Structure-Preserving Model Order Reduction for Network Systems

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von **M. Sc. Petar Mlinarić**

geb. am **25.04.1990** in Zagreb, Kroatien

genehmigt durch die Fakultät für Mathematik

der Otto-von-Guericke-Universität Magdeburg

Gutachter: **Prof. Dr. Peter Benner**

Prof. Dr. Harry L. Trentelman

eingereicht am: **30.10.2019**

Verteidigung am: **23.01.2020**

ABSTRACT

This thesis considers structure-preserving system-theoretic model order reduction for certain structured input-output systems, particularly network systems. In the first part, our focus lies on the clustering-based approach to reduce network systems. Therein, we begin by considering clustering-based model order reduction for linear multi-agent systems. This approach finds a reduced model whose dynamics evolve over a smaller network. To measure the reduction error, we use the $\mathcal{H}_2$-norm and consider the $\mathcal{H}_2$-optimal clustering problem. Since clustering is generally a difficult combinatorial problem, we propose a framework based on relaxing the discrete problem to find an $\mathcal{H}_2$-suboptimal clustering. Following on this, we directly extend the framework to a class of nonlinear multi-agent systems.

Next, we derive upper bounds for $\mathcal{H}_2$ and $\mathcal{H}_\infty$ clustering-based reduction errors for linear multi-agent systems based on almost equitable partitions. These results generalize work for multi-agent systems with single-integrator agents. Using a similar approach for power systems, which are a special class of nonlinear multi-agent systems, we find conditions for exact clustering-based reduction.

Additionally, we study subsystem reduction for network systems. We propose a balancing-based approach that guarantees stability preservation under a small-gain condition. Furthermore, we consider the $\mathcal{H}_2$-optimal subsystem reduction problem. We derive Gramian-based first-order necessary $\mathcal{H}_2$-optimality conditions and use a gradient-based optimization method to fulfill them.

Finally, we apply the structure-preserving $\mathcal{H}_2$-optimal model reduction approach for network systems to other structured systems. In particular, we consider $\mathcal{H}_2$-optimal model order reduction of second-order systems, port-Hamiltonian systems, time-delay systems and $\mathcal{H}_2 \otimes \mathcal{L}_2$-optimal model order reduction of parametric systems. Here, we also derive Gramian-based $\mathcal{H}_2$-optimality conditions and use an optimization approach to construct a reduced model. For some structured systems, we also derive interpolatory $\mathcal{H}_2$-optimality conditions under additional assumptions on the reduced model.

ZUSAMMENFASSUNG

Diese Arbeit beschäftigt sich mit der strukturerhaltenden systemtheoretischen Modellordnungsreduktion für bestimmte strukturierte Input-Output-Systeme, insbesondere Netzwerksysteme. Im ersten Teil liegt unser Fokus auf dem clusterbasierten Ansatz zur Reduktion von Netzwerksystemen. Zunächst betrachten wir dabei die clusterbasierte Modellreduktion für lineare Multiagentensysteme. Dieser Ansatz findet ein reduziertes Modell, dessen Dynamik sich über ein kleineres Netzwerk entwickelt. Zum Messen des Reduktionsfehlers verwenden wir die $\mathcal{H}_2$-Norm und betrachten das $\mathcal{H}_2$-optimale Clustering-Problem. Da das Clustering im Allgemeinen ein schwieriges kombinatorisches Problem ist, schlagen wir ein Framework vor, welches darauf basiert, das diskrete Problem zu relaxieren, um ein $\mathcal{H}_2$-suboptimales Clustering zu finden. Im Anschluss daran erweitern wir das Framework direkt auf eine Klasse nichtlinearer Multiagentensysteme.

Als nächstes leiten wir obere Schranken für $\mathcal{H}_2$ und $\mathcal{H}_\infty$ clusterbasierte Reduktionsfehler für lineare Multiagentensysteme her, basierend auf nahezu gerechten Partitionierungen. Diese Ergebnisse verallgemeinern die Arbeit für Multiagentensysteme mit Single-Integrator-Agenten. Unter Verwendung eines ähnlichen Ansatzes für Energiesysteme, die eine spezielle Klasse nichtlinearer Multiagentensysteme darstellen, finden wir Bedingungen für eine genaue clusterbasierte Reduktion.

Zusätzlich untersuchen wir die Subsystemreduktion für Netzwerksysteme. Wir schlagen einen balancierenden Ansatz vor, der die Erhaltung der Stabilität unter Small-Gain-Bedingungen gewährleistet. Weiterhin betrachten wir das $\mathcal{H}_2$-optimale Subsystemreduktionsproblem. Wir leiten gramschenbasierte notwendige $\mathcal{H}_2$-Optimalitätsbedingungen erster Ordnung her und verwenden eine gradientenbasierte Optimierungsmethode um diese zu erfüllen.

Schließlich wenden wir die strukturerhaltende $\mathcal{H}_2$-optimale Modellreduktion für Netzwerksysteme auf andere strukturierte Systeme an. Insbesondere betrachten wir $\mathcal{H}_2$-optimale Modellordnungsreduktion von Systemen zweiter Ordnung, Port-Hamiltonscher Systeme, zeitverzögerter Systeme und $\mathcal{H}_2 \otimes \mathcal{L}_2$-optimale Modellordnungsreduktion von parametrischen Systemen. Auch hier leiten wir gramschenbasierte $\mathcal{H}_2$-Optimalitätsbedingungen her und verwenden einen Optimierungsansatz um ein reduziertes Modell zu konstruieren. Für einige strukturierte Systeme leiten wir auch interpolatorische $\mathcal{H}_2$-Optimalitätsbedingungen unter zusätzlichen Annahmen an das reduzierte Modell

her.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

LIST OF ACRONYMS

AEP	almost equitable partition
BRBT	bounded real balanced truncation
BT	balanced truncation
FOM	full-order model
IRKA	iterative rational Krylov algorithm
LTI	linear time-invariant
MIMO	multiple-input multiple-output
MOR	model order reduction
ODE	ordinary differential equation
PLTI	parameterized LTI
ROM	reduced-order model
SISO	single-input single-output
TSIA	two-sided iteration algorithm

LIST OF SYMBOLS

$\mathbb{R}$, $\mathbb{C}$	fields of real and complex numbers
$\mathbb{C}_+$, $\mathbb{C}_-$	open right/left complex half plane
$\mathbb{R}_+$, $\mathbb{R}_-$	strictly positive/negative real line
$\mathbb{R}^n$, $\mathbb{C}^n$	vector space of real/complex n-tuples
$\mathbb{R}^{m\times n}$, $\mathbb{C}^{m\times n}$	set of real/complex $m \times n$ matrices
$\lvert S\rvert$	cardinality of set S
$\lvert\xi\rvert$	absolute value of real or complex scalar
$\arg \xi$	argument of complex scalar
$\boldsymbol{\imath}$	imaginary unit ($\boldsymbol{\imath}^2 = -1$)
$\mathrm{Re}(A)$, $\mathrm{Im}(A)$	real and imaginary part of a complex quantity $A = \mathrm{Re}(A) + \boldsymbol{\imath}\,\mathrm{Im}(A) \in \mathbb{C}^{m\times n}$
$\overline{A}$	$\coloneqq \mathrm{Re}(A) - \boldsymbol{\imath}\,\mathrm{Im}(A)$, complex conjugate of $A \in \mathbb{C}^{m\times n}$
a_{ij}	the (i,j)-th entry of A
$A_{i:j,:}$, $A_{:,k:\ell}$	rows $i,\ldots,j$ of A and columns $k,\ldots,\ell$ of A
$A_{i:j,k:\ell}$	rows $i,\ldots,j$ of columns $k,\ldots,\ell$ of A
$\mathrm{im}(A)$	$\coloneqq \mathrm{span}\{A_{:,1}, A_{:,2}, \ldots, A_{:,n}\}$, subspace generated by the columns of $A \in \mathbb{C}^{m\times n}$
$\ker(A)$	kernel of $A \in \mathbb{C}^{m\times n}$
A^{T}	the transpose of A
A^*	$\coloneqq (\overline{A})^{\mathrm{T}}$, the complex conjugate transpose
A^{-1}	the inverse of invertible A
$A^{-\mathrm{T}}$, A^{-*}	the inverse of A^{T}, A^*

A^+	the Moore-Penrose pseudoinverse of A
I_n, $I_{n,r}$	identity matrix of dimension n, first r columns of I_n
e_i	the ith column of the identity matrix
$\mathbb{1}_r$	vector of ones in $\mathbb{R}^r$
$\operatorname{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n)$	$n \times n$ diagonal matrix with the α_i's on the diagonal
$\operatorname{diag}(A_1, A_2, \ldots, A_n)$	block diagonal matrix with the A_i's as diagonal blocks
$\sigma(A)$, $\sigma(A, E)$	spectrum of matrix A/matrix pair (A, E)
$\lambda_i(A)$, $\lambda_i(A, E)$	i-th eigenvalue of A/(A, E)
$\rho(A, E)$	$:= \max_i \lvert\lambda_i(A, E)\rvert$, spectral radius of (A, E)
$\sigma_{\max}(A)$	the largest singular value of A
$\operatorname{tr}(A)$	$:= \sum_{i=1}^n a_{ii}$, trace of A
$\lVert u\rVert_2$	$:= \sqrt{\sum_{i=1}^n \lvert u_i\rvert^2}$ for $u \in \mathbb{C}^n$
$\lVert A\rVert_2$	$:= \sup\{\lVert Au\rVert_2 : \lVert u\rVert_2 = 1\}$, induced matrix 2-norm
$\lVert A\rVert_{\mathrm{F}}$	$:= \sqrt{\sum_{i,j} \lvert a_{ij}\rvert^2} = \sqrt{\operatorname{tr}(A^*A)}$, Frobenius norm of matrix $A \in \mathbb{C}^{m\times n}$
$\lVert u\rVert$, $\lVert A\rVert$	Euclidean vector or induced matrix norm $\lVert\cdot\rVert_2$
$A \succ (\succcurlyeq, \prec, \preccurlyeq)\ 0$	A is self-adjoint positive definite (positive semidefinite, negative definite, negative semidefinite)
$A > (\geqslant)\ B$	element-wise partial ordering $a_{ij} > (\geqslant)\ b_{ij}$, for all i, j
$A \circ B$	Hadamard (element-wise) product of A and B
$A \otimes B$	Kronecker product of A and B (see Definition 2.7)
$\operatorname{vec}(A)$	vectorization of A (see Definition 2.7)
$\operatorname{col}(a_1, a_2, \ldots, a_n)$	vector in $\mathbb{C}^{nm}$ formed by stacking $a_1, a_2, \ldots, a_n \in \mathbb{C}^m$
$\partial_{x_i} f := \frac{\partial}{\partial x_i} f$	partial derivative of f with respect to x_i

$\partial_{x_i x_j} f \coloneqq \frac{\partial^2}{\partial x_i \partial x_j} f$	$\coloneqq \partial_{x_i} \partial_{x_j} f$, second order partial derivative of f with respect to x_i and x_j
$\partial_{x_i}^2 f \coloneqq \partial_{x_i x_i} f$	second order partial derivative of f with respect to x_i
$\dot{f} \coloneqq \partial_t f \coloneqq \frac{\partial}{\partial t} f$	derivative of f with respect to time
$\ddot{f} \coloneqq \partial_t^2 f \coloneqq \frac{\partial^2}{\partial t^2} f$	second derivative of f with respect to time
$\mathrm{D} f(x)$	Fréchet differential of f at x (see Definition 2.10)
$\mathrm{D}^{\mathrm{G}} f(x)$	Gateaux differential of f at x (see Definition 2.14)
$\mathrm{d} f(x; h)$	directional derivative of f at x in direction h (see Definition 2.14)
$\nabla f(x)$	gradient of f at x (see Definition 2.16)

CHAPTER 1

INTRODUCTION

Contents

1.1 Motivation

Large-scale dynamical systems, consisting of many ordinary differential equations (ODEs), appear often in applications. Such systems typically arise from detailed discretizations of partial differential equations (PDEs). Another source are interconnected systems, forming large-scale network systems, see, e.g., [New10, EFHO10, BFF+14] for applications in complex networks, smart-grids, distributed systems, transportation networks, biological networks, and networked multi-agent systems. Due to the increasing demand for computational resources to analyze, simulate, or control such large-scale systems, there is an interest in using model order reduction (MOR) methods. The idea behind MOR is to find a reduced-order model (ROM) with a much smaller number of ODEs, making it easier to simulate, while reproducing the original model sufficiently accurately.

Direct application of established MOR techniques, such as balanced truncation, Hankel-norm approximation, and Krylov subspace methods, see, e.g., [Ant05, BMS05, BOCW17], to structured models generally leads to a loss of structure. Additionally, for multi-agent systems, properties such as consensus and synchrony are important to preserve in the reduced model (see [Mor05, OSM03, LDCH10, MZ10]). Structure-preserving MOR methods allow preserving the physical interpretation of the model, which can also improve accuracy. Additionally, analysis or optimization methods tailored to specific model structures allows reusing them for the reduced model.

MOR techniques specifically for networked multi-agent systems with first-order agents have been proposed in [IKIA14, IKG+15, CKS16]. Extensions to second-order agents have been considered in [II15, CKS17] and to more general higher-order agents in [IKI16a,

BSJ16, MTC13]. Some of these methods are based on clustering nodes in the network. With clustering, the idea is to partition the set of nodes in the network graph into disjoint sets called clusters, and to associate with each cluster a single, new, node in the reduced network, thus reducing the number of nodes and connections and the complexity of the network topology.

Complementarily, methods for subsystem reduction in more general network systems have been proposed in [RS07, RS08b, VVD08, SM09], with applications in, e.g., multibody systems. There, the network in the reduced model stays the same and only parts of the system in the nodes are reduced.

1.2 Outline of the thesis

This thesis is structured as follows. In Chapter 2, we review different topics used throughout the thesis. We begin with the necessary topics from linear algebra, particularly properties of eigenvalues and the Kronecker product. Then we give an overview of some topics from functional analysis. In particular, we will need differentiability concepts for functions between arbitrary Hilbert spaces in Chapter 5 and Chapter 6. Since the considered MOR methods are based on systems theory, we review system properties such as stability, controllability, and observability. Additionally, we cover Hardy spaces, particularly the $\mathcal{H}_2$ space. Next, we provide overview for some projection-based MOR methods, including balanced truncation and interpolatory methods. For multi-agents systems, we revise some basic concepts from graph theory. Then, we cover modeling and some properties of linear multi-agent systems used in Chapter 3 and Chapter 4.

In Chapter 3, we study clustering-based MOR of multi-agent systems. The goal is to find good partitioning of the nodes of the graph over which the dynamics is evolving. Clustering then produces a multi-agent system evolving over a smaller graph. First, we focus on linear multi-agent systems and consider the problem of $\mathcal{H}_2$-optimal clustering-based MOR. Since clustering is generally a very difficult combinatorial problem, we relax the discrete problem to a continuous $\mathcal{H}_2$-optimal MOR problem. To recover a discrete solution, we interpret it as a problem of approximating a subspace. We propose using a clustering algorithm over the sets of rows of a matrix used to project the original model. Then, we generalize this to a framework of combining a projection-based method with a clustering algorithm and apply it to nonlinear multi-agent systems. The results for linear multi-agent systems with single-integrator agents are published in

> P. Mlinarić, S. Grundel, and P. Benner, Efficient Model Order Reduction for Multi-Agent Systems Using QR Decomposition-Based Clustering, *Proceedings of the 54th IEEE Conference on Decision and Control (CDC)*, pp. 4794–4799, December 2015.

An extension for linear multi-agent systems with more general agents is published in

> P. Mlinarić, S. Grundel, and P. Benner, Clustering-Based Model Order Reduction for Multi-Agent Systems with General Linear Time-Invariant

Agents, *Proceedings of the 22nd International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, pp. 230–235, July 2016.

The results for nonlinear multi-agent systems will be published in a forthcoming paper.

In Chapter 4, we turn to more theoretical consideration of error due to clustering. First, we look at linear multi-agent systems and derive $\mathcal{H}_2$ and $\mathcal{H}_\infty$ error bounds when using an almost equitable partition (AEP), extending known $\mathcal{H}_2$-error expressions for multi-agent systems with single-integrator agents. We also propose an extension to arbitrary partitions using the distance to a graph for which the partition becomes almost equitable. These results are published in

H.-J. Jongsma, P. Mlinarić, S. Grundel, P. Benner, and H. L. Trentelman, Model Reduction of Linear Multi-Agent Systems by Clustering with $\mathcal{H}_2$ and $\mathcal{H}_\infty$ Error Bounds, *Mathematics of Control, Signals, and Systems*, vol. 30, April 2018.

Next, we consider power systems, which are a type of nonlinear multi-agent systems. There, we derive equivalent conditions for clustering with zero error. These conditions involve graph symmetries and equitable partitions. The results are published in

P. Mlinarić, T. Ishizaki, A. Chakrabortty, S. Grundel, P. Benner, and J. Imura, Synchronization and Aggregation of Nonlinear Power Systems with Consideration of Bus Network Structures, *Proceedings of the European Control Conference (ECC)*, pp. 2266–2271, June 2018.

In Chapter 5, we study the problem of subsystem reduction for linear network systems. This is a complementary approach to clustering, where the underlying graph is preserved and only the subsystems at the nodes are reduced. In the first part, we extend a balancing-based MOR method which preserves stability for network systems satisfying a certain small-gain condition. Using the known a priori $\mathcal{H}_\infty$ error bound, it allows automatic choice of the order of the reduced subsystems. The results are published in

P. Benner, S. Grundel, and P. Mlinarić, Stability Preserving Model Reduction for Linearly Coupled Linear Time-Invariant Systems, *Proceedings in Applied Mathematics and Mechanics*, vol. 16, pp. 817–818, October 2016.

The second part is about $\mathcal{H}_2$-optimal subsystem reduction. Using the Gramian-based formulation of the $\mathcal{H}_2$-error, we derive gradients with respect to matrices defining the ROM. Therefore, we also obtain Wilson-type necessary optimality conditions. These results are subject of a forthcoming article.

In Chapter 6, we use the ideas from the second part of Chapter 5 to other structure-preserving $\mathcal{H}_2$-optimal MOR problems. In particular, we consider structure-preserving MOR for second-order systems, port-Hamiltonian systems, and time-delay systems. Additionally, we also consider $\mathcal{H}_2 \otimes \mathcal{L}_2$-optimal MOR for parametric systems. We derive Wilson-type necessary optimality conditions and for some systems also the interpolatory optimality conditions. The preliminary results for parametric systems are published in

M. Hund, P. Mlinarić, and J. Saak, An $\mathcal{H}_2 \otimes \mathcal{L}_2$-Optimal Model Order Reduction Approach for Parametric Linear Time-Invariant Systems, *Proceedings in Applied Mathematics and Mechanics*, vol. 18, pp. e201800084, December 2018.

Finally, in Chapter 7 we summarize the results of the thesis and discuss possible future directions.

CHAPTER 2

MATHEMATICAL PRELIMINARIES

Contents

2.1 Linear algebra

Here, we recall some definitions and properties related to eigenvalues and matrix equations.

2.1.1 Eigenvalues and eigenvectors of matrices and matrix pairs

We begin with the definition of eigenvalues, right and left eigenvectors, spectrum, and spectral radius of a matrix.

Definition 2.1 ([GV13, Section 7.1.1]):
A scalar $\lambda \in \mathbb{C}$ is an *eigenvalue* of a matrix $A \in \mathbb{C}^{n\times n}$ if there exists a nonzero vector $x \in \mathbb{C}^n$ such that $Ax = \lambda x$. The vector x is called a *(right) eigenvector* of A corresponding to the eigenvalue λ. A nonzero vector $y \in \mathbb{C}^n$ is a *left eigenvector* of A corresponding to the eigenvalue λ if $y^*A = \lambda y^*$. The set of eigenvalues of A is called the *spectrum* and is denoted by $\sigma(A)$. The *spectral radius* of the matrix A is $\rho(A) \coloneqq \max_{\lambda\in\sigma(A)} |\lambda|$. ◊

Next, we continue with diagonalizability of a matrix and simultaneous diagonalizability of two or more matrices.

Definition 2.2 ([HJ85, Definition 1.3.6]):
The matrix $A \in \mathbb{C}^{n\times n}$ is said to be *diagonalizable* if there exists an invertible $T \in \mathbb{C}^{n\times n}$ such that $T^{-1}AT$ is a diagonal matrix. ◊

Definition 2.3 ([HJ85, Definition 1.3.11, 1.3.18]):
Two matrices $A, B \in \mathbb{C}^{n\times n}$ are said to be *simultaneously diagonalizable* if there exists an invertible $T \in \mathbb{C}^{n\times n}$ such that $T^{-1}AT$ and $T^{-1}BT$ are both diagonal matrices.

A *simultaneously diagonalizable* family $\mathcal{F} \subseteq \mathbb{C}^{n\times n}$ is a family for which there is a single invertible matrix $T \in \mathbb{C}^{n\times n}$ such that $T^{-1}AT$ is diagonal for every $A \in \mathcal{F}$. ◊

The following theorem gives necessary and sufficient conditions for simultaneous diagonalizability. Recall that two matrices A and B are said to commute if $AB = BA$. A family $\mathcal{F} \subseteq \mathbb{C}^{n\times n}$ of matrices is a commuting family if every pair of matrices from $\mathcal{F}$ commute [HJ85, Definition 1.3.16].

Theorem 2.4 ([HJ85, Theorem 1.3.12, 1.3.19]):
Let $A, B \in \mathbb{C}^{n\times n}$ be diagonalizable. Then A and B commute if and only if they are simultaneously diagonalizable.

Let $\mathcal{F} \subseteq \mathbb{C}^{n\times n}$ be a family of diagonalizable matrices. Then $\mathcal{F}$ is a commuting family if and only if it is a simultaneously diagonalizable family. ◊

Next definition is for eigenvalues and eigenvectors of a matrix pencil $A - \lambda B$ (or matrix pair (A, B)) with invertible B matrix.

Definition 2.5 ([GV13, Section 7.7]):
A scalar $\lambda \in \mathbb{C}$ is an *eigenvalue* of a matrix pencil $A - \lambda B$, $A, B \in \mathbb{C}^{n\times n}$, B invertible, if there exists a nonzero vector $x \in \mathbb{C}^n$ such that $Ax = \lambda Bx$. The vector x is called a *(right) eigenvector* of $A - \lambda B$ corresponding to the eigenvalue λ. A nonzero vector $y \in \mathbb{C}^n$ is a *left eigenvector* of $A - \lambda B$ corresponding to the eigenvalue λ if $y^*A = \lambda By^*$. The set of eigenvalues of $A - \lambda B$ is denoted by $\sigma(A, B)$. ◊

Notice that $\sigma(A, B) = \sigma(B^{-1}A)$.

The following result gives bounds for the eigenvalues of a "projected" symmetric matrix, where we use that the eigenvalues of a symmetric are real ([HJ85, Theorem 2.5.6]) to label them in increasing order.

Theorem 2.6 (Interlacing property, [MN99, Section 11.10]):
Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix with eigenvalues $\lambda_1(A) \leqslant \lambda_2(A) \leqslant \cdots \leqslant \lambda_n(A)$, $P \in \mathbb{R}^{n\times r}$ a matrix with orthonormal columns, and $B = P^{\mathrm{T}}AP$. Then the eigenvalues $\lambda_1(B) \leqslant \lambda_2(B) \leqslant \cdots \leqslant \lambda_r(B)$ of B satisfy

$$\lambda_i(A) \leqslant \lambda_i(B) \leqslant \lambda_{n-r+i}(A),$$

for $i = 1, 2, \ldots, r$. ◊

2.1.2 Kronecker product, vectorization, and matrix equations

We start with the definition of the Kronecker product and vectorization operator.

Definition 2.7 ([GV13, Section 1.3.6]):
For $A = [a_{ij}] \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$, the Kronecker product $A \otimes B \in \mathbb{R}^{mp\times nq}$ and vectorization $\mathrm{vec}(A) \in \mathbb{R}^{mn}$ are defined by

$$A \otimes B \coloneqq \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}, \quad \mathrm{vec}(A) \coloneqq \begin{bmatrix} A_{:,1} \\ A_{:,2} \\ \vdots \\ A_{:,n} \end{bmatrix}. \qquad ◊$$

Next, we state some properties of the Kronecker product and its relation to the vectorization operator.

Proposition 2.8 ([GV13, Sections 1.3.6 and 1.3.7]):
We have for any scalar α and all matrices A, B, C, D of compatible dimensions:

1. $(A \otimes B) \otimes C = A \otimes (B \otimes C)$,
2. $(\alpha A) \otimes B = A \otimes (\alpha B) = \alpha(A \otimes B)$,
3. $(A + B) \otimes C = A \otimes C + B \otimes C$,
4. $A \otimes (B + C) = A \otimes B + A \otimes C$,

5. $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$
6. $(A \otimes B)^{\mathrm{T}} = A^{\mathrm{T}} \otimes B^{\mathrm{T}}$,
7. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ (for invertible A and B),
8. $\operatorname{vec}(ABC) = (C^{\mathrm{T}} \otimes A) \operatorname{vec}(B)$,
9. $\sigma(A \otimes B) = \{\lambda\mu \mid \lambda \in \sigma(A),\ \mu \in \sigma(B)\}$ (for square matrices A and B). ◊

Note that the Kronecker product is not commutative, but the result is equal up to a permutation of rows and columns.

We will be interested in matrix equations of the form

$$AXB^{\mathrm{T}} + CXD^{\mathrm{T}} + E = 0, \tag{2.1}$$

where $A, C \in \mathbb{R}^{n\times n}$, $B, D \in \mathbb{R}^{m\times m}$, $X, E \in \mathbb{R}^{n\times m}$, and B and C are invertible. Vectorization of (2.1) gives

$$(B \otimes A + D \otimes C) \operatorname{vec}(X) = -\operatorname{vec}(E)\,.$$

Clearly, the above equation has a unique solution if and only if the matrix $B\otimes A+D\otimes C$ is invertible. The following theorem gives the equivalent condition.

Theorem 2.9 ([Chu87, Theorem 1]):
The matrix equation (2.1) has a unique solution if and only if $\lambda + \mu \neq 0$ for all $\lambda \in \sigma(A, C)$ and $\mu \in \sigma(D, B)$. ◊

Notice that if $\sigma(A, C) \subset \mathbb{C}_-$ and $\sigma(D, B) \subset \mathbb{C}_-$, then the condition in Theorem 2.9 is satisfied.

2.2 Functional analysis

Here, we give necessary basics of calculus in Banach spaces. In particular, we present differentiability, implicit function theorem, and Lagrange multiplier method. We base the presentation on the textbooks [Zei85a, Col12].

2.2.1 Fréchet and Gateaux differentiability

We only consider vector spaces over the field of real numbers $\mathbb{R}$. For normed vector spaces X, Y, we use $B(X, Y)$ to denote the space of bounded linear operators $A\colon X \to Y$. The dual space of X is denoted $X^* \coloneqq B(X, \mathbb{R})$.

We begin with the definition of Fréchet differentiability.

Definition 2.10 ([Col12, Section 2.2], [Zei85a, Definition 4.5]):
Let X and Y be normed vector spaces with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively, and $f\colon U \to Y$ a function, where U is an open subset of X and x an element of U.

The function f is *Fréchet differentiable at* x if there exists an operator $A \in B(X, Y)$ such that

$$\lim_{h\to 0} \frac{\|f(x+h) - f(x) - Ah\|_Y}{\|h\|_X} = 0.$$

If it exists, this A is called the *Fréchet differential of* f *at* x and is denoted by $\mathrm{D}f(x)$.◊

It can be seen that the operator A in the above definition is unique (see [Col12, Proposition 2.2]), which justifies using the notation $\mathrm{D}f(x)$.

For a function $r\colon U \subseteq X \to Y$, defined on some open neighborhood of zero, we write $r(h) = o(\|h\|^n)$ to mean

$$\lim_{h\to 0} \frac{r(h)}{\|h\|^n} = 0.$$

Then, for a function f which is Fréchet differentiable at x, we can write

$$f(x+h) = f(x) + \mathrm{D}f(x)h + o(\|h\|).$$

Similar properties as for finite-dimensional spaces hold here.

Proposition 2.11 ([Col12, Proposition 2.7], [Zei85a, Proposition 4.9]):
Let X and Y be normed vector spaces, U an open subset of X, and x an element of U. If $f, g\colon U \to Y$ are Fréchet differentiable at x, then $f+g$ is differentiable at x, as is αf, for any $\alpha \in \mathbb{R}$, and

$$\mathrm{D}(f+g)(x) = \mathrm{D}f(x) + \mathrm{D}g(x), \quad \mathrm{D}(\alpha f)(x) = \alpha\, \mathrm{D}f(x). \qquad \Diamond$$

Proposition 2.12 ([Col12, Theorem 2.1], [Zei85a, Proposition 4.10]):
Let X, Y, Z be normed vector spaces, U an open subset of X, V an open subset of Y, x an element of U, and $f\colon U \to Y$ and $g\colon V \to Z$ functions such that $f(U) \subseteq V$. If f is Fréchet differentiable at x and g is Fréchet differentiable at $f(x)$, then $g \circ f$ is Fréchet differentiable at x and

$$\mathrm{D}(g \circ f)(x) = \mathrm{D}g(f(x))\, \mathrm{D}f(x). \qquad \Diamond$$

The following is the extension of continuously differentiability to functions between normed vector spaces.

Definition 2.13 ([Col12, Section 2.4], [Zei85a, Definition 4.22]):
Let X and Y be normed vector spaces and $f\colon U \to Y$ a function, where U is an open subset of X. The function f is of *class* C^1 if it is Fréchet differentiable at every $x \in U$ and $\mathrm{D}f\colon U \to B(X, Y)$ is continuous. ◊

Next is the extension of directional derivatives.

Definition 2.14 ([Col12, Section 2.1], [Zei85a, Definition 4.5]):
Let X and Y be normed vector spaces and $f\colon U \to Y$ a function, where U is an open subset of X containing $x \in X$.

1. For $h \in X$, if the limit

$$\lim_{t\to 0} \frac{f(x+th) - f(x)}{t}$$

exists, we call it the *directional derivative of f at x in direction h* and denote it by $\mathrm{d}f(x;h)$.

2. The function f is *Gateaux differentiable at x* if $\mathrm{d}f(x;h)$ exists for all $h \in X$ and there exists an operator $A \in B(X,Y)$ such that $\mathrm{d}f(x;h) = Ah$ for all $h \in X$. If it exists, the operator A is called the *Gateaux differential of f at x* and is donated by $\mathrm{D}^{\mathrm{G}}f(x)$. ◊

Similarly as for Fréchet differential, the Gateaux differential of a function is unique if it exists.

The following result gives the relation between Fréchet and Gateaux differentials.

Proposition 2.15 ([Zei85a, Proposition 4.8]):
Let X and Y be normed vector spaces and $f\colon U \to Y$ a function, where U is an open subset of X containing $x \in X$. Then,

1. if f is Fréchet differentiable at x, then it is also Gateaux differentiable at x and $\mathrm{D}f(x) = \mathrm{D}^{\mathrm{G}}f(x)$,

2. if f is Gateaux differentiable in some neighborhood of x and $\mathrm{D}^{\mathrm{G}}f$ is continuous at x, then f is Fréchet differentiable at x and $\mathrm{D}f(x) = \mathrm{D}^{\mathrm{G}}f(x)$. ◊

This result provides an alternative way of proving a function is Fréchet differentiable and finding its differential. In particular, assuming $t \mapsto f(x+th)$ is continuously differentiable, we have that

$$\mathrm{d}f(x;h) = \lim_{t\to 0} \frac{f(x+th) - f(x)}{t} = \frac{\mathrm{d}}{\mathrm{d}t} f(x+th)\bigg|_{t=0}.$$

Therefore, methods for computing derivatives of functions of a real variable can be used to find a candidate for the Fréchet differential.

Notice that if $Y = \mathbb{R}$, then $\mathrm{D}f(x) \in B(X,\mathbb{R}) = X^*$. Therefore, if X is a Hilbert space, functional $\mathrm{D}f(x)$ can be identified with an element of X by the Riesz representation theorem ([Col12, Theorem 6.4]).

Definition 2.16 ([Col12, Section 6.4]):
Let X be a Hilbert space with inner product $\langle \cdot, \cdot \rangle_X$, U an open subset of X, and $f\colon U \to \mathbb{R}$ a function. Let f be Fréchet differentiable at $x \in U$. The element $a \in X$, such that $\mathrm{D}f(x)h = \langle a, h \rangle_X$ for all $h \in X$, is called the *gradient of f at x* and is denoted by $\nabla f(x)$. ◊

Next is the extension of partial differentials.

Definition 2.17 ([Zei85a, Definition 4.13]):
Let X, Y, Z be Banach spaces, U and open subset of $X \times Y$, (x_0, y_0) an element of U, and $f\colon U \to Z$ a function.

Define an open subset $U_{y_0} = \{x \in X : (x, y_0) \in U\} \subseteq X$ and a function $g\colon U_{y_0} \to Z$ with $g(x) = f(x, y_0)$ for all $x \in U_{y_0}$. Let g be Fréchet differentiable at x_0. Then $\mathrm{D}g(x_0)$ is called the *partial Fréchet differential of f at (x_0, y_0) with respect to the first variable* x and is denoted by $\mathrm{D}_x f(x_0, y_0)$. ◊

The differential $\mathrm{D}_y f(x_0, y_0)$ is defined similarly.

The following result states the relation between Fréchet and partial Fréchet differentials.

Proposition 2.18 ([Zei85a, Proposition 4.14]):
Let X, Y, Z be Banach spaces, U and open subset of $X \times Y$, (x_0, y_0) an element of U, and $f\colon U \to Z$ a function.

1. If f is Fréchet differentiable at (x_0, y_0), then its partial Fréchet differentials $\mathrm{D}_x f(x_0, y_0)$ and $\mathrm{D}_y f(x_0, y_0)$ also exist and

   $$\mathrm{D}f(x_0, y_0)(h, k) = \mathrm{D}_x f(x_0, y_0)h + \mathrm{D}_y f(x_0, y_0)k,$$

   for all $h \in X$ and $k \in Y$.

2. If f has partial Fréchet differentials $\mathrm{D}_x f$ and $\mathrm{D}_y f$ in some neighborhood of (x_0, y_0) and if these are continuous at (x_0, y_0), then f is Fréchet differentiable at (x_0, y_0).

3. The function f is continuously Fréchet differentiable in a neighborhood of (x_0, y_0) if and only if all partial Fréchet differentials are continuous in a neighborhood of (x_0, y_0). ◊

Partial differential of a function of more variables $f(x_1, x_2, \ldots, x_n)$ are defined similarly and the above result can be directly extended. Furthermore, partial Gateaux differential, partial directional derivatives, and partial gradients can be defined in a similar way.

