2 Application to the Example Problem

As a key ingredient to the extension of the above results to infinite-dimensional systems, we will establish the following lemma, which forms a basis for the results described in this section.

Lemma 11.2. Given $A, C \in \mathbb{R}^{m \times m}$ and a Laurent operator \bar{B} with scalar symbol, bounded on $l^2(\mathbb{Z})$, consider two infinite-dimensional Laurent operators $\bar{A} = \bar{I} \otimes A$ and $\bar{C} = \bar{B} \otimes C$, where \bar{I} stands for the infinite-dimensional matrix with ones on its diagonal and zeros elsewhere. Then the spectrum of $\bar{A} + \bar{C}$ can be characterized as:

$$\lambda(\bar{A} + \bar{C}) = \lambda(A + \lambda(\bar{B})C), \quad (49)$$

where $\lambda(\bar{B})$ is the spectrum of \bar{B} .

Proof. Let us denote the infinite-dimensional Laurent matrix under investigation as $\bar{X} = \bar{A} + \bar{C}$, and its symbol

$$\mathcal{F}\bar{X}\mathcal{F}^{-1} = X(z) = A + B(z)C = A + \lambda(B(z))C, \quad z = e^{j\theta} \in \mathbb{T} \quad (50)$$

since $B(z) = \lambda(B(z))$ is a scalar, where λ denotes the spectrum and \mathbb{T} is the unit circle.

Since there is an equivalence in terms of spectrum between the Laurent operator and its symbol [13], we can express the spectrum of \bar{X} in the following way:

$$\begin{aligned} \lambda(X(\mathbb{T})) &= \lambda(A + \lambda(B(\mathbb{T}))C) \\ &\Updownarrow \\ \lambda(\bar{X}) &= \lambda(A + \lambda(\bar{B})C) \end{aligned} \quad (51)$$

This proves the lemma, which can be readily extended to block Laurent matrices based on [13]. \square

This result can be interpreted as an extension of Proposition 2 in [30] to the case where the identity matrix and \bar{B} form infinite-dimensional “pattern” matrices. The following theorem will propose a procedure to construct stabilizing structured controllers for infinite-dimensional systems using the local LQR controller as building blocks.

Theorem 11.5. Consider the LQR problem (44) with N_L and weights chosen as in (47) and its solution $\tilde{K} = \mathbf{1} \otimes K_2 - I_{N_L} \otimes K_2 + I_{N_L} \otimes K_1$ and $\tilde{K} = \mathbf{1} \otimes \tilde{P}_2 + I_{N_L} \otimes (P - N_L \tilde{P}_2)$. (The matrix $\mathbf{1} \in \mathbb{R}^{N_L \times N_L}$ has ones in all its entries.)

Let \bar{M} be a bounded Laurent operator whose spectrum satisfies the following property¹:

$$\lambda(M) > \frac{N_L}{2}. \quad (52)$$

Construct a state feedback controller as the following Laurent matrix:

$$\tilde{K} = -\bar{I} \otimes R^{-1} B^T P + \bar{M} \otimes R^{-1} B^T \tilde{P}_2. \quad (53)$$

Then, the closed loop system

$$\bar{A}_{cl} = \bar{I} \otimes A + (\bar{I} \otimes B) \tilde{K} \quad (54)$$

is asymptotically stable.

Proof. Consider the spectrum of the closed-loop system matrix \bar{A}_{cl} :

$$\lambda(\bar{A}_{cl}) = \lambda(\bar{I} \otimes (A - XP) + \bar{M} \otimes (X \tilde{P}_2))$$

By Proposition 11.2:

$$\lambda(\bar{I} \otimes (A - XP) + \bar{M} \otimes (X \tilde{P}_2)) = \lambda(A - XP + \lambda(\bar{M}) X \tilde{P}_2). \quad (55)$$

From Corollary 11.1 and from condition (52), we conclude that $A - XP + \lambda(\bar{M}) X \tilde{P}_2$ is Hurwitz for all $\lambda(\bar{M})$ thus the closed-loop system is asymptotically stable. \square

Theorem 11.5 has several main consequences:

1. The controller \tilde{K} in (53) will have the same sparsity structure as the Laurent operator \bar{M} , which leads to an asymptotically stable *distributed* controller, if \bar{M} is sparse, banded or otherwise structured.
2. We can use *one local* LQR controller to compose distributed stabilizing controllers for an infinite collection of identical dynamically decoupled subsystems.
3. The first two consequences imply that we can not only find a stabilizing distributed controller with a desired sparsity pattern (which is in general a formidable task by itself), but it is enough to solve a small finite-dimensional problem. This attractive feature relies on the specific problem structure and LQR robustness properties defined in Sect. 5.1.

¹ This property can be relaxed to $\lambda(\bar{M}) \geq 0$ for a wide class of systems.

4. The result is independent of the local LQR tuning. Thus Q_1 , Q_2 and R in (47) can be used in order to influence the compromise between minimization of absolute and relative terms, and the control effort in the global performance.

For the platoon example, the following weighting matrices were used to solve a 3-by-3 local LQR problem ($N_L = 3$):

$$Q_1 = \begin{bmatrix} f_1^2 & 0 \\ 0 & 0 \end{bmatrix}, \quad Q_2 = \begin{bmatrix} \frac{f_2^2}{4} & 0 \\ 0 & 0 \end{bmatrix}, \quad R = 1.$$

The stabilizing controller having the same banded structure as the cost function is built in the following way. Using real parameters $b \geq 0$ and a , construct the global controller gain matrix \tilde{K} as

$$\tilde{K} = \begin{bmatrix} \ddots & \ddots & \ddots & & & \\ & \cdots & 0 & K_1 & 0 & \cdots \\ & & & \ddots & \ddots & \ddots \end{bmatrix} + \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots & \\ & \cdots & 0 & bK_2 & aK_2 & bK_2 & 0 & \cdots \\ & & & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \quad (56)$$

where the gain matrices K_1 and K_2 are obtained from the local LQR solution as explained in [30]. The closed-loop state matrix can be written as

$$\bar{A}_{cl} = \bar{I} \otimes A + (\bar{I} \otimes B)\tilde{K}. \quad (57)$$

The controller \tilde{K} has $R^{-1}B^T(\tilde{P}_1 - a\tilde{P}_2)$ on its diagonal entries and $-bR^{-1}B^T\tilde{P}_2$ on its two sub-diagonal entries. Using Theorem 11.4, \tilde{K} can be written as

$$\tilde{K} = -\bar{I} \otimes (R^{-1}B^T P) + \bar{M} \otimes (R^{-1}B^T \tilde{P}_2).$$

The Laurent operator \bar{M} , which defines the ‘‘pattern’’ of the distributed controller has $N_L - 1 - a = 2 - a$ on its diagonal and $-b$ on its two subdiagonals. This means that the spectrum of the symbol of \bar{M} on the unit circle is $2 - a - be^{j\theta} - be^{-j\theta}$. Thus \bar{M} has a purely real spectrum with elements belonging to the interval $[2 - a - 2b, 2 - a + 2b]$. Following similar arguments as in [30], this leads to the sufficient stability condition of $a + 2b \leq \frac{1}{2}$ (or $a + 2b \leq 2$ for a large class of systems as explained in the above reference). For the specific system under consideration the values of $a = 1.98$ and $b = 0.01$ were chosen.

6 Numerical Results of the Car Platoon Benchmark Problem

This section presents the numerical results of the application of the different controllers to the benchmark problem. We computed the controllers for different values of the weighting parameter f_2 while keeping $f_1 = 1$. A larger value of f_2 indicates

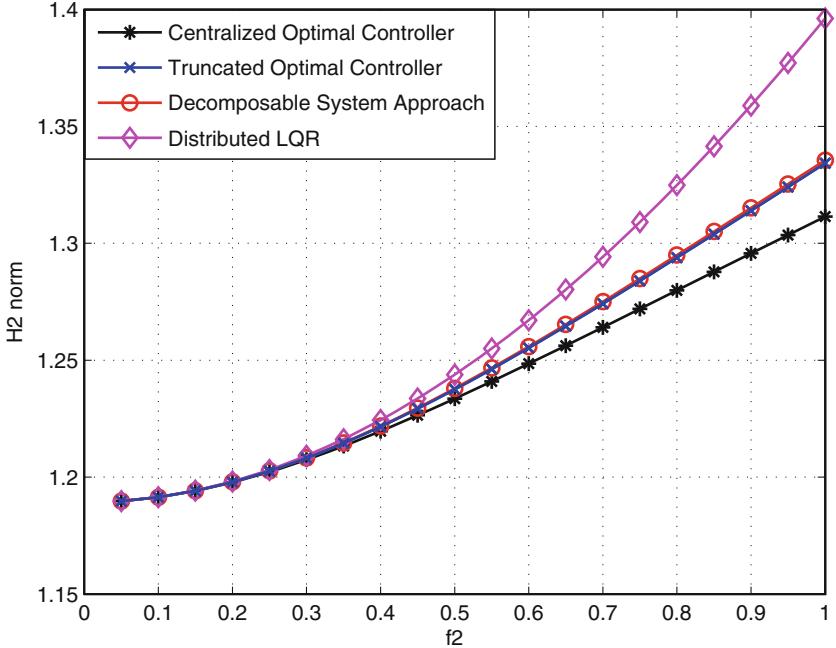


Fig. 3 Closed-loop performance of the four different controllers

an increased penalty on the relative positions of the vehicles in the platoon, while for $f_2 \rightarrow 0$ the problem turns into the control of each single vehicle independently from the others.

The results are depicted in Fig. 3, which shows the spatiotemporal \mathcal{H}_2 norm achieved with:

1. The optimal controller from Sect. 2.1 (identical in performance to the controller produced with the method of Sect. 3, to within 10^{-15});
2. The truncated version of this global controller, with only one off-diagonal band;
3. The decomposable system approach described in Sect. 4;
4. The distributed LQR controller of Sect. 5.

The centralized ϵ -suboptimal controller yields the best performance, while its truncated version is slightly better than the controller given by the decomposition approach, followed by the distributed LQR method, as could be expected since performance is not explicitly considered in that technique. Note as the coupling weight becomes small ($f_2 \rightarrow 0$), the four controllers tend to converge to the same performance.

Figure 4 illustrates the magnitude of the entries in the centralized optimal controller gain, providing some justification for the use of its truncated versions.

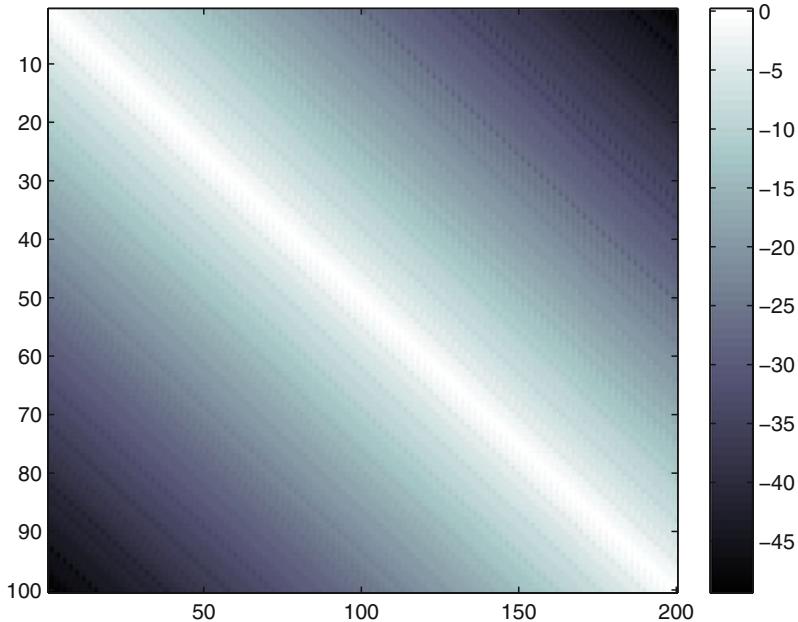


Fig. 4 A visualization of a section around the diagonal of the optimal centralized gain operator \bar{K} . The colors represent $\log_{10}(\cdot)$ of the magnitude of the entries of the matrix. The exponential spatial decay of the operator is apparent

7 Conclusions and Open Problems

We have presented three conceptually quite different approaches for distributed control design for large-scale systems. They all utilized special properties of the problem structure in order to arrive at various degrees of simplification for the large-scale control design problem.

Although they are applicable to wider classes of problems (as stated in the corresponding sections), the three approaches were applied for solving a benchmark control example taken from the literature involving an infinite platoon of vehicles. The resulting performance was compared with an ϵ -suboptimal centralized controller.

Despite the progress that can be observed in the development of various distributed control techniques for special classes of large-scale systems, several important challenges remain to be solved in this field. Below we list a few of them pertaining to the three approaches discussed in this work.

Although the iterative, Laurent operator structure preserving approach in Sect. 3 provides an attractive numerical solution for the class of systems under study, it does leave the question of designing optimal controllers with a specified width of their banded structure open. Optimal design for a fixed rational order remains a challenging problem as well. It would be of practical interest also to study suboptimal control design for which the sign iteration converges in only a few iterations, e.g., by specifying a certain region of closed-loop spectrum.

Using the decomposition-based method of Sect. 4, we have seen that conservatism is introduced by assuming a common Lyapunov matrix for all the modal sub-systems. An interesting research question would be whether it is possible to reduce this conservatism by employing robust control methods other than multi-objective optimization.

Although the local LQR solutions of Sect. 5 provide a very simple way of obtaining stabilizing solutions for a special class of systems under a wide variety of interconnections, obtaining near-optimal or guaranteed suboptimal controllers with this procedure remains an important open challenge. Addressing dynamically coupled systems has been attempted only in [31] for special symmetric interconnections. The powerful robustness properties of local LQR solutions could also allow the scheme to work with time-varying interconnections and a wider variety of weights in the cost function.

References

1. N. R. Sandell Jr., P. Varaiya, M. Athans, and M. G. Safonov. Survey of decentralized control methods for large scale systems. *IEEE Trans. Autom. Contr.*, 23:108–129, 1978
2. R. D’Andrea and G. E. Dullerud. Distributed control design for spatially interconnected systems. *IEEE Trans. Autom. Contr.*, 48:1478–1495, 2003
3. Roger W. Brockett and Jacques L. Willems. Discretized partial differential equations: Examples of control systems defined on modules. *Automatica*, 10:507–515, 1974
4. Morten Hovd, Richard D. Braatz, and Sigurd Skogestad. SVD controllers for H_2 -, H_∞ - and μ -optimal control. *Automatica*, 33:433–439, 1997
5. Jeremy G. VanAntwerp, Andrew P. Featherstone, and Richard D. Braatz. Robust cross-directional control of large scale sheet and film processes. *J. Process Contr.*, 11:149–177, 2001
6. M. R. Jovanović and B. Bamieh. On the ill-posedness of certain vehicular platoon control problems. *IEEE Trans. Autom. Contr.*, 50(9):1307–1321, September 2005
7. B. Bamieh, F. Paganini, and M. A. Dahleh. Distributed control of spatially invariant systems. *IEEE Trans. Autom. Contr.*, 47(7):1091–1107, 2002
8. E. W. Kamen. *Multidimensional systems Theory: Stabilization of Linear Spatially-Distributed Continuous-Time and Discrete-Time Systems*, pages 101–146 Norwell, MA: Kluwer, 1985
9. B. Bamieh, F. Paganini, and M. A. Dahleh. Distributed control of spatially invariant systems. *IEEE Trans. Autom. Contr.*, 47:1091–1106, 2002
10. G. A. de Castro and F. Paganini. Convex synthesis of localized controllers for spatially invariant systems. *Automatica*, 38:445–456, 2002
11. Rufus Fraanje. Private communication
12. A. Böttcher and B. Silberman. *Introduction to Large Truncated Toeplitz Matrices*. Springer, Berlin, 1991
13. I. Gohberg, S. Goldberg, and M. Kaashoek. *Classes of Linear Operators Vol. II*. Birkhauser, Basel, 1993
14. R. Curtain, O. Iftime, and H. Zwart. System theoretic properties of platoon-type systems. *Proceedings of the IEEE Conference on Decision and Control*, 2008
15. S. Boyd and V. Balakrishnan. A regularity result for the singular values of a transfer matrix and a quadratically convergent algorithm for computing its L_∞ norm. *Syst. Contr. Lett.*, 15:1–7, 1990
16. J. K. Rice and M. Verhaegen. Distributed control: A sequentially semi-separable approach for heterogeneous linear systems. *IEEE Trans. Autom. Contr.*, 54:1270–1283, 2009

17. J. Rice and M. Verhaegen. Distributed control of spatially invariant systems using fast iterative solutions to rationally parametric matrix problems. *Proceedings of IEEE Conference Decision and Control*, 2009
18. J. D. Roberts. Linear model reduction and solution of the algebraic Riccati equation by use of the sign function. *Int. J. Contr.*, 32:677–687, 1980
19. C. S. Kenney and A. J. Laub. The matrix sign function. *IEEE Trans. Autom. Contr.*, 40: 1330–1348, 1995
20. L. Grasedyck, W. Hackbusch, and B. N. Khoromskij. Solution of large scale algebraic matrix Riccati equations by use of hierarchical matrices. *Computing*, 70:121–165, 2003
21. C. Kenney, A. J. Laub, and E.A. Jonckheere. Positive and negative solutions of dual Riccati equations by matrix sign function iteration. *Syst. Contr. Lett.*, 13:109–116, 1989
22. J. K. Rice and M. Verhaegen. Distributed control: A sequentially semi-separable approach. *Proceedings of IEEE Conference Decision and Control*, 2008
23. T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice-Hall, Englewood Cliffs, 2000
24. K. Zhou, K. Glover, and J. C. Doyle. *Robust and Optimal Control*. Prentice-Hall, Englewood Cliffs, 1996
25. P. Massioni and M. Verhaegen. Distributed control for identical dynamically coupled systems: a decomposition approach. *IEEE Trans. Autom. Contr.*, 54(1):124–135, January 2009
26. C. Scherer and S. Weiland. Linear matrix inequalities in control. Online, <http://www.dcsc.tudelft.nl/~cscherer/2416/lmi05.pdf>, 2005, Lecture notes
27. C. Scherer, P. Gahinet, and M. Chilali. Multiobjective output-feedback control via LMI optimization. *IEEE Trans. Autom. Contr.*, 42(7):896–911, 1997
28. G.A. de Castro and F. Paganini. Convex synthesis of localized controllers for spatially invariant systems. *Automatica*, 38(3):445–456, 2002
29. A. Rantzer. On the Kalman–Yakubovich–Popov lemma. *Syst. Contr. Lett.*, 28(1):7–10, 1996
30. F. Borrelli and T. Keviczky. Distributed LQR design for identical dynamically decoupled systems. *IEEE Trans. Autom. Contr.*, 53(8):1901–1912, August 2008
31. M. K. Sundareshan and R. M. Elbanna. Qualitative analysis and decentralized controller synthesis for a class of large-scale systems with symmetrically interconnected subsystems. *Automatica*, 27(2):383–388, 1991
32. B. D. O. Anderson and J. B. Moore. *Optimal Control: Linear Quadratic Methods*. Prentice-Hall, Englewood Cliffs, 1990

Integrated Design of Large-Scale Collocated Structural System and Control Parameters Using a Norm Upper Bound Approach

Mona Meisami-Azad, Javad Mohammadpour Velni, Kazuhiko Hiramoto,
and Karolos M. Grigoriadis

1 Introduction

The traditional design optimization of system parameters and a feedback controller usually follows a sequential strategy, where the open-loop system parameters are optimized first, followed by the controller design. However, this design strategy will not lead to an optimal closed-loop performance due to the coupled nature of the plant and controller optimization problems. Specifically, this two-step design methodology does not utilize the full design freedom to achieve an optimal overall system. It has been shown that the overall system performance can be significantly improved if the design process of the plant and the control system is integrated [17, 18, 28–30]. The integrated design strategy corresponds to a simultaneous optimization of the design parameters of both the plant and the controller to satisfy desired set of design specifications and to optimize the closed-loop system performance. Past research work has demonstrated the improved performance achieved using the integrated design strategy compared to the sequential method of design. However, the general integrated plant/controller design optimization problem turns out to be a complex nonlinear nonconvex optimization problem and no solution is guaranteed to provide convergence to the global optimum of the design variables [1, 13, 21, 28, 32, 33]. This makes the integrated design strategy computationally challenging. Recently, several integrated \mathcal{H}^∞ -based plant/controller design approaches have been proposed using a Linear Matrix Inequality (LMI) formulation of the control problem to take advantage of systematic approaches developed in the past two decades for robust control design [12, 20, 27]. In principle, these formulations result in Bilinear Matrix

M. Meisami-Azad, J.M. Velni, and K.M. Grigoriadis
Department of Mechanical Engineering, University of Houston, Houston, TX 77204, USA
e-mail: mmeisami@mail.uh.edu; jmohammadpour@uh.edu; karolos@uh.edu

K. Hiramoto
Department of Mechanical and Production Engineering, Niigata University,
8050 Ikarashi-2-no-cho, Nishi-ku, Niigata 950-2181, Japan
e-mail: hiramoto@eng.niigata-u.ac.jp

Inequality (BMI) problems even if one assumes that the coefficient matrices of the plant state-space representation are linear functions of the system design parameters. In past attempts, the BMI formulation of the integrated plant/controller design has been solved using iterative LMI-based optimization schemes. However, these iterative methods are also unable to guarantee convergence to the optimum solution, and they are computationally intensive. Yang and Lee [38] proposed an analytical model for structural control optimization in which LQR feedback control parameters along with the non-collocated sensor/actuator placement were considered as independent design variables. Iwasaki et al. [15] proposed a design framework based on a finite frequency positive realness property and formulated the corresponding design condition in terms of LMIs. And recently Hiramoto and Grigoriadis [14] proposed an integrated design methodology using a homotopy approach to simultaneously optimize structural system and control design parameters. The approach taken in the latter work is an iterative one which introduces a small amount of perturbation in the coefficient matrices of the state-space representation of the plant and the matrices associated with the controller. They show that in case of small perturbations, the optimization problem can be cast as an LMI. However, the general integrated design problem is in form of a BMI and the number of iterations for the convergence of the algorithm significantly increases due to the linear approximations of the nonlinear problem.

The control design problem for structural systems with collocated sensors and actuators has been shown to provide great advantages from a stability, passivity, robustness and an implementation perspective [16]. For example, collocated control can easily be achieved in a space structure when an altitude rate sensor is placed at the same location as a torque actuator [4, 11]. Collocation of sensors and actuators leads to externally symmetric transfer functions [5]. Several other classes of engineering systems, such as circuit systems, chemical reactors and power networks, can be modeled as systems with symmetric transfer functions [2].

In this chapter, we present an efficient and computationally tractable design method in order to integrate the structural parameters and control gain design in collocated structural systems using \mathcal{H}^2 or \mathcal{H}^∞ norm closed-loop performance specifications. Specifically, the objective is to determine the optimal values of the damping parameters of the structural system and to simultaneously design optimal output feedback gains such that an upper-bound on the closed-loop system norm (either in the \mathcal{H}^2 or the \mathcal{H}^∞ setting) from the disturbance input signals to the desired outputs is minimized. The theoretical development of this chapter takes advantage of recently developed control-oriented algebraic tools to formulate the simultaneous damping and control gain design problem as an LMI optimization problem. LMI optimization problems involve the minimization of a linear objective function subject to matrix inequality constraints that are linear with respect to the parameter variables [6, 7]. An LMI optimization problem is essentially a generalization of the linear programming (LP) problem to the matrix case with positive definite matrix inequality constraints. The LMI problems are convex and can be efficiently solved using interior-point optimization solvers [10, 26].

A fundamental idea behind the work presented in this chapter is to demonstrate how to take advantage of the particular structure of collocated structural systems to develop explicit upper bound expressions for the \mathcal{H}^2 and \mathcal{H}^∞ norms of such systems. Following this idea, we demonstrate the formulation of the integrated damping parameters and output feedback control gain design as a convex optimization problem with respect to the unknown damping and control variables subject to practical constraints on the damping parameters and the feedback control gains. Hence, unlike past approaches, the proposed method in this chapter for integrated design is not based on an iterative procedure to determine the design parameters. The benefits of the proposed integrated design become apparent for large-scale structural systems, where the use of iterative methods for integrated design becomes intractable or even fails.

The notation used throughout this chapter is standard. The notation $>$ ($<$) is used to denote the positive (negative) definite symmetric matrices. The i th eigenvalue of a real symmetric matrix X will be denoted by $\lambda_i(X)$ where the ordering of the eigenvalues is defined as $\lambda_{\max}(X) = \lambda_1(X) \geq \lambda_2(X) \geq \dots \geq \lambda_n(X)$. The maximum singular value of a matrix Y will be denoted by $\sigma_{\max}(Y)$, which is also referred to as its spectral norm $\|Y\|$.