2.2.2 Implicit function theorem

We can now state the implicit function theorem.

Theorem 2.19 ([Zei85a, Theorem 4.B], [Col12, Theorem 8.2]):
Let X, Y, Z be Banach spaces, U an open subset of $X \times Y$, (x_0, y_0) an element of U, and $f\colon U \to Z$ a function. Suppose that

1. $f(x_0, y_0) = 0$,
2. $\mathrm{D}_y f$ exists on U and $\mathrm{D}_y f(x_0, y_0)$ is bijective,
3. f and $\mathrm{D}_y f$ are continuous at (x_0, y_0).

Then there exist open subsets $U_X \subseteq X$ and $U_Y \subseteq Y$ and a bijection $g\colon U_X \to U_Y$ such that $x_0 \in U_X$, $y_0 \in U_Y$, $g(x_0) = y_0$, and $f(x, g(x)) = 0$ for all $x \in U_X$.

Additionally, if f is of class C^1 on a neighborhood of (x_0, y_0), then g is of class C^1 on a neighborhood of x_0 and $\mathrm{D}g(x_0) = -\mathrm{D}_y f(x_0, g(x_0))^{-1} \mathrm{D}_x f(x_0, g(x_0))$. ◊

2.2.3 Constrained optimization and Lagrange multipliers

We conclude with the theorem about Lagrange multipliers. First, we need to define when is a function a submersion [Zei85b, Definition 43.15].

Definition 2.20:
Let X, Y be Banach spaces, U an open subset of X, x_0 an element of U, and $g\colon U \to Y$ a function. The function g is a *submersion* at x_0 if

1. g is of class C^1 in a neighborhood of x_0,
2. $\mathrm{D}g(x_0)\colon X \to Y$ is surjective,
3. the null space $\ker(\mathrm{D}g(x_0))$ splits X, i.e., there exists a continuous projection operator P of X on $\ker(\mathrm{D}g(x_0))$. ◊

Note that the third condition is immediately satisfied if X is a Hilbert space.

Next is the theorem about constrained local minima.

Theorem 2.21 ([Zei85b, Theorem 43.D]):
Let X, Y be real Banach spaces, U an open subset of X, x_0 an element of U, $f\colon U \to \mathbb{R}$ and $g\colon U \to Y$ functions.

Suppose f is Fréchet differentiable at x_0 and g is a submersion at x_0, and that f has a constrained local minimum at x_0 with respect to $\{x \in U : g(x) = 0\}$. Then there exists $\lambda \in Y^*$ such that

$$\mathrm{D}f(x_0)h - \lambda(\mathrm{D}g(x_0)h) = 0,$$

for all $h \in X$. ◊

For Hilbert spaces, we can use gradients.

Corollary 2.22:
Let X, Y be real Hilbert spaces, U an open subset X, x_0 an element of U, $f\colon U \to \mathbb{R}$ and $g\colon U \to Y$ functions.

Suppose f is Fréchet differentiable at x_0, g is of class C^1 in a neighborhood of x_0, $\mathrm{D}g(x_0)$ is surjective, and that f has a constrained local minimum at x_0 with respect to $\{x \in U : g(x) = 0\}$. Then there exists $\lambda \in Y$ such that

$$\langle \nabla f(x_0), h\rangle - \langle \lambda, \mathrm{D}g(x_0)h\rangle = 0,$$

for all $h \in X$. ◊

2.3 Linear time-invariant systems

We consider finite-dimensional, continuous-time, linear time-invariant (LTI) systems of the form

$$\begin{aligned} E\dot{x}(t) &= Ax(t) + Bu(t), \quad x(0) = x_0, \\ y(t) &= Cx(t) + Du(t), \end{aligned} \tag{2.2}$$

with *system matrices* $E, A \in \mathbb{R}^{n\times n}$, *input matrix* $B \in \mathbb{R}^{n\times m}$, *output matrix* $C \in \mathbb{R}^{p\times n}$, and *feedthrough matrix* $D \in \mathbb{R}^{p\times m}$. Furthermore, $t \in \mathbb{R}$ is *time*, $x_0 \in \mathbb{R}^n$ is the initial condition, while $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, and $y(t) \in \mathbb{R}^p$ are respectively the *state*, *input*, and *output* of the system. We call dimension n of the state $x(t)$ the *order* of the system (2.2). Throughout this thesis, we assume E to be invertible. Thereby, the state equation in (2.2) is a system of ODEs after multiplying from the left by E^{-1}.

If there is only one input and one output, i.e., if $m = p = 1$, we will refer to such a system as a *single-input single-output (SISO) system*. Otherwise, we will call it a *multiple-input multiple-output (MIMO) system*.

We will use $(E; A, B, C, D)$ to denote the system (2.2). If D is a zero matrix, we will also use $(E; A, B, C)$. Furthermore, we will use (A, B, C, D) and (A, B, C) to mean $(I_n; A, B, C, D)$ and $(I_n; A, B, C)$, respectively.

In this thesis, we focus on continuous-time systems. Results for discrete-time systems

$$\begin{aligned} Ex(k+1) &= Ax(k) + Bu(k), \quad x(0) = x_0, \\ y(k) &= Cx(k) + Du(k), \end{aligned}$$

should be directly extendable due to similarities to continuous-time systems (see [Ant05] for more details).

2.3.1 Solutions and stability

It can be verified that, for given initial condition x_0 and input u, the state and output of the system (2.2) satisfy

$$x(t) = e^{tE^{-1}A}x_0 + \int_0^t e^{\tau E^{-1}A}E^{-1}Bu(t-\tau)\,\mathrm{d}\tau,$$
$$y(t) = Ce^{tE^{-1}A}x_0 + \int_0^t Ce^{\tau E^{-1}A}E^{-1}Bu(t-\tau)\,\mathrm{d}\tau + Du(t).$$

The system (2.2) can also be solved in the frequency-domain. Assuming X, U, and Y exist as Laplace transforms of x, u, and y, by applying Laplace transform to the system (2.2), we find that

$$sEX(s) - Ex_0 = AX(s) + BU(s),$$
$$Y(s) = CX(s) + DU(s).$$

After eliminating $X(s)$, we obtain

$$Y(s) = C(sE - A)^{-1}Ex_0 + \big(C(sE - A)^{-1}B + D\big)\,U(s).$$

The function $H(s) = C(sE - A)^{-1}B + D$ is called the *transfer function*, and it characterizes the input-output relationship when $x_0 = 0$.

For any pair of invertible matrices $S, T \in \mathbb{R}^{n\times n}$, if we define a new state $\widetilde{x}(t) = T^{-1}x(t)$, we obtain an equivalent system

$$SET\dot{\widetilde{x}}(t) = SAT\widetilde{x}(t) + SBu(t), \quad \widetilde{x}(0) = T^{-1}x_0,$$
$$y(t) = CT\widetilde{x}(t) + Du(t),$$

in the sense that the input and output remain the same. Notice that also the transfer function H is independent of the choice of S and T. It is of interest to see which other system properties are invariant under this transformations.

Important properties of LTI systems are stability and asymptotic stability.

Definition 2.23 ([Ant05, Section 5.8]):
The autonomous system $E\dot{x}(t) = Ax(t)$ is called *stable* if all of its solutions x are bounded for positive time, i.e., the set $\{x(t)\,|\,t > 0\}$ is bounded. If additionally $\lim_{t\to\infty} x(t) = 0$, it is called *asymptotically stable*.

The system $(E; A, B, C, D)$ is *(asymptotically) stable* if the corresponding autonomous system $E\dot{x}(t) = Ax(t)$ is (asymptotically) stable. ◊

It is known that the system $(E; A, B, C, D)$ is

- asymptotically stable if and only if $\sigma(A, E) \subset \mathbb{C}_-$.
- stable if and only if $\sigma(A, E) \subset \overline{\mathbb{C}_-}$ and all purely imaginary eigenvalues are semi-simple (their geometric multiplicity and algebraic multiplicity are equal).

We call a matrix A (a matrix pair (A, E)) *asymptotically stable* or *Hurwitz* if $\sigma(A) \subset \mathbb{C}_-$ ($\sigma(A, E) \subset \mathbb{C}_-$). Additionally, we call a transfer function $H(s) = C(sE - A)^{-1}B + D$ *asymptotically stable* or *Hurwitz* if its poles are all in $\mathbb{C}_-$.

Notice that if the system $(E; A, B, C, D)$ is asymptotically stable, then its transfer function is also asymptotically stable, but the converse does not hold (e.g., if B or C is a zero matrix).

2.3.2 Controllability, observability, and Gramians

Further important properties are reachability and observability. We first discuss reachability and controllability (see [DP00, Section 2.2] and [Ant05, Section 4.2.1]).

Definition 2.24:
Let $(E; A, B, C, D)$ be a system with $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. A state $x_1 \in \mathbb{R}^n$ is *reachable by time* $t > 0$ if there exists an input $u \in \mathcal{L}_2(0, t)$ and a trajectory x such that $x(0) = 0$, $x(t) = x_1$, and $E\dot{x}(\tau) = Ax(\tau) + Bu(\tau)$ for almost all $\tau \in (0, t)$.

The set $\mathcal{R}(t) \subseteq \mathbb{R}^n$ of all reachable states by time $t > 0$ is called the *reachable set at time* t.

The $n \times nm$ matrix

$$\mathcal{C}(E; A, B) = \mathcal{C}(E^{-1}A, E^{-1}B) = \begin{bmatrix} E^{-1}B & E^{-1}AE^{-1}B & \cdots & (E^{-1}A)^{n-1}E^{-1}B \end{bmatrix}$$

is called the *controllability matrix*.

For $t > 0$, the *controllability Gramian* is the $n \times n$ matrix

$$P(t) = \int_0^t e^{E^{-1}At}E^{-1}BB^{\mathrm{T}}E^{-\mathrm{T}}e^{A^{\mathrm{T}}E^{-\mathrm{T}}t}\,\mathrm{d}t. \qquad \Diamond$$

We have the following result relating the reachable set, controllability matrix, and controllability Gramian.

Proposition 2.25 ([DP00, Theorem 2.2]):
For all $t > 0$, equality

$$\mathcal{R}(t) = \operatorname{im}(\mathcal{C}(E; A, B)) = \operatorname{im}(P(t))$$

holds. $\Diamond$

Since $\mathcal{C}(E; A, B)$ does not depend on t, we have that $\mathcal{R}(t)$ and $\operatorname{im}(P(t))$ are also constant.

Definition 2.26:
The system $(E; A, B, C, D)$ is *reachable* if $\mathcal{R}(t) = \mathbb{R}^n$, for some $t > 0$. $\Diamond$

For continuous-time systems, reachability is equivalent to controllability in the following sense (see [Ant05, Theorem 4.18]).

Definition 2.27:
The system $(E; A, B, C, D)$ is *controllable* if for any pair of states $x_0, x_1 \in \mathbb{R}^n$ there exist a time $t > 0$, an input $u \in \mathcal{L}_2(0, t)$, and trajectory x such that $x(0) = x_0$, $x(t) = x_1$, and $E\dot{x}(\tau) = Ax(\tau) + Bu(\tau)$ for almost all $\tau \in (0, t)$. ◊

Thus, we will use both terms. The following theorem gives equivalent conditions for controllability.

Theorem 2.28 ([Ant05, Theorem 4.15]):
The following are equivalent:

1. The system $(E; A, B, C, D)$ is controllable.
2. The controllability matrix is of full rank: $\operatorname{rank}(\mathcal{C}(E; A, B)) = n$.
3. The controllability Gramian $P(t)$ is positive definite for all $t > 0$.
4. If v is a left eigenvector of (A, E), then $v^*B \neq 0$.
5. $\operatorname{rank}\left(\begin{bmatrix}\lambda E - A & B\end{bmatrix}\right) = n$ for all $\lambda \in \mathbb{C}$. ◊

The following proposition gives an interesting property of the controllability matrix.

Proposition 2.29 ([DP00, Proposition 2.11]):
Suppose $E, A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. Then $\operatorname{im}(\mathcal{C}(E; A, B))$ is invariant under $E^{-1}A$.◊

In particular, we can see that $\operatorname{im}(\mathcal{C}(E; A, B))$ is the smallest subspace which contains $\operatorname{im}(E^{-1}B)$ and is invariant under $E^{-1}A$.

Next, we discuss observability (see [DP00, Section 2.4] and [Ant05, Section 4.2.2]).

Definition 2.30:
Let $(E; A, B, C, D)$ be a system with $A \in \mathbb{R}^{n\times n}$ and $C \in \mathbb{R}^{p\times n}$. A state $x_0 \in \mathbb{R}^n$ is *unobservable* if $Ce^{tE^{-1}A}x_0 = 0$ for all $t > 0$.

The set $\mathcal{N} \subseteq \mathbb{R}^n$ of all unobservable states is called the *unobservable set*. If $\mathcal{N} = \{0\}$, the system is called *observable*.

The $np \times n$ matrix

$$\mathcal{O}(E; A, C) = \mathcal{C}(E^{-1}A, C) = \begin{bmatrix} C \\ CE^{-1}A \\ \vdots \\ C\left(E^{-1}A\right)^{n-1} \end{bmatrix}$$

is called the *observability matrix*.

For $t > 0$, the *observability Gramian* is the $n \times n$ matrix

$$Q(t) = \int_0^t E^{-\mathrm{T}} e^{A^{\mathrm{T}}E^{-\mathrm{T}}t} C^{\mathrm{T}} C e^{E^{-1}At} E^{-1}\,\mathrm{d}t. \qquad ◊$$

Here is a result relating the unobservable set and observability matrix.

Proposition 2.31 ([DP00, Theorem 2.20]):
The unobservable set satisfies

$$\mathcal{N} = \ker(\mathcal{O}(E; A, C)) . \qquad \Diamond$$

The following gives equivalent conditions for observability. Notice the analogy with Theorem 2.28.

Theorem 2.32 ([Ant05, Theorem 4.26]):
The following are equivalent:

1. The system $(E; A, B, C, D)$ is observable.
2. The observability matrix is of full rank: $\operatorname{rank}(\mathcal{O}(E; A, C)) = n$.
3. The observability Gramian $Q(t)$ is positive definite for all $t > 0$.
4. If v is a right eigenvector of (A, E), then $Cv \neq 0$.
5. $\operatorname{rank}\left(\begin{bmatrix}\lambda E^{\mathrm{T}} - A^{\mathrm{T}} & C^{\mathrm{T}}\end{bmatrix}\right) = n$ for all $\lambda \in \mathbb{C}$. $\Diamond$

We introduced controllability and observability Gramians which depend on a finite time t. Important concepts are infinite Gramians [MS05, Section 3.2.4].

Definition 2.33:
For an asymptotically stable system $(E; A, B, C, D)$ the *(infinite) controllability Gramian* is

$$P := \int_0^\infty e^{E^{-1}At} E^{-1} B B^{\mathrm{T}} E^{-\mathrm{T}} e^{A^{\mathrm{T}} E^{-\mathrm{T}} t} \, \mathrm{d}t$$

and the *(infinite) observability Gramian* is

$$Q := \int_0^\infty E^{-\mathrm{T}} e^{A^{\mathrm{T}} E^{-\mathrm{T}} t} C^{\mathrm{T}} C e^{E^{-1}At} E^{-1} \, \mathrm{d}t. \qquad \Diamond$$

They can also be represented in the frequency domain.

Proposition 2.34 ([Ant05, Section 4.3], [MS05, Section 3.2.4]):
Under the assumptions of Definition 2.33, the Gramians satisfy

$$\begin{aligned} P &= \frac{1}{2\pi} \int_{-\infty}^{\infty} (\boldsymbol{\imath}\omega E - A)^{-1} B B^{\mathrm{T}} (-\boldsymbol{\imath}\omega E - A)^{-\mathrm{T}} \, \mathrm{d}\omega, \\ Q &= \frac{1}{2\pi} \int_{-\infty}^{\infty} (-\boldsymbol{\imath}\omega E - A)^{-\mathrm{T}} C^{\mathrm{T}} C (\boldsymbol{\imath}\omega E - A)^{-1} \, \mathrm{d}\omega. \end{aligned} \qquad \Diamond$$

Solving Lyapunov equations can be used to find the Gramians.

Proposition 2.35 ([Ant05, Proposition 4.27]):
Under the assumptions of Definition 2.33, the Gramians are the unique solutions to the following Lyapunov equations:

$$\begin{aligned} APE^{\mathrm{T}} + EPA^{\mathrm{T}} + BB^{\mathrm{T}} &= 0, \\ A^{\mathrm{T}}QE + E^{\mathrm{T}}QA + C^{\mathrm{T}}C &= 0. \end{aligned}$$ ◊

Compared to Definition 2.33 or Proposition 2.34, Proposition 2.35 provides an approach to efficient computation of Gramians (see surveys [BS13, Sim16]).

2.3.3 Hardy spaces and system norms

The transfer function $H(s) = C(sE - A)^{-1}B + D$ is a rational matrix function. Under some assumptions, H is an element of Hardy spaces $\mathcal{H}_2^{p\times m}$ and/or $\mathcal{H}_\infty^{p\times m}$.

Definition 2.36 ([Ant05, Section 5.1.3]):
Hardy space $\mathcal{H}_2^{p\times m}$:

$$\mathcal{H}_2^{p\times m} := \left\{ H : \mathbb{C}_+ \to \mathbb{C}^{p\times m} \,\middle|\, \begin{array}{l} H \text{ is analytic and} \\ \sup_{\xi>0} \int_{-\infty}^{\infty} \|H(\xi + \imath\omega)\|_{\mathrm{F}}^2 \,\mathrm{d}\omega < \infty \end{array} \right\},$$

with norm

$$\|H\|_{\mathcal{H}_2} := \sqrt{\sup_{\xi>0} \int_{-\infty}^{\infty} \|H(\xi + \imath\omega)\|_{\mathrm{F}}^2 \,\mathrm{d}\omega}.$$

Hardy space $\mathcal{H}_\infty^{p\times m}$:

$$\mathcal{H}_\infty^{p\times m} := \left\{ H : \mathbb{C}_+ \to \mathbb{C}^{p\times m} \,\middle|\, \begin{array}{l} H \text{ is analytic and} \\ \sup_{s\in\mathbb{C}_+} \|H(s)\|_2 < \infty \end{array} \right\}$$

with norm

$$\|H\|_{\mathcal{H}_\infty} := \sup_{s\in\mathbb{C}_+} \|H(s)\|_2.$$ ◊

For simplicity, we will often write $\mathcal{H}_2$ and $\mathcal{H}_\infty$ instead of $\mathcal{H}_2^{p\times m}$ and $\mathcal{H}_\infty^{p\times m}$.

It can be seen that if the system is asymptotically stable, then H is an element of $\mathcal{H}_\infty$. If additionally $D = 0$, then H is also an element of $\mathcal{H}_2$.

Remark 2.37:
Functions from Hardy spaces can be extended to the imaginary axis. Furthermore, $\mathcal{H}_2$ can be shown to be a Hilbert space with inner product

$$\langle H, G \rangle_{\mathcal{H}_2} := \frac{1}{2\pi} \int_{-\infty}^{\infty} \operatorname{tr}(H(\imath\omega)^* G(\imath\omega)) \,\mathrm{d}\omega$$

and $\mathcal{H}_\infty$ a Banach space with norm

$$\|H\|_{\mathcal{H}_\infty} := \sup_{\omega\in\mathbb{R}} \|H(\imath\omega)\|_2.$$

For more details, see [Ant05] and [ZDG96]. ◊

The following results states that the Gramians can be used to compute the $\mathcal{H}_2$-norm.

Proposition 2.38 ([Ant05, Section 5.5.1]):
For an asymptotically stable system (2.2),

$$\|H\|_{\mathcal{H}_2}^2 = \operatorname{tr}\big(CPC^{\mathrm{T}}\big) = \operatorname{tr}\big(B^{\mathrm{T}}QB\big)\,. \qquad \lozenge$$

2.4 Model order reduction

2.4.1 Projection-based model reduction

We consider the system (2.2). Here, we assume homogeneous initial conditions, i.e., $x_0 = 0$. For some recent work on MOR approaches for inhomogeneous initial conditions, see [HRA11, BGM17].

We can write the system (2.2) in variational form (see [ABG10, Section 2.2]):

> Find $x(t)$ contained in $\mathbb{R}^n$ such that
>
> $$E\dot{x}(t) - Ax(t) - Bu(t) \perp \mathbb{R}^n.$$
>
> Then the associated output is $y(t) = Cx(t) + Du(t)$.

Petrov-Galerkin projection consists of choosing two r-dimensional subspaces $\mathcal{V}, \mathcal{W} \subset \mathbb{R}^n$, where $r < n$ is the reduced order, and defining the ROM by

> Find $v(t)$ contained in $\mathcal{V}$ such that
>
> $$E\dot{v}(t) - Av(t) - Bu(t) \perp \mathcal{W}.$$
>
> Then the associated output is $\widehat{y}(t) = Cv(t) + Du(t)$.

Choosing $V, W \in \mathbb{R}^{n\times r}$ such that $\operatorname{im}(V) = \mathcal{V}$ and $\operatorname{im}(W) = \mathcal{W}$, we have $v(t) = V\widehat{x}(t)$ for some $\widehat{x}(t) \in \mathbb{R}^r$ and $W^{\mathrm{T}}(EV\dot{\widehat{x}}(t) - AV\widehat{x}(t) - Bu(t)) = 0$. A state-space form of the ROM is

$$\begin{aligned} \widehat{E}\dot{\widehat{x}}(t) &= \widehat{A}\widehat{x}(t) + \widehat{B}u(t), \quad \widehat{x}(0) = 0,\\ \widehat{y}(t) &= \widehat{C}\widehat{x}(t) + \widehat{D}u(t), \end{aligned}$$

with

$$\widehat{E} = W^{\mathrm{T}}EV, \quad \widehat{A} = W^{\mathrm{T}}AV, \quad \widehat{B} = W^{\mathrm{T}}B, \quad \widehat{C} = CV, \quad \widehat{D} = D.$$

In the following sections, we give an overview of some particular projection-based methods.

2.4.2 Balanced truncation

This method is based on the controllability and observability energy functionals

$$E_c(x_0) \coloneqq \inf \left\{ \|u\|^2_{\mathcal{L}_2(-\infty,0)} \,\middle|\, E\dot{x}(t) = Ax(t) + Bu(t),\ x(-\infty) = 0,\ x(0) = x_0 \right\},$$
$$E_o(x_0) \coloneqq \left\| Ce^{tE^{-1}A}x_0 \right\|^2_{\mathcal{L}_2(0,\infty)}.$$

We can interpret $E_c(x_0)$ as the energy necessary to reach state x_0 and $E_o(x_0)$ as the observed energy from the system when it starts from state x_0. The following result relates energy functionals to Gramians.

Proposition 2.39:
For a controllable system $(E; A, B, C, D)$, we have

$$E_c(x_0) = x_0^{\mathrm{T}} P^{-1} x_0 \quad \text{and} \quad E_o(x_0) = x_0^{\mathrm{T}} E^{\mathrm{T}} Q E x_0. \qquad \lozenge$$

For a controllable and observable system, it is possible to transform the system such that $\widetilde{P}$ and $\widetilde{E}^{\mathrm{T}}\widetilde{Q}\widetilde{E}$ become equal to a matrix $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$ where $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n > 0$ (see [Ant05, Lemma 7.3]). If there exists an r such that $\sigma_r \gg \sigma_{r+1}$, then we say that, in the new coordinate system, the states $e_1, \ldots, e_r$ are easy to reach and easy to observe, while the states $e_{r+1}, \ldots, e_n$ are difficult to reach and difficult to observe. The idea is then to truncate the latter states. Balanced truncation (BT), using the balancing-free square root method, is described in Algorithm 2.1. The following theorem gives the a priori error bound which allows automatic choice of the reduced order.

Theorem 2.40 ([Ant05, Theorem 7.9]):
Let $(E; A, B, C, D)$ be an asymptotically stable, controllable, and observable system of order n and $(\widehat{E}; \widehat{A}, \widehat{B}, \widehat{C}, \widehat{D})$ a ROM obtained by BT of order $r < n$. If $\sigma_r \neq \sigma_{r+1}$, then the ROM is asymptotically stable, controllable, observable, and

$$\|H - \widehat{H}\|_{\mathcal{H}_\infty} \leqslant 2 \sum_{i=r+1}^{n} \sigma_i,$$

where H and $\widehat{H}$ are the transfer functions of the two systems. $\lozenge$

2.4.3 Rational interpolation

MOR using rational interpolation, also called moment matching or Krylov methods, is described here. The main idea is formulated in the following theorem.

Theorem 2.41 ([ABG10, Theorem 2]):
Let $H(s) = C(sE - A)^{-1}B + D$ and $\widehat{H}(s) = \widehat{C}(s\widehat{E} - \widehat{A})^{-1}\widehat{B} + \widehat{D}$ be two transfer functions with $\widehat{E} = W^{\mathrm{T}}EV$, $\widehat{A} = W^{\mathrm{T}}AV$, $\widehat{B} = W^{\mathrm{T}}B$, $\widehat{C} = CV$, and $\widehat{D} = D$. Furthermore, let $\sigma, \mu \in \mathbb{C}$ be such that $sE - A$ and $s\widehat{E} - \widehat{A}$ are invertible for $s = \sigma, \mu$. If $b \in \mathbb{C}^m$ and $c \in \mathbb{C}^p$ are fixed nontrivial vectors, then

Algorithm 2.1: Balancing-free square root balanced truncation method

Input: Asymptotically stable system $(E; A, B, C, D)$ of order n, reduced order r.

Output: Reduced-order model $(\widehat{E}; \widehat{A}, \widehat{B}, \widehat{C}, \widehat{D})$.

1 For Gramians P and Q solving Lyapunov equations

$$APE^{\mathrm{T}} + EPA^{\mathrm{T}} + BB^{\mathrm{T}} = 0,$$
$$A^{\mathrm{T}}QE + E^{\mathrm{T}}QA + C^{\mathrm{T}}C = 0,$$

find Cholesky decompositions $P = Z_P Z_P^{\mathrm{T}}$ and $Q = Z_Q Z_Q^{\mathrm{T}}$, with $Z_P, Z_Q \in \mathbb{R}^{n\times n}$, or low-rank approximations $Z_P \in \mathbb{R}^{n\times r_P}$, $Z_Q \in \mathbb{R}^{n\times r_Q}$, with $r \leqslant \min(r_P, r_Q)$.

2 Compute a singular value decomposition

$$Z_Q^{\mathrm{T}} E Z_P = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^{\mathrm{T}} \\ V_2^{\mathrm{T}} \end{bmatrix}$$

where $\Sigma_1 \in \mathbb{R}^{r\times r}$.

3 Use thin QR decompositions to find $V, W \in \mathbb{R}^{n\times r}$ with orthonormal columns such that $\operatorname{im}(V) = \operatorname{im}(Z_P V_1)$ and $\operatorname{im}(W) = \operatorname{im}(Z_Q U_1)$.

4 Project to get reduced matrices

$$\widehat{E} = W^{\mathrm{T}} E V, \quad \widehat{A} = W^{\mathrm{T}} A V, \quad \widehat{B} = W^{\mathrm{T}} B, \quad \widehat{C} = CV, \quad \widehat{D} = D.$$

1. if

$$\left((\sigma E - A)^{-1} E\right)^i (\sigma E - A)^{-1} B b \in \operatorname{im}(V) \text{ for } i = 0, 1, \ldots, N-1, \tag{2.3}$$

then

$$H^{(k)}(\sigma) b = \widehat{H}^{(k)}(\sigma) b \text{ for } k = 0, 1, \ldots, N-1,$$

2. if

$$\left((\mu E - A)^{-\mathrm{T}} E^{\mathrm{T}}\right)^j (\mu E - A)^{-\mathrm{T}} C^{\mathrm{T}} c \in \operatorname{im}(W) \text{ for } j = 0, 1, \ldots, M-1, \tag{2.4}$$

then

$$c^{\mathrm{T}} H^{(k)}(\mu) = c^{\mathrm{T}} \widehat{H}^{(k)}(\mu) \text{ for } k = 0, 1, \ldots, M-1,$$

3. if both (2.3) and (2.4) hold and $\sigma = \mu$, then

$$c^{\mathrm{T}} H^{(k)}(\sigma) b = c^{\mathrm{T}} \widehat{H}^{(k)}(\sigma) b \text{ for } k = 1, 2, \ldots, M+N-1. \qquad \lozenge$$

In particular, if

$$\begin{aligned}\operatorname{im}(V) &= \operatorname{im}\left(\begin{bmatrix}(\sigma_1 E - A)^{-1}Bb_1 & (\sigma_2 E - A)^{-1}Bb_2 & \cdots & (\sigma_r E - A)^{-1}Bb_r\end{bmatrix}\right),\\ \operatorname{im}(W) &= \operatorname{im}\left(\begin{bmatrix}(\mu_1 E - A)^{-\mathrm{T}}C^{\mathrm{T}}c_1 & (\mu_2 E - A)^{-\mathrm{T}}C^{\mathrm{T}}c_2 & \cdots & (\mu_r E - A)^{-\mathrm{T}}C^{\mathrm{T}}c_r\end{bmatrix}\right),\end{aligned}$$

then we have

$$H(\sigma_i)b_i = \widehat{H}(\sigma_i)b_i \text{ and } c_i^{\mathrm{T}}H(\mu_i) = c_i^{\mathrm{T}}\widehat{H}(\mu_i), \text{ for } i = 1, 2, \ldots, r.$$

If additionally $\sigma_i = \mu_i$, then also

$$c_i^{\mathrm{T}}H'(\sigma_i)b_i = c_i^{\mathrm{T}}\widehat{H}'(\sigma_i)b_i, \text{ for } i = 1, 2, \ldots, r.$$

This case is particularly relevant for $\mathcal{H}_2$-optimal MOR, where the ROM necessarily satisfies such interpolatory conditions. More details follow in the next section.

2.4.4 $\mathcal{H}_2$-optimal model order reduction

There are a few approaches for $\mathcal{H}_2$-optimal MOR of first-order systems. We will in particular focus on two: interpolation-based and Gramian-based.

Meier and Luenberger [ML67] derived interpolatory necessary optimality conditions for SISO systems. Gugercin, Beattie, and Antoulas [GAB08, ABG10] generalized this to MIMO systems and proposed iterative rational Krylov algorithm (IRKA).

Wilson [Wil70] developed Gramian-based optimality conditions in form of coupled matrix equation. Xu and Zeng [XZ11] used this to propose two-sided iteration algorithm (TSIA). The same algorithm was proposed in parallel by Van Dooren, Gallivan, and Absil [VDGA08, VDGA10]. Benner, Köhler, and Saak [BKS11] investigate some implementation issues in TSIA related to Sylvester equations and orthonormalization.

Let us take the first-order system (2.2) as the full-order model, with E invertible and $\lambda E - A$ an asymptotically stable matrix pencil. The $\mathcal{H}_2$-optimal MOR problem is finding a ROM $(\widehat{E}; \widehat{A}, \widehat{B}, \widehat{C}, \widehat{D})$ with transfer function $\widehat{H}(s) = \widehat{C}(s\widehat{E} - \widehat{A})^{-1}\widehat{B} + \widehat{D}$, such that the $\mathcal{H}_2$-error $\|H - \widehat{H}\|_{\mathcal{H}_2}$ is locally minimized. Here, $\widehat{E}, \widehat{A} \in \mathbb{R}^{r\times r}$, $\widehat{B} \in \mathbb{R}^{r\times m}$, $\widehat{C} \in \mathbb{R}^{p\times r}$, and $\widehat{D} \in \mathbb{R}^{p\times m}$, with $r < n$. For the $\mathcal{H}_2$-error to be defined, it is necessary that $\widehat{D} = D$. In particular, we can assume $D = \widehat{D} = 0$.

2.4.4.1 Interpolation-based approach

The motivation for the interpolation-based approach are the interpolatory necessary optimality conditions, given in the following theorem. A transfer function H is called real if $\overline{(H(s))} = H(\overline{s})$ for all s in the domain of H.

Algorithm 2.2: Iterative rational Krylov algorithm (IRKA)

Input: System $(E; A, B, C)$, initial shifts σ_i and tangential directions b_i, c_i, $i = 1, 2, \ldots, r$.

Output: Reduced-order model $(\widehat{E}; \widehat{A}, \widehat{B}, \widehat{C})$.

1 **while** *not converged* **do**

2 Find $V, W \in \mathbb{R}^{n\times r}$ with orthonormal columns such that

$$\operatorname{im}(V) = \operatorname{im}\left(\begin{bmatrix}(\sigma_1 E - A)^{-1}Bb_1 & \cdots & (\sigma_r E - A)^{-1}Bb_r\end{bmatrix}\right),$$
$$\operatorname{im}(W) = \operatorname{im}\left(\begin{bmatrix}(\sigma_1 E - A)^{-\mathrm{T}}C^{\mathrm{T}}c_1 & \cdots & (\sigma_r E - A)^{-\mathrm{T}}C^{\mathrm{T}}c_r\end{bmatrix}\right).$$

3 Project $\widehat{E} = W^{\mathrm{T}}EV$, $\widehat{A} = W^{\mathrm{T}}AV$, $\widehat{B} = W^{\mathrm{T}}B$, $\widehat{C} = CV$.

4 Compute $\operatorname{diag}(\lambda_i), X, Y \in \mathbb{C}^{r\times r}$ such that $Y^{\mathrm{T}}\widehat{A}X = \operatorname{diag}(\lambda_i)$ and $Y^{\mathrm{T}}\widehat{E}X = I$.

5 Update interpolation points and tangential directions:

$$\sigma_i = -\lambda_i,\ b_i = \widehat{B}^{\mathrm{T}}Ye_i,\ c_i = \widehat{C}Xe_i,\ \text{for } i = 1, 2, \ldots, r.$$

Theorem 2.42 ([ABG10, Theorem 5]):
Suppose $H \in \mathcal{H}_2$ and $\widehat{H}(s) = \sum_{i=1}^{r} \frac{c_i b_i^{\mathrm{T}}}{s-\lambda_i}$, with λ_i pairwise distinct, are real transfer functions. Let $\widehat{H}$ be a locally $\mathcal{H}_2$-optimal ROM of order r. Then

$$H(-\lambda_i)b_i = \widehat{H}(-\lambda_i)b_i, \tag{2.5a}$$
$$c_i^{\mathrm{T}}H(-\lambda_i) = c_i^{\mathrm{T}}\widehat{H}(-\lambda_i), \tag{2.5b}$$
$$c_i^{\mathrm{T}}H'(-\lambda_i)b_i = c_i^{\mathrm{T}}\widehat{H}'(-\lambda_i)b_i, \tag{2.5c}$$

for $i = 1, 2, \ldots, r$. ◊

This results states that the transfer function $\widehat{H}$ of a locally $\mathcal{H}_2$-optimal ROM is a bitangential Hermite interpolant, at the reflected poles of $\widehat{H}$, of the full-order model's transfer function H.