2 Symmetric Output Feedback Control of Collocated Systems

Consider the following vector second-order representation of a structural system with collocated actuators and sensors

$$\begin{aligned} M\ddot{q}(t) + D\dot{q}(t) + Kq(t) &= Fu(t) + Ew(t) \\ y(t) &= F^T\dot{q}(t) \\ z(t) &= E^T\dot{q}(t) \end{aligned} \quad (1)$$

where $q(t) \in \mathbb{R}^n$ is the coordinate vector, $u(t) \in \mathbb{R}^m$ is the control input vector, $w(t) \in \mathbb{R}^k$ is the disturbance vector, $y(t) \in \mathbb{R}^m$ is the measured output vector, and $z(t) \in \mathbb{R}^k$ is the controlled output vector. The matrices M , D and K are assumed to be symmetric positive definite representing the structural system mass, damping and stiffness distribution, respectively. The above finite-dimensional representation is often encountered in the dynamics of structural systems resulting from a finite element approximation of distributed parameter structural systems [11]. It is noted that velocity feedback as in (1) is common in the collocated control of structural systems through a velocity sensor, a displacement sensor with a derivative controller or an accelerometer with an integral controller [5]. In smart structures with piezoelectric sensors velocity feedback can be readily achieved through direct strain rate feedback [8].

The symmetric static output feedback control problem is to design a symmetric static feedback gain G such that the output feedback control law

$$u(t) = -Gy(t) \quad (2)$$

renders the closed-loop system stable with appropriate closed-loop performance.

The closed-loop system has a state-space realization as follows

$$\begin{aligned}\dot{x} &= Ax + Bw \\ z &= Cx\end{aligned}\tag{3}$$

with

$$\begin{aligned}A &= \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}(D + FGF^T) \end{bmatrix} \\ B &= \begin{bmatrix} 0 \\ M^{-1}E \end{bmatrix}, \quad C = [0 \ E^T]\end{aligned}\tag{4}$$

where $x^T = [q^T \ \dot{q}^T]$. The system (3)–(4) is an *externally symmetric* state-space realization since its transfer function is symmetric. This implies that there exists a nonsingular matrix T such that

$$A^T T = TA, \quad C^T = TB\tag{5}$$

This class of systems is more general than the class of *internally* or *state-space symmetric* systems that satisfy the symmetry conditions (5) with a positive definite transformation matrix T . An analytical solution to the \mathcal{H}^∞ control problem for internally symmetric systems has been presented by Tan and Grigoriadis [37].

3 Upper Bounds on Collocated Structural System Norms

Recall that the \mathcal{H}^∞ norm of the system (3) is given by

$$\|H\|_\infty = \sup_{\omega \in \mathbb{R}} \sigma_{\max}\{H(j\omega)\}\tag{6}$$

where $H(s)$ is the transfer function of the system [39]. In a time-domain interpretation, the \mathcal{H}^∞ norm corresponds to the energy (or \mathcal{L}_2 norm) gain of the system from the input w to the output z . Hence, in this setting the \mathcal{H}^∞ norm defines a disturbance rejection property of the system.

The collocated \mathcal{H}^∞ control synthesis problem is to design a symmetric control law (2) that stabilizes the closed-loop system and guarantees an \mathcal{H}^∞ norm less than a prescribed bound $\gamma > 0$. The following result shows that for an open-loop vector second-order realization (3)–(4) (i.e., with $G = 0$), an upper bound on its \mathcal{H}^∞ norm can be computed using a simple explicit formula [3].

Theorem 1 Consider the open-loop ($u(t) = 0$) vector second-order system realization (1). The system has an \mathcal{H}^∞ norm γ from $w(t)$ to $z(t)$ that satisfies

$$\gamma < \bar{\gamma} = \lambda_{\max}(E^T D^{-1} E)\tag{7}$$

The explicit bound of the above result is obtained using the Bounded Real Lemma (BRL) characterization of the \mathcal{H}^∞ norm of an LTI system presented in the following lemma [6].

Lemma 1 *A stable system with a state-space realization (3) has an \mathcal{H}^∞ norm from input w to output z less than or equal to γ if and only if there exists a matrix $P \geq 0$ satisfying*

$$\begin{bmatrix} A^T P + PA & PB & C^T \\ B^T P & -\gamma I & 0 \\ C & 0 & -\gamma I \end{bmatrix} \leq 0 \quad (8)$$

Employing the BRL condition and the choice of a block diagonal matrix for the Lyapunov matrix P as a particular solution results in the bound of Theorem 1. Numerical examples demonstrate the validity and computational efficiency of the above analytical bound [3].

The \mathcal{H}^2 norm of a stable continuous-time system with transfer function $H(s) = C(sI - A)^{-1}B$ is defined as the root-mean-square (rms) of its impulse response [39], or equivalently

$$\|H\|_2 = \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} \text{trace}(H^H(j\omega)H(j\omega))d\omega}$$

In the following, an LMI formulation for computing the \mathcal{H}^2 norm of a system using its state-space data is recalled. This formulation enables us to use the efficient LMI solvers to solve for the Lyapunov matrix and compute the \mathcal{H}^2 norm μ .

The collocated \mathcal{H}^2 control synthesis problem is to design a symmetric static feedback gain G such that the output feedback control law (2) stabilizes the closed-loop system and guarantees an \mathcal{H}^2 norm less than a prescribed level $\mu > 0$.

Lemma 2 [34] *Suppose that the system (3) is asymptotically stable, and let $H(s)$ denote its transfer function. Then the following statements are equivalent:*

- $\|H\|_2 \leq \mu$
- There exist symmetric nonnegative definite matrices P and Z such that

$$\begin{bmatrix} PA + A^T P & PB \\ B^T P & -I \end{bmatrix} \leq 0 \quad (9)$$

$$\begin{bmatrix} P & C^T \\ C & Z \end{bmatrix} \geq 0 \quad (10)$$

$$\text{trace}(Z) \leq \mu^2 \quad (11)$$

In the following lemma, we recall the \mathcal{H}^2 norm calculation based on the solution of a Lyapunov equation.

Lemma 3 [34] *The \mathcal{H}^2 norm of the system (3) is given by*

$$\|H\|_2 = [\text{trace}(CPC^T)]^{\frac{1}{2}} \quad (12)$$

where P is determined by solving the following Lyapunov equation.

$$AP + PA^T + BB^T = 0 \quad (13)$$

To avoid the need for solving an LMI problem to determine \mathcal{H}^2 norm of a system in the collocated form, the following result provides a simple analytical explicit expression for an upper bound on the \mathcal{H}^2 norm of such systems [25].

Theorem 2 Consider the open-loop ($u(t) = 0$) collocated structural system in (1). This system has an \mathcal{H}^2 norm μ from the input $w(t)$ to the output $z(t)$ that satisfies the following bound

$$\mu \leq \bar{\mu} = \frac{[\lambda_{\max}(E^T D^{-1} E)]^{\frac{1}{2}}}{\sqrt{2}} [\text{trace}(E^T M^{-1} E)]^{\frac{1}{2}} \quad (14)$$

Numerical examples demonstrate the validity and computational efficiency of the above analytical bound on the \mathcal{H}^2 norm of collocated systems [25]. Meisami-Azad et al. [25] also developed an explicit parametrization of the suboptimal output feedback control gains that achieve a desired level of closed-loop \mathcal{H}^2 performance.

Next, we discuss the problem of computing an explicit expression for the mixed $\mathcal{H}^2/\mathcal{H}^\infty$ norm of the collocated structural systems. The mixed problem solves the minimization problem of an \mathcal{H}^2 norm criterion subject to an \mathcal{H}^∞ norm constraint. Presented next is an explicit upper bound on the mixed $\mathcal{H}^2/\mathcal{H}^\infty$ norm of the collocated structural system represented by (1). The following theorem provides such an explicit bound resulting in a computationally efficient calculation.

Theorem 3 Consider the unforced system in (1), i.e., $u(t) = 0$. For any given $\gamma \geq \bar{\gamma} = \lambda_{\max}(E^T D^{-1} E)$, an upper bound $\bar{\eta}$ on the \mathcal{H}^2 norm of the system, while the \mathcal{H}^∞ norm satisfies the condition $\|H_{zw}\|_\infty \leq \gamma$, can be computed from the following expression

$$\bar{\eta} = \frac{[\lambda_{\max}(F^T D^{-1} F)]^{\frac{1}{2}}}{\sqrt{2}} [\text{trace}(F^T M^{-1} F)]^{\frac{1}{2}} \quad (15)$$

if $\gamma \geq 1$. Otherwise, for $\gamma < 1$, an upper bound on the \mathcal{H}^2 norm of the system is determined from

$$\bar{\eta} = [\max(\delta, \sigma) \times \text{trace}(F^T M^{-1} F)]^{\frac{1}{2}} \quad (16)$$

where

$$\begin{aligned} \delta &= \frac{1}{2} \lambda_{\max}(F^T D^{-1} F) \\ \sigma &= \frac{\lambda_{\max}(F^T D^{-1} F)}{\gamma + [\gamma^2 - \lambda_{\max}^2(F^T D^{-1} F)]^{1/2}}. \end{aligned} \quad (17)$$

The detailed proof of this result can be found in [23]. Meisami-Azad [23] also provides an explicit expression for the output feedback control gain to guarantee the \mathcal{H}^∞ and \mathcal{H}^2 norms of the closed-loop to be less than given bounds γ and η , respectively.

4 Integrated Damping and Control Design Using the Analytical Bound Approach

In this section, we address the integrated design problem of simultaneously designing the damping parameters and the output feedback control gain of the collocated structural system (1) and (2) to satisfy \mathcal{H}^∞ norm, \mathcal{H}^2 norm or mixed $\mathcal{H}^2/\mathcal{H}^\infty$ norm closed-loop specifications. For lumped parameter systems, the damping matrix D can be expressed in terms of the elemental damping coefficients as follows

$$D = \sum_{i=1}^l c_i \mathfrak{T}_i \quad (18)$$

where c_i denotes the viscous damping constant of the i th damper and \mathfrak{T}_i represents the distribution matrix of the corresponding damper in the structural system. The distribution matrices \mathfrak{T}_i are known symmetric matrices with elements 0, 1 and -1 that define the structural connectivity of the damping elements in the structure. Our objective is to formulate the \mathcal{H}^2 and \mathcal{H}^∞ integrated damping parameter and control gain design problems as LMI optimization problems.

Practical structural system design specifications impose upper bound constraints on the values of the damping coefficients, that is

$$0 \leq c_i \leq c_{i\max}, \quad i = 1, \dots, l \quad (19)$$

Also, often an upper bound on the total available damping resources is enforced, that is

$$\sum_{i=1}^l c_i \leq c_{\text{cap}}. \quad (20)$$

Another useful constraint in the proposed integrated design method is a bound on the norm of the feedback gain matrix. This restriction is placed to limit the amount of control effort required by the controller. For this purpose, we include the following constraint in the integrated design problem.

$$\|G\| \leq g_{\text{bound}} \quad (21)$$

We assume that $c_{i\max}$, c_{cap} and g_{bound} are given scalar bounds determined by the physical constraints of the design problem.

4.1 Integrated Design Based on an \mathcal{H}^∞ Specification

Using the above formulation, the solution of the integrated design of the damping parameters and the symmetric output feedback controller to satisfy closed-loop \mathcal{H}^∞ specifications is obtained by the following result.

Theorem 4 Consider the collocated structural system (1) with the damping distribution (18). For a given positive scalar γ , the \mathcal{H}^∞ norm of the closed-loop system of the collocated structural system (1) and the output feedback controller (2) is less than γ if the following matrix inequalities with respect to the controller gain G and the damping coefficients c_i are feasible.

$$\begin{bmatrix} \sum_{i=1}^l c_i \mathfrak{T}_i + FGF^T & E \\ E^T & \gamma I \end{bmatrix} \geq 0 \quad (22a)$$

$$0 \leq c_i \leq c_{i\max} , \quad i = 1, \dots, l \quad (22b)$$

$$\sum_{i=1}^l c_i \leq c_{cap} \quad (22c)$$

$$\|G\| \leq g_{bound} \quad (22d)$$

Proof. Consider the closed-loop interconnection of the collocated structural system (1) and the output feedback law (2). We consider the Lyapunov matrix P as

$$P = \begin{bmatrix} K & 0 \\ 0 & M \end{bmatrix} \quad (23)$$

along with substituting the closed-loop system matrices (4) into the BRL condition (8) and using the Schur complement formula [34] results in the following inequality

$$\begin{bmatrix} D + FGF^T & E \\ E^T & \gamma I \end{bmatrix} \geq 0 \quad (24)$$

Substitution of the damping matrix expansion (18) results in the inequality (22a). The constraints (22b)–(22d) represent the physical constraints of the design problem as discussed earlier. \square

4.2 Integrated Design Based on an \mathcal{H}^2 Specification

Based on the proposed \mathcal{H}^2 upper bound results, the integrated design of the damping parameters of a collocated structural system and the symmetric output feedback control law to satisfy closed-loop \mathcal{H}^2 norm specifications is formulated as follows.