For given λ_i, b_i, c_i, the interpolation conditions in (2.5) can be achieved by a Petrov-Galerkin projection (see Theorem 2.41) where V and W span tangential rational Krylov subspaces

$$\operatorname{im}(V) = \operatorname{im}\left(\begin{bmatrix}(-\lambda_1 E - A)^{-1}Bb_1 & \cdots & (-\lambda_r E - A)^{-1}Bb_r\end{bmatrix}\right),$$
$$\operatorname{im}(W) = \operatorname{im}\left(\begin{bmatrix}(-\lambda_1 E - A)^{-\mathrm{T}}C^{\mathrm{T}}c_1 & \cdots & (-\lambda_r E - A)^{-\mathrm{T}}C^{\mathrm{T}}c_r\end{bmatrix}\right).$$

The difficulty is that λ_i, b_i, c_i are not given in advance. Gugercin et al. [GAB08, ABG10] proposed IRKA (see Algorithm 1 in [ABG10]). The pseudocode is presented in Algorithm 2.2.

2.4.4.2 Gramian-based approach

We use $\widehat{P}$ and $\widehat{Q}$ to denote the controllability and observability Gramians of $(\widehat{E}; \widehat{A}, \widehat{B}, \widehat{C})$, which solve reduced Lyapunov equations

$$\widehat{A}\widehat{P}\widehat{E}^{\mathrm{T}} + \widehat{E}\widehat{P}\widehat{A}^{\mathrm{T}} + \widehat{B}\widehat{B}^{\mathrm{T}} = 0, \tag{2.6a}$$

$$\widehat{A}^{\mathrm{T}}\widehat{Q}\widehat{E} + \widehat{E}^{\mathrm{T}}\widehat{Q}\widehat{A} + \widehat{C}^{\mathrm{T}}\widehat{C} = 0. \tag{2.6b}$$

Furthermore, we have that the controllability and observability Gramians of the error system

$$\underbrace{\begin{bmatrix} E & 0 \\ 0 & \widehat{E} \end{bmatrix}}_{E_{\mathrm{err}}} \begin{bmatrix} \dot{x}(t) \\ \dot{\widehat{x}}(t) \end{bmatrix} = \underbrace{\begin{bmatrix} A & 0 \\ 0 & \widehat{A} \end{bmatrix}}_{A_{\mathrm{err}}} \begin{bmatrix} x(t) \\ \widehat{x}(t) \end{bmatrix} + \underbrace{\begin{bmatrix} B \\ \widehat{B} \end{bmatrix}}_{B_{\mathrm{err}}} u(t),$$

$$y(t) - \widehat{y}(t) = \underbrace{\begin{bmatrix} C & -\widehat{C} \end{bmatrix}}_{C_{\mathrm{err}}} \begin{bmatrix} x(t) \\ \widehat{x}(t) \end{bmatrix},$$

are

$$P_{\mathrm{err}} = \begin{bmatrix} P & \widetilde{P} \\ \widetilde{P}^{\mathrm{T}} & \widehat{P} \end{bmatrix} \text{ and } Q_{\mathrm{err}} = \begin{bmatrix} Q & \widetilde{Q} \\ \widetilde{Q}^{\mathrm{T}} & \widehat{Q} \end{bmatrix},$$

where $\widetilde{P}, \widetilde{Q} \in \mathbb{R}^{n \times r}$ solve Sylvester equations

$$A\widetilde{P}\widehat{E}^{\mathrm{T}} + E\widetilde{P}\widehat{A}^{\mathrm{T}} + B\widehat{B}^{\mathrm{T}} = 0, \tag{2.7a}$$

$$A^{\mathrm{T}}\widetilde{Q}\widehat{E} + E^{\mathrm{T}}\widetilde{Q}\widehat{A} - C^{\mathrm{T}}\widehat{C} = 0. \tag{2.7b}$$

Gramian-based approach uses Wilson conditions [Wil70], given in the following theorem. We include a proof because the result is generalized to include invertible matrices E and $\widehat{E}$. Additionally, the proof uses Lagrange multiplier method, instead of computing the derivatives directly as in [Wil70] or [VDGA08]. We will use the following lemma in the proof.

Lemma 2.43:
Let $A \in \mathbb{R}^{n \times m}$, $B \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{m \times m}$, and $f, g \colon \mathbb{R}^{n \times m} \to \mathbb{R}$ such that $f(X) = \operatorname{tr}\left(AX^{\mathrm{T}}\right)$ and $g(X) = \operatorname{tr}\left(BXCX^{\mathrm{T}}\right)$. Then $\nabla f(X_0) = A$ and $\nabla g(X_0) = BX_0C + B^{\mathrm{T}}X_0C^{\mathrm{T}}$ for arbitrary $X_0 \in \mathbb{R}^{n \times m}$. ◊

Proof. We have

$$f(X_0 + H) = \operatorname{tr}\left(A(X_0 + H)^{\mathrm{T}}\right) = \operatorname{tr}\left(AX_0^{\mathrm{T}}\right) + \operatorname{tr}\left(AH^{\mathrm{T}}\right) = f(X_0) + \langle A, H \rangle,$$

from which it directly follows that f is Fréchet differentiable at X_0 and $\nabla f(X_0) = A$.

For g we have

$$\begin{aligned} g(X_0 + H) &= \operatorname{tr}\Big(B(X_0 + H)C(X_0 + H)^{\mathrm{T}}\Big) \\ &= \operatorname{tr}\big(BX_0CX_0^{\mathrm{T}}\big) + \operatorname{tr}\big(BX_0CH^{\mathrm{T}}\big) + \operatorname{tr}\big(BHCX_0^{\mathrm{T}}\big) + \operatorname{tr}\big(BHCH^{\mathrm{T}}\big) \\ &= g(X_0) + \big\langle BX_0C + B^{\mathrm{T}}X_0C^{\mathrm{T}}, H\big\rangle + o(\|H\|), \end{aligned}$$

so g is also Fréchet differentiable at X_0 and $\nabla g(X_0) = BX_0C + B^{\mathrm{T}}X_0C^{\mathrm{T}}$. □

Theorem 2.44:
Let $(\widehat{E}; \widehat{A}, \widehat{B}, \widehat{C})$ be a locally $\mathcal{H}_2$-optimal ROM for $(E; A, B, C)$. Then

$$\begin{aligned} \widetilde{Q}^{\mathrm{T}}E\widetilde{P} + \widehat{Q}\widehat{E}\widehat{P} &= 0, \\ \widetilde{Q}^{\mathrm{T}}A\widetilde{P} + \widehat{Q}\widehat{A}\widehat{P} &= 0, \\ \widetilde{Q}^{\mathrm{T}}B + \widehat{Q}\widehat{B} &= 0, \\ C\widetilde{P} - \widehat{C}\widehat{P} &= 0. \end{aligned} \tag{2.8}$$

♢

Proof. The proof consists of applying the Lagrange multiplier method to the optimization problem

$$\begin{aligned} \text{minimize} \quad & \operatorname{tr}\big(C_{\mathrm{err}}P_{\mathrm{err}}C_{\mathrm{err}}^{\mathrm{T}}\big), \\ \text{subject to} \quad & \text{(2.7a)}, \ \text{(2.6a)}. \end{aligned}$$

The Lagrange function is

$$\begin{aligned} \mathcal{L}(\widehat{E}, \widehat{A}, \widehat{B}, \widehat{C}, \widetilde{P}, \widehat{P}, \widetilde{\Lambda}, \widehat{\Lambda}) = \operatorname{tr}&\Big(CPC^{\mathrm{T}} - 2C\widetilde{P}\widehat{C}^{\mathrm{T}} + \widehat{C}\widehat{P}\widehat{C}^{\mathrm{T}}\Big) \\ &+ \operatorname{tr}\Big(\widetilde{\Lambda}^{\mathrm{T}}\Big(A\widetilde{P}\widehat{E}^{\mathrm{T}} + E\widetilde{P}\widehat{A}^{\mathrm{T}} + B\widehat{B}^{\mathrm{T}}\Big)\Big) \\ &+ \operatorname{tr}\Big(\widehat{\Lambda}^{\mathrm{T}}\Big(\widehat{A}\widehat{P}\widehat{E}^{\mathrm{T}} + \widehat{E}\widehat{P}\widehat{A}^{\mathrm{T}} + \widehat{B}\widehat{B}^{\mathrm{T}}\Big)\Big), \end{aligned}$$

where $\widetilde{\Lambda} \in \mathbb{R}^{n\times r}$ and $\widehat{\Lambda} \in \mathbb{R}^{r\times r}$ are the Lagrange multipliers. The gradients of $\mathcal{L}$ with respect to the Gramians are

$$\begin{aligned} \nabla_{\widetilde{P}}\mathcal{L} &= -2C^{\mathrm{T}}\widehat{C} + A^{\mathrm{T}}\widetilde{\Lambda}\widehat{E} + E^{\mathrm{T}}\widetilde{\Lambda}\widehat{A}, \\ \nabla_{\widehat{P}}\mathcal{L} &= \widehat{C}^{\mathrm{T}}\widehat{C} + \widehat{A}^{\mathrm{T}}\widehat{\Lambda}\widehat{E} + \widehat{E}^{\mathrm{T}}\widehat{\Lambda}\widehat{A}. \end{aligned}$$

From $\nabla_{\widetilde{P}}\mathcal{L} = 0$ and $\nabla_{\widehat{P}}\mathcal{L} = 0$ it follows that $\widetilde{\Lambda} = 2\widetilde{Q}$ and $\widehat{\Lambda} = \widehat{Q}$. The Lagrange function now simplifies to

$$\begin{aligned} \mathcal{L} = \operatorname{tr}&\Big(CPC^{\mathrm{T}} - 2C\widetilde{P}\widehat{C}^{\mathrm{T}} + \widehat{C}\widehat{P}\widehat{C}^{\mathrm{T}}\Big) \\ &+ \operatorname{tr}\Big(2\widetilde{Q}^{\mathrm{T}}A\widetilde{P}\widehat{E}^{\mathrm{T}} + 2\widetilde{Q}^{\mathrm{T}}E\widetilde{P}\widehat{A}^{\mathrm{T}} + 2\widetilde{Q}^{\mathrm{T}}B\widehat{B}^{\mathrm{T}}\Big) \\ &+ \operatorname{tr}\Big(2\widehat{Q}\widehat{A}\widehat{P}\widehat{E}^{\mathrm{T}} + \widehat{Q}\widehat{B}\widehat{B}^{\mathrm{T}}\Big). \end{aligned}$$

Algorithm 2.3: Two-sided iteration algorithm (TSIA)

Input: System (E, A, B, C) and initial reduced-order model $(\widehat{E}, \widehat{A}, \widehat{B}, \widehat{C})$.
Output: Reduced-order model $(\widehat{E}, \widehat{A}, \widehat{B}, \widehat{C})$ approximately satisfying (2.8).

1 **while** *not converged* **do**
2 Solve $A\widetilde{P}\widehat{E}^{\mathrm{T}} + E\widetilde{P}\widehat{A}^{\mathrm{T}} + B\widehat{B}^{\mathrm{T}} = 0$ and $A^{\mathrm{T}}\widetilde{Q}\widehat{E} + E^{\mathrm{T}}\widetilde{Q}\widehat{A} - C^{\mathrm{T}}\widehat{C} = 0$.
3 Find $V, W \in \mathbb{R}^{n\times r}$ with orthonormal columns such that $\operatorname{im}(V) = \operatorname{im}(\widetilde{P})$ and $\operatorname{im}(W) = \operatorname{im}(\widetilde{Q})$.
4 Project $\widehat{E} = W^{\mathrm{T}}EV$, $\widehat{A} = W^{\mathrm{T}}AV$, $\widehat{B} = W^{\mathrm{T}}B$, $\widehat{C} = CV$.

Finally, the gradients of the Lagrange function with respect to the reduced matrices are

$$\begin{aligned}
\nabla_{\widehat{E}}\mathcal{L} &= 2\widetilde{Q}^{\mathrm{T}}A\widetilde{P} + 2\widehat{Q}\widehat{A}\widehat{P},\\
\nabla_{\widehat{A}}\mathcal{L} &= 2\widetilde{Q}^{\mathrm{T}}E\widetilde{P} + 2\widehat{Q}\widehat{E}\widehat{P},\\
\nabla_{\widehat{B}}\mathcal{L} &= 2\widetilde{Q}^{\mathrm{T}}B + 2\widehat{Q}\widehat{B},\\
\nabla_{\widehat{C}}\mathcal{L} &= -2C\widetilde{P} + 2\widehat{C}\widehat{P},
\end{aligned}$$

which completes the proof. □

Based on these optimality conditions, Xu and Zeng [XZ11] proposed TSIA, given in Algorithm 2.3.

Let $(\widehat{A}, \widehat{E})$ be diagonalizable and $\Lambda, X, Y \in \mathbb{C}^{r\times r}$ such that $Y^{\mathrm{T}}\widehat{E}X = I_r$ and $Y^{\mathrm{T}}\widehat{A}X = \Lambda$. Then also $\widehat{A}x_i = \lambda_i\widehat{E}x_i$ and $\widehat{A}^{\mathrm{T}}y_i = \lambda_i\widehat{E}^{\mathrm{T}}y_i$, where $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$, $X = \begin{bmatrix} x_1 & x_2 & \cdots & x_r \end{bmatrix}$, and $Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_r \end{bmatrix}$. Then we have

$$\begin{aligned}
A\widetilde{P}\widehat{E}^{\mathrm{T}}y_i + E\widetilde{P}\widehat{A}^{\mathrm{T}}y_i + B\widehat{B}^{\mathrm{T}}y_i &= 0,\\
A^{\mathrm{T}}\widetilde{Q}\widehat{E}x_i + E^{\mathrm{T}}\widetilde{Q}\widehat{A}x_i - C^{\mathrm{T}}\widehat{C}x_i &= 0,
\end{aligned}$$

which implies

$$\begin{aligned}
(A + \lambda_i E)\widetilde{P}\widehat{E}^{\mathrm{T}}y_i + B\widehat{B}^{\mathrm{T}}y_i &= 0,\\
(A^{\mathrm{T}} + \lambda_i E^{\mathrm{T}})\widetilde{Q}\widehat{E}x_i - C^{\mathrm{T}}\widehat{C}x_i &= 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\widetilde{P}\widehat{E}^{\mathrm{T}}y_i &= (-\lambda_i E - A)^{-1}B\widehat{B}^{\mathrm{T}}y_i,\\
\widetilde{Q}\widehat{E}x_i &= -(-\lambda_i E - A)^{-\mathrm{T}}C^{\mathrm{T}}\widehat{C}x_i,
\end{aligned}$$

which shows the connection to IRKA. This can also be used to prove interpolatory conditions in Theorem 2.42 using the Wilson conditions from Theorem 2.44, as given in the following theorem (see [VDGA08, Theorem 4.1]).

Theorem 2.45:
Let $\widehat{H}(s) = \widehat{C}(s\widehat{E} - \widehat{A})^{-1}\widehat{B} = \sum_{i=1}^{r} \frac{c_i b_i^{\mathrm{T}}}{s-\lambda_i}$, with λ_i pairwise distinct, $S^{\mathrm{T}}\widehat{E}T = I_r$, $S^{\mathrm{T}}\widehat{A}T = \Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$, $b_i^{\mathrm{T}} = s_i^{\mathrm{T}}\widehat{B}$, $s_i = Se_i$, $c_i = Ct_i$, and $t_i = Te_i$. Then

$$
\begin{aligned}
\left(C\widetilde{P} - \widehat{C}\widehat{P}\right)\widehat{E}^{\mathrm{T}}s_i &= \left(H(-\lambda_i) - \widehat{H}(-\lambda_i)\right)b_i,\\
-t_i^{\mathrm{T}}\widehat{E}^{\mathrm{T}}\left(\widetilde{Q}^{\mathrm{T}}B + \widehat{Q}\widehat{B}\right) &= c_i^{\mathrm{T}}\left(H(-\lambda_i) - \widehat{H}(-\lambda_i)\right),\\
t_i^{\mathrm{T}}\widehat{E}^{\mathrm{T}}\left(\widetilde{Q}^{\mathrm{T}}E\widetilde{P} + \widehat{Q}\widehat{E}\widehat{P}\right)\widehat{E}^{\mathrm{T}}s_i &= c_i^{\mathrm{T}}\left(H'(-\lambda_i) - \widehat{H}'(-\lambda_i)\right)b_i,\\
t_i^{\mathrm{T}}\widehat{E}^{\mathrm{T}}\left(\widetilde{Q}^{\mathrm{T}}E\widetilde{P} + \widehat{Q}\widehat{E}\widehat{P}\right)\widehat{E}^{\mathrm{T}}s_j &= c_i^{\mathrm{T}}\left(\frac{H(-\lambda_i) - H(-\lambda_j)}{(-\lambda_i) - (-\lambda_j)} - \frac{\widehat{H}(-\lambda_i) - \widehat{H}(-\lambda_j)}{(-\lambda_i) - (-\lambda_j)}\right)b_j,\\
t_i^{\mathrm{T}}\widehat{E}^{\mathrm{T}}\left(\widetilde{Q}^{\mathrm{T}}A\widetilde{P} + \widehat{Q}\widehat{A}\widehat{P}\right)\widehat{E}^{\mathrm{T}}s_i &= c_i^{\mathrm{T}}\left([sH(s)]'\big|_{s=-\lambda_i} - [s\widehat{H}(s)]'\big|_{s=-\lambda_i}\right)b_i,\\
t_i^{\mathrm{T}}\widehat{E}^{\mathrm{T}}\left(\widetilde{Q}^{\mathrm{T}}A\widetilde{P} + \widehat{Q}\widehat{A}\widehat{P}\right)\widehat{E}^{\mathrm{T}}s_j &= c_i^{\mathrm{T}}\Bigg(\frac{(-\lambda_i)H(-\lambda_i) - (-\lambda_j)H(-\lambda_j)}{(-\lambda_i) - (-\lambda_j)}\\
&\qquad - \frac{(-\lambda_i)\widehat{H}(-\lambda_i) - (-\lambda_j)\widehat{H}(-\lambda_j)}{(-\lambda_i) - (-\lambda_j)}\Bigg)b_j,
\end{aligned}
$$

for $i, j = 1, 2, \ldots, r$, $i \neq j$. ◊

Proof. From $S^{\mathrm{T}}\widehat{E}T = I_r$ and $S^{\mathrm{T}}\widehat{A}T = \Lambda$, we get $\widehat{A}t_i = \lambda_i\widehat{E}t_i$ and $s_i^{\mathrm{T}}\widehat{A} = \lambda_i s_i^{\mathrm{T}}\widehat{E}$. Next, from

$$
\begin{aligned}
A\widetilde{P}\widehat{E}^{\mathrm{T}}s_i + E\widetilde{P}\widehat{A}^{\mathrm{T}}s_i + B\widehat{B}^{\mathrm{T}}s_i &= 0,\\
\widehat{A}\widehat{P}\widehat{E}^{\mathrm{T}}s_i + \widehat{E}\widehat{P}\widehat{A}^{\mathrm{T}}s_i + \widehat{B}\widehat{B}^{\mathrm{T}}s_i &= 0,\\
A^{\mathrm{T}}\widetilde{Q}\widehat{E}t_i + E^{\mathrm{T}}\widetilde{Q}\widehat{A}t_i - C^{\mathrm{T}}\widehat{C}t_i &= 0,\\
\widehat{A}^{\mathrm{T}}\widehat{Q}\widehat{E}t_i + \widehat{E}^{\mathrm{T}}\widehat{Q}\widehat{A}t_i + \widehat{C}^{\mathrm{T}}\widehat{C}t_i &= 0,
\end{aligned}
$$

we find

$$
\begin{aligned}
\widetilde{P}\widehat{E}^{\mathrm{T}}s_i &= (-\lambda_i E - A)^{-1}Bb_i,\\
\widehat{P}\widehat{E}^{\mathrm{T}}s_i &= \left(-\lambda_i\widehat{E} - \widehat{A}\right)^{-1}\widehat{B}b_i,\\
\widetilde{Q}\widehat{E}t_i &= -(-\lambda_i E - A)^{-\mathrm{T}}C^{\mathrm{T}}c_i,\\
\widehat{Q}\widehat{E}t_i &= \left(-\lambda_i\widehat{E} - \widehat{A}\right)^{-\mathrm{T}}\widehat{C}^{\mathrm{T}}c_i.
\end{aligned}
$$

Now it follows that

$$
\begin{aligned}
\left(C\widetilde{P} - \widehat{C}\widehat{P}\right)\widehat{E}^{\mathrm{T}}s_i &= C(-\lambda_i E - A)^{-1}Bb_i - \widehat{C}\left(-\lambda_i\widehat{E} - \widehat{A}\right)^{-1}\widehat{B}b_i\\
&= \left(H(-\lambda_i) - \widehat{H}(-\lambda_i)\right)b_i,
\end{aligned}
$$

and

$$\begin{aligned}-t_i^{\mathrm{T}}\widehat{E}^{\mathrm{T}}\left(\widetilde{Q}^{\mathrm{T}}B+\widehat{Q}\widehat{B}\right) &= c_i^{\mathrm{T}}C(-\lambda_i E-A)^{-1}B - c_i^{\mathrm{T}}\widehat{C}\left(-\lambda_i\widehat{E}-\widehat{A}\right)^{-1}\widehat{B}\\ &= c_i^{\mathrm{T}}\left(H(-\lambda_i)-\widehat{H}(-\lambda_i)\right).\end{aligned}$$

From

$$\begin{aligned}H'(s) &= -C(sE-A)^{-1}E(sE-A)^{-1}B,\\ H(s_1)-H(s_2) &= -(s_1-s_2)\,C(s_1E-A)^{-1}E(s_2E-A)^{-1}B,\\ [sH(s)]' &= -C(sE-A)^{-1}A(sE-A)^{-1}B,\\ s_1H(s_1)-s_2H(s_2) &= -(s_1-s_2)\,C(s_1E-A)^{-1}A(s_2E-A)^{-1}B,\end{aligned}$$

we find

$$\begin{aligned}&t_i^{\mathrm{T}}\widehat{E}^{\mathrm{T}}\left(\widetilde{Q}^{\mathrm{T}}E\widetilde{P}+\widehat{Q}\widehat{E}\widehat{P}\right)\widehat{E}^{\mathrm{T}}s_i\\ &= -c_i^{\mathrm{T}}C(-\lambda_iE-A)^{-1}E(-\lambda_iE-A)^{-1}Bb_i + c_i^{\mathrm{T}}\widehat{C}(-\lambda_i\widehat{E}-\widehat{A})^{-1}\widehat{E}(-\lambda_i\widehat{E}-\widehat{A})^{-1}\widehat{B}b_i\\ &= c_i^{\mathrm{T}}\left(H'(-\lambda_i)-\widehat{H}'(-\lambda_i)\right)b_i,\\ &t_i^{\mathrm{T}}\widehat{E}^{\mathrm{T}}\left(\widetilde{Q}^{\mathrm{T}}E\widetilde{P}+\widehat{Q}\widehat{E}\widehat{P}\right)\widehat{E}^{\mathrm{T}}s_j\\ &= -c_i^{\mathrm{T}}C(-\lambda_iE-A)^{-1}E(-\lambda_jE-A)^{-1}Bb_j + c_i^{\mathrm{T}}\widehat{C}(-\lambda_i\widehat{E}-\widehat{A})^{-1}\widehat{E}(-\lambda_j\widehat{E}-\widehat{A})^{-1}\widehat{B}b_j\\ &= c_i^{\mathrm{T}}\left(\frac{H(-\lambda_i)-H(-\lambda_j)}{(-\lambda_i)-(-\lambda_j)}-\frac{\widehat{H}(-\lambda_i)-\widehat{H}(-\lambda_j)}{(-\lambda_i)-(-\lambda_j)}\right)b_j,\\ &t_i^{\mathrm{T}}\widehat{E}^{\mathrm{T}}\left(\widetilde{Q}^{\mathrm{T}}A\widetilde{P}+\widehat{Q}\widehat{A}\widehat{P}\right)\widehat{E}^{\mathrm{T}}s_i\\ &= -c_i^{\mathrm{T}}C(-\lambda_iE-A)^{-1}A(-\lambda_iE-A)^{-1}Bb_i + c_i^{\mathrm{T}}\widehat{C}(-\lambda_i\widehat{E}-\widehat{A})^{-1}\widehat{A}(-\lambda_i\widehat{E}-\widehat{A})^{-1}\widehat{B}b_i\\ &= c_i^{\mathrm{T}}\left([sH(s)]'\big|_{s=-\lambda_i}-[s\widehat{H}(s)]'\big|_{s=-\lambda_i}\right)b_i,\\ &t_i^{\mathrm{T}}\widehat{E}^{\mathrm{T}}\left(\widetilde{Q}^{\mathrm{T}}A\widetilde{P}+\widehat{Q}\widehat{A}\widehat{P}\right)\widehat{E}^{\mathrm{T}}s_j\\ &= -c_i^{\mathrm{T}}C(-\lambda_iE-A)^{-1}A(-\lambda_jE-A)^{-1}Bb_j + c_i^{\mathrm{T}}\widehat{C}(-\lambda_i\widehat{E}-\widehat{A})^{-1}\widehat{A}(-\lambda_j\widehat{E}-\widehat{A})^{-1}\widehat{B}b_j\\ &= c_i^{\mathrm{T}}\left(\frac{(-\lambda_i)H(-\lambda_i)-(-\lambda_j)H(-\lambda_j)}{(-\lambda_i)-(-\lambda_j)}-\frac{(-\lambda_i)\widehat{H}(-\lambda_i)-(-\lambda_j)\widehat{H}(-\lambda_j)}{(-\lambda_i)-(-\lambda_j)}\right)b_j,\end{aligned}$$

which concludes the proof. □

2.4.4.3 Comparison of the two approaches

The interpolation-based and Gramian-based approaches differ in two important ways.

The first is the assumption of diagonalizability. Notice that Theorem 2.42 has the assumption that the ROM has pairwise distinct poles, while Theorem 2.44 does not. Diagonalizability of $\widehat{E}^{-1}\widehat{A}$ is a generic property, in the sense that the set of non-diagonalizable matrices forms a set of measure zero in the set of all matrices, which justifies this assumptions for unstructured first-order systems. But for structured systems, where the assumption of simultaneous diagonalizability of two or more matrices appears, will no longer be a generic property. Additionally, as illustrated by Van Dooren, Gallivan, and Absil [VDGA10], even in the case of first-order systems, if the optimal ROM has a higher-order pole, this can cause numerical issues in IRKA, while TSIA will still converge. The difference in the algorithms is in solving sparse-dense Sylvester equations, where the (generalized) Schur decomposition of $\widehat{A}$ and $\widehat{E}$ is sufficient (details in [BKS11]).

The second difference is that the Gramian-based approach gives necessary optimality conditions which are equations that the reduced matrices need to satisfy. In particular, for first-order systems, it is clear from Theorem 2.44 that the ROM is necessarily obtained by Petrov-Galerkin projection, assuming $\widehat{P}$ and $\widehat{Q}$ are invertible. This is not immediately clear from Theorem 2.42, where the interpolatory necessary optimality conditions do not explicitly give the equations for the reduced matrices.

One advantage of the interpolation-based approach is its direct application to MOR of infinite-dimensional systems where it is possible to evaluate the transfer function and its derivative. This was used to develop the *transfer function IRKA* in [BG12]. In the Gramian-based approach, access to matrices E, A, B, C is necessary. One possibility could be to extend these results to infinite-dimensional systems where E, A, B, C become operators. Then the corresponding Sylvester equations could be solved in a similar way as Lyapunov equations in [ORW13], but this is outside the scope of this thesis.

2.4.5 Model order reduction of unstable systems

So far, we assumed the system to be asymptotically stable. Since we will also need to consider systems which are not asymptotically stable, we need to see if and how the previously discussed MOR methods could be applied to such systems. Therefore, here we consider a system $(E; A, B, C, D)$ with an invertible E, but which is not asymptotically stable. We will not assume minimality (i.e., both controllability and observability), which means that its transfer function H may be asymptotically stable.

In the following chapters, we will be interested in finding ROMs for which the $\mathcal{H}_2$-error $\|H - \widehat{H}\|_{\mathcal{H}_2}$ or the $\mathcal{H}_\infty$-error $\|H - \widehat{H}\|_{\mathcal{H}_\infty}$ are small. Particularly, for these quantities to be defined, it is necessary that the transfer function $H - \widehat{H}$ is asymptotically stable. This implies that the ROM's transfer function $\widehat{H}$ needs to have the same unstable part as the original H. Therefore, we will consider additive decomposition into the asymptotically stable and unstable part as in [Enn85].

There are works ([YCDAGX93, Zil91, BNBG10]) about MOR for unstable systems

using shifting, but this approach does not guarantee preservation of unstable poles. The work in [ZSW99] gives an extension for BT, assuming there are no poles on the imaginary axis. There is also work extending $\mathcal{H}_2$-optimal MOR to unstable systems ([MBG10, BBG19]), which also assumes there are no poles on the imaginary axis or cannot preserve unstable poles.

Let $H = H_- + H_+$, where H_- has poles in $\mathbb{C}_-$ and H_+ in $\overline{\mathbb{C}_+}$. Furthermore, let S and T be such that

$$S^{\mathrm{T}}ET = \begin{bmatrix} E_- & 0 \\ 0 & E_+ \end{bmatrix}, \; S^{\mathrm{T}}AT = \begin{bmatrix} A_- & 0 \\ 0 & A_+ \end{bmatrix}, \; S^{\mathrm{T}}B = \begin{bmatrix} B_- \\ B_+ \end{bmatrix}, \text{ and } CT = \begin{bmatrix} C_- & C_+ \end{bmatrix},$$

where

$$T = \begin{bmatrix} T_- & T_+ \end{bmatrix} \in \mathbb{C}^{n\times n} \text{ and } S = \begin{bmatrix} S_- & S_+ \end{bmatrix} \in \mathbb{C}^{n\times n},$$

with $\sigma(A_-, E_-) \subset \mathbb{C}_-$ and $\sigma(A_+, E_+) \subset \overline{\mathbb{C}_+}$. Notice that $H_-(s) = C_-(sE_- - A_-)^{-1}B_-$ and $H_+(s) = C_+(sE_+ - A_+)^{-1}B_+$. To find a ROM $\widehat{H} = \widehat{H}_- + H_+$, we can apply Petrov-Galerkin projection to the asymptotically stable part $(E_-; A_-, B_-, C_-)$ to get $(\widehat{E}_-; \widehat{A}_-, \widehat{B}_-, \widehat{C}_-)$ with

$$\widehat{E}_- = W_-^{\mathrm{T}}E_-V_-, \quad \widehat{A}_- = W_-^{\mathrm{T}}A_-V_-, \quad \widehat{B}_- = W_-^{\mathrm{T}}B_-, \quad \widehat{C}_- = C_-V_-.$$

Then we have

$$\begin{aligned} \widehat{E} &= \begin{bmatrix} \widehat{E}_- & 0 \\ 0 & E_+ \end{bmatrix} \\ &= \begin{bmatrix} W_-^{\mathrm{T}} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} S_- & S_+ \end{bmatrix}^{\mathrm{T}} E \begin{bmatrix} T_- & T_+ \end{bmatrix} \begin{bmatrix} V_- & 0 \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} S_-W_- & S_+ \end{bmatrix}^{\mathrm{T}} E \begin{bmatrix} T_-V_- & T_+ \end{bmatrix}. \end{aligned}$$

In a similar way, we find $\widehat{A} = W^{\mathrm{T}}AV$, $\widehat{B} = W^{\mathrm{T}}B$, $\widehat{C} = CV$ for $V = \begin{bmatrix} T_-V_- & T_+ \end{bmatrix}$ and $W = \begin{bmatrix} S_-W_- & S_+ \end{bmatrix}$. Therefore, Petrov-Galerkin projection of $(E_-; A_-, B_-, C_-)$ can be represented as a Petrov-Galerkin projection of $(E; A, B, C)$. BT can also be applied to the asymptotically stable part and the $\mathcal{H}_\infty$-error bound will hold since $\|H - \widehat{H}\|_{\mathcal{H}_\infty} = \|H_- - \widehat{H}_-\|_{\mathcal{H}_\infty}$. Alternatively, IRKA can be applied to H_- to find $\widehat{H}_-$ which is a local minimum for $\|H_- - \widehat{H}_-\|_{\mathcal{H}_2}$. This implies that $\widehat{H} = \widehat{H}_- + H_+$ is a local minimum for $\|H - \widehat{H}\|_{\mathcal{H}_2}$. Notice that $H - \widehat{H} = H_- - \widehat{H}_-$, which can be used to compute the $\mathcal{H}_2$ or $\mathcal{H}_\infty$ error.

2.5 Graph theory

Parts of this thesis are about systems defined over networks. Thus, we present some basic concepts from graph theory here. Notation is based on [ME10] and [GR01].

2.5.1 Basic concepts

A *graph* $\mathfrak{G}$ consists of a *vertex set* $\mathfrak{V}$ and an *edge set* $\mathfrak{E}$ encoding the relation between vertices. *Undirected* graphs are those for which the edge set is a subset of the set of all unordered pairs of vertices, i.e., $\mathfrak{E} \subseteq \{\{i,j\} : i,j \in \mathfrak{V},\ i \neq j\}$. On the other hand, a graph is *directed* if $\mathfrak{E} \subseteq \{(i,j) : i,j \in \mathfrak{V},\ i \neq j\}$. We think of an edge (i,j) as an arrow starting from vertex i and ending at j.

Remark 2.46:
Notice that we exclude graphs with multiple copies of the same edge, i.e., *multigraphs*. Furthermore, we exclude graphs containing *self-loops*, i.e., edges of the form $\{i\}$ or (i,i). Therefore, we will only consider *simple* graphs. ◊

We will only consider *finite graphs*, i.e., graphs with a finite number of vertices $\mathfrak{n} \coloneqq |\mathfrak{V}|$. Without loss of generality, let $\mathfrak{V} = \{1, 2, \dots, \mathfrak{n}\}$.

For an undirected graph, a *path* of length ℓ is a sequence of distinct vertices $i_0, i_1, \dots, i_\ell$ such that $\{i_k, i_{k+1}\} \in \mathfrak{E}$ for $k = 0, 1, \dots, \ell - 1$. For a directed graph, a *directed path* of length ℓ is a sequence of distinct vertices $i_0, i_1, \dots, i_\ell$ such that $(i_k, i_{k+1}) \in \mathfrak{E}$ for $k = 0, 1, \dots, \ell - 1$. An undirected graph is *connected* if there is a path between any two distinct vertices $i, j \in \mathfrak{V}$. A directed graph is *strongly connected* if there is a directed path between any two distinct vertices $i, j \in \mathfrak{V}$.