Theorem 5 Consider the collocated structural system (1) with the damping distribution (18). For a given positive scalar μ , the \mathcal{H}^2 norm of the closed-loop system of the collocated structural system (1) and the output feedback controller (2) is less

than μ if the following matrix inequalities with respect to the controller gain G , the damping coefficients c_i , and the positive scalar α are feasible.

$$-2\left(\sum_{i=1}^l c_i \mathfrak{T}_i + FGF^T\right) + \alpha FF^T \leq 0 \quad (25a)$$

$$\begin{bmatrix} \alpha M & F \\ F^T & Z \end{bmatrix} \geq 0 \quad (25b)$$

$$\text{trace}(Z) \leq \mu^2 \quad (25c)$$

$$0 \leq c_i \leq c_{i\max}, \quad i = 1, \dots, l \quad (25d)$$

$$\sum_{i=1}^l c_i \leq c_{cap} \quad (25e)$$

$$\|G\| \leq g_{bound} \quad (25f)$$

Proof. We now consider a Lyapunov matrix P as follows

$$P = \alpha \begin{bmatrix} K & 0 \\ 0 & M \end{bmatrix} \quad (26)$$

where α is a positive scalar. Substituting the matrix P and the closed-loop system matrices (4) into the \mathcal{H}^2 inequality conditions (9)–(11) results in the following inequalities

$$\begin{bmatrix} -2\alpha(D + FGF^T) & \alpha F \\ \alpha F^T & -I \end{bmatrix} \leq 0 \quad (27a)$$

$$\begin{bmatrix} \alpha M & F \\ F^T & Z \end{bmatrix} \geq 0 \quad (27b)$$

$$\text{trace}(Z) \leq \mu^2 \quad (27c)$$

The scalar α in the selected Lyapunov matrix (26) is an unknown parameter that can be used as an additional degree of freedom in our formulation in order to reduce the conservativeness of the \mathcal{H}^2 norm bound. Note that due to the cross product of α and D in (27a), this inequality is not an LMI. However, application of the Schur complement formula to (27a) yields

$$-2(D + FGF^T) + \alpha FF^T \leq 0$$

which is an LMI with respect to α , G , and D . Substitution of the damping matrix expansion (18) completes the results. \square

4.3 Integrated Design Based on a Mixed $\mathcal{H}^2/\mathcal{H}^\infty$ Specification

The solution of the integrated design problem of the damping parameters and the symmetric output feedback controller to satisfy a closed-loop mixed $\mathcal{H}^2/\mathcal{H}^\infty$ specifications is obtained from the following result. The proof of this result follows similar lines as the proofs of Theorems 4 and 5.

Theorem 6 Consider the collocated structural system (1) with the damping distribution (18). For given positive scalars γ and μ , the \mathcal{H}^∞ norm of the closed-loop system of the collocated structural system (1) and the output feedback controller (2) is less than γ and the \mathcal{H}^2 norm of the closed-loop system is less than μ if the following matrix inequalities with respect to the damping coefficients c_i , the controller gain G , and the parameters α and Z are satisfied.

$$-2\left(\sum_{i=1}^l c_i \mathfrak{T}_i + FGF^T\right) + \alpha EE^T \leq 0 \quad (28a)$$

$$\begin{bmatrix} \alpha M & E \\ E^T & Z \end{bmatrix} \geq 0 \quad (28b)$$

$$\begin{bmatrix} \sum_{i=1}^l c_i \mathfrak{T}_i + FGF^T - \frac{\alpha}{2\gamma} EE^T & E \\ E^T & 2\alpha\gamma \end{bmatrix} \geq 0 \quad (28c)$$

$$\text{trace}(Z) \leq \mu^2 \quad (28d)$$

$$0 \leq c_i \leq c_{i\max}, \quad i = 1, \dots, l \quad (28e)$$

$$\sum_{i=1}^l c_i \leq c_{cap} \quad (28f)$$

$$\|G\| \leq g_{bound} \quad (28g)$$

It should be noted that the integrated design formulation presented here provides a non-iterative and computationally tractable method for simultaneous design of structural damping parameters and control gain for collocated structural systems. This is in contrast to the available methods in the literature, that seek to solve complex nonlinear optimization problems to address the integrated plant/control design. It is emphasized that the proposed upper bound design approach is only capable of optimizing the values of damping coefficients and control gains. Simultaneous design of stiffness and damping parameters and control gains as in [12,27] would result in a non-convex optimization problem of much greater computational complexity.

4.4 Decentralized Control Using the Norm Upper Bound Formulation

Complex large-scale systems often consist of smaller interconnected interacting subsystems. A need for decentralized control naturally arises in such systems due to communication and hardware implementation constraints. The goal of decentralized control is to design local controllers that use local measurements to generate a local control action for each subsystem. At the same time global objectives for stability and performance of the closed-loop large-scale system should be satisfied [35]. Decentralized \mathcal{H}^∞ or \mathcal{H}^2 control design problem for large-scale systems has been shown to result in a non-convex optimization problem [40], and several iterative algorithms to achieve a locally optimal control have been proposed [22, 31, 40]. The proposed norm upper bound method of this chapter results in an LMI problem even if the decentralized control structure is fixed. Imposing a particular structure for the control matrix G is dictated by the structure of the subsystems in the large-scale system. Using the method of this chapter, we are able to obtain the optimal static rate-feedback gain matrices and damping parameters for externally symmetric systems when an upper bound on the closed-loop system norm is sought. The procedure to formulate the decentralized control design problem for the systems with collocated sensors and actuators in the form presented by (1) is described below. It is noted that stability conditions for such systems using decentralized dynamic displacement feedback are provided in [9].

We consider l structural subsystems with collocated sensors and actuators interconnected by flexible links which can be modeled as springs and dampers and described by

$$\begin{aligned} M_i \ddot{q}_i + D_i \dot{q}_i + K_i q_i &= F_i u_i + E_i w_i \\ &+ \sum_{j=1}^l N_{ij} \{ K_{Cij} (N_{ji}^T q_j - N_{ij}^T q_i) + D_{Cij} (N_{ji}^T \dot{q}_j - N_{ij}^T \dot{q}_i) \} \\ y_i &= F_i^T \dot{q}_i \\ z_i &= E_i^T \dot{q}_i, \quad i = 1, \dots, l \end{aligned}$$

where $q_i \in R^{n_i}$, $u_i \in R^{r_i}$, and $y_i \in R^{r_i}$ are the displacement (translational displacement and rotational angle), control input (force and torque), and the measured velocity output of the i th subsystem, respectively. Also, $w_i \in R^{t_i}$ and $z_i \in R^{t_i}$ represent the external disturbance input, and the controlled output of the i th subsystem, respectively. The matrices M_i , D_i and K_i are the mass, damping, and stiffness matrices corresponding to the i th subsystem. Also, N_{ij} is a matrix representing the locations and directions of the springs and dampers, which connect the i th subsystem with the j th subsystem. The matrices K_{Cij} and D_{Cij} denote spring and damper matrices, respectively, which are positive definite. The effects of the springs and dampers are bilateral, and hence $K_{Cij} = K_{Cji}$ and $D_{Cij} = D_{Cji}$.

We consider local velocity output feedback control laws as

$$u_i = -G_i y_i \quad (29)$$

The overall closed-loop system is then described by

$$\begin{aligned} M\ddot{q} + D\dot{q} + Kq &= Fu + Ew \\ y &= F^T \dot{q} \\ z &= E^T \dot{q} \end{aligned} \quad (30)$$

where the control law is

$$u = -Gy \quad (31)$$

with

$$G = \text{diag}\{G_1, \dots, G_l\}$$

and

$$q = [q_1^T, \dots, q_l^T]^T, w = [w_1^T, \dots, w_l^T]^T, z = [z_1^T, \dots, z_l^T]^T, u = [u_1^T, \dots, u_l^T]^T$$

$$E = \text{diag}\{E_1, \dots, E_l\}$$

$$M = \text{diag}\{M_1, \dots, M_l\}$$

$$D = \text{diag}\{C_1, \dots, C_l\} + \sum_{i=1}^{l-1} \sum_{j=i+1}^l \bar{N}_{ij} D_{Cij} \bar{N}_{ij}^T$$

$$K = \text{diag}\{K_1, \dots, K_l\} + \sum_{i=1}^{l-1} \sum_{j=i+1}^l \bar{N}_{ij} K_{Cij} \bar{N}_{ij}^T$$

where

$$\bar{N}_{ij} = \begin{bmatrix} \vdots \\ N_{ij} \\ \vdots \\ -N_{ji} \\ \vdots \end{bmatrix} \quad (32)$$

in which all the elements except the submatrices N_{ij} and $-N_{ji}$ are zero.

The design of the decentralized control law (31) then follows a similar formulation as the prior centralized control design results with the additional constraints that G has a block diagonal structure. Hence, such constraint can be easily incorporated in the corresponding upper bound LMI formulation.

4.5 Additional Remarks

In this section, we provide some additional remarks and comments on the integrated design method proposed in this chapter.

Remark 1 *The conditions of Theorems 4 and 5 establish an LMI feasibility problem with respect to the damping coefficients c_i and the controller gain G . Given the scalar bounds $c_{i\max}$, c_{cap} and g_{bound} , the optimum values of the damping coefficients and the controller gain that minimize the \mathcal{H}^2 norm bound can be obtained by solving the following LMI optimization problem*

$$\begin{cases} \min_{\alpha, c_i, G} & \mu^2 \\ \text{subject to} & (25a) - (25f) \end{cases} \quad (33)$$

Remark 2 *The total number of unknown parameters in the LMI optimization problems of Theorems 4 and 5 is $\frac{m(m+1)}{2} + l$ and $m(m+1) + l + 1$, respectively. These correspond to the independent elements of the symmetric feedback gain matrix G and the damping parameters c_i . It is also noted that the total number of LMI constraints is $l + 3$ and $l + 5$, respectively.*

Remark 3 *The control gain matrix norm bound condition in (22d) and (25f) can be expressed in an LMI form as follows [6]*

$$\begin{bmatrix} g_{bound}^2 I & G^T \\ G & I \end{bmatrix} \geq 0 \quad (34)$$

Remark 4 *Following similar lines as above, the results of Theorems 4 and 5 can be used to minimize the available damping resources c_{cap} or the control gain norm g_{bound} subject to a given bound on the \mathcal{H}^∞ or the \mathcal{H}^2 norm of the closed-loop system. For example, minimization of the control gain norm g_{bound} subject to a given bound γ of the \mathcal{H}^∞ norm of the closed-loop system can be achieved by solving the following LMI optimization problem.*

$$\begin{cases} \min_{c_i, G} & g_{bound} \\ \text{subject to} & (22a) - (22d) \end{cases} \quad (35)$$

5 Simulation Results

In this section, we validate the proposed integrated damping parameters and control gain design methods using the different norm specifications for a collocated structural system and the corresponding static output feedback control gain computation by providing illustrative examples. The MATLAB Robust Control Toolbox is used for the computational solution of the corresponding LMI optimization problems that are involved in the integrated design procedures.

Example 1

As the first application example, we consider the lumped model of a ten-story base isolated building structure as shown in Fig. 1. The mass of each floor, including that of the base, is assumed to be 250 tons. The stiffness of the structure varies in steps of 10^7 N/m between floors from 10^8 N/m for the first floor to 1.9×10^8 N/m for the tenth floor. The design objective is to optimize the values of the damping coefficients c_i , $i = 1, \dots, 10$, as well as, the output feedback control gain G such that the \mathcal{H}^∞ norm of the closed-loop system consisting of the collocated structure and the output feedback from the disturbance forces $w_1(t)$ and $w_2(t)$ (acting on the first and forth floors) to the velocities of the masses m_1 and m_4 is minimized.