We can associate weights to edges of a graph by a *weight function* $\mathfrak{w} \colon \mathfrak{E} \to \mathbb{R}$. If $\mathfrak{w}(\mathfrak{e}) > 0$ for all $\mathfrak{e} \in \mathfrak{E}$, the tuple $\mathfrak{G} = (\mathfrak{V}, \mathfrak{E}, \mathfrak{w})$ is called a *weighted graph*. In the following, we will focus on weighted graphs. In particular, we will directly generalize concepts for unweighted graphs from [ME10, GR01], as was done in [MTC14].

The *adjacency matrix* $\mathfrak{A} = [\mathfrak{a}_{ij}]_{i,j \in \mathfrak{V}} \in \mathbb{R}^{\mathfrak{n} \times \mathfrak{n}}$ of an undirected graph is defined component-wise by

$$\mathfrak{a}_{ij} \coloneqq \begin{cases} \mathfrak{w}(\{i,j\}), & \text{if } \{i,j\} \in \mathfrak{E}, \\ 0, & \text{otherwise,} \end{cases}$$

and for a directed graph as

$$\mathfrak{a}_{ij} \coloneqq \begin{cases} \mathfrak{w}((j,i)), & \text{if } (j,i) \in \mathfrak{E}, \\ 0, & \text{otherwise.} \end{cases}$$

For every vertex $i \in \mathfrak{V}$, its *in-degree* is $\mathfrak{d}_i \coloneqq \sum_{j=1}^{\mathfrak{n}} \mathfrak{a}_{ij}$. The diagonal matrix $\mathfrak{D} \coloneqq \operatorname{diag}(\mathfrak{d}_1, \mathfrak{d}_2, \dots, \mathfrak{d}_\mathfrak{n})$ is called the *in-degree matrix*. Notice that $\mathfrak{D} = \operatorname{diag}(\mathfrak{A}\mathbb{1})$.

Let $\mathfrak{e}_1, \mathfrak{e}_2, \dots, \mathfrak{e}_{|\mathfrak{E}|}$ be all the edges of $\mathfrak{G}$ in some order. The *incidence matrix* $\mathfrak{R} \in \mathbb{R}^{\mathfrak{n} \times |\mathfrak{E}|}$ of a directed graph $\mathfrak{G}$ is defined component-wise

$$[\mathfrak{R}]_{ik} \coloneqq \begin{cases} -1, & \text{if } \mathfrak{e}_k = (i,j) \text{ for some } j \in \mathfrak{V}, \\ 1, & \text{if } \mathfrak{e}_k = (j,i) \text{ for some } j \in \mathfrak{V}, \\ 0, & \text{otherwise.} \end{cases}$$

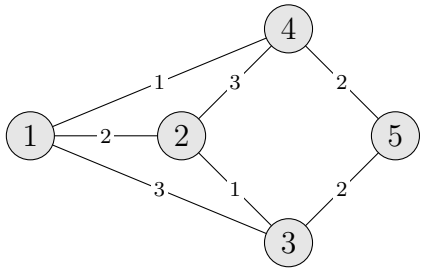

Figure 2.1: An undirected, weighted, connected graph

If $\mathfrak{G}$ is undirected, we assign some orientation to every edge to define a directed graph $\mathfrak{G}^o$, and define the incidence matrix of $\mathfrak{G}$ to be the incidence matrix of $\mathfrak{G}^o$. The *weight matrix* is defined as $\mathfrak{W} \coloneqq \operatorname{diag}\big(\mathfrak{w}(\mathfrak{e}_1), \mathfrak{w}(\mathfrak{e}_2), \ldots, \mathfrak{w}(\mathfrak{e}_{|\mathfrak{E}|})\big)$.

The *(in-degree) Laplacian matrix* $\mathfrak{L}$ is defined by $\mathfrak{L} \coloneqq \mathfrak{D} - \mathfrak{A}$. For undirected graphs, it can be checked that $\mathfrak{L} = \mathfrak{R}\mathfrak{W}\mathfrak{R}^{\mathrm{T}}$, using

$$\mathfrak{R}\mathfrak{W}\mathfrak{R}^{\mathrm{T}} = \sum_{\{i,j\}\in\mathfrak{E}} \mathfrak{a}_{ij}(e_i - e_j)(e_i - e_j)^{\mathrm{T}},$$

which is independent of the order of edges defining $\mathfrak{R}$ and $\mathfrak{W}$ or the orientation of edges in $\mathfrak{G}^o$. From the definition of $\mathfrak{L}$, it directly follows that the sum of each row in $\mathfrak{L}$ is zero, i.e., $\mathfrak{L}\mathbb{1} = 0$. From $\mathfrak{L} = \mathfrak{R}\mathfrak{W}\mathfrak{R}^{\mathrm{T}}$, we immediately see that, for undirected weighted graphs, the Laplacian matrix $\mathfrak{L}$ is symmetric positive semidefinite.

Example 2.1:
For the graph in Figure 2.1, the adjacency matrix is

$$\mathfrak{A} = \begin{bmatrix} 0 & 2 & 3 & 1 & 0 \\ 2 & 0 & 1 & 3 & 0 \\ 3 & 1 & 0 & 0 & 2 \\ 1 & 3 & 0 & 0 & 2 \\ 0 & 0 & 2 & 2 & 0 \end{bmatrix},$$

degree matrix $\mathfrak{D} = \operatorname{diag}(6, 6, 6, 6, 4)$, Laplacian and incidence matrix

$$\mathfrak{L} = \begin{bmatrix} 6 & -2 & -3 & -1 & 0 \\ -2 & 6 & -1 & -3 & 0 \\ -3 & -1 & 6 & 0 & -2 \\ -1 & -3 & 0 & 6 & -2 \\ 0 & 0 & -2 & -2 & 4 \end{bmatrix}, \quad \mathfrak{R} = \begin{bmatrix} -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix},$$

and weight matrix $\mathfrak{W} = \operatorname{diag}(2, 3, 1, 1, 3, 2, 2)$. ◊

The following theorem states that connectedness of a graph is related to the spectral properties of $\mathfrak{L}$.

Theorem 2.47 ([ME10, Theorem 2.8]):
Let $\mathfrak{G} = (\mathfrak{V}, \mathfrak{E}, \mathfrak{w})$ be an undirected weighted graph, $\mathfrak{L}$ its Laplacian matrix, and $0 = \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_\mathfrak{n}$ the eigenvalues of $\mathfrak{L}$. Then the following statements are equivalent:

1. $\mathfrak{G}$ is connected,
2. $\lambda_2 > 0$,
3. $\ker(\mathfrak{L}) = \mathrm{im}(\mathbb{1})$. ◊

2.5.2 Graph partitions

A nonempty subset $\mathfrak{C} \subseteq \mathfrak{V}$ is called a *cluster* or *cell* of $\mathfrak{V}$. A *graph partition* π is a partition of the vertex set $\mathfrak{V}$. Vertices i and j are called *cellmates* in π if they belong to the same cell of π. The *characteristic vector* of a cluster $\mathfrak{C} \subseteq \mathfrak{V}$ is the $\mathfrak{n}$-dimensional column vector $\mathfrak{p}(\mathfrak{C})$ defined as

$$[\mathfrak{p}(\mathfrak{C})]_i = \begin{cases} 1 & \text{if } i \in \mathfrak{C}, \\ 0 & \text{otherwise.} \end{cases}$$

The *characteristic matrix of a partition* $\pi = \{\mathfrak{C}_1, \mathfrak{C}_2, \ldots, \mathfrak{C}_\mathfrak{r}\}$ of the graph $\mathfrak{G}$ is the matrix $\mathfrak{P} \in \mathbb{R}^{\mathfrak{n}\times\mathfrak{r}}$ defined by

$$\mathfrak{P} = \begin{bmatrix} \mathfrak{p}(\mathfrak{C}_1) & \mathfrak{p}(\mathfrak{C}_2) & \cdots & \mathfrak{p}(\mathfrak{C}_\mathfrak{r}) \end{bmatrix}.$$

For a given vertex $i \in \mathfrak{V}$ and a cluster $\mathfrak{C}_q \in \pi$, we define the *in-degree of vertex i with respect to cluster $\mathfrak{C}_q$* by

$$\mathfrak{d}(i, \mathfrak{C}_q) = \sum_{k \in \mathfrak{C}_q} \mathfrak{a}_{ik}.$$

The partition π is an *equitable partition* if $\mathfrak{d}(i, \mathfrak{C}_q) = \mathfrak{d}(j, \mathfrak{C}_q)$ for all vertices $i, j \in \mathfrak{C}_p$ and any two clusters $\mathfrak{C}_p, \mathfrak{C}_q \in \pi$ (they can be the same cluster). The partition π is an *AEP* if $\mathfrak{d}(i, \mathfrak{C}_q) = \mathfrak{d}(j, \mathfrak{C}_q)$ all vertices $i, j \in \mathfrak{C}_p$ and for any two different clusters $\mathfrak{C}_p, \mathfrak{C}_q \in \pi$.

Example 2.2:
For the partition $\pi = \{\{1, 2\}, \{3, 4\}, \{5\}\}$, the characteristic matrix is

$$\mathfrak{P} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Notice that π is an equitable partition for the graph in Figure 2.1. Furthermore, notice that $\{\{1, 2, 5\}, \{3, 4\}\}$ is an AEP, but not an equitable partition. ◊

The following theorem gives necessary and sufficient conditions for equitableness and almost-equitableness (see [ME10, Lemma 2.24], [MTC14, Lemma 5], and [CDR07, Proposition 1]).

Lemma 2.48:
Let $\mathfrak{G}$ be an undirected weighted graph with adjacency matrix $\mathfrak{A}$ and Laplacian matrix $\mathfrak{L}$. Furthermore, let π be a partition of the graph $\mathfrak{G}$. Then,

- π is equitable if and only if $\mathrm{im}(\mathfrak{P})$ is $\mathfrak{A}$-invariant,
- π is almost equitable if and only if $\mathrm{im}(\mathfrak{P})$ is $\mathfrak{L}$-invariant. ◊

Notice that $\{\mathfrak{V}\}$ and $\{\{1\},\{2\},\ldots,\{\mathfrak{n}\}\}$ are always AEPs. We will refer to them as *trivial AEPs*.

Remark 2.49:
To the best of our knowledge, there is no known polynomial-time algorithm for finding nontrivial AEPs of a given graph. There is a polynomial-time algorithm for finding the coarsest AEP which is finer than a given partition (see [ZCC14]), but there is no guarantee that it will find a nontrivial AEP.

Furthermore, it is not clear whether a given graph has any nontrivial AEPs at all. On the other hand, a graph can have many AEPs, e.g., every partition of a complete unweighted graph is an AEP. ◊

2.6 Linear multi-agent systems

2.6.1 System description

Our goal here is to unify definitions from Besselink et al. [BSJ16], Cheng et al. [CKS16, CKS18], Ishizaki et al. [IKIA14, IKG+15, IKI16a], and Monshizadeh et al. [MTC13, MTC14]. In particular, we focus on LTI multi-agent systems. Additionally, we restrict to multi-agent system on a undirected, weighted, and connected graph $\mathfrak{G} = (\mathfrak{V}, \mathfrak{E}, \mathfrak{w})$.

We base the derivation on [BSJ16] and [CKS18]. The dynamics of the ith agent, for $i \in \mathfrak{V} = \{1, 2, \ldots, \mathfrak{n}\}$, is

$$\begin{aligned} E\dot{x}_i(t) &= Ax_i(t) + Bv_i(t), \\ z_i(t) &= Cx_i(t), \end{aligned}$$

with system matrices $E, A \in \mathbb{R}^{n\times n}$, input matrix $B \in \mathbb{R}^{n\times m}$, output matrix $C \in \mathbb{R}^{p\times n}$, state $x_i(t) \in \mathbb{R}^n$, input $v_i(t) \in \mathbb{R}^m$, and output $z_i(t) \in \mathbb{R}^p$. We assume the matrix E to be invertible. The interconnections are

$$\mathfrak{m}_i v_i(t) = K \sum_{j=1}^{\mathfrak{n}} \mathfrak{a}_{ij}(z_j(t) - z_i(t)) + \sum_{k=1}^{\mathfrak{m}} \mathfrak{b}_{ik} u_k(t),$$

for $i = 1, 2, \dots, \mathfrak{n}$, with inertia $\mathfrak{m}_i > 0$, coupling matrix $K \in \mathbb{R}^{m \times p}$, external inputs $u_k(t) \in \mathbb{R}^m$, $k = 1, 2, \dots, \mathfrak{m}$, where $\mathfrak{A} = [\mathfrak{a}_{ij}]$ is the adjacency matrix of the graph $\mathfrak{G}$. Here, we allow m and p to be different, and for this reason also need a (possibly non-square) coupling matrix K. The outputs are

$$y_\ell(t) = \sum_{j=1}^{\mathfrak{n}} \mathfrak{c}_{\ell j} z_j(t)$$

for $\ell = 1, 2, \dots, \mathfrak{p}$. Define

$$\begin{gathered}\mathfrak{M} \coloneqq \operatorname{diag}(\mathfrak{m}_i) \in \mathbb{R}^{\mathfrak{n} \times \mathfrak{n}},\ \mathfrak{B} \coloneqq [\mathfrak{b}_{ik}] \in \mathbb{R}^{\mathfrak{n} \times \mathfrak{m}},\ \mathfrak{C} \coloneqq [\mathfrak{c}_{\ell j}] \in \mathbb{R}^{\mathfrak{p} \times \mathfrak{n}}, \\ x(t) \coloneqq \operatorname{col}(x_i(t)) \in \mathbb{R}^{\mathfrak{n}n},\ v(t) \coloneqq \operatorname{col}(v_i(t)) \in \mathbb{R}^{\mathfrak{n}m},\ z(t) \coloneqq \operatorname{col}(z_i(t)) \in \mathbb{R}^{\mathfrak{n}p}, \\ u(t) \coloneqq \operatorname{col}(u_k(t)) \in \mathbb{R}^{\mathfrak{m}m},\ \text{and}\ y(t) \coloneqq \operatorname{col}(y_\ell(t)) \in \mathbb{R}^{\mathfrak{p}p}.\end{gathered}$$

Then the agent dynamics can be rewritten as

$$\begin{aligned}(I_\mathfrak{n} \otimes E)\dot{x}(t) &= (I_\mathfrak{n} \otimes A)x(t) + (I_\mathfrak{n} \otimes B)v(t), \\ z(t) &= (I_\mathfrak{n} \otimes C)x(t),\end{aligned}$$

interconnection as

$$(\mathfrak{M} \otimes I_n)v(t) = (-\mathfrak{L} \otimes K)z(t) + (\mathfrak{B} \otimes I_m)u(t),$$

and output as

$$y(t) = (\mathfrak{C} \otimes I_p)z(t).$$

Therefore, we have

$$\begin{aligned}(\mathfrak{M} \otimes E)\dot{x}(t) &= (\mathfrak{M} \otimes A - \mathfrak{L} \otimes BKC)x(t) + (\mathfrak{B} \otimes B)u(t), \\ y(t) &= (\mathfrak{C} \otimes C)x(t).\end{aligned}$$

This can be generalized further by replacing B in the input matrix $\mathfrak{B} \otimes B$ or C in the output matrix $\mathfrak{C} \otimes C$. For example, output $y(t) = (\mathfrak{C} \otimes I_n)x(t)$ is used in [IKI16a], with $\mathfrak{C} = I_\mathfrak{n}$. We can consider

$$\begin{aligned}(\mathfrak{M} \otimes E)\dot{x}(t) &= (\mathfrak{M} \otimes A - \mathfrak{L} \otimes BKC)x(t) + (\mathfrak{B} \otimes F)u(t), \\ y(t) &= (\mathfrak{C} \otimes G)x(t),\end{aligned} \tag{2.9}$$

where F and G can be different from B and C respectively (including the dimensions).

Of particular interest are *leader-follower multi-agent systems* where only some agents (*leaders*) receive external input, while other agents (*followers*) receive no inputs. Let $\mathfrak{m} \in \{1, 2, \dots, \mathfrak{n}\}$ be the number of leaders, $\mathfrak{V}_\mathrm{L} = \{\mathfrak{v}_1, \mathfrak{v}_2, \dots, \mathfrak{v}_\mathfrak{m}\} \subseteq \mathfrak{V}$ the set of leaders, and $\mathfrak{V}_\mathrm{F} = \mathfrak{V} \setminus \mathfrak{V}_\mathrm{L}$ the set of followers. Then, with $\mathfrak{B}$ defined by

$$\mathfrak{b}_{ij} \coloneqq \begin{cases} 1, & \text{if } i = \mathfrak{v}_j, \\ 0, & \text{otherwise,} \end{cases}$$

the system (2.9) becomes a *leader-follower multi-agent system*. One important class are multi-agent systems with *single-integrator agents*, i.e., with $n = 1$, $A = 0$, $B = C = 1$, and $K = 1$. Thus, system (2.9) becomes

$$\begin{aligned} \mathfrak{M}\dot{x}(t) &= -\mathfrak{L}x(t) + \mathfrak{B}u(t), \\ y(t) &= \mathfrak{C}x(t). \end{aligned} \tag{2.10}$$

In particular, [MTC14] uses $\mathfrak{M} = I_{\mathfrak{n}}$ and $\mathfrak{C} = \mathfrak{W}^{\frac{1}{2}}\mathfrak{R}^{\mathrm{T}}$, while [IKG+15] and [CKS16] use $\mathfrak{C} = I_{\mathfrak{n}}$.

The property of interest for multi-agent systems is *synchronization*.

Definition 2.50:
The system $(\mathfrak{M} \otimes E)\dot{x}(t) = (\mathfrak{M} \otimes A - \mathfrak{L} \otimes B)x(t)$ is *synchronized* if

$$\lim_{t\to\infty} (x_i(t) - x_j(t)) = 0,$$

for all $i, j \in \mathfrak{V}$ and all initial conditions $x(0) = x_0$. ◊

For system (2.10), this is equivalent to $(\mathfrak{L} \otimes I_n)x(t) \to 0$, because the multi-agent system is defined on a connected graph. The following results gives another equivalent condition.

Proposition 2.51 ([LDCH10, Theorem 1], [MTC13, Lemma 4.2]):
The system $(\mathfrak{M}\otimes E)\dot{x}(t) = (\mathfrak{M}\otimes A-\mathfrak{L}\otimes B)x(t)$ is synchronized if and only if $(A-\lambda B, E)$ is Hurwitz for all nonzero eigenvalues λ of $(\mathfrak{L}, \mathfrak{M})$. ◊

2.6.2 Clustering-based model order reduction

By choosing some matrices $V, W \in \mathbb{R}^{\mathfrak{n}\times\mathfrak{r}}$, we get the ROM for (2.10)

$$\begin{aligned} W^{\mathrm{T}}\mathfrak{M}V\dot{\widehat{x}}(t) &= -W^{\mathrm{T}}\mathfrak{L}V\widehat{x}(t) + W^{\mathrm{T}}\mathfrak{B}u(t), \\ \widehat{y}(t) &= \mathfrak{C}V\widehat{x}(t), \end{aligned} \tag{2.11}$$

or, for (2.9),

$$\begin{aligned} \left(W^{\mathrm{T}}\mathfrak{M}V \otimes E\right)\dot{\widehat{x}}(t) &= \left(W^{\mathrm{T}}\mathfrak{M}V \otimes A - W^{\mathrm{T}}\mathfrak{L}V \otimes BKC\right)\widehat{x}(t) + \left(W^{\mathrm{T}}\mathfrak{B} \otimes F\right)u(t), \\ \widehat{y}(t) &= (\mathfrak{C}V \otimes G)\widehat{x}(t), \end{aligned} \tag{2.12}$$

which is not necessarily a multi-agent system. Different projection methods are suggested in literature:

1. $V = \mathfrak{P}$, $W = \mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}$ ([MTC14]),
2. $V = W = \mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}}$ ([IKIA14]),
3. $V = W = \mathfrak{P}$ ([CKS16]),

where $\mathfrak{P}$ is a characteristic matrix of a partition π of the vertex set $\mathfrak{V}$. Since $\mathrm{im}(V) = \mathrm{im}(W) = \mathrm{im}(\mathfrak{P})$ in all cases, all methods give equivalent ROMs. They each have their advantages. The first preserves structure when $\mathfrak{M} = I_{\mathfrak{n}}$, in the sense that the ROM represents a multi-agent system defined on a directed, symmetric graph [MTC14]. The second gives the realization for which it is easy to apply the interlacing property (Theorem 2.6), from which it follows that all these methods preserve synchronization. The third preserves the structure in the general case.

CHAPTER 3

SUBOPTIMAL CLUSTERING-BASED MODEL ORDER REDUCTION

Contents

3.1 Introduction

In this chapter, we study methods for clustering-based MOR for multi-agent systems (see Section 2.6 for an overview). Clustering was proposed in the literature as a way to preserve the multi-agent structure. Here, we will be interested in finding optimal (in a specific error measure) partitions. Since clustering is generally a difficult combinatorial problem (see, e.g., [Sch07]), we will propose a heuristic approach for finding suboptimal partitions.

References [MEB08, RJME09, MEB10] introduce *leader-invariant equitable partitions* and corresponding *quotient graphs*, which can be used for eliminating some uncontrollable states, thus performing no model reduction error. Reference [CM11] proposes *leader partitions*, which introduce "only small model errors". The authors of [IKIA14] develop a clustering-based $\mathcal{H}_\infty$ MOR method based on *positive tridiagonalization* and

θ-reducible clusters, applicable to LTI systems with asymptotically stable and symmetric dynamics matrices.

In [MTC14], the authors focus on *leader-follower linearly diffusively coupled multi-agent systems* with agents having single-integrator dynamics. In particular, these systems have Laplacian-based dynamics, which means they are not asymptotically stable. The authors demonstrate how using partitions for MOR can be transformed to using Petrov-Galerkin projections, while preserving network structure and consensus in the ROM. Further, they derive a simple expression for the relative $\mathcal{H}_2$-error when using an AEP and establish a lower bound based on AEPs when using any partition.

The authors of [BSJ16] study *networks of identical passive systems over weighted and directed graphs*, but confine to interconnections with tree structures. Otherwise, these systems would be more general then those in [MTC14]. They present a clustering method relying on the analysis of the corresponding *edge system* to find adjacent subsystems to cluster. Furthermore, they prove that this method preserves the consensus property. Nonetheless, clustering only adjacent subsystems might be restrictive.

The paper [IKG+12] presents an efficient clustering-based method for $\mathcal{H}_2$ MOR of *positive networks*, which include systems with Laplacian-based dynamics. Similarly to [IKIA14], the method is based on θ-reducible clusters and an $\mathcal{H}_2$-error bound. However, it is not clear how the ROMs resulting from this method compare to the $\mathcal{H}_2$-optimal ones.

To the best of our knowledge, there is no efficient method for finding an $\mathcal{H}_2$-optimal ROM using graph partitions. In Section 3.2, we propose an efficient $\mathcal{H}_2$-suboptimal method, originating from the $\mathcal{H}_2$-optimal MOR problem for LTI systems, and illustrate it on an example. Results indicate that the method finds a partition close to the optimal. In Section 3.3, we extend the approach to a class of nonlinear multi-agent systems.

3.2 $\mathcal{H}_2$-suboptimal clustering

The outline of the section is as follows. In Section 3.2.2, we motivate and describe our MOR method, together with the issue of computing the $\mathcal{H}_2$-error for systems with Laplacian-based dynamics. We illustrate the method on an example in Section 3.2.6 and conclude with Section 3.4.

3.2.1 Single-integrator agents

We first consider multi-agent systems with single-integrator agents as in (2.10). Let

$$H(s) = \mathfrak{C}(s\mathfrak{M} + \mathfrak{L})^{-1}\mathfrak{B},$$
$$\widehat{H}(s) = \mathfrak{C}V\left(sW^{\mathrm{T}}\mathfrak{M}V + W^{\mathrm{T}}\mathfrak{L}V\right)^{-1}W^{\mathrm{T}}\mathfrak{B}$$

be the transfer functions of systems (2.10) and (2.11), respectively. We consider the following $\mathcal{H}_2$-optimal MOR problem:

$$\begin{aligned} \underset{V,W\in\mathbb{R}^{\mathfrak{n}\times\mathfrak{r}}}{\text{minimize}} \quad & \|H-\widehat{H}\|_{\mathcal{H}_2}\,, && \text{(3.1a)}\\ \text{subject to} \quad & V=\mathfrak{P}, && \text{(3.1b)}\\ & W=\mathfrak{P}, && \text{(3.1c)}\\ & \pi\in\Pi,\ |\pi|=\mathfrak{r}, && \text{(3.1d)} \end{aligned}$$

where Π is a set of all partitions of the vertex set $\mathfrak{V}$.

To the best of our knowledge, there is no efficient method to exactly solve the optimization problem (3.1). Additionally, since many graph clustering problems are NP-hard (see [Sch07]), we assume the same is true for the above problem.

The problem (3.1) is actually a discrete optimization problem over the set of partitions of the set $\{1,2,\ldots,\mathfrak{n}\}$ with $\mathfrak{r}$ clusters. Thus, our idea is to solve a relaxed, continuous optimization problem, and use that solution to find a feasible solution for (3.1). We hope that this feasible solution is then close to the optimal solution as it is close to the optimal solution of the relaxed problem.

We relax the problem (3.1) by dropping all constraints (3.1b), (3.1c), and (3.1d). Thus, we obtain the $\mathcal{H}_2$-optimal MOR problem for an LTI system. Note that the system (2.10) is not asymptotically stable. See also Section 3.2.4 below for some elaboration on the technicalities associated to this problem.

Applying an $\mathcal{H}_2$-optimal method, such as IRKA, to (2.10) will not in general solve the original problem (3.1). Therefore, IRKA will return as a result matrices V and W which will not (in general) satisfy the constraints. However, note that in Petrov-Galerkin projection, the subspaces $\mathcal{V}$ and $\mathcal{W}$ are enough to determine the transfer function of the ROM. Therefore, the constraints (3.1b) and (3.1c) can be replaced by

$$\begin{aligned} \operatorname{im}(V) &= \operatorname{im}(\mathfrak{P})\,, && \text{(3.2a)}\\ \operatorname{im}(W) &= \operatorname{im}(\mathfrak{P})\,, && \text{(3.2b)} \end{aligned}$$

without changing $\widehat{H}$ and the cost (3.1a). The expressions (3.2) motivate us to look for a partition π such that $\operatorname{im}(\mathfrak{P})$ is "close" to $\operatorname{im}(V)$ and/or $\operatorname{im}(W)$.

3.2.2 QR decomposition-based clustering

We know that the condition (3.2a) is equivalent to the existence of a invertible matrix Z such that $V=\mathfrak{P}Z$. In general, condition (3.2a) will not be satisfied, so there will be an error Δ such that $V=\mathfrak{P}Z+\Delta$. A very similar problem of finding a partition π was encountered in [ZHD+01, §3], where a proposed solution is a clustering algorithm based on the *QR decomposition with column pivoting*. Algorithm 3.1 outlines the procedure. The motivation behind it is the result of the following lemma, which says that if $V=\mathfrak{P}Z$, the procedure will return the correct result. Therefore, the idea is that if the error Δ is small, then Algorithm 3.1 should still return the correct partition.

Algorithm 3.1: Clustering using QR decomposition with column pivoting [ZHD+01, §3]

Input: Matrix $V \in \mathbb{R}^{\mathfrak{n}\times\mathfrak{r}}$ of rank $\mathfrak{r}$.
Output: Partition π such that $\operatorname{im}(\mathfrak{P}) \approx \operatorname{im}(V)$.

1 Compute QR decomposition with column pivoting for the matrix V^{T}, i.e., find orthogonal $Q \in \mathbb{R}^{\mathfrak{r}\times\mathfrak{r}}$, upper-trapezoidal $R \in \mathbb{R}^{\mathfrak{r}\times\mathfrak{n}}$, and a permutation matrix $P \in \mathbb{R}^{\mathfrak{n}\times\mathfrak{n}}$ such that $V^{\mathrm{T}}P = QR$.
2 Partition R as $\begin{bmatrix} R_{11} & R_{12} \end{bmatrix}$, with $R_{11} \in \mathbb{R}^{\mathfrak{r}\times\mathfrak{r}}$ upper-triangular and $R_{12} \in \mathbb{R}^{\mathfrak{r}\times(\mathfrak{n}-\mathfrak{r})}$.
3 Solve the triangular system $R_{11}X = R_{12}$.
4 Compute $Y = P \begin{bmatrix} I_{\mathfrak{r}} & X \end{bmatrix}^{\mathrm{T}} = [y_{ij}] \in \mathbb{R}^{\mathfrak{n}\times\mathfrak{r}}$.
5 Find a partition $\pi = \{\mathfrak{C}_1, \mathfrak{C}_2, \ldots, \mathfrak{C}_{\mathfrak{r}}\}$ such that $i \in \mathfrak{C}_j$ if and only if $j = \operatorname{argmax}_k |y_{ik}|$.

Lemma 3.1:
Algorithm 3.1 returns π with $\mathfrak{P}Z$ as input, for an arbitrary partition π and a invertible matrix Z. ◊

Proof. Let the number of clusters in π be $\mathfrak{r}$, i.e., $|\pi| = \mathfrak{r}$. Let us denote the rows of Z with $z_1^{\mathrm{T}}, z_2^{\mathrm{T}}, \ldots, z_{\mathfrak{r}}^{\mathrm{T}}$. Furthermore, without loss of generality we can assume that $\mathfrak{P} = \operatorname{diag}\big(\mathbb{1}_{|C_1|}, \mathbb{1}_{|C_2|}, \ldots, \mathbb{1}_{|C_{\mathfrak{r}}|}\big)$, where $\mathbb{1}_k \in \mathbb{R}^k$ is a vector of all ones (this structure can be achieved by relabeling the vertices of the graph). Then we have

$$Z^{\mathrm{T}}\mathfrak{P}^{\mathrm{T}} = \begin{bmatrix} z_1\mathbb{1}_{|C_1|}^{\mathrm{T}} & z_2\mathbb{1}_{|C_2|}^{\mathrm{T}} & \cdots & z_{\mathfrak{r}}\mathbb{1}_{|C_{\mathfrak{r}}|}^{\mathrm{T}} \end{bmatrix}. \tag{3.3}$$

Therefore, $Z^{\mathrm{T}}\mathfrak{P}^{\mathrm{T}}$ has repeating columns in blocks. If we perform the QR decomposition with column pivoting on (3.3) (ignoring the orthogonal matrix) and then undo the permutation of columns, the result is

$$\begin{bmatrix} \widetilde{z}_1\mathbb{1}_{|C_1|}^{\mathrm{T}} & \widetilde{z}_2\mathbb{1}_{|C_2|}^{\mathrm{T}} & \cdots & \widetilde{z}_{\mathfrak{r}}\mathbb{1}_{|C_{\mathfrak{r}}|}^{\mathrm{T}} \end{bmatrix}, \tag{3.4}$$

where a permutation of the columns of $\begin{bmatrix} \widetilde{z}_1 & \widetilde{z}_2 & \cdots & \widetilde{z}_{\mathfrak{r}} \end{bmatrix}$ gives the upper-triangular matrix R_{11} from Algorithm 3.1. We see that multiplying (3.4) on the left by R_{11}^{-1} ($\mathfrak{P}Z$ having full rank implies that R_{11} is invertible) and transposing produces $\mathfrak{P}$, possibly with permuted columns. Thus we conclude that Algorithm 3.1 returns the partition π. □

Furthermore, note that the number of floating point operations in Algorithm 3.1 is $\mathcal{O}(\mathfrak{n}\mathfrak{r}^2)$ and the size of additional storage is $\mathcal{O}(\mathfrak{n}\mathfrak{r})$. Therefore, it is efficient in the large-scale setting if $\mathfrak{r}$ is small compared to $\mathfrak{n}$.

3.2.3 Clustering by k-means algorithm

The motivation for using the k-means algorithm ([HW79]) is the following result about how a change in the Petrov-Galerkin subspaces affects the ROM.