The damping matrix D of this system is given by (18), where the elemental distribution matrices \mathfrak{T}_i are easily defined based on the system configuration. To examine integrated design trade-offs, we consider a family of optimal designs using the result of Theorem 4. We examine two scenarios. First, we fix the upper bound on the feedback control gain matrix norm to be $g_{\text{bound}} = 1$. We consider different designs corresponding to different values of the total damping capacity c_{cap} ranging from 0.1 to 10^4 Ns/m. The results of the corresponding integrated designs (using Theorem 4) are shown in Figs. 2, 3 and 4. Figure 2 shows the values of the \mathcal{H}^∞ norm bound obtained from solving the convex optimization problem for each design, as well as,

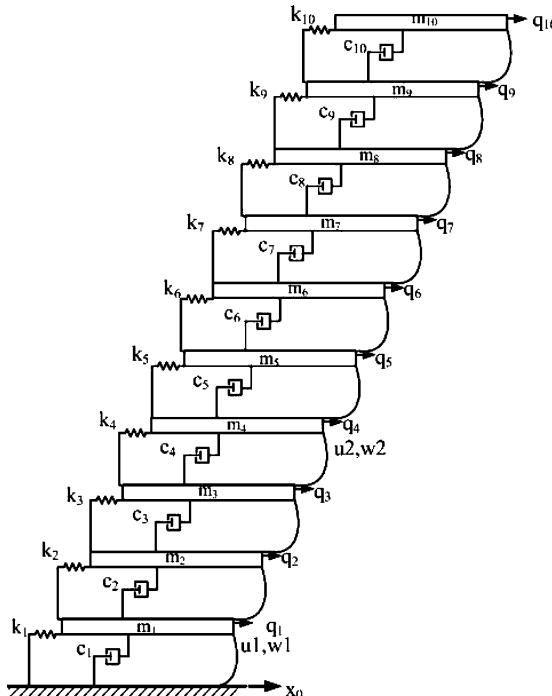


Fig. 1 Ten-DOF system

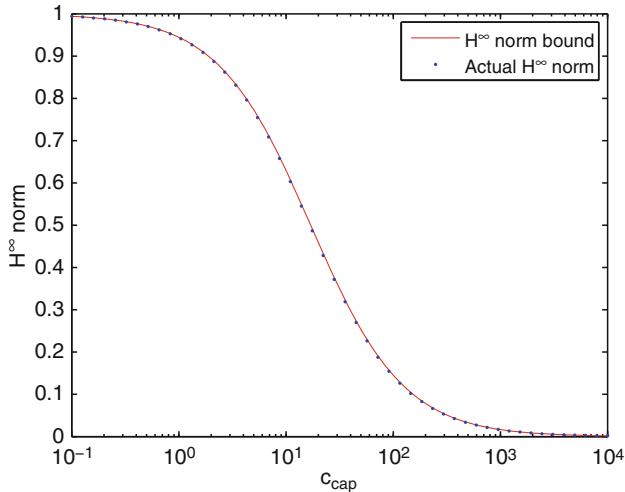


Fig. 2 Profiles of the \mathcal{H}^∞ norm upper bound and the actual \mathcal{H}^∞ norm for the optimized structure vs. the total damping capacity

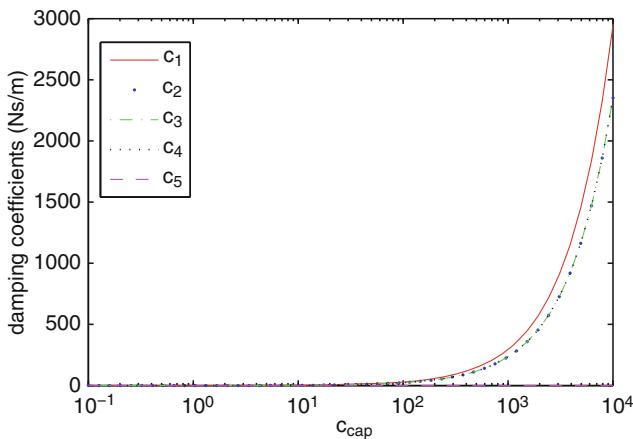


Fig. 3 Values of the optimized structural damping coefficients vs. the total damping capacity

the exact \mathcal{H}^∞ norm for each design as the total damping capacity c_{cap} changes. As we expect, the optimal closed-loop \mathcal{H}^∞ norm of the integrated design decreases as the overall damping allowed in the system increases. This plot also illustrates the accuracy of the \mathcal{H}^∞ norm bound calculated using Theorem 4. Figures 3 and 4 show the values of the structural damping parameters and the closed-loop damping ratios (in %) corresponding to each design, respectively. It should be noted that the damping ratios corresponding to all the floors are comparable even though the dominant damping parameters are c_1, c_2, c_3 and c_4 . Indeed, the rest of the damping parameters

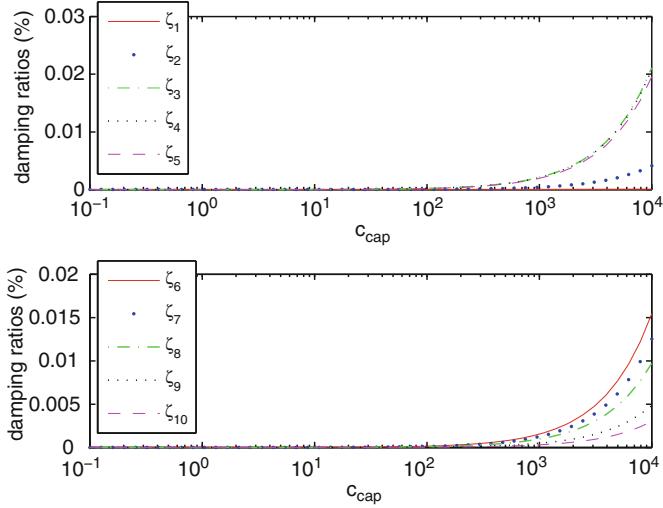


Fig. 4 The optimized closed-loop damping ratios vs. the total damping capacity

remain very small (close to zero) and the damping parameters corresponding to the four lower floors become significantly larger as c_{cap} increases. The reason for this behavior is the location of the sensors (and actuators) on the first and forth floors. It is indeed expected that the first four dampers will be the dominant ones to damp the structure's velocity response measured by the sensors on the first and fourth floors.

As a second design scenario, we consider a given bound for the total damping capacity $c_{cap} = 50$ Ns/m, and we minimize the \mathcal{H}^∞ norm of the closed-loop system with respect to the damping coefficients and the control gain. We compare different designs obtained by varying the upper bound on the controller gain norm g_{bound} . Figure 5 depicts the optimal \mathcal{H}^∞ norm bound (obtained from solving the optimization problem of Theorem 4), as well as, the exact \mathcal{H}^∞ norm corresponding to each design versus g_{bound} . As expected the closed-loop \mathcal{H}^∞ norm decreases when the allowable magnitude of the control gain matrix norm increases. Figure 6 shows the optimized structural damping coefficients (associated with the first five floors) obtained from the integrated designs. Note that the damping parameters associated with the 6th to 10th floors are close to zero. From Fig. 6, it is observed that the damping parameters do not change significantly as g_{bound} varies, and as expected, c_1 , c_2 , c_3 and c_4 are the dominant damping parameters. However, the closed-loop damping ratios vary as the allowed control gain increases (not shown here).

Figures 2 and 5 verify the accuracy of the obtained closed-loop \mathcal{H}^∞ norm of the system. Note that for calculating the actual \mathcal{H}^∞ norm of the closed-loop system, the feedback interconnection of the open-loop structure and the controller is considered, where the structure includes the designed values of the damping parameters c_i , and the controller is constructed using the feedback control gain G obtained from the solution to the convex optimization problems discussed in Sect. 4

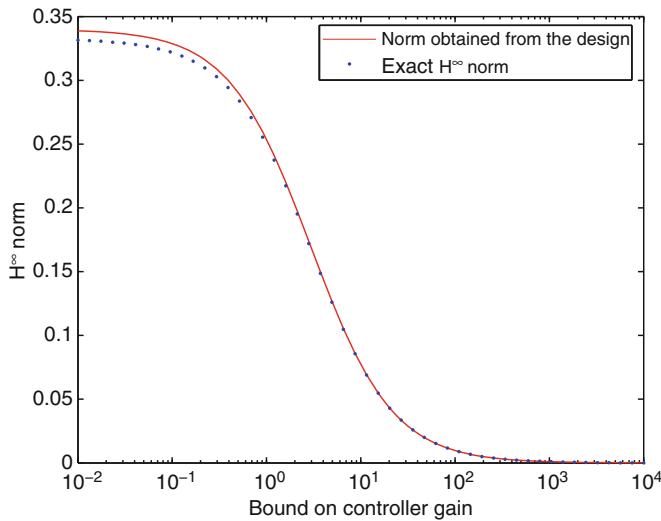


Fig. 5 Profiles of the closed-loop \mathcal{H}^∞ norm obtained by solving the LMI optimization problem and the actual norm calculated based on the optimized structure vs. bound on feedback control gain

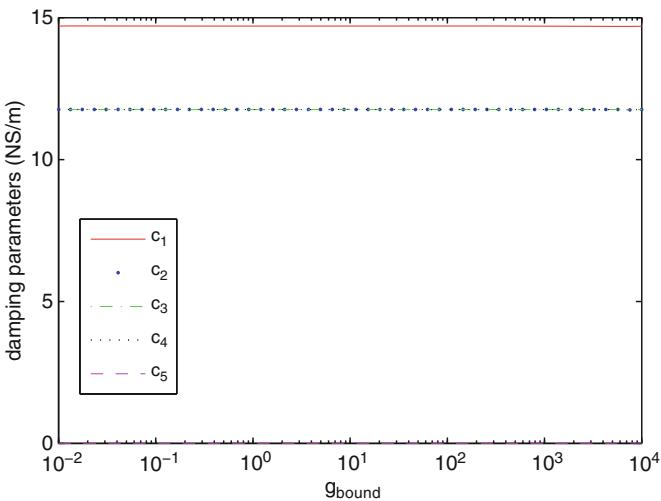


Fig. 6 Values of the optimized structural damping coefficients vs. bound on the norm of the feed-back control gain

Example 2

Next, we consider the single-input/single-output structural model of a cantilevered beam controlled by pairs of piezoelectric patches as collocated sensors and actuators. This is the model of a thin aluminum flexible beam with low damping

characteristics with properties listed in Table 1 [23]. We derive a 40-element finite element structural model for this structural system that has the vector second-order form (1). We seek to examine the application of the proposed integrated design method for the damping distribution and control gain design of this large-scale collocated structural system. For this example, we are interested in the optimum design of the damping parameters and the output feedback control gain from the closed-loop \mathcal{H}^2 norm performance viewpoint. We assume given bounds on the total damping capacity and the norm of the controller gain as follows: $c_{cap} = 1 \text{ Ns/m}$ and $g_{bound} = 20$. Solving the LMI optimization problem of Theorem 5 for the unknown damping parameters and control gains results in an optimal closed-loop \mathcal{H}^2 norm bound of $\mu = 18.542$. It is noted that the actual \mathcal{H}^2 norm of the system for the designed parameters is $\mu = 18.360$. The frequency responses of the undamped open-loop system, the open-loop system damped by optimized damping parameters only, and the optimal closed-loop system designed using the \mathcal{H}^2 upper bound approach are shown in Fig. 7. The results demonstrate that the simultaneous design of

Table 1 Aluminum beam properties

Quality	Description	Units	Value
L	Beam length	mm	736.5
w_b	Beam width	mm	53.1
t_b	Beam thickness	mm	1
ρ_b	Beam density	kg/m^3	2,690
E	Modulus of elasticity	N/m^2	7.03×10^{10}

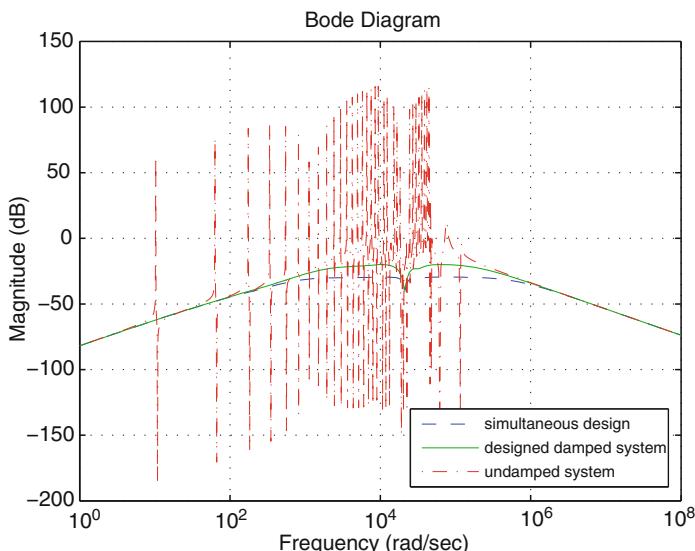


Fig. 7 Frequency responses of the undamped system, damped system constructed by designed damping parameters, and the closed-loop system of the damped system and \mathcal{H}^2 controller, from $w_1(t)$ to $\dot{q}_1(t)$