Theorem 3.2 ([BGW12, Theorem 3.3]):
Let $V_1, V_2, W_1, W_2 \in \mathbb{R}^{\mathfrak{n}\times\mathfrak{r}}$,

$$\mathcal{V}_i = \operatorname{im}(V_i)\,,\ \mathcal{W}_i = \operatorname{im}(W_i)\,,\ \widehat{H}_i(s) = CV_i\left(sW_i^{\mathrm{T}}EV_i - W_i^{\mathrm{T}}AV_i\right)^{-1}W_i^{\mathrm{T}}B,$$

for $i = 1, 2$. Then

$$\frac{\|\widehat{H}_1 - \widehat{H}_2\|_{\mathcal{H}_\infty}}{\frac{1}{2}(\|\widehat{H}_1\|_{\mathcal{H}_\infty} + \|\widehat{H}_2\|_{\mathcal{H}_\infty})} \leqslant M \max(\sin\Theta(\mathcal{V}_1, \mathcal{V}_2), \sin\Theta(\mathcal{W}_1, \mathcal{W}_2)),$$

where

$$\begin{aligned}
M &= 2\max(M_1, M_2)\,,\\
M_1 &= \frac{\max_{\omega\in\mathbb{R}}\|C\|_2\left\|V_1\left(\imath\omega W_1^{\mathrm{T}}EV_1 - W_1^{\mathrm{T}}AV_1\right)^{-1}W_1^{\mathrm{T}}B\right\|_2\|\widehat{H}_1(\imath\omega)\|_2^{-1}}{\min_{\omega\in\mathbb{R}}\cos\Theta(\ker\left(W_2^{\mathrm{T}}\left(\imath\omega E - A\right)^{-1}\right)^{\perp}, \mathcal{V}_2)},\\
M_2 &= \frac{\max_{\omega\in\mathbb{R}}\left\|CV_2\left(\imath\omega W_2^{\mathrm{T}}EV_2 - W_2^{\mathrm{T}}AV_2\right)^{-1}W_2^{\mathrm{T}}\right\|_2\|B\|_2\|\widehat{H}_2(\imath\omega)\|_2^{-1}}{\min_{\omega\in\mathbb{R}}\cos\Theta(\operatorname{im}\left(\left(\imath\omega E - A\right)^{-1}V_1\right), \mathcal{W}_1)},
\end{aligned}$$

and $\Theta(\mathcal{M}, \mathcal{N})$ is the largest principal angle between subspaces $\mathcal{M}, \mathcal{N} \subseteq \mathbb{R}^n$. ◊

The angle between two subspaces $\mathcal{V}_1, \mathcal{V}_2 \subseteq \mathbb{R}^n$ is defined by (see [BGW12, Section 3.1])

$$\sin\Theta(\mathcal{V}_1, \mathcal{V}_2) := \sup_{v_1\in\mathcal{V}_1}\inf_{v_2\in\mathcal{V}_2}\frac{\|v_2 - v_1\|_2}{\|v_1\|_2}.$$

If $\dim\mathcal{V}_1 = \dim\mathcal{V}_2$, then we have

$$\sin\Theta(\mathcal{V}_1, \mathcal{V}_2) = \sin\Theta(\mathcal{V}_2, \mathcal{V}_1) = \left\|\left(I - V_1V_1^{\mathrm{T}}\right)V_2\right\|_2,$$

where $\mathcal{V}_1 = \operatorname{im}(V_1)$, $\mathcal{V}_2 = \operatorname{im}(V_2)$, and both V_1 and V_2 have orthonormal columns. If additionally $V_1 = \mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}}$ and $V_2 = V$ from IRKA, then

$$\begin{aligned}
(\sin\Theta(\mathcal{V}_1, \mathcal{V}_2))^2 &\leqslant \left\|\left(I - \mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\right)V\right\|_{\mathrm{F}}^2\\
&= \left\|\left(I - \begin{bmatrix}\mathfrak{p}(\mathfrak{C}_1) & \cdots & \mathfrak{p}(\mathfrak{C}_\mathfrak{r})\end{bmatrix}\begin{bmatrix}|\mathfrak{C}_1|^{-1} & & \\ & \ddots & \\ & & |\mathfrak{C}_\mathfrak{r}|^{-1}\end{bmatrix}\begin{bmatrix}\mathfrak{p}(\mathfrak{C}_1)^{\mathrm{T}}\\ \vdots \\ \mathfrak{p}(\mathfrak{C}_\mathfrak{r})^{\mathrm{T}}\end{bmatrix}\right)V\right\|_{\mathrm{F}}^2\\
&= \left\|\left(I - \sum_{i=1}^{\mathfrak{r}}\frac{1}{|\mathfrak{C}_i|}\mathfrak{p}(\mathfrak{C}_i)\mathfrak{p}(\mathfrak{C}_i)^{\mathrm{T}}\right)V\right\|_{\mathrm{F}}^2\\
&= \sum_{i=1}^{\mathfrak{r}}\left\|V_{\mathfrak{C}_i,:} - \mathfrak{p}(\mathfrak{C}_i)\frac{1}{|\mathfrak{C}_i|}\mathbb{1}_{|\mathfrak{C}_i|}^{\mathrm{T}}V_{\mathfrak{C}_i,:}\right\|_{\mathrm{F}}^2\\
&= \sum_{i=1}^{\mathfrak{r}}\sum_{p\in\mathfrak{C}_i}\left\|V_{p,:} - \frac{1}{|\mathfrak{C}_i|}\sum_{q\in\mathfrak{C}_i}V_{q,:}\right\|_{\mathrm{F}}^2,
\end{aligned}$$

which is equal to the k-means cost functional. Therefore, applying the k-means algorithm to the rows of V will minimize an upper bound on the largest principal angle between $\operatorname{im}(V)$ and $\operatorname{im}(\mathfrak{P})$.

The advantage of using k-means compared to QR decomposition-based in that the latter can only, given $V \in \mathbb{R}^{\mathfrak{n}\times\mathfrak{r}}$, return a partition with $\mathfrak{r}$ clusters. On the other hand, k-means clustering can return a partition with any number of clusters, which makes it more efficient if the number of clusters is relatively large.

3.2.4 Computing the $\mathcal{H}_2$-error

When using an AEP, the relative $\mathcal{H}_2$-error can be computed directly [MTC14, Theorem 6]. In other cases, we need to solve Lyapunov equations to compute the $\mathcal{H}_2$-norms.

As discussed in Section 2.4.5, we can use the eigenspaces to find the asymptotically stable part of the system and use it to compute the $\mathcal{H}_2$-norm. We are looking for an invertible matrix T such that

$$T^{\mathrm{T}}\mathfrak{M}T = \begin{bmatrix} \mathfrak{M}_- & 0 \\ 0 & \mathfrak{M}_+ \end{bmatrix} \text{ and } T^{\mathrm{T}}\mathfrak{L}T = \begin{bmatrix} \mathfrak{L}_- & 0 \\ 0 & \mathfrak{L}_+ \end{bmatrix},$$

where $\sigma(-\mathfrak{L}_-, \mathfrak{M}_-) \subset \mathbb{C}_-$ and $\sigma(-\mathfrak{L}_+, \mathfrak{M}_+) \subset \overline{\mathbb{C}_+}$. We see that if

$$T = \begin{bmatrix} T_- & \mathbb{1}_{\mathfrak{n}} \end{bmatrix},$$

then

$$T^{\mathrm{T}}\mathfrak{M}T = \begin{bmatrix} T_-^{\mathrm{T}}\mathfrak{M}T_- & T_-^{\mathrm{T}}\mathfrak{M}\mathbb{1}_{\mathfrak{n}} \\ \mathbb{1}_{\mathfrak{n}}^{\mathrm{T}}\mathfrak{M}T_- & \mathbb{1}_{\mathfrak{n}}^{\mathrm{T}}\mathfrak{M}\mathbb{1}_{\mathfrak{n}} \end{bmatrix} \text{ and } T^{\mathrm{T}}\mathfrak{L}T = \begin{bmatrix} T_-^{\mathrm{T}}\mathfrak{L}T_- & 0 \\ 0 & 0 \end{bmatrix}.$$

To have $T_-^{\mathrm{T}}\mathfrak{M}\mathbb{1}_{\mathfrak{n}} = 0$, we need for the columns of T_- to be orthogonal to $\mathfrak{M}\mathbb{1}_{\mathfrak{n}}$. Additionally, T_- should be such that both $T_-^{\mathrm{T}}\mathfrak{M}T_-$ and $T_-^{\mathrm{T}}\mathfrak{L}T_-$ are sparse. We choose

$$T_- = \begin{bmatrix} \alpha_1 & & & \\ -\beta_1 & \alpha_2 & & \\ & -\beta_2 & \ddots & \\ & & \ddots & \alpha_{\mathfrak{n}-1} \\ & & & -\beta_{\mathfrak{n}-1} \end{bmatrix}$$

for some $\alpha_i, \beta_i > 0$, $i = 1, 2, \ldots, \mathfrak{n}-1$, which we determine next. From $e_i^{\mathrm{T}}T_-^{\mathrm{T}}\mathfrak{M}\mathbb{1}_{\mathfrak{n}} = 0$, we find $\alpha_i\mathfrak{m}_i = \beta_i\mathfrak{m}_{i+1}$. If we additionally set $\alpha_i^2 + \beta_i^2 = 1$, we get

$$\alpha_i = \frac{\mathfrak{m}_{i+1}}{\sqrt{\mathfrak{m}_i^2 + \mathfrak{m}_{i+1}^2}} \text{ and } \beta_i = \frac{\mathfrak{m}_i}{\sqrt{\mathfrak{m}_i^2 + \mathfrak{m}_{i+1}^2}}.$$

Algorithm 3.2: QR decomposition with column pivoting for matrices with block-columns

Input: Matrix $X \in \mathbb{R}^{n\mathfrak{r}\times n\mathfrak{n}}$ of full rank and $\mathfrak{r} < \mathfrak{n}$.

Output: Orthogonal matrix Q, upper-triangular matrix R, and permutation matrix P such that $XP = QR$

1 Denote $X = \begin{bmatrix} X_1 & X_2 & \cdots & X_{\mathfrak{n}} \end{bmatrix}$, where $X_i \in \mathbb{R}^{n\mathfrak{r}\times n}$.
2 Find a block-column X_i with the largest Frobenius norm and swap it with X_1.
3 Perform QR decomposition with column pivoting on X_1, i.e. find an orthogonal $Q_1 \in \mathbb{R}^{n\mathfrak{r}\times n\mathfrak{r}}$, an upper-triangular $R_1 \in \mathbb{R}^{n\mathfrak{r}\times n}$, and a permutation matrix $P_1 \in \mathbb{R}^{n\times n}$ such that $X_1P_1 = Q_1R_1$.
4 Multiply all block-columns in X from the right by P_1.
5 Multiply X from the left by Q_1^{T}.
6 Repeat the procedure for $X_{n+1:n\mathfrak{r},n+1:n\mathfrak{n}}$, which computes the matrices Q_i, R_i, and P_i, for $i \in \{2, 3, \ldots, \mathfrak{r}\}$.
7 Return $Q = Q_1Q_2\cdots Q_{\mathfrak{r}}$, $R = X$, and P with all of the column permutations recorded.

3.2.5 Extension to higher-order agents

For multi-agent systems (2.9) with agents of order n, the constraints (3.2) become

$$\operatorname{im}(V) = \operatorname{im}(\mathfrak{P} \otimes I_n)\,, \tag{3.5a}$$
$$\operatorname{im}(W) = \operatorname{im}(\mathfrak{P} \otimes I_n)\,. \tag{3.5b}$$

QR decomposition-based clustering can then be extended as in Algorithm 3.2 by clustering the block-columns of V^{T} (or W^{T}). For the k-means algorithm, we can show in a similar way as in the single-integrator case that clustering the block-rows leads to minimizing an upper bound of the largest principal angle. Therefore, k-means can be directly applied to the vectorized block-rows of V or W.

Additionally, we need to consider synchronization preservation. In the single-integrator case, clustering using any partition preserves synchronization. In the general case, using Proposition 2.51, we need that $(A - \widehat{\lambda}, E)$ is Hurwitz for all nonzero eigenvalues $\widehat{\lambda}$ of $(\widehat{\mathfrak{L}}, \widehat{\mathfrak{M}})$. This is true for AEP because $\sigma(\widehat{\mathfrak{L}}, \widehat{\mathfrak{M}})$ is a subset of $\sigma(\mathfrak{L}, \mathfrak{M})$. Note that if $(A - \lambda, E)$ is Hurwitz for all $\lambda \in [\lambda_2, \ldots, \lambda_{\mathfrak{n}}]$, where $0 = \lambda_1 < \lambda_2 \leqslant \cdots \leqslant \lambda_{\mathfrak{n}}$ are the eigenvalues of $(\mathfrak{L}, \mathfrak{M})$, then from Theorem 2.6 we get that every partition preserves synchronization.

3.2.6 Numerical example

We use the example from [MTC14]. It is a leader-follower multi-agent system, defined on a undirected, weighted, connected graph with 10 vertices, shown in Figure 3.1. Agents 6 and 7 are the leaders of the multi-agent system. The Laplacian and input

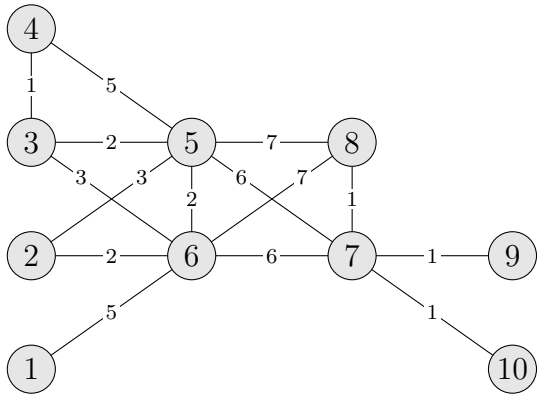

Figure 3.1: The undirected, weighted graph from [MTC14]

matrices are

$$\mathfrak{L} = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 & -5 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & -3 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & -1 & -2 & -3 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 6 & -5 & 0 & 0 & 0 & 0 & 0 \\ 0 & -3 & -2 & -5 & 25 & -2 & -6 & -7 & 0 & 0 \\ -5 & -2 & -3 & 0 & -2 & 25 & -6 & -7 & 0 & 0 \\ 0 & 0 & 0 & 0 & -6 & -6 & 15 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & -7 & -7 & -1 & 15 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathfrak{B} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

We chose an edge ordering and orientation such that the incidence and edge-weights matrices are

$$\mathfrak{R} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and $\mathfrak{W} = \operatorname{diag}(5, 3, 2, 1, 2, 3, 5, 2, 6, 7, 6, 7, 1, 1, 1)$.

For this example, we focus on partitions with five clusters. There are in total $42,525$ such partitions. Table 3.1 shows the best 20 partitions of size 5 and their relative $\mathcal{H}_2$-errors. Figure 3.2 shows all relative $\mathcal{H}_2$-errors for all partitions with five clusters.

Table 3.1: Top 20 partitions with 5 clusters for reducing the multi-agent system in Figure 3.1

Rank	Relative $\mathcal{H}_2$-error	Partition
1	0.128053	$\{\{1,8\},\{2,3,4,9,10\},\{5\},\{6\},\{7\}\}$
2	0.131311	$\{\{1,2,3,4\},\{5,8\},\{6\},\{7\},\{9,10\}\}$
3	0.137466	$\{\{1,2,3,4,9,10\},\{5\},\{6\},\{7\},\{8\}\}$
4	0.137473	$\{\{1,3,8\},\{2,4,9,10\},\{5\},\{6\},\{7\}\}$
5	0.143700	$\{\{1,5,8\},\{2,3,4\},\{6\},\{7\},\{9,10\}\}$
6	0.145900	$\{\{1,2,3\},\{4,9,10\},\{5,8\},\{6\},\{7\}\}$
7	0.146196	$\{\{1,8\},\{2,3,4,9\},\{5,10\},\{6\},\{7\}\}$
8	0.146196	$\{\{1,8\},\{2,3,4,10\},\{5,9\},\{6\},\{7\}\}$
9	0.147022	$\{\{1,2,3,8\},\{4,9,10\},\{5\},\{6\},\{7\}\}$
10	0.149240	$\{\{1,8,9\},\{2,3,4,10\},\{5\},\{6\},\{7\}\}$
11	0.149240	$\{\{1,8,10\},\{2,3,4,9\},\{5\},\{6\},\{7\}\}$
12	0.149654	$\{\{1,8\},\{2,4,9,10\},\{3,5\},\{6\},\{7\}\}$
13	0.150440	$\{\{1,5\},\{2,3,4,9,10\},\{6\},\{7\},\{8\}\}$
14	0.150654	$\{\{1,3\},\{2,4,9,10\},\{5,8\},\{6\},\{7\}\}$
15	0.151684	$\{\{1,2,8\},\{3,4,9,10\},\{5\},\{6\},\{7\}\}$
16	0.153100	$\{\{1,2,3,4,9\},\{5,8\},\{6\},\{7\},\{10\}\}$
17	0.153100	$\{\{1,2,3,4,10\},\{5,8\},\{6\},\{7\},\{9\}\}$
18	0.153819	$\{\{1\},\{2,3,4,9,10\},\{5,8\},\{6\},\{7\}\}$
19	0.154374	$\{\{1,3,8,9\},\{2,4,10\},\{5\},\{6\},\{7\}\}$
20	0.154374	$\{\{1,3,8,10\},\{2,4,9\},\{5\},\{6\},\{7\}\}$

The partition

$$\{\{1,2,3,4\},\{5,6\},\{7\},\{8\},\{9,10\}\} \tag{3.6}$$

is an AEP of the graph in Figure 3.1 (see Figure 3.3) used in [MTC14]. It has five clusters, so we compare the relative $\mathcal{H}_2$-errors of different ROMs of order $\mathfrak{r} = 5$. The partition (3.6) is actually the only AEP with five clusters, among the total of five AEPs of the graph in Figure 3.1.

From [MTC14, Theorem 6], it follows that for the AEP in (3.6), the relative $\mathcal{H}_2$-error is

$$\frac{\|H - \widehat{H}_{\text{AEP}}\|_{\mathcal{H}_2}}{\|H\|_{\mathcal{H}_2}} = \sqrt{\frac{\left(1-\frac{1}{2}\right)+\left(1-\frac{1}{1}\right)}{2\left(1-\frac{1}{10}\right)}} \approx 0.527046, \tag{3.7}$$

where $\widehat{H}_{\text{AEP}}$ is the transfer function of the ROM using the graph partition (3.6).

Next, we used IRKA to find a ROM of order $\mathfrak{r} = 5$. It found a ROM with relative

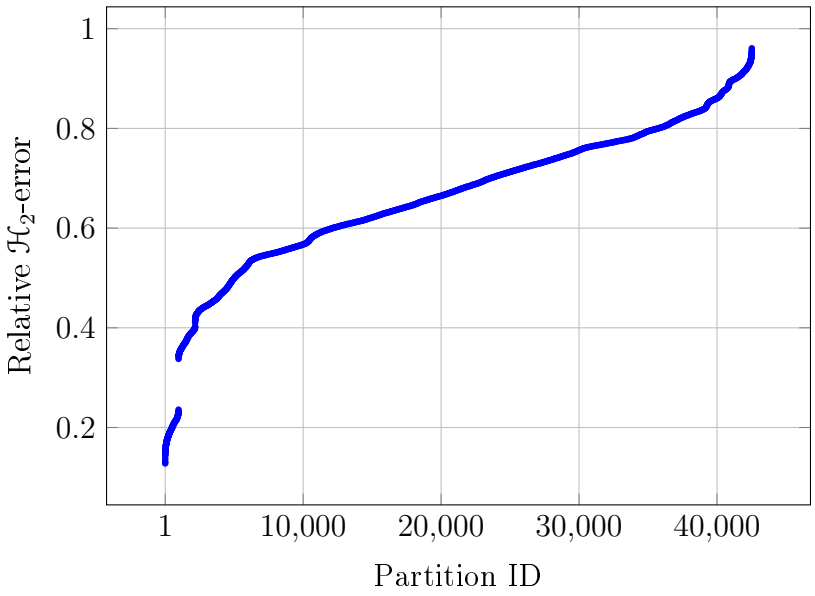

Figure 3.2: Relative $\mathcal{H}_2$-errors for all partitions with five clusters

$\mathcal{H}_2$-error of

$$\frac{\|H - \widehat{H}_{\text{IRKA}}\|_{\mathcal{H}_2}}{\|H\|_{\mathcal{H}_2}} \approx 0.0330412, \tag{3.8}$$

which is almost 16 times better than (3.7).

The partition resulting from Algorithm 3.1 applied to IRKA's V matrix is

$$\{\{1,3\},\{2,4,9,10\},\{5,8\},\{6\},\{7\}\}, \tag{3.9}$$

and is shown in Figure 3.4. The relative $\mathcal{H}_2$-error of a ROM using partition (3.9) is

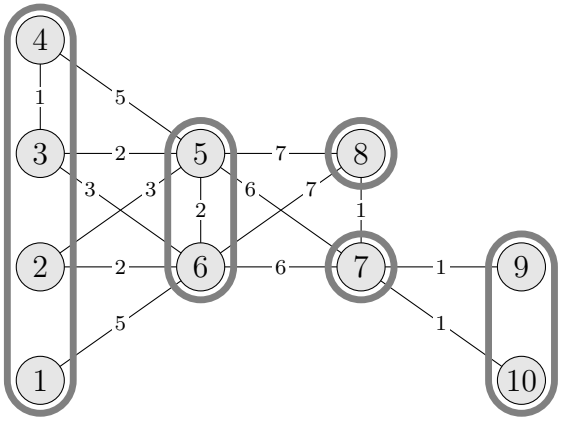

Figure 3.3: Graph with partition

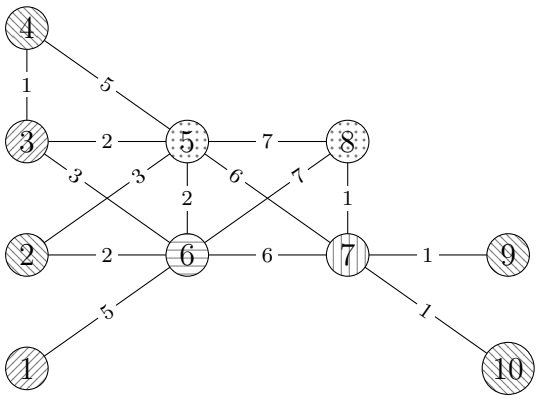

Figure 3.4: Graph partition (3.9) with five clusters. The assignment of vertices to clusters is represented with different patterns.

$$\frac{\|H - \widehat{H}_V\|_{\mathcal{H}_2}}{\|H\|_{\mathcal{H}_2}} \approx 0.150654, \tag{3.10}$$

which is more than 4 times worse than (3.8), but 3 times better than (3.7).

We notice by (3.2b) that W can also be used to find a good partition. In this example, Algorithm 3.1 returns the partition

$$\{\{1, 2, 3, 9, 10\}, \{4, 8\}, \{5\}, \{6\}, \{7\}\}. \tag{3.11}$$

The relative $\mathcal{H}_2$-error when using the partition (3.11) is

$$\frac{\|H - \widehat{H}_W\|_{\mathcal{H}_2}}{\|H\|_{\mathcal{H}_2}} \approx 0.179746,$$

which is worse than (3.10).

We notice that partition (3.9) is the 14th partition and that the best partition produces about 1.18 times better error. Table 3.1 does not show that partition (3.11) is 192nd and partition (3.6) is 5996th. Thus, the new method gets a lot closer to the optimal partition than the AEP in this example. Using the k-means algorithm to cluster the rows of V and W results in the $6th$ and 27th partition, respectively, which is in this case an improvement over QR decomposition-based clustering.

Furthermore, from Table 3.1 we see that, in all of the top 20 partitions, leaders 6 and 7 appear in singletons. This makes sense, because this way no input is diffused over more agents. However, further research is needed to see if, in general, the best partition has leaders appearing in singletons.

3.3 Clustering for nonlinear multi-agent systems

In this section, we extend the approach from the previous section to certain nonlinear multi-agent systems. We describe the class of multi-agent systems in Section 3.3.1. Next, in Section 3.3.2, we extend clustering by projection to this class of systems.

3.3.1 Nonlinear multi-agent systems

Here, we consider a class of nonlinear multi-agent systems. In particular, let the dynamics of the ith agent, for $i = 1, 2, \ldots, \mathfrak{n}$, be

$$\begin{aligned} \dot{x}_i(t) &= f(x_i(t)) + g(x_i(t))v_i(t), \\ z_i(t) &= h(x_i(t)), \end{aligned}$$

with functions $f\colon \mathbb{R}^n \to \mathbb{R}^n$, $g\colon \mathbb{R}^n \to \mathbb{R}^{n\times m}$, $h\colon \mathbb{R}^n \to \mathbb{R}^{p\times n}$, state $x_i(t) \in \mathbb{R}^n$, input $v_i(t) \in \mathbb{R}^m$, and output $z_i(t) \in \mathbb{R}^p$. The interconnections are

$$v_i(t) = \sum_{j=1}^{\mathfrak{n}} \mathfrak{a}_{ij} k(z_j(t) - z_i(t)) + \sum_{\ell=1}^{\mathfrak{m}} \mathfrak{b}_{i\ell} u_\ell(t),$$

for $i = 1, 2, \ldots, \mathfrak{n}$, with coupling $k\colon \mathbb{R}^p \to \mathbb{R}^m$, external input $u_\ell(t) \in \mathbb{R}^m$, $\ell = 1, 2, \ldots, \mathfrak{m}$, where $\mathfrak{A} = [\mathfrak{a}_{ij}]$ is the adjacency matrix of the graph $\mathfrak{G}$, and $\mathfrak{B} = [\mathfrak{b}_{i\ell}]$. Additionally, let the external output be

$$y_i(t) = \sum_{j=1}^{\mathfrak{n}} \mathfrak{c}_{ij} z_j(t),$$

with $\mathfrak{C} = [\mathfrak{c}_{ij}]$. We assume functions f, g, h, k are continuous and that there is a unique global solution $x(t) = \operatorname{col}(x_1(t), x_2(t), \ldots, x_{\mathfrak{n}}(t))$ for any admissible $u(t)$.

3.3.2 Clustering by projection

We want to find the form of the reduced order model obtained from Galerkin projection with $V = \mathfrak{P} \otimes I_n$. We can write

$$\begin{aligned} \dot{x}(t) &= F(x(t), u(t)), \\ y(t) &= G(x(t)), \end{aligned}$$

for some functions F and G. The reduced model is

$$\begin{aligned} \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P} \otimes I_n\right) \dot{\widehat{x}}(t) &= \left(\mathfrak{P}^{\mathrm{T}} \otimes I_n\right) F((\mathfrak{P} \otimes I_n)\widehat{x}(t), u(t)), \\ \widehat{y}(t) &= G((\mathfrak{P} \otimes I_n)\widehat{x}(t)). \end{aligned}$$

Premultiplying the first equation with $e_p^{\mathrm{T}} \otimes I_n$, we find

$$\begin{aligned}
|\mathfrak{C}_p|\,\dot{\widehat{x}}_p(t) &= \sum_{i\in\mathfrak{C}_p}\left(f(\widehat{x}_p(t)) + g(\widehat{x}_p(t))\left(\sum_{j=1}^{\mathfrak{n}}\mathfrak{a}_{ij}k(h(\widehat{x}_{C(j)}(t)) - h(\widehat{x}_p(t))) + \sum_{\ell=1}^{\mathfrak{m}}\mathfrak{b}_{i\ell}u_\ell(t)\right)\right)\\
&= |\mathfrak{C}_p|\,f(\widehat{x}_p(t))\\
&\quad + g(\widehat{x}_p(t))\left(\sum_{i\in\mathfrak{C}_p}\sum_{j=1}^{\mathfrak{n}}\mathfrak{a}_{ij}k(h(\widehat{x}_{C(j)}(t)) - h(\widehat{x}_p(t))) + \sum_{i\in\mathfrak{C}_p}\sum_{\ell=1}^{\mathfrak{m}}\mathfrak{b}_{i\ell}u_\ell(t)\right)\\
&= |\mathfrak{C}_p|\,f(\widehat{x}_p(t))\\
&\quad + g(\widehat{x}_p(t))\left(\sum_{q=1}^{\mathfrak{r}}\sum_{i\in\mathfrak{C}_p}\sum_{j\in\mathfrak{C}_q}\mathfrak{a}_{ij}k(h(\widehat{x}_q(t)) - h(\widehat{x}_p(t))) + \sum_{\ell=1}^{\mathfrak{m}}\sum_{i\in\mathfrak{C}_p}\mathfrak{b}_{i\ell}u_\ell(t)\right)\\
&= |\mathfrak{C}_p|\,f(\widehat{x}_p(t))\\
&\quad + g(\widehat{x}_p(t))\left(\sum_{q=1}^{\mathfrak{r}}\widehat{\mathfrak{a}}_{pq}k(h(\widehat{x}_q(t)) - h(\widehat{x}_p(t))) + \sum_{\ell=1}^{\mathfrak{m}}\widehat{\mathfrak{b}}_{p\ell}u_\ell(t)\right),
\end{aligned}$$

for

$$\widehat{\mathfrak{a}}_{pq} = \sum_{i\in\mathfrak{C}_p}\sum_{j\in\mathfrak{C}_q}\mathfrak{a}_{ij}, \quad \widehat{\mathfrak{b}}_{p\ell} = \sum_{i\in\mathfrak{C}_p}\mathfrak{b}_{i\ell}.$$

Defining $\widehat{\mathfrak{A}} \coloneqq [\widehat{\mathfrak{a}}_{pq}]$ and $\widehat{\mathfrak{B}} \coloneqq [\widehat{\mathfrak{b}}_{p\ell}]$, we see that $\widehat{\mathfrak{A}} = \mathfrak{P}^{\mathrm{T}}\mathfrak{A}\mathfrak{P}$ and $\widehat{\mathfrak{B}} = \mathfrak{P}^{\mathrm{T}}\mathfrak{B}$. For the output we have

$$\widehat{y}_i(t) = \sum_{j=1}^{\mathfrak{n}}\mathfrak{c}_{ij}h(\widehat{x}_{C(j)}(t)) = \sum_{j=1}^{\mathfrak{r}}\sum_{j\in\mathfrak{C}_q}\mathfrak{c}_{ij}h(\widehat{x}_q(t)) = \sum_{j=1}^{\mathfrak{r}}\widehat{\mathfrak{c}}_{iq}h(\widehat{x}_q(t)),$$

where

$$\widehat{\mathfrak{c}}_{iq} = \sum_{j\in\mathfrak{C}_q}\mathfrak{c}_{ij}.$$

Thus, for $\widehat{\mathfrak{C}} \coloneqq [\widehat{\mathfrak{c}}_{iq}]$, we have $\widehat{\mathfrak{C}} = \mathfrak{C}\mathfrak{P}$. Therefore, we showed how to construct a ROM of the same structure as the original multi-agent system. Based on this, to find a good partition, we can apply any projection-based MOR method for nonlinear systems (e.g., proper orthogonal decomposition [HV05]) and cluster the (block-)rows of the matrix used to project the system.

3.4 Conclusions

In Section 3.2, we presented our method, combining IRKA and a clustering algorithm, for MOR of multi-agent systems using graph partitions. It seems heuristically that

this method is able to create a partition whose $\mathcal{H}_2$-error is small. We furthermore elaborated that this method is scalable to large-scale systems. Our numerical test for a small network shows that, among 42 525 partitions, our algorithm found the 14th best approximation whose error is of the same order of magnitude as the error of the best partition. A theoretical foundation that the algorithm always finds a partition in some sense close to optimal remains an open problem for future work.

In Section 3.3, we showed how to extend the approach to nonlinear multi-agent systems by combining a projection-based method and a clustering algorithm.

CHAPTER 4

GRAPH SYMMETRIES AND EQUITABLE PARTITIONS IN CLUSTERING-BASED MODEL ORDER REDUCTION

Contents

4.1 Introduction

Here, we consider theoretical aspects of clustering-based MOR, using graph symmetries and equitable partitions.

The work in [MTC14] showed upper bounds for $\mathcal{H}_2$ model reduction error using AEP for leader-follower multi-agent systems with single-integrator agents. In Section 4.2, we extend these results to more general agents, including also upper bounds for $\mathcal{H}_\infty$ error and arbitrary partitions.

The results of [RJME09] on clustering using graph symmetries and equitable partitions for linear mutli-agent systems are extended in Section 4.3 nonlinear power systems.

4.2 Error bounds for clustering-based model order reduction of linear multi-agent systems

4.2.1 Introduction

In [MTC14], MOR by clustering was put in the context of MOR by Petrov-Galerkin projection. The results in [MTC14] provide explicit expressions for the $\mathcal{H}_2$ model reduction error if a leader-follower network with *single integrator agent dynamics* is clustered using an almost equitable partition of the graph. Here, our aim is to generalize and extend the results in [MTC14] to networks where the agent dynamics is given by an *arbitrary multivariable input-state-output system*. We also aim at finding explicit formulas and a priori upper bounds for the model reduction error measured in the $\mathcal{H}_\infty$-norm. Finally, we will consider the problem of clustering a network according to arbitrary, not necessarily almost equitable, graph partitions. The main contributions of this section are the following:

1. We derive an a priori upper bound for the $\mathcal{H}_2$ model reduction error for the case that the agents are represented by an arbitrary input-state-output system.

2. We extend the results in [MTC14] for single integrator dynamics by giving an explicit expression for the $\mathcal{H}_\infty$ model reduction error in terms of properties of the given graph partition.

3. We establish an a priori upper bound for the $\mathcal{H}_\infty$ model reduction error for the case that the agents are represented by an arbitrary but *symmetric* input-state-output system.

4. We establish some preliminary results on the model reduction error in case of clustering using an arbitrary, possibly non almost equitable, partition.

The outline of this section is as follows. In Section 4.2.3, we formulate our problem of MOR of leader-follower multi-agent networks. Section 4.2.4 reviews some theory on graph partitions and MOR by clustering, and relates this method to Petrov-Galerkin

projection of the original network. Also preservation of synchronization is discussed here. In Section 4.2.5, we provide a priori error bounds on the $\mathcal{H}_2$ model reduction error for networks with arbitrary agent dynamics, clustered using almost equitable partitions. In Section 4.2.6, we complement these results by providing upper bounds on the $\mathcal{H}_\infty$ model reduction error. In Section 4.2.7, the problem of clustering networks according to general partitions is considered and the first steps towards a priori error bounds on both the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ model reduction errors are made. Numerical examples for which we compare the actual errors with the a priori bounds established here are presented in Section 4.2.8.

4.2.2 Preliminaries

In this section, we briefly discuss some basic facts on finite-dimensional linear systems and how they extend to unstable systems.

Consider the input-state-output system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t), \end{aligned} \tag{4.1}$$

with $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}^p$, and transfer function $H(s) = C(sI - A)^{-1}B$. If A is Hurwitz, then the $\mathcal{H}_2$-norm can be computed as

$$\|H\|_{\mathcal{H}_2}^2 = \operatorname{tr}\left(B^{\mathrm{T}} X B\right),$$

where X is the unique positive semi-definite solution of the Lyapunov equation

$$A^{\mathrm{T}} X + XA + C^{\mathrm{T}} C = 0. \tag{4.2}$$

For the purposes of the work in this part, we also need to deal with the situation when A is not Hurwitz. Let $\mathcal{X}_+(A)$ denote the unstable subspace of A, i.e., the direct sum of the generalized eigenspaces of A corresponding to its eigenvalues in the closed right half plane. We state the following proposition.

Proposition 4.1:
Assume that $\mathcal{X}_+(A) \subseteq \ker(C)$. Then the Lyapunov equation (4.2) has at least one positive semi-definite solution. Among all positive semi-definite solutions, there is exactly one solution, say X, with the property $\mathcal{X}_+(A) \subseteq \ker(X)$. For this particular solution X, we have $\|H\|_{\mathcal{H}_2}^2 = \operatorname{tr}\left(B^{\mathrm{T}} X B\right)$. ◊

Proof. Without loss of generality, assume that

$$A = \begin{bmatrix} A_- & 0 \\ 0 & A_+ \end{bmatrix}, \quad B = \begin{bmatrix} B_- \\ B_+ \end{bmatrix}, \quad C = \begin{bmatrix} C_- & 0 \end{bmatrix},$$

where A_- is Hurwitz, and A_+ has all its eigenvalues in the closed right half plane. Let X_- be the unique solution to the reduced Lyapunov equation

$$A_-^{\mathrm{T}} X_- + X_- A_- + C_-^{\mathrm{T}} C_- = 0. \tag{4.3}$$

Then $X_- = \int_0^\infty e^{A_-^{\mathrm{T}} t} C_-^{\mathrm{T}} C_- e^{A_- t}\,\mathrm{d}t \succcurlyeq 0$. Obviously then, $X = \operatorname{diag}(X_-, 0)$ is a positive semi-definite solution of (4.2). Now, let X be a positive semi-definite solution to (4.2) with the property that $\mathcal{X}_+(A) \subseteq \ker(X)$. Then X must be of the form $X = \operatorname{diag}(X_1, 0)$, and X_1 must satisfy the reduced Lyapunov equation (4.3). Thus $X = \operatorname{diag}(X_-, 0)$. Finally, H is asymptotically stable since $\mathcal{X}_+(A) \subseteq \ker(C)$. Moreover,

$$\begin{aligned}
\|H\|_{\mathcal{H}_2}^2 &= \operatorname{tr}\left(B^{\mathrm{T}} \int_0^\infty e^{A^{\mathrm{T}} t} C^{\mathrm{T}} C e^{At}\,\mathrm{d}t\; B\right) \\
&= \operatorname{tr}\left(B_-^{\mathrm{T}} \int_0^\infty e^{A_-^{\mathrm{T}} t} C_-^{\mathrm{T}} C_- e^{A_- t}\,\mathrm{d}t\; B_-\right) \\
&= \operatorname{tr}\left(B_-^{\mathrm{T}} X_- B_-\right) \\
&= \operatorname{tr}\left(B^{\mathrm{T}} X B\right).
\end{aligned}$$

□

We will now deal with computing the $\mathcal{H}_\infty$-norm. The result is a generalization of Lemma 4 in [IKIA14].

Lemma 4.2:
Consider the system (4.1). Assume that its transfer function H has all its poles in the open left half plane. If there exists $X \in \mathbb{R}^{p\times p}$ such that $X = X^{\mathrm{T}}$ and $CA = XC$, then $\|H\|_{\mathcal{H}_\infty} = \|H(0)\|_2$. ◊

Proof. For the first part of the proof, let us assume that (A, B, C) is minimal. Then, in particular, A is Hurwitz and (A, B) is controllable.

Clearly, the inequality $\|S\|_{\mathcal{H}_\infty} \geqslant \|S(0)\|_2$ is always satisfied. We will prove that $\|S\|_{\mathcal{H}_\infty} \leqslant \|S(0)\|_2$ using the bounded real lemma [Ran96], which states that $\|S\|_{\mathcal{H}_\infty} \leqslant \gamma$ if and only if there exists $P \in \mathbb{R}^{n\times n}$ such that $P = P^{\mathrm{T}}$ and

$$A^{\mathrm{T}} P + PA + C^{\mathrm{T}} C + \frac{1}{\gamma^2} PBB^{\mathrm{T}} P \preccurlyeq 0.$$

Let us take $\gamma = \|S(0)\|_2 = \|CA^{-1}B\|_2$. This implies that

$$CA^{-1}BB^{\mathrm{T}} A^{-\mathrm{T}} C^{\mathrm{T}} \preccurlyeq \gamma^2 I_p. \tag{4.4}$$

Defining $P \coloneqq -A^{-\mathrm{T}} C^{\mathrm{T}} X C A^{-1}$ and using (4.4) yields

$$\begin{aligned}
&A^{\mathrm{T}} P + PA + C^{\mathrm{T}} C + \frac{1}{\gamma^2} PBB^{\mathrm{T}} P \\
&\quad = -C^{\mathrm{T}} X C A^{-1} - A^{-\mathrm{T}} C^{\mathrm{T}} X C + C^{\mathrm{T}} C \\
&\qquad + \frac{1}{\gamma^2} A^{-\mathrm{T}} C^{\mathrm{T}} X C A^{-1} BB^{\mathrm{T}} A^{-\mathrm{T}} C^{\mathrm{T}} X C A^{-1} \\
&\quad \preccurlyeq -C^{\mathrm{T}} X C A^{-1} - A^{-\mathrm{T}} C^{\mathrm{T}} X C + C^{\mathrm{T}} C + A^{-\mathrm{T}} C^{\mathrm{T}} X X C A^{-1} \\
&\quad = \left(XCA^{-1} - C\right)^{\mathrm{T}} \left(XCA^{-1} - C\right) \\
&\quad = 0.
\end{aligned}$$

From the bounded real lemma, we conclude that $\|S\|_{\mathcal{H}_\infty} \leqslant \|S(0)\|_2$.

For a non-minimal representation (A, B, C), applying the Kalman decomposition, let T be a invertible matrix such that

$$T^{-1}AT = \begin{bmatrix} A_1 & 0 & A_6 & 0 \\ A_2 & A_3 & A_4 & A_5 \\ 0 & 0 & A_7 & 0 \\ 0 & 0 & A_8 & A_9 \end{bmatrix}, \quad T^{-1}B = \begin{bmatrix} B_1 \\ B_2 \\ 0 \\ 0 \end{bmatrix}, \quad CT = \begin{bmatrix} C_1 & 0 & C_2 & 0 \end{bmatrix},$$

where (A_1, B_1, C_1) is a minimal representation of (A, B, C) with A_1 Hurwitz. From,

$$\begin{aligned} (CT)\left(T^{-1}AT\right) &= CAT = XCT = X(CT), \\ (CT)\left(T^{-1}AT\right) &= \begin{bmatrix} C_1 & 0 & C_2 & 0 \end{bmatrix} \begin{bmatrix} A_1 & 0 & A_6 & 0 \\ A_2 & A_3 & A_4 & A_5 \\ 0 & 0 & A_7 & 0 \\ 0 & 0 & A_8 & A_9 \end{bmatrix} \\ &= \begin{bmatrix} C_1A_1 & 0 & C_1A_6 + C_2A_7 & 0 \end{bmatrix}, \\ X(CT) = X \begin{bmatrix} C_1 & 0 & C_2 & 0 \end{bmatrix} &= \begin{bmatrix} XC_1 & 0 & XC_2 & 0 \end{bmatrix}, \end{aligned}$$

we find that $C_1A_1 = XC_1$. Therefore, the minimal representation satisfies the sufficient condition and using the result obtained above the proof is completed. □

Continuing our effort to compute the $\mathcal{H}_\infty$-norm, we now formulate a lemma that will be instrumental in evaluating a transfer function at the origin. Recall that for a given matrix A, its Moore-Penrose inverse is denoted by A^+.

Lemma 4.3:
Consider the system (4.1). If A is symmetric and $\ker(A) \subseteq \ker(C)$, then 0 is not a pole of the transfer function H and we have $H(0) = -CA^+B$. ◊

Proof. If A is invertible, then the conclusion follows immediately. Otherwise, let $A = U\Lambda U^{\mathrm{T}}$ be an eigenvalue decomposition with orthogonal U and $\Lambda = \operatorname{diag}(0, \Lambda_2)$, where $\Lambda_2 \in \mathbb{R}^{r\times r}$ and r is the rank of A. We denote $U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$, with $U_2 \in \mathbb{R}^{n\times r}$. Then

$$A^+ = U\Lambda^+U^{\mathrm{T}} = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_2^{-1} \end{bmatrix} \begin{bmatrix} U_1^{\mathrm{T}} \\ U_2^{\mathrm{T}} \end{bmatrix} = U_2\Lambda_2^{-1}U_2^{\mathrm{T}}.$$

Note that $CU_1 = 0$. We have

$$\begin{aligned} H(s) &= CU(sI - \Lambda)^{-1}U^{\mathrm{T}}B \\ &= C \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} s^{-1}I & 0 \\ 0 & (sI - \Lambda_2)^{-1} \end{bmatrix} \begin{bmatrix} U_1^{\mathrm{T}} \\ U_2^{\mathrm{T}} \end{bmatrix} B \\ &= CU_2(sI - \Lambda_2)^{-1}U_2^{\mathrm{T}}B. \end{aligned}$$

Hence, $H(s)$ is defined at $s = 0$ and $H(0) = -CU_2\Lambda_2^{-1}U_2^{\mathrm{T}}B = -CA^+B$. □

4.2.3 Problem formulation

We consider leader-follower multi-agent systems of the form

$$\dot{x}(t) = (I_{\mathfrak{n}} \otimes A - \mathfrak{L} \otimes B)x(t) + (\mathfrak{B} \otimes F)u(t), \tag{4.5a}$$
$$y(t) = (\mathfrak{L} \otimes I_n)x(t), \tag{4.5b}$$

defined on an undirected, weighted, connected graph $\mathfrak{G} = (\mathfrak{V}, \mathfrak{E}, \mathfrak{w})$ with vertex set $\mathfrak{V} = \{1, 2, \ldots, \mathfrak{n}\}$, adjacency matrix $\mathfrak{A}$, Laplacian matrix $\mathfrak{L}$, leader set $\mathfrak{V}_{\mathrm{L}} = \{\mathfrak{v}_1, \mathfrak{v}_2, \ldots, \mathfrak{v}_{\mathfrak{m}}\} \subseteq \mathfrak{V}$, and follower set $\mathfrak{V}_{\mathrm{F}} = \mathfrak{V} \setminus \mathfrak{V}_{\mathrm{L}}$. Furthermore, $x_i(t) \in \mathbb{R}^n$ is the state of agent i, and $u_\ell(t) \in \mathbb{R}^m$ is the external input to the leader v_ℓ. Finally, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times n}$, and $F \in \mathbb{R}^{n\times m}$ are real matrices. Denote $x(t) = \mathrm{col}(x_1(t), x_2(t), \ldots, x_{\mathfrak{n}}(t))$ and $u(t) = \mathrm{col}(u_1(t), u_2(t), \ldots, u_{\mathfrak{m}}(t))$.

The goal of this section is to find a reduced order networked system, whose dynamics is a good approximation of the networked system (4.5a). Following [MTC14], the idea to obtain such an approximation is to *cluster* groups of agents in the network, and to treat each of the resulting clusters as a vertex in a new, reduced order, network. The reduced order network will again be a leader-follower network, and by the clustering procedure, essential interconnection features of the network will be preserved. We will also require that the *synchronization* properties of the network are preserved after reduction. We assume that the original network is synchronized, meaning that if the external inputs satisfy $u_\ell = 0$ for $\ell = 1, 2, \ldots, \mathfrak{m}$, then for all $i, j \in \mathfrak{V}$, we have

$$x_i(t) - x_j(t) \to 0$$

as $t \to \infty$. We impose that the reduction procedure preserves this property.

Being a measure for the disagreement between the states of the agents in (4.5a), we choose $y(t) = (\mathfrak{L} \otimes I_n)x(t)$ as the output of the original network. Indeed, this output y can be considered a measure of the disagreement in the network, in the sense that $y(t)$ is small if and only if the network is close to being synchronized.

Note that the state space dimension of (4.5) is equal to $\mathfrak{n}n$, its number of inputs equals $\mathfrak{m}m$, and the number of outputs is $\mathfrak{n}n$.

Here, we will use clustering to obtain a reduced order network, i.e., a network with a reduced number of agents, as an approximation of the original network (4.5).

4.2.4 Graph partitions and reduction by clustering

We consider networks whose interaction topologies are represented by weighted graphs $\mathfrak{G}$ with vertex set $\mathfrak{V}$. The graph of the original network (4.5a) is undirected, however, our reduction procedure will lead to networks on directed graphs.

Next, we will construct a reduced order approximation of (4.5) by clustering the agents in the network using a partition of $\mathfrak{G}$. Extending the main idea in [MTC14], we take as reduced order system the Petrov-Galerkin projection of the original system (4.5), with the following choice for the matrices V and W:

$$W = \mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1} \otimes I_n \in \mathbb{R}^{\mathfrak{n}n\times\mathfrak{t}n}, \quad V = \mathfrak{P} \otimes I_n \in \mathbb{R}^{\mathfrak{n}n\times\mathfrak{t}n}.$$

The dynamics of the resulting reduced order model is then given by

$$\begin{aligned}\dot{\widehat{x}}(t) &= (I_{\mathfrak{r}} \otimes A - \widehat{\mathfrak{L}} \otimes B)\widehat{x}(t) + (\widehat{\mathfrak{B}} \otimes F)u(t),\\ \widehat{y}(t) &= (\mathfrak{L}\mathfrak{P} \otimes I_n)\widehat{x}(t),\end{aligned} \tag{4.6}$$

where

$$\begin{aligned}\widehat{\mathfrak{L}} &= \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{L}\mathfrak{P} \in \mathbb{R}^{\mathfrak{r}\times\mathfrak{r}},\\ \widehat{\mathfrak{B}} &= \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{B} \in \mathbb{R}^{\mathfrak{r}\times\mathfrak{m}}.\end{aligned}$$

It can be seen by inspection that the matrix $\widehat{\mathfrak{L}}$ is the Laplacian of a weighted *directed* graph with vertex set $\{1, 2, \ldots, \mathfrak{r}\}$, with $\mathfrak{r}$ equal to the number of clusters in the partition π, and adjacency matrix $\widehat{\mathfrak{A}} = [\widehat{\mathfrak{a}}_{pq}]$, with

$$\widehat{\mathfrak{a}}_{pq} = \frac{1}{|\mathfrak{C}_p|}\sum_{i\in\mathfrak{C}_p, j\in\mathfrak{C}_q}\mathfrak{a}_{ij},$$

where $|\mathfrak{C}_p|$ the cardinality of $\mathfrak{C}_p$. In other words: in the reduced graph, the edge from vertex q to vertex p is obtained by summing over all $j \in \mathfrak{C}_q$ the weights of all edges to $i \in \mathfrak{C}_p$ and dividing this sum by the cardinality of $\mathfrak{C}_p$. The row sums of $\widehat{\mathfrak{L}}$ are indeed equal to zero since $\widehat{\mathfrak{L}}\mathbb{1}_{\mathfrak{r}} = 0$. The matrix $\widehat{\mathfrak{B}} \in \mathbb{R}^{\mathfrak{r}\times\mathfrak{m}}$ satisfies

$$[\widehat{\mathfrak{B}}]_{pj} = \begin{cases}\frac{1}{|\mathfrak{C}_p|} & \text{if } \mathfrak{v}_j \in \mathfrak{C}_p,\\ 0 & \text{otherwise,}\end{cases}$$

where $\mathfrak{v}_1, \mathfrak{v}_2, \ldots, \mathfrak{v}_{\mathfrak{m}}$ are the leader vertices, $p = 1, 2, \ldots, \mathfrak{r}$, and $j = 1, 2, \ldots, \mathfrak{m}$.

Clearly, the state space dimension of the reduced order network (4.6) is equal to $\mathfrak{r}n$, whereas the dimensions $\mathfrak{m}m$ and $\mathfrak{n}n$ of the input and output have remained unchanged. Thus we can investigate the error between the original and reduced order network by looking at the difference of their transfer functions. In the sequel, we will investigate both the $\mathcal{H}_2$-norm as well as the $\mathcal{H}_\infty$-norm of this difference.

Before doing this, we will now first study the question whether our reduction procedure preserves synchronization. It is important to note that since, by assumption, the original undirected graph is connected, it has a directed spanning tree. It is easily verified that this property is preserved by our clustering procedure. Then, since the property of having a directed spanning tree is equivalent with 0 being a simple eigenvalue of the Laplacian (see [ME10, Proposition 3.8]), the reduced order Laplacian $\widehat{\mathfrak{L}}$ has again 0 as a simple eigenvalue.

Now assume that the original network (4.5) is synchronized. It is well known, see e.g. [TTM13], that this is equivalent with the condition that for each nonzero eigenvalue λ of the Laplacian $\mathfrak{L}$ the matrix $A - \lambda B$ is Hurwitz. Thus, synchronization is preserved if and only if for each nonzero eigenvalue $\widehat{\lambda}$ of the reduced order Laplacian $\widehat{\mathfrak{L}}$ the matrix $A - \widehat{\lambda}B$ is Hurwitz.

Unfortunately, in general $A - \lambda B$ Hurwitz for all nonzero $\lambda \in \sigma(\mathfrak{L})$ does *not* imply that $A - \widehat{\lambda}B$ Hurwitz for all nonzero $\lambda \in \sigma(\widehat{\mathfrak{L}})$. An exception is the 'single integrator'

case $A = 0$ and $B = 1$, where this condition is trivially satisfied, so in this special case synchronization is preserved. Also if we restrict ourselves to a special type of graph partitions, namely *almost equitable partitions*, then synchronization turns out to be preserved.

As an immediate consequence, the reduced Laplacian $\widehat{\mathfrak{L}}$ resulting from an AEP satisfies $\mathfrak{L}\mathfrak{P} = \mathfrak{P}\widehat{\mathfrak{L}}$. Indeed, since $\mathrm{im}(\mathfrak{P})$ is $\mathfrak{L}$-invariant we have $\mathfrak{L}\mathfrak{P} = \mathfrak{P}X$ for some matrix X. Obviously, we must then have $X = \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{L}\mathfrak{P} = \widehat{\mathfrak{L}}$. From this, it follows that $\sigma(\widehat{\mathfrak{L}}) \subseteq \sigma(\mathfrak{L})$. It then readily follows that synchronization is preserved if we cluster according to an AEP:

Theorem 4.4:
Assume that the network (4.5) is synchronized. Let π be an AEP. Then the reduced order network (4.6) obtained by clustering according to π is synchronized. ◊

4.2.5 $\mathcal{H}_2$-error bounds

In this section, we will formulate the first main theorem. The theorem gives an a priori upper bound for the $\mathcal{H}_2$-norm of the approximation error in the case that we cluster according to an AEP. After formulating the theorem, in the remainder of this section we will establish a proof.

Before stating the theorem, we will now first discuss some important ingredients. Let H and $\widehat{H}$ denote the transfer functions of the original (4.5) and reduced order network (4.6), respectively. We will measure the approximation error by the $\mathcal{H}_2$-norm $\|H - \widehat{H}\|_{\mathcal{H}_2}$ of these transfer functions. An important role will be played by the $\mathfrak{n} - 1$ auxiliary input-state-output systems

$$\begin{aligned} \dot{x}(t) &= (A - \lambda B)x(t) + Fd(t), \\ z(t) &= \lambda x(t), \end{aligned} \tag{4.7}$$

where λ ranges over the $\mathfrak{n} - 1$ nonzero eigenvalues of the Laplacian $\mathfrak{L}$. Let $H_\lambda(s) = \lambda(sI_n - A + \lambda B)^{-1}F$ be the transfer matrices of these systems. We assume that the original network (4.5) is synchronized, so that all of the $A - \lambda B$ are Hurwitz. Let $\|H_\lambda\|_{\mathcal{H}_2}$ denote the $\mathcal{H}_2$-norm of H_λ. Recall that the set of leader vertices is $\mathfrak{V}_{\mathrm{L}} = \{\mathfrak{v}_1, \mathfrak{v}_2, \ldots, \mathfrak{v}_{\mathfrak{m}}\}$. Vertex $\mathfrak{v}_i$ will be called leader i. This leader is an element of cluster $\mathfrak{C}_{k_i}$ for some $k_i \in \{1, 2, \ldots, \mathfrak{r}\}$. We now have the following theorem.

Theorem 4.5:
Assume that the network (4.5) is synchronized. Let π be an AEP of the graph $\mathfrak{G}$. The absolute approximation error when clustering $\mathfrak{G}$ according to π then satisfies

$$\|H - \widehat{H}\|_{\mathcal{H}_2}^2 \leqslant H_{\max,\mathcal{H}_2}^2 \sum_{i=1}^{\mathfrak{m}} \left(1 - \frac{1}{|\mathfrak{C}_{k_i}|}\right),$$

where $\mathfrak{C}_{k_i}$ is the cluster containing the leader i, and

$$H_{\max,\mathcal{H}_2} := \max_{\lambda \in \sigma(\mathfrak{L}) \setminus \sigma(\widehat{\mathfrak{L}})} \|H_\lambda\|_{\mathcal{H}_2}.$$

Furthermore, the relative approximation error satisfies

$$\frac{\|H-\widehat{H}\|_{\mathcal{H}_2}^2}{\|H\|_{\mathcal{H}_2}^2} \leqslant \left(\frac{H_{\max,\mathcal{H}_2}}{H_{\min,\mathcal{H}_2}}\right)^2 \frac{\sum_{i=1}^{\mathfrak{m}} \left(1-\frac{1}{|\mathfrak{C}_{k_i}|}\right)}{\mathfrak{m}\left(1-\frac{1}{\mathfrak{n}}\right)},$$

where

$$H_{\min,\mathcal{H}_2} := \min_{\lambda\in\sigma(\mathfrak{L})\setminus\{0\}} \|H_\lambda\|_{\mathcal{H}_2}. \qquad \Diamond$$

Remark 4.6:
We see that, with fixed number of agents and fixed number of leaders, the approximation error is equal to 0 if in each cluster that contains a leader, the leader is the only vertex in that cluster. In general, the upper bound increases if the numbers of cellmates of the leaders increase. The upper bound also depends multiplicatively on the maximal $\mathcal{H}_2$-norm of the auxiliary systems (4.7) over all Laplacian eigenvalues in the complement of the spectrum of the reduced Laplacian $\widehat{\mathfrak{L}}$. The relative error in addition depends on the minimal $\mathcal{H}_2$-norm of the auxiliary systems (4.7) over all nonzero eigenvalues of the Laplacian $\mathfrak{L}$. $\Diamond$

Remark 4.7:
For the special case that the agents are single integrators (so $n=1$, $A=0$, $B=1$, and $F=1$) it is easily seen that $H_{\max,\mathcal{H}_2} = \frac{1}{2}\max\{\lambda : \lambda\in\sigma(L)\setminus\sigma(\widehat{L})\}$ and $H_{\min,\mathcal{H}_2} = \frac{1}{2}\min\{\lambda : \lambda\in\sigma(L),\ \lambda\neq 0\}$. Thus, in the single integrator case the corresponding a priori upper bounds explicitly involve the Laplacian eigenvalues. As already noted in Section 4.2.1, the single integrator case was also studied in [MTC14] for the slightly different set up that the output equation in the original network (4.5) is taken as $y(t) = (\mathfrak{W}^{\frac{1}{2}}\mathfrak{R}^{\mathrm{T}}\otimes I_n)x(t)$ instead of $y(t) = (\mathfrak{L}\otimes I_n)x(t)$. Here, $\mathfrak{R}$ is the incidence matrix of the graph and $\mathfrak{W}$ the diagonal matrix with the edge weights on the diagonal (in other words, $\mathfrak{L} = \mathfrak{R}\mathfrak{W}\mathfrak{R}^{\mathrm{T}}$). It was shown in [MTC14] that in that case the absolute and relative approximation errors even admit the explicit formulas

$$\|H-\widehat{H}\|_{\mathcal{H}_2}^2 = \frac{1}{2}\sum_{i=1}^{\mathfrak{m}}\left(1-\frac{1}{|\mathfrak{C}_{k_i}|}\right),$$

and

$$\frac{\|H-\widehat{H}\|_{\mathcal{H}_2}^2}{\|H\|_{\mathcal{H}_2}^2} = \frac{\sum_{i=1}^{\mathfrak{m}}\left(1-\frac{1}{|\mathfrak{C}_{k_i}|}\right)}{\mathfrak{m}\left(1-\frac{1}{\mathfrak{n}}\right)}. \qquad \Diamond$$

In the remainder of this section, we will establish a proof of Theorem 4.5. As a first step, we establish the following lemma (see also [MTC14], where only the single integrator case was treated).

Lemma 4.8:
Let π be an AEP of the graph $\mathfrak{G}$. The approximation error when clustering $\mathfrak{G}$ according to π then satisfies

$$\|H - \widehat{H}\|_{\mathcal{H}_2}^2 = \|H\|_{\mathcal{H}_2}^2 - \|\widehat{H}\|_{\mathcal{H}_2}^2 . \qquad \Diamond$$

Proof. First, note that the columns of $\mathfrak{P}$ are orthogonal. We construct a matrix $\mathfrak{T} = \begin{bmatrix}\mathfrak{P} & \mathfrak{P}_\perp\end{bmatrix}$, where the $\mathfrak{n} \times (\mathfrak{n} - \mathfrak{r})$ matrix $\mathfrak{P}_\perp$ is chosen such that the columns of $\mathfrak{T}$ form an orthogonal basis for $\mathbb{R}^{\mathfrak{n}}$. In this case, we have $\mathfrak{P}^{\mathrm{T}}\mathfrak{P}_\perp = 0$. Next, we apply the state space transformation $x(t) = (\mathfrak{T} \otimes I_n)\widetilde{x}(t)$ to system (4.5). We obtain

$$\begin{aligned}\begin{bmatrix}\dot{\widetilde{x}}_1(t)\\ \dot{\widetilde{x}}_2(t)\end{bmatrix} &= \widetilde{A}\begin{bmatrix}\widetilde{x}_1(t)\\ \widetilde{x}_2(t)\end{bmatrix} + \widetilde{B}u(t),\\ y(t) &= \widetilde{C}\begin{bmatrix}\widetilde{x}_1(t)\\ \widetilde{x}_2(t)\end{bmatrix},\end{aligned} \tag{4.8}$$

where the matrices $\widetilde{A}$, $\widetilde{B}$, and $\widetilde{C}$ are given by

$$\widetilde{A} = \begin{bmatrix} I_{\mathfrak{r}} \otimes A - \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{L}\mathfrak{P} \otimes B & -\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{L}\mathfrak{P}_\perp \otimes B \\ -\left(\mathfrak{P}_\perp^{\mathrm{T}}\mathfrak{P}_\perp\right)^{-1}\mathfrak{P}_\perp^{\mathrm{T}}\mathfrak{L}\mathfrak{P} \otimes B & I_{\mathfrak{n}-\mathfrak{r}} \otimes A - \left(\mathfrak{P}_\perp^{\mathrm{T}}\mathfrak{P}_\perp\right)^{-1}\mathfrak{P}_\perp^{\mathrm{T}}\mathfrak{L}\mathfrak{P}_\perp \otimes B \end{bmatrix},$$
$$\widetilde{B} = \begin{bmatrix}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{B} \otimes F\\ \left(\mathfrak{P}_\perp^{\mathrm{T}}\mathfrak{P}_\perp\right)^{-1}\mathfrak{P}_\perp^{\mathrm{T}}\mathfrak{B} \otimes F\end{bmatrix}, \quad \widetilde{C} = \begin{bmatrix}\mathfrak{L}\mathfrak{P} \otimes I_n & \mathfrak{L}\mathfrak{P}_\perp \otimes I_n\end{bmatrix}.$$

Obviously, in (4.8) the transfer function from u to y is equal to H. Furthermore, if the state component $\widetilde{x}_2$ is truncated from (4.8), what we are left with is the reduced order model (4.6). Since π is an AEP of $\mathfrak{G}$, by Lemma 2.48, $\mathrm{im}(\mathfrak{P})$ is invariant under $\mathfrak{L}$. From this, it follows that not only $\mathfrak{P}_\perp^{\mathrm{T}}\mathfrak{P} = 0$, but also

$$\mathfrak{P}_\perp^{\mathrm{T}}\mathfrak{L}\mathfrak{P} = 0 \text{ and } \mathfrak{P}_\perp^{\mathrm{T}}\mathfrak{L}^2\mathfrak{P} = 0. \tag{4.9}$$

It is easily checked that

$$H(s) = \widehat{H}(s) + H_{\mathrm{err}}(s),$$

where $H_{\mathrm{err}}(s)$ is given by

$$\begin{aligned}H_{\mathrm{err}}(s) = (\mathfrak{L}\mathfrak{P}_\perp \otimes I_n)\left(sI - \left(I_{\mathfrak{n}-\mathfrak{r}} \otimes A - \left(\mathfrak{P}_\perp^{\mathrm{T}}\mathfrak{P}_\perp\right)^{-1}\mathfrak{P}_\perp^{\mathrm{T}}\mathfrak{L}\mathfrak{P}_\perp \otimes B\right)\right)^{-1}\\ \cdot \left(\left(\mathfrak{P}_\perp^{\mathrm{T}}\mathfrak{P}_\perp\right)^{-1}\mathfrak{P}_\perp^{\mathrm{T}}\mathfrak{B} \otimes F\right).\end{aligned} \tag{4.10}$$

From (4.9) and (4.10), we have $\widehat{H}(-s)^{\mathrm{T}}H_{\mathrm{err}}(s) = 0$. Thus, we find that

$$\|H\|_{\mathcal{H}_2}^2 = \|\widehat{H}\|_{\mathcal{H}_2}^2 + \|H_{\mathrm{err}}\|_{\mathcal{H}_2}^2 ,$$

which concludes the proof. □

Recall that, since π is an AEP, we have $\sigma(\widehat{\mathfrak{L}}) \subseteq \sigma(\mathfrak{L})$. Label the eigenvalues of $\mathfrak{L}$ as $0, \lambda_2, \lambda_3, \ldots, \lambda_{\mathfrak{n}}$ in such a way that $0, \lambda_2, \lambda_3, \ldots, \lambda_{\mathfrak{r}}$ are the eigenvalues of $\widehat{\mathfrak{L}}$. Also, without loss of generality, we assume that π is *regularly formed*, i.e., all ones in each of the columns of $\mathfrak{P}$ are consecutive. One can always relabel the agents in the graph in such a way that this is achieved. Consider now the symmetric matrix

$$\overline{\mathfrak{L}} \coloneqq \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}} \widehat{\mathfrak{L}} \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} = \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \mathfrak{P}^{\mathrm{T}}\mathfrak{L}\mathfrak{P} \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}}. \tag{4.11}$$

Note that the eigenvalues of $\overline{\mathfrak{L}}$ and $\widehat{\mathfrak{L}}$ coincide. Let $\widehat{U}$ be an orthogonal matrix that diagonalizes $\overline{\mathfrak{L}}$. We then have

$$\widehat{U}^{\mathrm{T}}\overline{\mathfrak{L}}\widehat{U} = \operatorname{diag}(0, \lambda_2, \ldots, \lambda_{\mathfrak{r}}) =: \widehat{\Lambda}. \tag{4.12}$$

Next, take $U_1 = \mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \widehat{U}$. The columns of U_1 are orthonormal:

$$U_1^{\mathrm{T}}U_1 = \widehat{U}^{\mathrm{T}} \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \mathfrak{P}^{\mathrm{T}}\mathfrak{P} \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \widehat{U} = \widehat{U}^{\mathrm{T}}\widehat{U} = I.$$

Furthermore, we have that

$$U_1^{\mathrm{T}}\mathfrak{L}U_1 = \widehat{U}^{\mathrm{T}}\overline{\mathfrak{L}}\widehat{U} = \widehat{\Lambda}.$$

Now choose U_2 such that $U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$ is an orthogonal matrix and

$$\Lambda \coloneqq U^{\mathrm{T}}\mathfrak{L}U = \begin{bmatrix} \widehat{\Lambda} & 0 \\ 0 & \overline{\Lambda} \end{bmatrix}, \tag{4.13}$$

where $\overline{\Lambda} = \operatorname{diag}(\lambda_{\mathfrak{r}+1}, \ldots, \lambda_{\mathfrak{n}})$. It is easily verified that the first column of U_1, and thus the first column of U, is given by $\frac{1}{\sqrt{\mathfrak{n}}}\mathbb{1}_{\mathfrak{n}}$, a fact that we will use in the remainder of this section.

Using the above, we will now first establish explicit formulas for the $\mathcal{H}_2$-norms of H and $\widehat{H}$ separately. The following lemma gives a formula for the $\mathcal{H}_2$-norm of the original transfer function H.

Lemma 4.9:
Let U be as in (4.13). For $i = 2, \ldots, \mathfrak{n}$, let X_i be the observability Gramian of the auxiliary system $(A - \lambda_i B, F, \lambda_i I)$ in (4.7), i.e., the unique solution of the Lyapunov equation $(A - \lambda_i B)^{\mathrm{T}} X_i + X_i(A - \lambda_i B) + \lambda_i^2 I = 0$. Then the $\mathcal{H}_2$-norm of H is given by:

$$\|H\|_{\mathcal{H}_2}^2 = \operatorname{tr}\Big(\left(U^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U \otimes I\right) \operatorname{diag}\big(0, F^{\mathrm{T}}X_2F, \ldots, F^{\mathrm{T}}X_{\mathfrak{n}}F\big)\Big). \tag{4.14}$$

$\Diamond$

Proof. It can be verified, using the fact that $A - \lambda_i B$ is Hurwitz for $i = 2, 3, \ldots, \mathfrak{n}$, that

$$\mathcal{X}_+(I \otimes A - \mathfrak{L} \otimes B) = \mathbb{1}_{\mathfrak{n}} \otimes \mathcal{X}_+(A).$$

This immediately implies that $\mathcal{X}_+(I \otimes A - \mathfrak{L} \otimes B) \subseteq \ker(\mathfrak{L} \otimes I)$. As a consequence of Proposition 4.1, we have

$$\|H\|_{\mathcal{H}_2}^2 = \operatorname{tr}\left(\left(\mathfrak{B}^{\mathrm{T}} \otimes F^{\mathrm{T}}\right) X (\mathfrak{B} \otimes F)\right),$$

where X is the unique positive semi-definite solution to the Lyapunov equation

$$\left(I \otimes A^{\mathrm{T}} - \mathfrak{L} \otimes B^{\mathrm{T}}\right) X + X(I \otimes A - \mathfrak{L} \otimes B) + \mathfrak{L}^2 \otimes I = 0 \tag{4.15}$$

with the property that $\mathcal{X}_+(I \otimes A - \mathfrak{L} \otimes B) \subseteq \ker(X)$. In order to compute this solution X, premultiply (4.15) by $U^{\mathrm{T}} \otimes I$ and postmultiply by $U \otimes I$, and substitute $Z = (U^{\mathrm{T}} \otimes I)X(U \otimes I)$ to obtain

$$\left(I \otimes A^{\mathrm{T}} - \Lambda \otimes B^{\mathrm{T}}\right) Z + Z(I \otimes A - \Lambda \otimes B) + \Lambda^2 \otimes I = 0. \tag{4.16}$$

Solving (4.16), we take Z as

$$Z = \operatorname{diag}(0, X_2, \ldots, X_{\mathfrak{n}}),$$

where X_i, for $i = 2, \ldots, \mathfrak{n}$, is the observability Gramian of the auxiliary system $(A - \lambda_i B, F, \lambda_i I)$ in (4.7). Next, $X := (U \otimes I)Z(U^{\mathrm{T}} \otimes I)$ is a solution of the original Lyapunov equation, and it is easily verified that indeed $\mathcal{X}_+(I \otimes A - \mathfrak{L} \otimes B) \subseteq \ker(X)$. Thus, we obtain the following expression for the $\mathcal{H}_2$-norm of H:

$$\begin{aligned}\|H\|_{\mathcal{H}_2}^2 &= \operatorname{tr}\left(\left(\mathfrak{B}^{\mathrm{T}} U \otimes F^{\mathrm{T}}\right) \operatorname{diag}(0, X_2, \ldots, X_{\mathfrak{n}}) \left(U^{\mathrm{T}} \mathfrak{B} \otimes F\right)\right) \\ &= \operatorname{tr}\left(\left(U^{\mathrm{T}} \mathfrak{B} \mathfrak{B}^{\mathrm{T}} U \otimes I\right) \operatorname{diag}\left(0, F^{\mathrm{T}} X_2 F, \ldots, F^{\mathrm{T}} X_{\mathfrak{n}} F\right)\right).\end{aligned}$$ □

We proceed with finding a formula for the $\mathcal{H}_2$-norm for the reduced system. This will be dealt with in the following lemma.