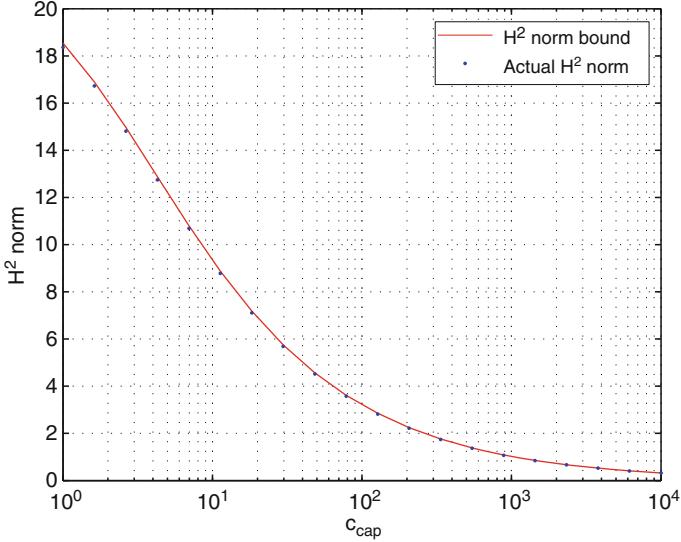


Fig. 8 Profiles of the \mathcal{H}^2 norm obtained by solving the convex optimization problem of Theorem 5 and the actual \mathcal{H}^2 norm of the closed-loop system of the optimized structure and output feedback control vs. the total damping capacity

damping parameters and controller provides improved disturbance rejection compared to past work that seeks to optimize only damping parameters [24]. Note that using traditional methods for simultaneous control and damping parameter design for this system could easily become prohibitive due to the high dimensionality of the system.

Finally, Fig. 8 shows the \mathcal{H}^2 norm bound obtained by solving the integrated damping parameter and control gain optimization problem presented in Theorem 5 for different values of the total damping capacity c_{cap} and the actual \mathcal{H}^2 norm of the structural system for each design. It is observed that the value of the \mathcal{H}^2 norm bound and the achievable \mathcal{H}^2 norm are indeed extremely close.

Example 3

As a last example, we consider the model of an interconnected structure composed of two substructures, denoted by S_1 and S_2 as depicted in Fig. 9. The structural parameters are listed in Table 2, and the two subsystems S_1 and S_2 are connected with a damper d_c and a spring k_c at the third floor. We assume a decentralized control scheme for each subsystem given as the follows.

$$u_d(t) = -Gy_d(t), \quad u_d(t) = \begin{bmatrix} u_1^1(t) \\ u_2^1(t) \\ u_1^2(t) \\ u_2^2(t) \end{bmatrix}, \quad y_d(t) = \begin{bmatrix} \dot{q}_1^1(t) \\ \dot{q}_3^1(t) \\ \dot{q}_1^2(t) \\ \dot{q}_3^2(t) \end{bmatrix}, \quad G = \text{diag}(G_1, G_2) \quad (36)$$

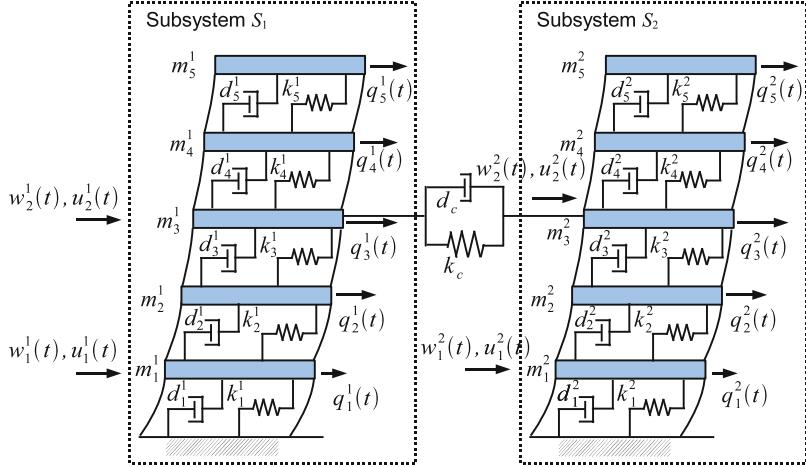


Fig. 9 Interconnected ten-DOF system consisting of two subsystems S_1 and S_2

Table 2 Structural parameters of the decentralized control problem

Structural parameters	Unit	Value
m_i^j ($i = 1, \dots, 5, j = 1, 2$)	kg	6.0×10^5
k_1^j ($j = 1, 2$)	N/m	7.0×10^8
k_2^j ($j = 1, 2$)	N/m	7.5×10^8
k_3^j ($j = 1, 2$)	N/m	8.0×10^8
k_4^j ($j = 1, 2$)	N/m	8.5×10^8
k_5^j ($j = 1, 2$)	N/m	9.0×10^8

Table 3 Optimized damping parameters using the proposed decentralized control method

Optimized damping parameters (Ns/m)	Value
$d_1^1 = d_1^2$	11.765
$d_2^1 = d_2^2$	17.647
$d_3^1 = d_3^2$	17.647
$d_4^1 = d_4^2$	23.529
$d_5^1 = d_5^2$	23.529
$\sum_{i=1}^2 \sum_{j=1}^5 d_j^i$	$188.24 \leq c_{cap}$
Upper bound of the closed-loop \mathcal{H}^∞ norm	0.02
Exact closed-loop \mathcal{H}^∞ norm	0.02

Let $d_c = 100$ Ns/m and $k_c = 8 \times 10^7$ N/m. Using the results in Theorem 4, the output feedback gain matrix G and damping coefficients d_j^i ($i = 1, 2, j = 1, \dots, 5$) are obtained by minimizing an upper bound on the closed-loop \mathcal{H}^∞ norm considering $g_{bound} = 50$ and $c_{cap} = \sum_{i=1}^2 \sum_{j=1}^5 d_j^i = 200$. The result is given in Table 3 with the exact closed-loop \mathcal{H}^∞ norm obtained from the solution to the integrated design problem. The open and closed-loop singular value plots and the impulse responses are shown in Figs. 10 and 11, respectively. It is observed that the obtained

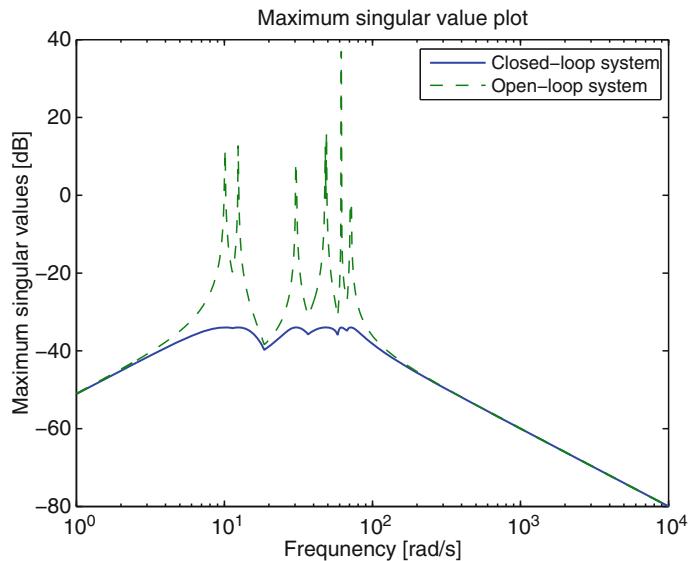


Fig. 10 Maximum singular values of the open-loop and closed-loop systems

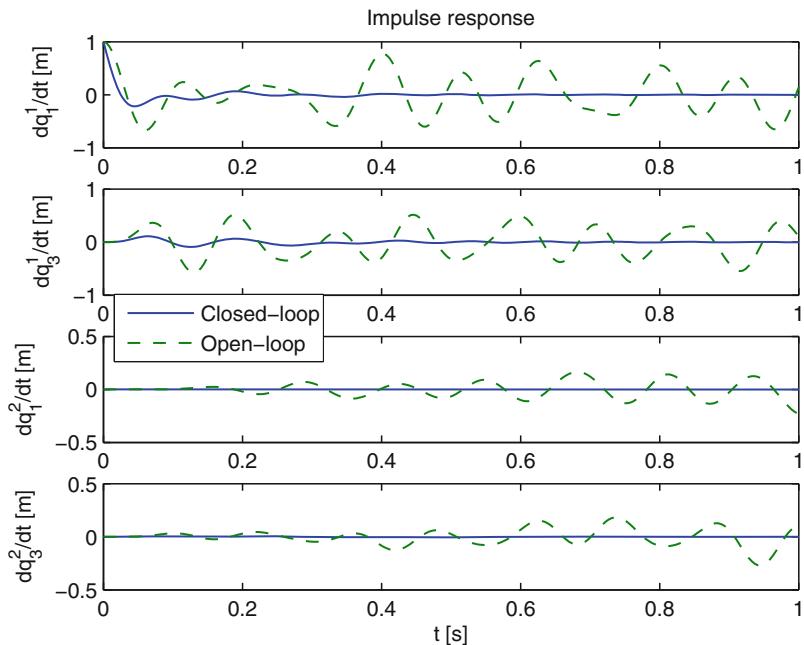


Fig. 11 Impulse responses of the open-loop and closed-loop systems

upper bound on the closed-loop \mathcal{H}^∞ norm is accurate and that closed-loop vibration suppression is achieved using the decentralized control scheme of this chapter.

6 Concluding Remarks

Presented in this chapter is an efficient computational methodology for the simultaneous design of the damping parameters and the output feedback control gain of a collocated structural system with velocity feedback such that the closed-loop system satisfies \mathcal{H}^2 , \mathcal{H}^∞ or mixed $\mathcal{H}^2/\mathcal{H}^\infty$ performance specifications. The proposed integrated design approach is based on an LMI formulation of the design problem that can be efficiently solved for the design variables using available semi-definite programming optimization solvers. Despite the fact that the method is based on an upper bound formulation of the norm performance specifications of the closed-loop system, computational examples demonstrate that the bounds result in a close approximation of the actual norms of the system and are effective for structural parameter and control design. The integrated design method is also shown to be efficient for design of large-scale interconnected structural systems with a large number of subsystems and states. As demonstrated, the proposed method is especially suitable for large-scale systems where existing nonlinear optimization approaches used as the standard tools to determine system parameters and control gains are computationally prohibitive.

References

1. K. Adachi, K. Sakamoto, and T. Iwatsubo, "Redesign of Closed-Loop System for Integrated Design of Structure and its Vibration Control System," *Proceedings IEEE International Conference on Control Applications*, pp. 80–85, 1999
2. B. Anderson and S. Vongpanitlerd, *Network Analysis and Synthesis: A Modern Systems Theory*, Prentice-Hall, Englewood Cliffs, 1973
3. Y. Bai and K.M. Grigoriadis, " \mathcal{H}^∞ Collocated Control of Structural Systems: An Analytical Bound Approach," *AIAA Journal of Guidance, Control, and Dynamics*, vol. 28, pp. 850–853, 2005
4. A.V. Balakrishnan, "Compensator Design for Stability Enhancement with Collocated Controllers," *IEEE Transactions on Automatic Control*, vol. 36, pp. 994–1007, 1991
5. M.J. Balas, "Direct Velocity Feedback Control of Large Space Structures," *Journal of Guidance and Control*, vol. 2, pp. 252–253, 1979
6. S.P. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, vol. 15, Studies in Applied Mathematics, SIAM, Philadelphia, PA, 1994
7. S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004
8. E.F. Crawley and J. de Luis, "Use of Piezoelectric Actuators as Elements of Intelligent Structures," *AIAA Journal*, vol. 25, pp. 1373–1385, 1987
9. Y. Fujisaki, M. Ikeda, and K. Miki, "Robust Stabilization of Large Space Structures via Displacement Feedback," *IEEE Transactions on Automatic Control*, vol. 46, pp. 1993–1996, Dec. 2001

10. P. Gahinet, A. Nemirovskii, A.J. Laub, and M. Chilali, *LMI Control Toolbox Users Guide*, The Mathematical Works, Natick, MA 1995
11. W.K. Gawronski, *Advanced Structural Dynamics and Active Control of Structures*, Mechanical Engineering Series, Springer, New York, 2004
12. K.M. Grigoriadis and F. Wu, "Integrated \mathcal{H}^∞ Plant/Controller Design using Linear Matrix Inequalities," *Proceedings 36th IEEE Conference on Decision and Control*, pp. 789–790, San Diego, CA, Dec. 1997
13. K.M. Grigoriadis, G. Zhu, and R.E. Skelton, "Optimal Redesign of Linear Systems," *ASME Transactions on Dynamic Systems, Measurements and Control*, vol. 118, pp. 598–605, 1996
14. K. Hiramoto and K.M. Grigoriadis, "Integrated Design of Structural and Control Systems with a Homotopy Like Iterative Method," *International Journal of Control*, vol. 79, pp. 1062–1073, 2006
15. T. Iwasaki, S. Hara, and H. Yamauchi, "Structure/Control Design Integration with Finite Frequency Positive Real Property," *Proceedings of American Control Conference*, pp. 549–553, 2000
16. S.P. Joshi, "Control for Energy Dissipation in Structures," *Journal of Guidance, Control, and Dynamics*, vol. 13, pp. 751–753, 1990
17. N.S. Khot, "Structure/Control Optimization to Improve the Dynamic Response of Space Structures," *Computational Mechanics*, vol. 3, pp. 179–186, 1988
18. N.S. Khot, V.B. Venkayya, H. Oz, R.V. Grandhi, and F.E. Eastep, "Optimal Structural Design with Control Gain Norm Constraint," *AIAA Journal*, vol. 26, pp. 604–611, 1988
19. Y. Kobayashi, M. Ikeda, and Y. Fujisaki, "Stability of Large Space Structures Preserved under Failures of Local Controllers," *IEEE Transactions on Automatic Control*, vol. 52, pp. 318–322, Feb. 2007
20. F. Liao, K.Y. Lum, and J.L. Wang, "An LMI-Based Optimization Approach for Integrated Plant/Output-Feedback Controller Design," *Proceedings of 24th American Control Conference*, pp. 4480–4485, Portland, OR, 2005
21. J.B. Lu and R.E. Skelton, "Integrating Structure and Control Design to Achieve Mixed $\mathcal{H}^2/\mathcal{H}^\infty$ Performance," *International Journal of Control*, vol. 73, pp. 1449–1462, 2000
22. C.S. Mehendale and K.M. Grigoriadis, "A Double Homotopy Method for Decentralised Control Design," *International Journal of Control*, vol. 81, pp. 1600–1608, Oct. 2008
23. M. Meisami-Azad, *Control Design Methods for Large-Scale Systems Using an Upper Bound Formulation*, Masters Thesis, University of Houston, TX, Aug. 2007
24. J. Mohammadpour, M. Meisami-Azad, and K.M. Grigoriadis, "An Efficient Approach for Damping Parameter Design in Collocated Structural Systems Using An \mathcal{H}^2 Upper Bound," *Proceedings of the 26th American Control Conference*, pp. 3015–3016, New York, NY, Jul. 2007
25. M. Meisami-Azad, J. Mohammadpour, and K.M. Grigoriadis, "Explicit Solutions for Collocated Structural Control with Guaranteed \mathcal{H}^2 Norm Performance Specifications," *Smart Materials and Structures*, 18 035004 (9pp), Mar. 2009
26. Y. Nesterov and A. Nemirovsky, *Interior-point Polynomial Methods in Convex Programming*, vol. 13, Studies in Applied Mathematics, SIAM, Philadelphia, PA, 1994
27. R.J. Niewoehner and I.I. Kaminer, "Integrated Aircraft-Controller Design Using Linear Matrix Inequalities," *AIAA Journal of Guidance, Control and Dynamics*, vol. 19, pp. 445–452, 1996
28. J. Onoda and R.T. Haftka, "An Approach to Structure/Control Simultaneous Optimization for Large Flexible Spacecraft," *AIAA Journal*, vol. 25, pp. 1133–1138, 1987
29. S.S. Rao, V.B. Venkayya, and N.S. Khot, "Game Theory Approach for the Integrated Design of Structures and Controls," *AIAA Journal*, vol. 26, pp. 463–469, 1988
30. M. Salama, J. Garba, L. Demsetz, and F. Udwadia, "Simultaneous Optimization of Controlled Structures," *Computational Mechanics*, vol. 3, pp. 275–282, 1988
31. G. Scoletti and G. Duc, "An LMI Approach to Decentralized \mathcal{H}^∞ Control," *International Journal of Control*, vol. 74, pp. 211–224, Feb. 2001
32. G. Shi and R.E. Skelton, "An Algorithm for Integrated Structure and Control Design with Variance Bounds," *Proceedings of 35th IEEE Conference on Decision and Control*, pp. 167–172, Kobe, Japan, Dec. 1996

33. R.E. Skelton, "Integrated Plant and Controller Design," *Proceedings of American Control Conference*, Seattle, WA, Jun. 1995
34. R.E. Skelton, T. Iwasaki, and K.M. Grigoriadis, *A Unified Algebraic Approach to Linear Control Design*, Taylor & Francis, 1998
35. D.D. Siljak, *Decentralized Control of Complex Systems*, Academic, Boston, MA, 1991
36. S. Sidi-Ali-Cherif, K.M. Grigoriadis, and M. Subramaniam, "Model Reduction of Large Space Structures Using Approximate Component Cost Analysis," *AIAA Journal of Guidance, Control and Dynamics*, vol. 22, pp. 551–558, 1999
37. K. Tan and K.M. Grigoriadis, "Stabilization and \mathcal{H}^∞ Control of Symmetric Systems: An Explicit Solution," *Systems and Control Letters*, vol. 44, pp. 57–72, 2001
38. S.M. Yang and Y.J. Lee, "Optimization of Non-collocated Sensor/actuator Location and Feedback Gain in Control Systems," *Smart Materials and Structures*, vol. 2, pp. 96–102, 1993
39. K. Zhou, J.C. Doyle, and K. Glover, *Robust and Optimal Control*, Prentice-Hall, Upper Saddle River, 1996
40. G. Zhai, M. Ikeda, and Y. Fujisaki, "Decentralized \mathcal{H}^∞ Controller Design: A Matrix Inequality Approach Using A Homotopy Method," *Automatica*, vol. 37, pp. 565–572, 2001

Index

A

Adaptive dual dynamic high-gain scaling paradigm
assumptions, 137–139
observer and controller designs
 η_m dynamics, 141
observer errors and scaled observer errors, 140
stability analysis
closed-loop stability properties, 150–151
composite Lyapunov function, 142, 145
observer and controller Lyapunov functions, 142
parameter estimator dynamics, 146

Aero-engine model
application
first-order model, 105–109
second-order model, 109–110

gas turbine system
nonlinear dynamic model, 92–93
nonlinear static model, 91–92
reheat bypass turbojet engine, 90–91

OE parameter identification
least-squares estimate (LSE) algorithm, 99
 $R(k)$ and Hessian approximation, 101–103
 $\partial V(\theta_k)/\partial \theta$ and Jacobian calculation, 100–101

reduced order data driven modeling
criterion selection, 94–96
model selection: EE vs. OE, 96–99

Air vehicle model, 59

Auto-regressive exogeneous (ARX) model, 98

B

Balancing method, reduced order model
balancing over a disk
asymptotic balancing process, 68, 70

bilinear mapping, 68–69
infinite bandwidth, 68
Cholesky factors, 65
Hankel minimum degree approximate algorithm, 66–68
observability and reachability mapping, 64

Bilinear matrix inequality (BMI), 307

Brute force technique, 130

C

Canonical models
linear and time-invariant systems, 220
linear state-space dynamics, 224

NDS
coupled at input, 226–227
coupled at output, 225–226
coupled at state, 227–228
coupled by state, input and output, 228

Cascading upper diagonal dominance (CUDD), 136

Closed-loop stability properties, 150–151

Collocated structural system
advantages, 306
symmetric output feedback control, 307–308
upper bound, 308–310

Composite Lyapunov function, 142, 145

Continuous-time Lyapunov equation, 119

Control input level consensus
local dynamic output feedback control laws
aggregation relations, 201
impulse response, 200–203
LQG methodology, 203
vector Lyapunov functions, 205
vector-matrix differential equation, 202

local static feedback control laws
design, 207
impulse response, 205–207

- Coordination strategy. *See* Multi-agent control
- CUDD. *See* Cascading upper diagonal dominance
- D**
- Decentralized control. *See also* Decentralized output feedback control
- local controller design, 315
 - overall closed-loop system, 316
 - UAVs formations
 - decentralized state estimation, 213–214
 - design, 210
 - experiments, 215–217
 - formation model, 210–211
 - global LQ optimal state feedback, 212–213
- Decentralized output feedback control
- adaptive dual dynamic high-gain scaling paradigm
 - assumptions, 137–139
 - observer and controller designs, 140–141
 - stability analysis, 141–151
 - cascading upper diagonal dominance, 136
 - example
 - admissible uncertainties, 185
 - local operation modes and global operation mode, 184
 - uncertain large-scale system, 183
 - generalised scaling
 - assumptions, 152–154
 - controller design, 155–158
 - observer design, 154–155
 - stability analysis, 158–164
 - guaranteed cost controller design
 - design methodology, 173–176
 - global mode dependent controllers, 176–180
 - local mode dependent controllers, 180–182
 - procedure, 182–183
 - problem formulation
 - IQC-type descriptions, 172–173
 - operation modes, 170–171
 - stationary ergodic continuous-time Markov process, 170
 - vector mode process, 170
- Decomposable systems, distributed control
- design
 - applications
 - infinite dimensional system generalization, 292–293
 - platoon, 293–294
 - LMIs, 280, 291
 - Lyapunov shaping method, 291
- multiobjective optimization, 291
- state-space matrices, 290
- static state feedback, 290
- symmetric decomposable systems, 289
- Directed weighted graphs, 211
- Distributed control methods
- Car platoon benchmark problem
 - controllers, 299
 - optimal centralized gain operator, 301–302
 - spatiotemporal H_2 norm, 300
 - decomposable systems
 - application, 292–294
 - LMIs, 280, 291
 - Lyapunov shaping method, 291
 - multiobjective optimization, 291
 - state-space matrices, 290
 - static state feedback, 290
 - symmetric decomposable systems, 289
 - identical systems, LQR
 - closed-loop state matrix, 299
 - closed-loop system matrix, 298
 - dynamically decoupled systems, 295–297
 - infinite-dimensional Laurent matrix, 297
 - Laurent operator, rational
 - applications, 286–287
 - convergence, 285–286
 - definition, 285
 - doubly infinite one-dimensional strings, 282–283
 - H_2 optimal state feedback problem, 288
 - L-operator sign function, 284–285
 - posteriori closed loop stability, 287
 - string interconnection, 284
 - problem statement
 - controlling absolute and relative distances, 280
 - H_2 problem and exact solution, 281–282
 - structure preserving iterative algorithms, 279
 - DNLS. *See* Dynamic nonlinear least squares
 - Dynamic gradient descent (DGD) algorithm, 106
 - Dynamic nonlinear least squares (DNLS) algorithm, 103
- E**
- Equation error (EE) model
- auto-regressive exogeneous (ARX) model, 98
 - definition, 96–97
- Equations of motion, 78–79
- Euler–Lagrange equation, 84

F

FACTS. *See* Flexible alternating current transmission systems
 Flexible alternating current transmission systems (FACTS), 272

G

Gas turbine system
 first order model
 DGD algorithm, 106
 DNLS algorithm, 107
 objective function, 105–106
 nonlinear dynamic model, 92–93
 nonlinear static model, 91–92
 reheat bypass turbojet engine, 90–91
 second-order model, 109
 Gaussian white noise, 234, 237
 Gauss–Newton methods, 102
 Global LQ optimal state feedback, 212–213
 Global mode dependent controllers
 definition, 168
 guaranteed cost controller design, 176–180
 Graph-theoretic bounds and analysis
 H_2 performance, NDS
 coupled at the output, 233–235
 coupled at the state, 236–241
 observability and controllability of NDS
 coupled at input, 232
 coupled at output, 230–231
 properties of linear system, 229
 Guaranteed cost controller design
 design procedure, 182–183
 global mode dependent controllers, 176–180
 local mode dependent controllers, 180–182
 methodology, 173–176

H

Hadamard product, 221, 231
 Hankel minimum degree approximate algorithm, 66–68
 H_2 norms
 graph-theoretic bounds performance
 NDS coupled at output, 233–235
 NDS coupled at state, 236–241
 topology design
 NDS coupled at output, 242–244
 NDS coupled at state, sensor placement, 244–246
 H^∞ norm upper bound approach, 318–319
 H^2 upper bound approach, 322