Lemma 4.10:
Let $\widehat{U}$ be as in (4.12) above. For $i = 2, \ldots, \mathfrak{r}$, let X_i be the observability Gramian of the auxiliary system $(A - \lambda_i B, F, \lambda_i I)$ in (4.7), i.e., the unique solution of the Lyapunov equation $(A - \lambda_i B)^{\mathrm{T}} X_i + X_i(A - \lambda_i B) + \lambda_i^2 I = 0$. Then the $\mathcal{H}_2$-norm of $\widehat{H}$ is given by

$$\|\widehat{H}\|_{\mathcal{H}_2}^2 = \operatorname{tr}\left(\left(\widehat{U}^{\mathrm{T}} \left(\mathfrak{P}^{\mathrm{T}} \mathfrak{P}\right)^{\frac{1}{2}} \widehat{\mathfrak{B}} \widehat{\mathfrak{B}}^{\mathrm{T}} \left(\mathfrak{P}^{\mathrm{T}} \mathfrak{P}\right)^{\frac{1}{2}} \widehat{U} \otimes I\right) \operatorname{diag}\left(0, F^{\mathrm{T}} X_2 F, \ldots, F^{\mathrm{T}} X_{\mathfrak{r}} F\right)\right). \tag{4.17}$$

◊

Proof. Firstly, it can be verified that

$$\mathcal{X}_+(I \otimes A - \widehat{\mathfrak{L}} \otimes B) = \mathbb{1}_{\mathfrak{r}} \otimes \mathcal{X}_+(A).$$

This implies that $\mathcal{X}_+(I \otimes A - \widehat{\mathfrak{L}} \otimes B) \subseteq \ker(\mathfrak{L}\mathfrak{P} \otimes I)$. By Proposition 4.1, we then have

$$\|\widehat{H}\|_{\mathcal{H}_2}^2 = \operatorname{tr}\left(\left(\widehat{\mathfrak{B}}^{\mathrm{T}} \otimes F^{\mathrm{T}}\right) \widehat{X} \left(\widehat{\mathfrak{B}} \otimes F\right)\right),$$

where $\widehat{X}$ is the unique positive semi-definite solution to the Lyapunov equation

$$\left(I \otimes A^{\mathrm{T}} - \widehat{\mathfrak{L}}^{\mathrm{T}} \otimes B^{\mathrm{T}}\right) \widehat{X} + \widehat{X}(I \otimes A - \widehat{\mathfrak{L}} \otimes B) + \mathfrak{P}^{\mathrm{T}} \mathfrak{L}^2 \mathfrak{P} \otimes I = 0 \tag{4.18}$$

satisfying the property that $\mathcal{X}_+(I \otimes A - \widehat{\mathfrak{L}} \otimes B) \subseteq \ker(\widehat{X})$. In order to compute this solution, pre- and postmultiply (4.18) by $\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \otimes I$ and substitute

$$\widehat{Y} = \left(\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \otimes I\right) \widehat{X} \left(\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \otimes I\right)$$

to obtain

$$\left(I \otimes A^{\mathrm{T}} - \overline{\mathfrak{L}} \otimes B^{\mathrm{T}}\right) \widehat{Y} + \widehat{Y} \left(I \otimes A - \overline{\mathfrak{L}} \otimes B\right) + \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \mathfrak{P}^{\mathrm{T}} \mathfrak{L}^2 \mathfrak{P} \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \otimes I = 0. \tag{4.19}$$

Recall from Section 4.2.4 that $\mathfrak{L}\mathfrak{P} = \mathfrak{P}\widehat{\mathfrak{L}}$. From this it follows that

$$\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \mathfrak{P}^{\mathrm{T}} \mathfrak{L}^2 \mathfrak{P} \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} = \overline{\mathfrak{L}}^2.$$

Consequently, we can diagonalize the corresponding term in (4.19) by premultiplying by $\widehat{U}^{\mathrm{T}} \otimes I$ and postmultiplying by $\widehat{U} \otimes I$, where $\widehat{U}$ is as in (4.12). Next, we denote $\widehat{Z} = (\widehat{U}^{\mathrm{T}} \otimes I)\widehat{Y}(\widehat{U} \otimes I)$ so that (4.19) reduces to

$$\left(I \otimes A^{\mathrm{T}} - \widehat{\Lambda} \otimes B^{\mathrm{T}}\right) \widehat{Z} + \widehat{Z} \left(I \otimes A - \widehat{\Lambda} \otimes B\right) + \widehat{\Lambda}^2 \otimes I = 0,$$

which can be solved by taking

$$\widehat{Z} = \operatorname{diag}(0, X_2, \ldots, X_{\mathfrak{r}}),$$

where again X_i, for $i = 2, \ldots, \mathfrak{r}$, is the observability Gramian of the auxiliary system $(A - \lambda_i B, F, \lambda_i I)$ in (4.7). Next,

$$\widehat{X} = \left(\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}} \widehat{U} \otimes I\right) \widehat{Z} \left(\widehat{U}^{\mathrm{T}} \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}} \otimes I\right)$$

then satisfies (4.18), and it can be verified that $\mathcal{X}_+(I \otimes A - \widehat{\mathfrak{L}} \otimes B) \subseteq \ker(\widehat{X})$. Thus, the $\mathcal{H}_2$-norm of $\widehat{H}$ is given by

$$\begin{aligned} \|\widehat{H}\|_{\mathcal{H}_2}^2 &= \operatorname{tr}\left(\left(\widehat{\mathfrak{B}}^{\mathrm{T}} \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}} \widehat{U} \otimes F^{\mathrm{T}}\right) \operatorname{diag}(0, X_2, \ldots, X_{\mathfrak{r}}) \left(\widehat{U}^{\mathrm{T}} \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}} \widehat{\mathfrak{B}} \otimes F\right)\right) \\ &= \operatorname{tr}\left(\left(\widehat{U}^{\mathrm{T}} \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}} \widehat{\mathfrak{B}}\widehat{\mathfrak{B}}^{\mathrm{T}} \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}} \widehat{U} \otimes I\right) \operatorname{diag}\left(0, F^{\mathrm{T}}X_2F, \ldots, F^{\mathrm{T}}X_{\mathfrak{r}}F\right)\right). \quad \square \end{aligned}$$

We will now combine the previous lemmas, and give a proof of Theorem 4.5.

Proof of Theorem 4.5. Using Lemma 4.8, and formulas (4.14) and (4.17), we compute

$$\begin{aligned}
\|H-\widehat{H}\|_{\mathcal{H}_2}^2 &= \operatorname{tr}\big(\big(U^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U\otimes I\big)\operatorname{diag}\big(0,F^{\mathrm{T}}X_2F,\dots,F^{\mathrm{T}}X_{\mathfrak{n}}F\big)\big)\\
&\quad-\operatorname{tr}\Big(\Big(\widehat{U}^{\mathrm{T}}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}}\widehat{\mathfrak{B}}\widehat{\mathfrak{B}}^{\mathrm{T}}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}}\widehat{U}\otimes I\Big)\operatorname{diag}\big(0,F^{\mathrm{T}}X_2F,\dots,F^{\mathrm{T}}X_{\mathfrak{r}}F\big)\Big)\\
&= \operatorname{tr}\Bigg(\Bigg(\begin{bmatrix}U_1^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U_1 & U_1^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U_2\\ U_2^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U_1 & U_2^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U_2\end{bmatrix}\otimes I\Bigg)\operatorname{diag}\big(0,F^{\mathrm{T}}X_2F,\dots,F^{\mathrm{T}}X_{\mathfrak{n}}F\big)\Bigg)\\
&\quad-\operatorname{tr}\big(\big(U_1^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U_1\otimes I\big)\operatorname{diag}\big(0,F^{\mathrm{T}}X_2F,\dots,F^{\mathrm{T}}X_{\mathfrak{r}}F\big)\big)\\
&= \operatorname{tr}\big(\big(U_2^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U_2\otimes I\big)\operatorname{diag}\big(F^{\mathrm{T}}X_{\mathfrak{r}+1}F,\dots,F^{\mathrm{T}}X_{\mathfrak{n}}F\big)\big)\,,
\end{aligned}\tag{4.20}$$

where the second equality follows from the fact that

$$\begin{aligned}
\widehat{\mathfrak{B}}^{\mathrm{T}}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}}\widehat{U} &= \mathfrak{B}^{\mathrm{T}}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}}\widehat{U}\\
&= \mathfrak{B}^{\mathrm{T}}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}}\widehat{U}\\
&= \mathfrak{B}^{\mathrm{T}}U_1.
\end{aligned}$$

Next, observe that (4.20) can be rewritten as

$$\begin{aligned}
\|H-\widehat{H}\|_{\mathcal{H}_2}^2 &= \operatorname{tr}\big(\big(U_2^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U_2\otimes I\big)\operatorname{diag}\big(F^{\mathrm{T}}X_{\mathfrak{r}+1}F,\dots,F^{\mathrm{T}}X_{\mathfrak{n}}F\big)\big)\\
&= \operatorname{tr}\big(\big(U_2^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U_2\big)\operatorname{diag}\big(\operatorname{tr}\big(F^{\mathrm{T}}X_{\mathfrak{r}+1}F\big),\dots,\operatorname{tr}\big(F^{\mathrm{T}}X_{\mathfrak{n}}F\big)\big)\big)\\
&= \operatorname{tr}\Big(\big(U_2^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U_2\big)\operatorname{diag}\Big(\big\|H_{\lambda_{\mathfrak{r}+1}}\big\|_{\mathcal{H}_2}^2,\dots,\|H_{\lambda_{\mathfrak{n}}}\|_{\mathcal{H}_2}^2\Big)\Big)\,,
\end{aligned}$$

where H_{λ_i} for $i=\mathfrak{r}+1,\dots,\mathfrak{n}$ is the transfer function of the auxiliary system (4.7). An upper bound for this expression is given by

$$\operatorname{tr}\Big(\big(U_2^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U_2\big)\operatorname{diag}\Big(\big\|H_{\lambda_{\mathfrak{r}+1}}\big\|_{\mathcal{H}_2}^2,\dots,\|H_{\lambda_{\mathfrak{n}}}\|_{\mathcal{H}_2}^2\Big)\Big)\leqslant H_{\max,\mathcal{H}_2}^2\operatorname{tr}\big(U_2^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U_2\big)\,,$$

where $H_{\max,\mathcal{H}_2}=\max_{\mathfrak{r}+1\leqslant j\leqslant\mathfrak{n}}\big\|H_{\lambda_j}\big\|_{\mathcal{H}_2}$. Furthermore, we have

$$\begin{aligned}
\operatorname{tr}\big(U_2^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U_2\big) &= \operatorname{tr}\big(U^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U\big)-\operatorname{tr}\big(U_1^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U_1\big)\\
&= \mathfrak{m}-\operatorname{tr}\Big(\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}\Big)\,.
\end{aligned}$$

Since, by assumption, the partition π is regularly formed, $\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}$ is a block diagonal matrix of the form

$$\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}=\operatorname{diag}(\mathfrak{P}_1,\mathfrak{P}_2,\dots,\mathfrak{P}_k)\,.$$

It is easily verified that each $\mathfrak{P}_i$ is a $|\mathfrak{C}_i|\times|\mathfrak{C}_i|$ matrix whose elements are all equal to $\frac{1}{|\mathfrak{C}_i|}$. The matrix $\mathfrak{B}\mathfrak{B}^{\mathrm{T}}$ is a diagonal matrix whose diagonal entries are either 0 or 1. We then have that the ith column of $\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}$ is either equal to the ith

column of $\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}$ if agent i is a leader, or zero otherwise. It then follows that the diagonal elements of $\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}$ are either zero or $\frac{1}{|\mathfrak{C}_{k_i}|}$ if i is part of the leader set, where $\mathfrak{C}_{k_i}$ is the cluster containing agent i. Hence, we have

$$\operatorname{tr}\big(U_1^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U_1\big) = \sum_{i=1}^{\mathfrak{m}} \frac{1}{|\mathfrak{C}_{k_i}|},$$

and consequently

$$\operatorname{tr}\big(U_2^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U_2\big) = \mathfrak{m} - \sum_{i=1}^{\mathfrak{m}} \frac{1}{|\mathfrak{C}_{k_i}|}.$$

In conclusion, we have

$$\|H - \widehat{H}\|_{\mathcal{H}_2}^2 \leqslant H_{\max,\mathcal{H}_2}^2 \sum_{i=1}^{\mathfrak{m}} \left(1 - \frac{1}{|\mathfrak{C}_{k_i}|}\right),$$

which completes the proof of the first part of the theorem.

We now prove the statement about the relative error. For this, we will establish a lower bound for $\|H\|_{\mathcal{H}_2}^2$. By (4.14), we have

$$\begin{aligned} \|H\|_{\mathcal{H}_2}^2 &= \operatorname{tr}\big(\big(U^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U \otimes I\big)\operatorname{diag}\big(0, F^{\mathrm{T}}X_2F, \ldots, F^{\mathrm{T}}X_{\mathfrak{n}}F\big)\big) \\ &= \operatorname{tr}\big(\big(U^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}U\big)\operatorname{diag}\big(0, \operatorname{tr}\big(F^{\mathrm{T}}X_2F\big), \ldots, \operatorname{tr}\big(F^{\mathrm{T}}X_{\mathfrak{n}}F\big)\big)\big). \end{aligned} \tag{4.21}$$

The first column of U spans the eigenspace corresponding to the eigenvalue 0 of $\mathfrak{L}$ and hence must be equal to $u_1 = \frac{1}{\sqrt{\mathfrak{n}}}\mathbb{1}_{\mathfrak{n}}$. Let $\overline{U}$ be such that $U = \begin{bmatrix} u_1 & \overline{U} \end{bmatrix}$. It is then easily verified using (4.21) that

$$\begin{aligned} \|H\|_{\mathcal{H}_2}^2 &= \operatorname{tr}\Big(\Big(\overline{U}^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}\overline{U}\Big)\operatorname{diag}\big(\operatorname{tr}\big(F^{\mathrm{T}}X_2F\big), \ldots, \operatorname{tr}\big(F^{\mathrm{T}}X_{\mathfrak{n}}F\big)\big)\Big) \\ &= \operatorname{tr}\Big(\Big(\overline{U}^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}\overline{U}\Big)\operatorname{diag}\big(\|H_{\lambda_2}\|_{\mathcal{H}_2}^2, \ldots, \|H_{\lambda_{\mathfrak{n}}}\|_{\mathcal{H}_2}^2\big)\Big). \end{aligned}$$

Finally, since

$$\operatorname{tr}\Big(\overline{U}^{\mathrm{T}}\mathfrak{B}\mathfrak{B}^{\mathrm{T}}\overline{U}\Big) = \operatorname{tr}\Big(\mathfrak{B}^{\mathrm{T}}\overline{U}\overline{U}^{\mathrm{T}}\mathfrak{B}\Big) = \operatorname{tr}\big(\mathfrak{B}^{\mathrm{T}}\left(UU^{\mathrm{T}} - u_1u_1^{\mathrm{T}}\right)\mathfrak{B}\big) = \mathfrak{m} - \frac{\mathfrak{m}}{\mathfrak{n}},$$

we obtain that $\|H\|_{\mathcal{H}_2}^2 \geqslant \mathfrak{m}\left(1 - \frac{1}{\mathfrak{n}}\right)\left(H_{\min,\mathcal{H}_2}\right)^2$. This then yields the upper bound for the relative error as claimed. □

Remark 4.11:
Note that by our labeling of the eigenvalues of $\mathfrak{L}$, in the formulation of Theorem 4.5, we have that $\sigma(\mathfrak{L}) \setminus \sigma(\widehat{\mathfrak{L}})$ is equal to $\{\lambda_{\mathfrak{r}+1}, \ldots, \lambda_{\mathfrak{n}}\}$ used in the proof. We stress that this should not be confused with the notation often used in the literature, where the λ_i's are labeled in increasing order. ◊

4.2.6 $\mathcal{H}_\infty$-error bounds

Whereas in the previous section we studied a priori upper bounds for the approximation error in terms of the $\mathcal{H}_2$-norm, the present section aims at expressing the approximation error in terms of the $\mathcal{H}_\infty$-norm. This section consists of two subsections. In the first subsection, we consider the special case that the agent dynamics is a single integrator system. Here we obtain an explicit formula for the $\mathcal{H}_\infty$-norm of the error. In the second subsection, we find an upper bound for the $\mathcal{H}_\infty$-error for symmetric systems.

4.2.6.1 The single integrator case

Here we consider the special case that the agent dynamics is a single integrator system. In this case, we have $A = 0$, $B = 1$, and $F = 1$ and the original system (4.5) reduces to

$$\begin{aligned} \dot{x}(t) &= -\mathfrak{L}x(t) + \mathfrak{B}u(t), \\ y(t) &= \mathfrak{L}x(t). \end{aligned} \tag{4.22}$$

The state space dimension of (4.22) is then simply $\mathfrak{n}$, the number of agents. For a given partition $\pi = \{\mathfrak{C}_1, \mathfrak{C}_2, \ldots, \mathfrak{C}_\mathfrak{r}\}$, the reduced system (4.6) is now given by

$$\begin{aligned} \dot{\widehat{x}}(t) &= -\widehat{\mathfrak{L}}\widehat{x}(t) + \widehat{\mathfrak{B}}u(t), \\ \widehat{y}(t) &= \mathfrak{L}\mathfrak{P}\widehat{x}(t), \end{aligned}$$

where $\mathfrak{P}$ is again the characteristic matrix of π and $\widehat{x}(t) \in \mathbb{R}^\mathfrak{r}$. The transfer functions H and $\widehat{H}$, of the original and reduced system respectively, are given by

$$\begin{aligned} H(s) &= \mathfrak{L}(sI_\mathfrak{n} + \mathfrak{L})^{-1}\mathfrak{B}, \\ \widehat{H}(s) &= \mathfrak{L}\mathfrak{P}\left(sI_\mathfrak{r} + \widehat{\mathfrak{L}}\right)^{-1}\widehat{\mathfrak{B}}. \end{aligned}$$

The first main result of this section is the following explicit formula for the $\mathcal{H}_\infty$ model reduction error. It complements the formula for the $\mathcal{H}_2$-error obtained in [MTC14] (see also Remark 4.7).

Theorem 4.12:
Let π be an AEP of the graph $\mathfrak{G}$. If the network with single integrator agent dynamics (4.22) is clustered according to π, then the $\mathcal{H}_\infty$-error is given by

$$\|H - \widehat{H}\|_{\mathcal{H}_\infty}^2 = \begin{cases} \max\limits_{1 \leqslant i \leqslant \mathfrak{m}} \left(1 - \frac{1}{|\mathfrak{C}_{k_i}|}\right) & \text{if the leaders are in different clusters,} \\ 1 & \text{otherwise,} \end{cases}$$

where, for some $k_i \in \{1, 2, \ldots, \mathfrak{m}\}$, $\mathfrak{C}_{k_i}$ is the cluster containing the leader i. Furthermore, $\|H\|_{\mathcal{H}_\infty} = 1$, hence the relative and absolute $\mathcal{H}_\infty$-errors coincide. ◊

Remark 4.13:
We see that the $\mathcal{H}_\infty$-error lies in the interval $[0, 1]$. The error is maximal ($= 1$) if and only if two or more leader vertices occupy one and the same cluster. The error is minimal ($= 0$) if and only if each leader vertex occupies a different cluster, and is the only vertex in this cluster. In general, the error increases if the number of cellmates of the leaders increases. ◊

Proof of Theorem 4.12. To simplify notation, denote $H_{\mathrm{err}}(s) = H(s) - \widehat{H}(s)$. Note that both H and $\widehat{H}$ have all poles in the open left half plane. We now first show that, since π is an AEP, we have

$$\|H_{\mathrm{err}}\|_{\mathcal{H}_\infty} = \|H_{\mathrm{err}}(0)\|_2 . \tag{4.23}$$

First note that $\widehat{H}(s) = \mathfrak{L}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}}\left(sI_{\mathfrak{r}} + \overline{\mathfrak{L}}\right)^{-1}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}}\widehat{\mathfrak{B}}$, where the symmetric matrix $\overline{\mathfrak{L}}$ is given by (4.11). Thus, a state space representation for the error system is given by

$$\begin{aligned} \dot{x}_e(t) &= \begin{bmatrix} -\mathfrak{L} & 0 \\ 0 & -\overline{\mathfrak{L}} \end{bmatrix} x_e(t) + \begin{bmatrix} \mathfrak{B} \\ \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}}\widehat{\mathfrak{B}} \end{bmatrix} u(t), \\ e(t) &= \begin{bmatrix} \mathfrak{L} & -\mathfrak{L}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \end{bmatrix} x_e(t). \end{aligned} \tag{4.24}$$

Next, we show that (4.23) holds by applying Lemma 4.2 to system (4.24). Indeed, with $X = -\mathfrak{L}$, we have

$$\begin{aligned} \begin{bmatrix} \mathfrak{L} & -\mathfrak{L}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} -\mathfrak{L} & 0 \\ 0 & -\overline{\mathfrak{L}} \end{bmatrix} &= \begin{bmatrix} -\mathfrak{L}^2 & \mathfrak{L}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}}\overline{\mathfrak{L}} \end{bmatrix} \\ &= \begin{bmatrix} -\mathfrak{L}^2 & \mathfrak{L}\mathfrak{P}\widehat{\mathfrak{L}}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \end{bmatrix} \\ &= \begin{bmatrix} -\mathfrak{L}^2 & \mathfrak{L}^2\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \end{bmatrix} \\ &= X \begin{bmatrix} \mathfrak{L} & -\mathfrak{L}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \end{bmatrix}, \end{aligned}$$

and from Lemma 4.2 it then immediately follows that $\|H_{\mathrm{err}}\|_{\mathcal{H}_\infty} = \|H_{\mathrm{err}}(0)\|_2$. To compute $\|H_{\mathrm{err}}(0)\|_2$, we apply Lemma 4.3 to system (4.24). First, it is easily verified that

$$\ker \begin{bmatrix} -\mathfrak{L} & 0 \\ 0 & -\overline{\mathfrak{L}} \end{bmatrix} \subseteq \ker \begin{bmatrix} \mathfrak{L} & -\mathfrak{L}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \end{bmatrix}.$$

By applying Lemma 4.3 we then obtain

$$\begin{aligned} H_{\mathrm{err}}(0) &= \begin{bmatrix} \mathfrak{L} & -\mathfrak{L}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \mathfrak{L} & 0 \\ 0 & \overline{\mathfrak{L}} \end{bmatrix}^{+} \begin{bmatrix} \mathfrak{B} \\ \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}}\widehat{\mathfrak{B}} \end{bmatrix} \\ &= \mathfrak{L}\left(\mathfrak{L}^{+} - \mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}}\overline{\mathfrak{L}}^{+}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}}\mathfrak{P}^{\mathrm{T}}\right)\mathfrak{B}. \end{aligned} \tag{4.25}$$

Recall that $\widehat{U}$ in (4.12) is an orthogonal matrix that diagonalizes $\overline{\mathfrak{L}}$ and that $U_1 = \mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}}\widehat{U}$. Then $\overline{\mathfrak{L}}^{+} = \widehat{U}\widehat{\Lambda}^{+}\widehat{U}^{\mathrm{T}}$. Thus, we have

$$\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}}\overline{\mathfrak{L}}^{+}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}}\mathfrak{P}^{\mathrm{T}} = U_1\widehat{\Lambda}^{+}U_1^{\mathrm{T}}.$$

Next, we compute

$$\begin{aligned}\mathfrak{L}\mathfrak{L}^{+} &= U\Lambda U^{\mathrm{T}}U\Lambda^{+}U^{\mathrm{T}}\\ &= U\Lambda\Lambda^{+}U^{\mathrm{T}}\\ &= I_{\mathfrak{n}} - \frac{1}{\mathfrak{n}}\mathbb{1}_{\mathfrak{n}}\mathbb{1}_{\mathfrak{n}}^{\mathrm{T}},\end{aligned} \tag{4.26}$$

where the last equality follows from the fact that the first column of U is $\frac{1}{\sqrt{\mathfrak{n}}}\mathbb{1}_{\mathfrak{n}}$. Now observe that

$$\begin{aligned}\mathfrak{L}U_1\widehat{\Lambda}^{+}U_1^{\mathrm{T}} &= U\Lambda U^{\mathrm{T}}U_1\widehat{\Lambda}^{+}U_1^{\mathrm{T}}\\ &= U_1\widehat{\Lambda}\widehat{\Lambda}^{+}U_1^{\mathrm{T}}\\ &= U_1U_1^{\mathrm{T}} - \frac{1}{\mathfrak{n}}\mathbb{1}_{\mathfrak{n}}\mathbb{1}_{\mathfrak{n}}^{\mathrm{T}}\\ &= \mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}} - \frac{1}{\mathfrak{n}}\mathbb{1}_{\mathfrak{n}}\mathbb{1}_{\mathfrak{n}}^{\mathrm{T}}.\end{aligned} \tag{4.27}$$

Combining (4.26) and (4.27) with (4.25), we obtain

$$H_{\mathrm{err}}(0) = \left(I_{\mathfrak{n}} - \mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\right)\mathfrak{B}.$$

From (4.23) then, we have that the $\mathcal{H}_\infty$-error is given by

$$\begin{aligned}\|H - \widehat{H}\|_{\mathcal{H}_\infty}^2 &= \lambda_{\max}\left(H_{\mathrm{err}}(0)^{\mathrm{T}}H_{\mathrm{err}}(0)\right)\\ &= \lambda_{\max}\left(\mathfrak{B}^{\mathrm{T}}\left(I_{\mathfrak{n}} - \mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\right)^2\mathfrak{B}\right)\\ &= \lambda_{\max}\left(I_{\mathfrak{m}} - \mathfrak{B}^{\mathrm{T}}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{B}\right)\\ &= 1 - \lambda_{\min}\left(\mathfrak{B}^{\mathrm{T}}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{B}\right).\end{aligned} \tag{4.28}$$

All that is left is to compute the minimal eigenvalue of $\mathfrak{B}^{\mathrm{T}}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{B}$. Again, let $\{\mathfrak{v}_1, \mathfrak{v}_2, \ldots, \mathfrak{v}_{\mathfrak{m}}\}$ be the set of leaders and note that $\mathfrak{B}$ satisfies

$$\mathfrak{B} = \begin{bmatrix} e_{\mathfrak{v}_1} & e_{\mathfrak{v}_2} & \cdots & e_{\mathfrak{v}_{\mathfrak{m}}}\end{bmatrix}.$$

As before, without loss of generality, assume that π is regularly formed. Then the matrix $\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}$ is block diagonal where each diagonal block $\mathfrak{P}_i$ is a $|\mathfrak{C}_i| \times |\mathfrak{C}_i|$ matrix whose entries are all $\frac{1}{|\mathfrak{C}_i|}$. Let $k_i \in \{1, 2, \ldots, \mathfrak{r}\}$ be such that $\mathfrak{v}_i \in \mathfrak{C}_{k_i}$. If all the leaders are in different clusters, then

$$\mathfrak{B}^{\mathrm{T}}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{B} = \operatorname{diag}\left(\frac{1}{|\mathfrak{C}_{k_1}|}, \frac{1}{|\mathfrak{C}_{k_2}|}, \ldots, \frac{1}{|\mathfrak{C}_{k_{\mathfrak{m}}}|}\right),$$

and so

$$\lambda_{\min}\left(\mathfrak{B}^{\mathrm{T}}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{B}\right)=\min_{1\leqslant i\leqslant m}\frac{1}{|\mathfrak{C}_{k_i}|}. \tag{4.29}$$

Now suppose that two leaders $\mathfrak{v}_i$ and $\mathfrak{v}_j$ are cellmates. Then we have

$$\mathfrak{B}^{\mathrm{T}}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{B}(e_i-e_j)=\mathfrak{B}^{\mathrm{T}}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}(e_{\mathfrak{v}_i}-e_{\mathfrak{v}_j})=0.$$

which together with $\mathfrak{B}^{\mathrm{T}}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{B}\succcurlyeq 0$ implies

$$\lambda_{\min}\left(\mathfrak{B}^{\mathrm{T}}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{B}\right)=0. \tag{4.30}$$

From (4.28), (4.29), and (4.30), we find the absolute $\mathcal{H}_\infty$-error. To find the relative $\mathcal{H}_\infty$-error, we compute $\|H\|_{\mathcal{H}_\infty}$ by applying Lemma 4.2 and Lemma 4.3 to the original system (4.22). Combined with (4.26), this results in the $\mathcal{H}_\infty$-norm of the original system:

$$\|H\|_{\mathcal{H}_\infty}^2=\lambda_{\max}\left(H(0)^{\mathrm{T}}H(0)\right)=\lambda_{\max}\left(\mathfrak{B}^{\mathrm{T}}\left(I_{\mathfrak{n}}-\frac{1}{\mathfrak{n}}\mathbb{1}_{\mathfrak{n}}\mathbb{1}_{\mathfrak{n}}^{\mathrm{T}}\right)\mathfrak{B}\right)=1.$$

This completes the proof. □

4.2.6.2 The general case with symmetric agent dynamics

In this subsection, we return to the general case that the agent dynamics is given by an arbitrary multivariable input-state-output system. Thus, the original and reduced networks are again given by (4.5) and (4.6), respectively. As in the proof of Theorem 4.12, we will rely heavily on Lemma 4.3 to compute the $\mathcal{H}_\infty$-error. Since Lemma 4.3 relies on a symmetry argument, we will need to assume that the matrices A and B are both symmetric, which will be a standing assumption in the remainder of this section.

We will now establish an a priori upper bound for the $\mathcal{H}_\infty$-norm of the approximation error in the case that we cluster according to an AEP. Again, an important role is played by the $\mathfrak{n}-1$ auxiliary systems (4.7) with λ ranging over the nonzero eigenvalues of the Laplacian $\mathfrak{L}$. Again, let $H_\lambda(s)=\lambda(sI-A+\lambda B)^{-1}F$ be their transfer functions. We assume that the original network (4.5) is synchronized, so that all of the $A-\lambda B$ are Hurwitz. We again use H, $\widehat{H}$, and H_{err} to denote the relevant transfer functions.

The following is the second main theorem.