I

Integral quadratic constraints (IQCs)
 controller uncertainty, 169

Markovian jump parameter systems, 168,
 169

uncertainties and interconnections, 171

Integrated design, large-scale collocated structural system
 BMI formulation, 306
 damping coefficients, 317–318
 decentralized control, 315–316
 H^∞ specification, 311–312
 H^2 specification, 312–313
 mixed H^2/H^∞ specification, 314
 validation
 interconnected structure, 323–326
 lumped model, 318–321
 single-input/single-output structural model, 321–323

Interpolatory model reduction

advantages, 4
 approximation theory, 4
 coprime factorizations
 driven cavity flow, 33–35
 isotropic incompressible viscoelastic solid, 30
 second-order dynamical systems, 35–37
 error measurement
 full-order system, 14
 H_∞ norm, 15
 H_2 norms, 15–16
 matrix-valued meromorphic function, 16
 transfer functions $G(s)$ and $H(s)$, 16, 17
 framework
 approximation, 6
 low order transfer function, 6
 system matrices, 8
 tangential interpolation, 6–7
 goals, 5

input-output map
 input-output data, 8
 reduced system, 5
 state-space realization, 4–5
 interpolatory projections
 differential algebraic equation (DAE), 13
 full-order dynamical system, 10
 Hermite interpolation problem, 11
 interpolant construction, 13–14
 optimal point selection strategy, 13
 Petrov–Galerkin projective approximation, 8–9
 rational tangential interpolation problem, 12
 tangential direction selection, 12
 measurements
 coupled mechanical system, 49–51
 four-pole band-pass filter, 52

- Loewner and Pick matrices, 46–47
 Loewner matrix pair and construction
 of interpolants, 42–46
 simple low-order example, 47–49
 S-parameter representation, 42
 optimal H_2 approximation
 interpolation-based H_2 -optimality conditions, 18–24 (*see also* Iterative rational Krylov algorithm)
 Lyapunov-based optimal H_2 method, 18
 parametric systems, 37–41
 passive systems
 rational square matrix function $H(s)$, 25
 RLC circuit, 27–30
 spectral zeros (λ), 25–27
 state space system data, 7
 Inverse dynamics
 bounded-input-bounded state (BIBS), 152
 input-to-state practically stable (ISpS), 137
 Iterative rational Krylov algorithm (IRKA)
 bode plots of error systems, 24–25
 convergence behavior, 22–23
 first-order conditions, 18–20
 MIMO H_2 optimal tangential interpolation method, 22
 numerical results, 23–24
 reduced model, error norms, 24
- K**
 Kruskal's algorithm, 243
- L**
 Laurent operator, rational
 applications, 286–287
 convergence, 285–286
 definition, 285
 doubly infinite one-dimensional strings, 282–283
 H_2 optimal state feedback problem, 288
 L-operator sign function, 284–285
 posteriori closed loop stability, 287
 string interconnection, 284
 Least-squares estimate (LSE) algorithm, 99
 Linear matrix inequalities (LMIs), 280, 291
 Linear time-invariant (LTI) control systems, 115
LMIs. *See* Linear matrix inequalities
 Local dynamic output feedback control laws
 aggregation relations, 201
 impulse response, 200–203
 LQG methodology, 203
 vector Lyapunov functions, 205
 vector-matrix differential equation, 202
- Local mode dependent controllers
 definition, 169
 guaranteed cost controller design, 180–182
 Local static feedback control laws
 design, 207
 impulse response, 205–208
 Loewner and Pick matrices
 construction of interpolants
 generalized tangential reachability and observability matrices, 43–44
 shifted Loewner matrix, 44
 singular value decomposition (SVD), 46
 coupled mechanical system, 49–51
 four-pole band-pass filter, 52
 simple low-order example, 47–49
 Long-term prediction, DNLS
 first order model, 107–108
 second order model, 109
 LQR solutions
 closed-loop state matrix, 299
 closed-loop system matrix, 298
 dynamically decoupled systems, 295–297
 infinite-dimensional Laurent matrix, 297
 Lyapunov-based optimal H_2 method, 18
 Lyapunov shaping method, 291
- M**
 Markovian jump large-scale system model, 167
 MATLAB Robust Control Toolbox, 317
 Matrix sign function, 284
 Minimum spanning tree (MST) problem, 242–244
 Mode dependent controllers
 global, 168
 local, 169
 Model reduction techniques
 balancing method
 balancing over a disk, 68–69
 Cholesky factors, 65
 Hankel minimum degree approximate algorithm, 66–68
 observability and reachability mapping, 64
 frequency response characteristics, 59
 large-scale systems
 applications, 69–71
 asymptotic balancing process, 70
 non-minimum phase zero effect, 70
 residualization effect, 71
 simultaneous gradient error reduction
 conjugate gradient search routine, 62
 fit error, 63
 schematic representation, 63

- spectral decomposition process
 distinct frequency range groupings, 60
 eigenvalue and eigenvector calculation,
 60–62
 schematic representation, 61
- Multi-agent control
 control input level consensus
 local dynamic output feedback control
 laws, 200–205
 local static feedback control laws, 205–207
 dynamic consensus methodology, 198
 goals, 256
 network characteristics and control
 objectives, 256–257
 optimal power flow control, 255
 overlapping subnetworks
 common nodes, 263–264
 control problem formulation for one agent,
 264–267
 control scheme for multiple agents, 267
- parameters
 control objectives, 273
 setting up control problems, 273–274
 simulations, 274–276
 steady-state characteristics, 268–273
- problem formulation, 199–200
- state estimation level consensus
 globally LQ optimal controller, 209, 210
 Luenberger form, 208
 overlapping decentralized estimators, 208
 stability analysis, 209
- subnetworks
 control, 253–255
 definition, 257
 non-overlapping, touching/overlapping,
 254
- touching subnetworks
 control problem formulation for one agent,
 259–261
 control scheme for multiple agents,
 262–263
 internal and external nodes, 258–259
- UAVs formations, decentralized control
 decentralized state estimation, 213–214
 design, 210
 experiments, 215–217
 formation model, 210–211
 global LQ optimal state feedback, 212–213
- Multi-input/multi-output (MIMO) system, 4
- N**
- NDS. *See* Networked dynamic systems
- Networked dynamic systems (NDS)
 canonical models
- continuous linear time-invariant systems,
 220
- linear state-space dynamics, 224
- NDS coupled at input, 226–227
- NDS coupled at output, 225–226
- NDS coupled at state, 227–228
- NDS coupled by state, input and output,
 228
- graph-centric analysis, 220
- graph-theoretic bounds and analysis
 H_2 performance, 233–241
 observability and controllability, 229–232
- H_2 topology design
 NDS coupled at output, 242–244
 NDS coupled at state, sensor placement,
 244–246
- mathematical preliminaries and notations
 edge and graph Laplacian relation, 222
 K_{10} complete graph, 223–224
 k -regular graph, 223–224
 Kronecker product, 221
 singular value decomposition, 221
 subsystem dynamics, 220
- Non-minimum phase zeros effect, 70
- Non-overlapping power subnetwork, 254
- Norm upper bound approach
 collocated structural system
 mixed H^2/H^∞ , 310
 stable continuous-time system, 309
 transfer function, 308
- control design problem, 306
- convex optimization problem, 307
- integrated damping and control design
 damping matrix, 311
 decentralized control, 315–316
 integrated design, 311–314
 LMI, 317
- LMI-based optimization schemes, 306
- simulation
 aluminum beam properties, 321–322
 cantilevered beam, 321
- closed-loop damping ratios vs. total
 damping capacity, 319–320
- closed-loop H^∞ norm, 320–321
- decentralized control problem, 323–324
- H^∞ norm upper bound, 318–319
- H^2 upper bound approach, 322
- interconnected structure, 323
- MATLAB Robust Control Toolbox, 317
- open-loop and closed-loop systems,
 324–325
- structural damping coefficients vs. bound,
 320–321

structural damping coefficients *vs.* total damping capacity, 319
ten-DOF system, 318
symmetric output feedback control
closed-loop system, 308
symmetric static feedback gain, 307
Nyquist stability criteria, 63

O

Observer and controller Lyapunov functions, 142
Observer errors and scaled observer errors, 140
Open-loop system, 308–310
Optimal and robust control system design, 59–60
Output error (OE) model
definition, 96–97
parameter identification
 $R(k)$ and Hessian approximation, 101–103
 $\partial V(\theta_k)/\partial \theta$ and Jacobian calculation, 100–101
pseudo-regression vector, 98
Overlapping power subnetwork
coordination strategy
control, 253–255
goals, 256
network characteristics and control objectives, 256–257
optimal flow control, 268
optimal power flow control, 255
parameters, 268
power networks, control structure, 252–253
definition, 257
multi-agent control
overlapping, 263–267
touching, 258–263

P

Petrov–Galerkin approximation, 8–9
Pick matrices. *See* Loewner and Pick matrices
Polya's theorem, 119
Power networks
control objectives, 273
multi agent control, 252–253
parameters, 268
setting up the control problems, 273–274
simulations, 274–276
steady-state characteristics
FACTS devices, 272
generators, 271
loads, 271

power balance, 272–273
transmission lines, 270–271
Putinar's theorem, 119

R

Reduced order data driven model
criterion selection
long-step prediction *vs.* one-step-ahead prediction, 94–96
parallel-series model, 95
model selection: EE *vs.* OE
equation error (EE) model, 97–98
output error (OE) model, 98–99
Robust closed-loop performance, 116
Robust control
applications, 116–117
continuous-time Lyapunov equation, 119
controllability and observability, 116
LTI uncertain fourth-order system, 129
non-polynomial rational function, 120
polytopic region
homogeneous matrix polynomial, 125–126
optimization problem, 124
positive semi-definite matrix polynomial, 121
problem formulation
polynomially uncertain system, 117
sum-of-squares, 118–119
robust stability, 115
SOS matrix polynomials, 122–123

S

SADPA. *See* Subspace accelerated dominant pole algorithm
Scalar real-valued continuous functions, 136
Second-order tangential MIMO order reduction, 36
Shifted Loewner matrix, 44
Simultaneous gradient error reduction
conjugate gradient search routine, 62
fit error, 63
schematic representation, 63
Single-input/single-output (SISO) system, 4
Single rigid rod dynamics
configuration matrix, 75
equations of motion, 78–79
generalized forces and torques, 78
illustration, 74–75
kinetic energy, 75
nodes, 76–77
string forces, 77–78
SOS. *See* Sum-of-squares
Spanning trees, 222, 223, 238–240, 242–244
Spectral decomposition process

- distinct frequency range groupings, 60
eigenvalue and eigenvector calculation, 60–62
schematic representation, 61
- State estimation level consensus
globally LQ optimal controller, 209, 210
Luenberger form, 208
overlapping decentralized estimators, 208
stability analysis, 209
- Static var compensators (SVC)
flexible alternating current transmission systems, 272
simulation control, 274–276
- Stochastic projection method, 180
- String connectivity, 82
- String forces, 77–78
- Structural damping coefficients, 319–321
- Structure-preserving model reduction, coprime factorizations
driven cavity flow, 33–35
isotropic incompressible viscoelastic solid, 31
second-order dynamical systems, 35–37
- Subspace accelerated dominant pole algorithm (SADPA), 27
- Sum-of-squares (SOS)
Polya’s theorem, 119
Putinar’s theorem, 119
scalar polynomial, 118
- T**
- Tangential interpolation, 6–7
- Ten-DOF system, 318
- Tensegrity systems
class k
class 2 tensegrity cable model, 82–87
 N interconnected rigid rods, 82–83
- class 1 tensegrity systems
compact matrix expression, 80
configuration matrix, 79
Lagrangian approach, 79
prism, 81
string connectivity, 82
definition, 73
dynamic models, 74
single rigid rod dynamics
configuration matrix, 75
equations of motion, 78–79
generalized forces and torques, 78
kinetic energy, 75
nodes, 76–77
string forces, 77–78
- Thyristor controlled series compensators (TCSC)
flexible alternating current transmission systems, 272
simulation control, 275–276
- U**
- Unmanned aerial vehicles (UAVs) formations
decentralized state estimation, 213–214
design, 210
experiments
consensus based controllers, 215–216
expansion/contraction paradigm, 215–217
inclusion principle, 215
formation model
Kronecker’s product, 211
linear double integrator model, 210
global LQ optimal state feedback, 212–213
- W**
- Weak cascading upper diagonal dominance (w-CUDD), 159