Theorem 4.14:
Assume the network (4.5) is synchronized and that A and B are symmetric matrices. Let π be an AEP of the graph $\mathfrak{G}$. The $\mathcal{H}_\infty$-error when clustering $\mathfrak{G}$ according to π then satisfies

$$\|H-\widehat{H}\|_{\mathcal{H}_\infty}^2\leqslant\begin{cases}H_{\max,\mathcal{H}_\infty}^2\max\limits_{1\leqslant i\leqslant\mathfrak{m}}\left(1-\frac{1}{|\mathfrak{C}_{k_i}|}\right) & \text{if the leaders are in different clusters,}\\ H_{\max,\mathcal{H}_\infty}^2 & \text{otherwise,}\end{cases}$$

and

$$\frac{\|H-\widehat{H}\|_{\mathcal{H}_\infty}^2}{\|H\|_{\mathcal{H}_\infty}^2} \leqslant \begin{cases} \left(\frac{H_{\max,\mathcal{H}_\infty}}{H_{\min,\mathcal{H}_\infty}}\right)^2 \max\limits_{1\leqslant i\leqslant \mathfrak{m}} \left(1-\frac{1}{|\mathfrak{C}_{k_i}|}\right) & \text{if the leaders are in different clusters,} \\ \left(\frac{H_{\max,\mathcal{H}_\infty}}{H_{\min,\mathcal{H}_\infty}}\right)^2 & \text{otherwise,} \end{cases}$$

where

$$H_{\max,\mathcal{H}_\infty} \coloneqq \max_{\lambda\in\sigma(\mathfrak{L})\setminus\sigma(\widehat{\mathfrak{L}})} \|H_\lambda\|_{\mathcal{H}_\infty}, \tag{4.31}$$

and

$$H_{\min,\mathcal{H}_\infty} \coloneqq \min_{\lambda\in\sigma(\mathfrak{L})\setminus\{0\}} \sigma_{\min}\left(H_\lambda(0)\right), \tag{4.32}$$

with H_λ the transfer functions of the auxiliary systems (4.7). ◊

Remark 4.15:
The absolute $\mathcal{H}_\infty$-error thus lies in the interval $[0, H_{\max,\mathcal{H}_\infty}]$ with $H_{\max,\mathcal{H}_\infty}$ the maximum over the $\mathcal{H}_\infty$-norms of the transfer functions H_λ with $\lambda \in \sigma(\mathfrak{L}) \setminus \sigma(\widehat{\mathfrak{L}})$. The error is minimal ($= 0$) if each leader vertex occupies a different cluster, and is the only vertex in this cluster. In general, the upper bound increases if the number of cellmates of the leaders increases. ◊

Proof of Theorem 4.14. First note that the transfer function $\widehat{H}$ of the reduced network (4.6) is equal to

$$\widehat{H}(s) = \left(\mathfrak{L}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}} \otimes I_n\right)\left(sI - I_{\mathfrak{r}}\otimes A + \overline{\mathfrak{L}}\otimes B\right)^{-1}\left(\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}}\widehat{\mathfrak{B}}\otimes F\right), \tag{4.33}$$

with the symmetric matrix $\overline{\mathfrak{L}}$ given by (4.11). Analogous to the proof of Theorem 4.12, we first apply Lemma 4.2 to the error system

$$\begin{aligned} \dot{x}_e(t) &= \begin{bmatrix} I_{\mathfrak{n}}\otimes A - \mathfrak{L}\otimes B & 0 \\ 0 & I_{\mathfrak{r}}\otimes A - \overline{\mathfrak{L}}\otimes B \end{bmatrix} x_e(t) + \begin{bmatrix} \mathfrak{B}\otimes F \\ \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}}\widehat{\mathfrak{B}}\otimes F \end{bmatrix} u(t), \\ e(t) &= \begin{bmatrix} \mathfrak{L}\otimes I_n & -\mathfrak{L}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}}\otimes I_n \end{bmatrix} x_e(t), \end{aligned}$$

with transfer function H_{err}. Take $X = I_{\mathfrak{n}}\otimes A - \mathfrak{L}\otimes B$. We then have

$$\begin{aligned} &\begin{bmatrix} \mathfrak{L}\otimes I_n & -\mathfrak{L}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}}\otimes I_n \end{bmatrix}\begin{bmatrix} I_{\mathfrak{n}}\otimes A - \mathfrak{L}\otimes B & 0 \\ 0 & I_{\mathfrak{r}}\otimes A - \overline{\mathfrak{L}}\otimes B \end{bmatrix} \\ &\quad = X\begin{bmatrix} \mathfrak{L}\otimes I_n & -\mathfrak{L}\mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}}\otimes I_n \end{bmatrix}. \end{aligned}$$

From Lemma 4.2, we thus obtain that

$$\|H_{\mathrm{err}}\|_{\mathcal{H}_\infty} = \|H_{\mathrm{err}}(0)\|_2 = \lambda_{\max}\left(H_{\mathrm{err}}(0)^{\mathrm{T}}H_{\mathrm{err}}(0)\right)^{\frac{1}{2}}.$$

In the proof of Lemma 4.8, it was shown that

$$\widehat{H}(-s)^{\mathrm{T}} H_{\mathrm{err}}(s) = \widehat{H}(-s)^{\mathrm{T}}(H(s) - \widehat{H}(s)) = 0.$$

Since all transfer functions involved are asymptotically stable, in particular this holds for $s = 0$. We then have that $\widehat{H}(0)^{\mathrm{T}}(H(0) - \widehat{H}(0)) = 0$, i.e., $\widehat{H}(0)^{\mathrm{T}} H(0) = \widehat{H}(0)^{\mathrm{T}} \widehat{H}(0)$. By transposing, we also have $H(0)^{\mathrm{T}} \widehat{H}(0) = \widehat{H}(0)^{\mathrm{T}} \widehat{H}(0)$. Therefore,

$$\begin{aligned} H_{\mathrm{err}}(0)^{\mathrm{T}} H_{\mathrm{err}}(0) &= (H(0) - \widehat{H}(0))^{\mathrm{T}}(H(0) - \widehat{H}(0)) \\ &= H(0)^{\mathrm{T}} H(0) - H(0)^{\mathrm{T}} \widehat{H}(0) - \widehat{H}(0)^{\mathrm{T}} H(0) + \widehat{H}(0)^{\mathrm{T}} \widehat{H}(0) \\ &= H(0)^{\mathrm{T}} H(0) - \widehat{H}(0)^{\mathrm{T}} \widehat{H}(0). \end{aligned}$$

By applying Lemma 4.3 to system (4.5), we obtain

$$\begin{aligned} H(0)^{\mathrm{T}} H(0) &= \left(\mathfrak{B}^{\mathrm{T}} \otimes F^{\mathrm{T}}\right) (I_{\mathfrak{n}} \otimes A - \mathfrak{L} \otimes B)^{+} \left(\mathfrak{L}^2 \otimes I_n\right) (I_{\mathfrak{n}} \otimes A - \mathfrak{L} \otimes B)^{+} (\mathfrak{B} \otimes F) \\ &= \left(\mathfrak{B}^{\mathrm{T}} \otimes F^{\mathrm{T}}\right) (U \otimes I_n)(I_{\mathfrak{n}} \otimes A - \Lambda \otimes B)^{+} \left(\Lambda^2 \otimes I_n\right) \\ &\quad \cdot (I_{\mathfrak{n}} \otimes A - \Lambda \otimes B)^{+} \left(U^{\mathrm{T}} \otimes I_n\right) (\mathfrak{B} \otimes F) \\ &= \left(\mathfrak{B}^{\mathrm{T}} U \otimes F^{\mathrm{T}}\right) \operatorname{diag}\left(0, \lambda_2^2 (A - \lambda_2 B)^{-2}, \ldots, \lambda_{\mathfrak{n}}^2 (A - \lambda_{\mathfrak{n}} B)^{-2}\right) (U^{\mathrm{T}} \mathfrak{B} \otimes F) \\ &= \left(\mathfrak{B}^{\mathrm{T}} U \otimes I_m\right) \operatorname{diag}\Big(0, H_{\lambda_2}(0)^{\mathrm{T}} H_{\lambda_2}(0), \ldots, H_{\lambda_{\mathfrak{n}}}(0)^{\mathrm{T}} H_{\lambda_{\mathfrak{n}}}(0)\Big) \left(U^{\mathrm{T}} \mathfrak{B} \otimes I_m\right), \end{aligned} \tag{4.34}$$

where H_λ is again given by (4.7). Recall that $\widehat{\mathfrak{B}} = \left(\mathfrak{P}^{\mathrm{T}} \mathfrak{P}\right)^{-1} \mathfrak{P}^{\mathrm{T}} M$ and $U_1 = \mathfrak{P} \left(\mathfrak{P}^{\mathrm{T}} \mathfrak{P}\right)^{-\frac{1}{2}} \widehat{U}$. Now apply Lemma 4.3 to the transfer function (4.33) of the system (4.6):

$$\begin{aligned} \widehat{H}(0)^{\mathrm{T}} \widehat{H}(0) &= \left(\mathfrak{B}^{\mathrm{T}} \mathfrak{P} \left(\mathfrak{P}^{\mathrm{T}} \mathfrak{P}\right)^{-\frac{1}{2}} \otimes F^{\mathrm{T}}\right) \left(I_{\mathfrak{n}} \otimes A - \overline{\mathfrak{L}} \otimes B\right)^{+} \\ &\quad \cdot \left(\left(\mathfrak{P}^{\mathrm{T}} \mathfrak{P}\right)^{-\frac{1}{2}} \mathfrak{P}^{\mathrm{T}} \mathfrak{L}^2 \mathfrak{P} \left(\mathfrak{P}^{\mathrm{T}} \mathfrak{P}\right)^{-\frac{1}{2}} \otimes I_n\right) \\ &\quad \cdot \left(I_{\mathfrak{n}} \otimes A - \overline{\mathfrak{L}} \otimes B\right)^{+} \left(\left(\mathfrak{P}^{\mathrm{T}} \mathfrak{P}\right)^{-\frac{1}{2}} \mathfrak{P}^{\mathrm{T}} \mathfrak{B} \otimes F\right) \\ &= \left(\mathfrak{B}^{\mathrm{T}} \mathfrak{P} \left(\mathfrak{P}^{\mathrm{T}} \mathfrak{P}\right)^{-\frac{1}{2}} \otimes F^{\mathrm{T}}\right) \left(\widehat{U} \otimes I_n\right) \left(I_{\mathfrak{n}} \otimes A - \widehat{\Lambda} \otimes B\right)^{+} \\ &\quad \cdot \left(\widehat{\Lambda}^2 \otimes I_n\right) \left(I_{\mathfrak{n}} \otimes A - \widehat{\Lambda} \otimes B\right)^{+} \\ &\quad \cdot \left(\widehat{U}^{\mathrm{T}} \otimes I_n\right) \left(\left(\mathfrak{P}^{\mathrm{T}} \mathfrak{P}\right)^{-\frac{1}{2}} \mathfrak{P}^{\mathrm{T}} \mathfrak{B} \otimes F\right) \\ &= \left(\mathfrak{B}^{\mathrm{T}} U_1 \otimes F^{\mathrm{T}}\right) \operatorname{diag}\left(0, \lambda_2^2 (A - \lambda_2 B)^{-2}, \ldots, \lambda_{\mathfrak{r}}^2 (A - \lambda_{\mathfrak{r}} B)^{-2}\right) \left(U_1^{\mathrm{T}} \mathfrak{B} \otimes F\right) \\ &= \left(\mathfrak{B}^{\mathrm{T}} U_1 \otimes I_m\right) \operatorname{diag}\Big(0, H_{\lambda_2}(0)^{\mathrm{T}} H_{\lambda_2}(0), \ldots, H_{\lambda_{\mathfrak{r}}}(0)^{\mathrm{T}} H_{\lambda_{\mathfrak{r}}}(0)\Big) \left(U_1^{\mathrm{T}} \mathfrak{B} \otimes I_m\right). \end{aligned}$$

Combining the two expressions above, it immediately follows that

$$\begin{aligned} H_{\mathrm{err}}(0)^{\mathrm{T}} H_{\mathrm{err}}(0) &= H(0)^{\mathrm{T}} H(0) - \widehat{H}(0)^{\mathrm{T}} \widehat{H}(0) \\ &= \left(\mathfrak{B}^{\mathrm{T}} U_2 \otimes I_m\right) \operatorname{diag}\Big(H_{\lambda_{\mathfrak{r}+1}}(0)^{\mathrm{T}} H_{\lambda_{\mathfrak{r}+1}}(0), \ldots, H_{\lambda_N}(0)^{\mathrm{T}} H_{\lambda_N}(0)\Big) \\ &\quad \cdot \left(U_2^{\mathrm{T}} \mathfrak{B} \otimes I_m\right). \end{aligned}$$

By taking $H_{\max,\mathcal{H}_\infty}$ as defined by (4.31) it then follows that

$$\begin{aligned} H_{\mathrm{err}}(0)^{\mathrm{T}} H_{\mathrm{err}}(0) &\preccurlyeq \left(\mathfrak{B}^{\mathrm{T}} U_2 \otimes I_m\right) \operatorname{diag}\left(H_{\max,\mathcal{H}_\infty}^2 I_m, \ldots, H_{\max,\mathcal{H}_\infty}^2 I_m\right) \left(U_2^{\mathrm{T}} \mathfrak{B} \otimes I_m\right) \\ &= H_{\max,\mathcal{H}_\infty}^2 \left(\mathfrak{B}^{\mathrm{T}} U_2 U_2^{\mathrm{T}} \mathfrak{B} \otimes I_m\right) \\ &= H_{\max,\mathcal{H}_\infty}^2 \left(\mathfrak{B}^{\mathrm{T}} (I_{\mathfrak{n}} - U_1 U_1^{\mathrm{T}}) \mathfrak{B} \otimes I_m\right) \\ &= H_{\max,\mathcal{H}_\infty}^2 \left(\left(I_{\mathfrak{m}} - \mathfrak{B}^{\mathrm{T}} \mathfrak{P} \left(\mathfrak{P}^{\mathrm{T}} \mathfrak{P}\right)^{-1} \mathfrak{P}^{\mathrm{T}} \mathfrak{B}\right) \otimes I_m\right). \end{aligned}$$

Continuing as in the proof of Theorem 4.12, we find an upper bound for the $\mathcal{H}_\infty$-error:

$$\|H_{\mathrm{err}}\|_{\mathcal{H}_\infty}^2 \leqslant H_{\max,\mathcal{H}_\infty}^2 \lambda_{\max}\left(I_{\mathfrak{m}} - \mathfrak{B}^{\mathrm{T}} \mathfrak{P} \left(\mathfrak{P}^{\mathrm{T}} \mathfrak{P}\right)^{-1} \mathfrak{P}^{\mathrm{T}} \mathfrak{B}\right).$$

To compute an upper bound for the relative $\mathcal{H}_\infty$-error, we bound the $\mathcal{H}_\infty$-norm of system (4.5) from below. Again, let $\overline{U}$ be such that $U = \begin{bmatrix} u_1 & \overline{U} \end{bmatrix}$ and let $H_{\min,\mathcal{H}_\infty}$ be as defined by (4.32). From (4.34) it now follows that

$$\begin{aligned} H(0)^{\mathrm{T}} H(0) &= \left(\mathfrak{B}^{\mathrm{T}} \overline{U} \otimes I_m\right) \operatorname{diag}\left(H_{\lambda_2}(0)^{\mathrm{T}} H_{\lambda_2}(0), \ldots, H_{\lambda_{\mathfrak{n}}}(0)^{\mathrm{T}} H_{\lambda_{\mathfrak{n}}}(0)\right) \left(\overline{U}^{\mathrm{T}} \mathfrak{B} \otimes I_m\right) \\ &\succcurlyeq \left(\mathfrak{B}^{\mathrm{T}} \overline{U} \otimes I_m\right) \operatorname{diag}\left(H_{\min,\mathcal{H}_\infty}^2 I_m, \ldots, H_{\min,\mathcal{H}_\infty}^2 I_m\right) \left(\overline{U}^{\mathrm{T}} \mathfrak{B} \otimes I_m\right) \\ &= \left(H_{\min,\mathcal{H}_\infty}\right)^2 \left(\mathfrak{B}^{\mathrm{T}} \overline{U} \overline{U}^{\mathrm{T}} \mathfrak{B} \otimes I_m\right) \\ &= \left(H_{\min,\mathcal{H}_\infty}\right)^2 \left(\mathfrak{B}^{\mathrm{T}} \left(I_{\mathfrak{n}} - \frac{1}{\mathfrak{n}} \mathbb{1}_{\mathfrak{n}} \mathbb{1}_{\mathfrak{n}}^{\mathrm{T}}\right) \mathfrak{B} \otimes I_m\right). \end{aligned}$$

Again using Lemma 4.3, we find a lower bound to the $\mathcal{H}_\infty$-norm of H:

$$\|H\|_{\mathcal{H}_\infty}^2 = \lambda_{\max}\left(H(0)^{\mathrm{T}} H(0)\right) \geqslant H_{\min,\mathcal{H}_\infty}^2,$$

which concludes the proof of the theorem. □

4.2.7 Towards a priori error bounds for general graph partitions

Up to now, we have only dealt with establishing error bounds for network reduction by clustering using almost equitable partitions of the network graph. Of course, we would also like to obtain error bounds for *arbitrary*, possibly non almost equitable, partitions. In this section, we present some ideas to address this more general problem. We will first study the single integrator case. Subsequently, we will look at the general case.

4.2.7.1 The single integrator case

Consider the multi-agent network

$$\begin{aligned} \dot{x}(t) &= -\mathfrak{L} x(t) + \mathfrak{B} u(t), \\ y(t) &= \mathfrak{L} x(t). \end{aligned} \tag{4.35}$$

As before, assume that the underlying graph $\mathfrak{G}$ is connected. The network is then synchronized. Let $\pi = \{\mathfrak{C}_1, \mathfrak{C}_2, \ldots, \mathfrak{C}_\mathfrak{r}\}$ be a graph partition, not necessarily an AEP, and let $\mathfrak{P} \in \mathbb{R}^{\mathfrak{n}\times\mathfrak{r}}$ be its characteristic matrix. As before, the reduced order network is taken to be the Petrov-Galerkin projection of (4.35), and is represented by

$$\begin{aligned} \dot{\widehat{x}}(t) &= -\widehat{\mathfrak{L}}\widehat{x}(t) + \widehat{\mathfrak{B}}u(t), \\ \widehat{y}(t) &= \mathfrak{L}\mathfrak{P}\widehat{x}(t), \end{aligned} \tag{4.36}$$

Again, let H and $\widehat{H}$ be the transfer functions of (4.35) and (4.36), respectively. We will address the problem of obtaining a priori upper bounds for $\|H - \widehat{H}\|_{\mathcal{H}_2}$ and $\|H - \widehat{H}\|_{\mathcal{H}_\infty}$. We will pursue the following idea: as a first step we will approximate the original Laplacian matrix $\mathfrak{L}$ (of the original network graph $\mathfrak{G}$) by a new Laplacian matrix, denoted by $\mathfrak{L}_{\text{AEP}}$ (corresponding to a 'nearby' graph $\mathfrak{G}_{\text{AEP}}$) such that the given partition π is an AEP for this new graph $\mathfrak{G}_{\text{AEP}}$. This new graph $\mathfrak{G}_{\text{AEP}}$ defines a new multi-agent system with transfer function $H_{\text{AEP}}(s) = \mathfrak{L}_{\text{AEP}}(sI_\mathfrak{n} + \mathfrak{L}_{\text{AEP}})^{-1}\mathfrak{B}$. The reduced order network of H_{AEP} (using the AEP π) has transfer function $\widehat{H}_{\text{AEP}}(s) = \mathfrak{L}_{\text{AEP}}\mathfrak{P}\left(sI_\mathfrak{r} + \widehat{\mathfrak{L}}_{\text{AEP}}\right)^{-1}\widehat{\mathfrak{B}}$. Then using the triangle inequality, both for $p = 2$ and $p = \infty$, we have

$$\begin{aligned} \|H - \widehat{H}\|_{\mathcal{H}_p} &= \|H - H_{\text{AEP}} + H_{\text{AEP}} - \widehat{H}_{\text{AEP}} + \widehat{H}_{\text{AEP}} - \widehat{H}\|_{\mathcal{H}_p} \\ &\leqslant \|H - H_{\text{AEP}}\|_{\mathcal{H}_p} + \|H_{\text{AEP}} - \widehat{H}_{\text{AEP}}\|_{\mathcal{H}_p} + \|\widehat{H}_{\text{AEP}} - \widehat{H}\|_{\mathcal{H}_p}. \end{aligned} \tag{4.37}$$

The idea is to obtain a priori upper bounds for all three terms in (4.37). We first propose an approximating Laplacian matrix $\mathfrak{L}_{\text{AEP}}$, and subsequently study the problems of establishing upper bounds for the three terms in (4.37) separately.

In the following, denote $\mathcal{P} \coloneqq \mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}$. Note that $\mathcal{P}$ is the orthogonal projector onto $\operatorname{im}(\mathfrak{P})$. As approximation for $\mathfrak{L}$, we compute the unique solution to the convex optimization problem

$$\begin{aligned} \underset{\mathfrak{L}_{\text{AEP}}}{\text{minimize}} \quad & \|\mathfrak{L} - \mathfrak{L}_{\text{AEP}}\|_{\mathrm{F}}^2, \\ \text{subject to} \quad & (I_\mathfrak{n} - \mathcal{P})\mathfrak{L}_{\text{AEP}}\mathfrak{P} = 0, \\ & \mathfrak{L}_{\text{AEP}} = \mathfrak{L}_{\text{AEP}}^{\mathrm{T}}, \\ & \mathfrak{L}_{\text{AEP}} \succcurlyeq 0, \\ & \mathfrak{L}_{\text{AEP}}\mathbb{1}_\mathfrak{n} = 0. \end{aligned} \tag{4.38}$$

In other words, we want to compute a positive semi-definite matrix $\mathfrak{L}_{\text{AEP}}$ with row sums equal to zero, and with the property that $\operatorname{im}(\mathfrak{P})$ is invariant under $\mathfrak{L}_{\text{AEP}}$ (equivalently, the given partition π is an AEP for the new graph). We will show that such an $\mathfrak{L}_{\text{AEP}}$ may correspond to an undirected graph *with negative weights*. However, it is constrained to be positive semi-definite, so the results of Sections 4.2.4 to 4.2.6 in this section will remain valid.

Theorem 4.16:
The matrix $\mathfrak{L}_{\text{AEP}} \coloneqq \mathcal{P}\mathfrak{L}\mathcal{P} + (I_\mathfrak{n} - \mathcal{P})\mathfrak{L}(I_\mathfrak{n} - \mathcal{P})$ is the unique solution to the convex optimization problem (4.38). If $\mathfrak{L}$ corresponds to a connected graph, then, in fact, $\ker(\mathfrak{L}_{\text{AEP}}) = \operatorname{im}(\mathbb{1}_\mathfrak{n})$. ◊

Proof. Clearly, $\mathfrak{L}_{\mathrm{AEP}}$ is symmetric and positive semi-definite since $\mathfrak{L}$ is. Also, $(I_\mathfrak{n} - \mathcal{P})\mathfrak{L}_{\mathrm{AEP}}\mathfrak{P} = 0$ since $(I_\mathfrak{n} - \mathcal{P})\mathfrak{P} = 0$. It is also obvious that $\mathfrak{L}_{\mathrm{AEP}}\mathbb{1}_\mathfrak{n} = 0$ since $\mathcal{P}\mathbb{1}_\mathfrak{n} = \mathbb{1}_\mathfrak{n}$. We now show that $\mathfrak{L}_{\mathrm{AEP}}$ uniquely minimizes the distance to $\mathfrak{L}$. Let X satisfy the constraints and define $\Delta = \mathfrak{L}_{\mathrm{AEP}} - X$. Then we have

$$\|\mathfrak{L} - X\|_{\mathrm{F}}^2 = \|\mathfrak{L} - \mathfrak{L}_{\mathrm{AEP}}\|_{\mathrm{F}}^2 + \|\Delta\|_{\mathrm{F}}^2 + 2\,\mathrm{tr}((\mathfrak{L} - \mathfrak{L}_{\mathrm{AEP}})\Delta)\,.$$

It can be verified that $\mathfrak{L} - \mathfrak{L}_{\mathrm{AEP}} = (I_\mathfrak{n} - \mathcal{P})\mathfrak{L}\mathcal{P} + \mathcal{P}\mathfrak{L}(I_\mathfrak{n} - \mathcal{P})$. Thus,

$$\mathrm{tr}((\mathfrak{L} - \mathfrak{L}_{\mathrm{AEP}})\Delta) = \mathrm{tr}((I_\mathfrak{n} - \mathcal{P})\mathfrak{L}\mathcal{P}\Delta) + \mathrm{tr}(\mathcal{P}\mathfrak{L}(I_\mathfrak{n} - \mathcal{P})\Delta)\,.$$

Now, since both X and $\mathfrak{L}_{\mathrm{AEP}}$ satisfy the first constraint, we have $(I_\mathfrak{n} - \mathcal{P})\Delta\mathcal{P} = 0$. Using this we have

$$\mathrm{tr}((I_\mathfrak{n} - \mathcal{P})\mathfrak{L}\mathcal{P}\Delta) = \mathrm{tr}(\mathcal{P}\Delta(I_\mathfrak{n} - \mathcal{P})\mathfrak{L}) = \mathrm{tr}(\mathfrak{L}(I_\mathfrak{n} - \mathcal{P})\Delta\mathcal{P}) = 0.$$

Also,

$$\mathrm{tr}(\mathcal{P}\mathfrak{L}(I_\mathfrak{n} - \mathcal{P})\Delta) = \mathrm{tr}(\mathfrak{L}(I_\mathfrak{n} - \mathcal{P})\Delta\mathcal{P}) = 0.$$

Thus, we obtain

$$\|\mathfrak{L} - X\|_{\mathrm{F}}^2 = \|\mathfrak{L} - \mathfrak{L}_{\mathrm{AEP}}\|_{\mathrm{F}}^2 + \|\Delta\|_{\mathrm{F}}^2\,,$$

from which it follows that $\|\mathfrak{L} - X\|_{\mathrm{F}}$ is minimal if and only if $\Delta = 0$, equivalently, $X = \mathfrak{L}_{\mathrm{AEP}}$.

To prove the second statement, let $x \in \ker(\mathfrak{L}_{\mathrm{AEP}})$, so $x^{\mathrm{T}}\mathfrak{L}_{\mathrm{AEP}}x = 0$. Then both $x^{\mathrm{T}}\mathcal{P}\mathfrak{L}\mathcal{P}x = 0$ and $x^{\mathrm{T}}(I_\mathfrak{n} - \mathcal{P})\mathfrak{L}(I_\mathfrak{n} - \mathcal{P})x = 0$. This clearly implies $\mathfrak{L}\mathcal{P}x = 0$ and $\mathfrak{L}(I_\mathfrak{n} - \mathcal{P})x = 0$. Since $\mathfrak{L}$ corresponds to a connected graph, we must have $\mathcal{P}x \in \mathrm{im}(\mathbb{1}_\mathfrak{n})$ and $(I_\mathfrak{n} - \mathcal{P})x \in \mathrm{im}(\mathbb{1}_\mathfrak{n})$. We conclude that $x \in \mathrm{im}(\mathbb{1}_\mathfrak{n})$, as desired. □

As announced above, $\mathfrak{L}_{\mathrm{AEP}}$ may have positive off-diagonal elements, corresponding to a graph with some of its edge weights being negative. For example, for

$$\mathfrak{L} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}, \quad \mathfrak{P} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix},$$

we have

$$\mathfrak{L}_{\mathrm{AEP}} = \begin{bmatrix} \frac{11}{9} & -\frac{7}{9} & -\frac{1}{9} & 0 & -\frac{1}{3} \\ -\frac{7}{9} & \frac{20}{9} & -\frac{10}{9} & 0 & -\frac{1}{3} \\ -\frac{1}{9} & -\frac{10}{9} & \frac{14}{9} & -\frac{1}{2} & \frac{1}{6} \\ 0 & 0 & -\frac{1}{2} & \frac{3}{2} & -1 \\ -\frac{1}{3} & -\frac{1}{3} & \frac{1}{6} & -1 & \frac{3}{2} \end{bmatrix},$$

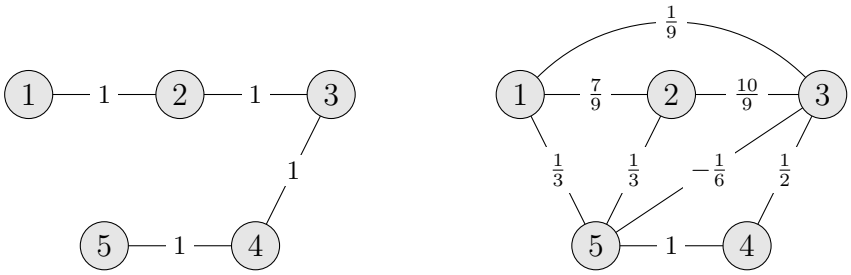

Figure 4.1: A path graph on 5 vertices and its closest graph such that the partition $\{\{1,2,3\},\{4,5\}\}$ is almost equitable.

so the edge between vertices 3 and 5 has a negative weight. Figure 4.1 shows the graphs corresponding to $\mathfrak{L}$ and $\mathfrak{L}_{\mathrm{AEP}}$. Although $\mathfrak{L}_{\mathrm{AEP}}$ is not necessarily a Laplacian matrix with only nonpositive off-diagonal elements, it has all the properties we associate with a Laplacian matrix. Specifically, it can be checked that all results in this section remain valid, since they only depend on the symmetric positive semi-definiteness of the Laplacian matrix.

Using the approximating Laplacian $\mathfrak{L}_{\mathrm{AEP}} = \mathcal{P}\mathfrak{L}\mathcal{P} + (I_{\mathfrak{n}} - \mathcal{P})\mathfrak{L}(I_{\mathfrak{n}} - \mathcal{P})$ as above, we will now deal with establishing upper bounds for the three terms in (4.37). We start off with the middle term $\|H_{\mathrm{AEP}} - \widehat{H}_{\mathrm{AEP}}\|_{\mathcal{H}_p}$ in (4.37).

According to Remark 4.7, for $p = 2$ this term has an upper bound depending on the maximal $\lambda \in \sigma(\mathfrak{L}_{\mathrm{AEP}}) \setminus \sigma(\widehat{\mathfrak{L}}_{\mathrm{AEP}})$, and on the number of cellmates of the leaders with respect to the partitioning π. For $p = \infty$, in Theorem 4.12 this term was expressed in terms of the maximal number of cellmates with respect to the partitioning π (noting that it is equal to 1 in case two or more leaders share the same cluster).

Next, we will take a look at the first and third term in (4.37), i.e., $\|H - H_{\mathrm{AEP}}\|_{\mathcal{H}_p}$ and $\|\widehat{H} - \widehat{H}_{\mathrm{AEP}}\|_{\mathcal{H}_p}$. Let us denote $\Delta\mathfrak{L} = \mathfrak{L} - \mathfrak{L}_{\mathrm{AEP}}$. We find

$$\begin{aligned}
H(s) - H_{\mathrm{AEP}}(s) &= \mathfrak{L}(sI_{\mathfrak{n}} + \mathfrak{L})^{-1}\mathfrak{B} - \mathfrak{L}_{\mathrm{AEP}}(sI_{\mathfrak{n}} + \mathfrak{L}_{\mathrm{AEP}})^{-1}\mathfrak{B} \\
&= \mathfrak{L}(sI_{\mathfrak{n}} + \mathfrak{L})^{-1}\mathfrak{B} \\
&\quad - \mathfrak{L}_{\mathrm{AEP}}\left[(sI_{\mathfrak{n}} + \mathfrak{L})^{-1} + (sI_{\mathfrak{n}} + \mathfrak{L}_{\mathrm{AEP}})^{-1}\Delta\mathfrak{L}(sI_{\mathfrak{n}} + \mathfrak{L})^{-1}\right]\mathfrak{B} \\
&= \mathfrak{L}(sI_{\mathfrak{n}} + \mathfrak{L})^{-1}\mathfrak{B} - \mathfrak{L}_{\mathrm{AEP}}(sI_{\mathfrak{n}} + \mathfrak{L})^{-1}\mathfrak{B} \\
&\quad - \mathfrak{L}_{\mathrm{AEP}}(sI_{\mathfrak{n}} + \mathfrak{L}_{\mathrm{AEP}})^{-1}\Delta\mathfrak{L}(sI_{\mathfrak{n}} + \mathfrak{L})^{-1}\mathfrak{B} \\
&= \Delta\mathfrak{L}(sI_{\mathfrak{n}} + \mathfrak{L})^{-1}\mathfrak{B} - \mathfrak{L}_{\mathrm{AEP}}(sI_{\mathfrak{n}} + \mathfrak{L}_{\mathrm{AEP}})^{-1}\Delta\mathfrak{L}(sI_{\mathfrak{n}} + \mathfrak{L})^{-1}\mathfrak{B} \\
&= \left[I_{\mathfrak{n}} - \mathfrak{L}_{\mathrm{AEP}}(sI_{\mathfrak{n}} + \mathfrak{L}_{\mathrm{AEP}})^{-1}\right]\Delta\mathfrak{L}(sI_{\mathfrak{n}} + \mathfrak{L})^{-1}\mathfrak{B}.
\end{aligned}$$

Thus, both for $p = 2$ and $p = \infty$, we have

$$\|H - H_{\mathrm{AEP}}\|_{\mathcal{H}_p} \leqslant \left\|I_{\mathfrak{n}} - \mathfrak{L}_{\mathrm{AEP}}(sI_{\mathfrak{n}} + \mathfrak{L}_{\mathrm{AEP}})^{-1}\right\|_{\mathcal{H}_\infty} \left\|\Delta\mathfrak{L}(sI_{\mathfrak{n}} + \mathfrak{L})^{-1}\mathfrak{B}\right\|_{\mathcal{H}_p}$$

Using

$$\left\|I_{\mathfrak{n}} - \mathfrak{L}_{\mathrm{AEP}}(sI_{\mathfrak{n}} + \mathfrak{L}_{\mathrm{AEP}})^{-1}\right\|_{\mathcal{H}_\infty} = \left\|s(sI_{\mathfrak{n}} + \mathfrak{L}_{\mathrm{AEP}})^{-1}\right\|_{\mathcal{H}_\infty} = \sup_{\omega\in\mathbb{R}} \left\|\imath\omega(\imath\omega I_{\mathfrak{n}} + \mathfrak{L}_{\mathrm{AEP}})^{-1}\right\|_2$$

$$= \sup_{\omega\in\mathbb{R}} \frac{|\omega|}{\sqrt{\lambda_{\min}(\omega^2 I_{\mathfrak{n}} + \mathfrak{L}_{\mathrm{AEP}}^2)}} = 1,$$

we get

$$\|H - H_{\mathrm{AEP}}\|_{\mathcal{H}_p} \leqslant \left\|\Delta\mathfrak{L}(sI_{\mathfrak{n}} + \mathfrak{L})^{-1}\mathfrak{B}\right\|_{\mathcal{H}_p}. \tag{4.39}$$

It is also easily seen that $\widehat{\mathfrak{L}}_{\mathrm{AEP}} = \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{L}_{\mathrm{AEP}}\mathfrak{P} = \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{L}\mathfrak{P} = \widehat{\mathfrak{L}}$ and $\mathfrak{L}_{\mathrm{AEP}}\mathfrak{P} = \mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1}\mathfrak{P}^{\mathrm{T}}\mathfrak{L}\mathfrak{P} = \mathfrak{P}\widehat{\mathfrak{L}}$. Therefore,

$$\begin{aligned}
\widehat{H}(s) - \widehat{H}_{\mathrm{AEP}}(s) &= \mathfrak{L}\mathfrak{P}\left(sI_{\mathfrak{n}} + \widehat{\mathfrak{L}}\right)^{-1}\widehat{\mathfrak{B}} - \mathfrak{L}_{\mathrm{AEP}}\mathfrak{P}\left(sI_{\mathfrak{n}} + \widehat{\mathfrak{L}}_{\mathrm{AEP}}\right)^{-1}\widehat{\mathfrak{B}} \\
&= \mathfrak{L}\mathfrak{P}\left(sI_{\mathfrak{n}} + \widehat{\mathfrak{L}}\right)^{-1}\widehat{\mathfrak{B}} - \mathfrak{P}\widehat{\mathfrak{L}}\left(sI_{\mathfrak{n}} + \widehat{\mathfrak{L}}\right)^{-1}\widehat{\mathfrak{B}} \\
&= \left(\mathfrak{L}\mathfrak{P} - \mathfrak{P}\widehat{\mathfrak{L}}\right)\left(sI_{\mathfrak{n}} + \widehat{\mathfrak{L}}\right)^{-1}\widehat{\mathfrak{B}}.
\end{aligned}$$

Since $\left(\mathfrak{L}\mathfrak{P} - \mathfrak{P}\widehat{\mathfrak{L}}\right)^{\mathrm{T}}\left(\mathfrak{L}\mathfrak{P} - \mathfrak{P}\widehat{\mathfrak{L}}\right) = \mathfrak{P}^{\mathrm{T}}(\Delta\mathfrak{L})^2\mathfrak{P}$, for $p = 2$ and $p = \infty$, we obtain

$$\|\widehat{H} - \widehat{H}_{\mathrm{AEP}}\|_{\mathcal{H}_p} = \left\|\Delta\mathfrak{L}\mathfrak{P}(sI_{\mathfrak{n}} + \widehat{\mathfrak{L}})^{-1}\widehat{\mathfrak{B}}\right\|_{\mathcal{H}_p}. \tag{4.40}$$

Finally, combining (4.37), (4.39) and (4.40), we get

$$\|H - \widehat{H}\|_{\mathcal{H}_p} \leqslant \left\|\Delta\mathfrak{L}(sI_{\mathfrak{n}} + \mathfrak{L})^{-1}\mathfrak{B}\right\|_{\mathcal{H}_p} + \|H_{\mathrm{AEP}} - \widehat{H}_{\mathrm{AEP}}\|_{\mathcal{H}_p} + \left\|\Delta\mathfrak{L}\mathfrak{P}(sI_{\mathfrak{n}} + \widehat{\mathfrak{L}})^{-1}\widehat{\mathfrak{B}}\right\|_{\mathcal{H}_p}.$$

Thus, both in (4.39) and (4.40) the upper bound involves the difference $\Delta\mathfrak{L} = \mathfrak{L} - \mathfrak{L}_{\mathrm{AEP}}$ between the original Laplacian and its optimal approximation in the set of Laplacian matrices for which the given partition π is an AEP. In a sense, the difference $\Delta\mathfrak{L}$ measures how far π is away from being an AEP for the original graph $\mathfrak{G}$. Obviously, $\Delta\mathfrak{L} = 0$ if and only if π is an AEP for $\mathfrak{G}$. In that case only the middle term in (4.37) is present.

4.2.7.2 The general case

In this final subsection, we will put forward some ideas to deal with the case that the agent dynamics is a general linear input-state-output system and the given graph partition π, with characteristic matrix $\mathfrak{P}$, is not almost equitable. In this case, the original network is given by (4.5) and the reduced network by (4.6). Their transfer functions are H and $\widehat{H}$, respectively. Let $\mathfrak{L}_{\mathrm{AEP}}$ and $\widehat{\mathfrak{L}}_{\mathrm{AEP}}$ as in the previous subsection and let

$$H_{\mathrm{AEP}}(s) = (\mathfrak{L}_{\mathrm{AEP}} \otimes I_n)(sI - I_{\mathfrak{n}} \otimes A + \mathfrak{L}_{\mathrm{AEP}} \otimes B)^{-1}(\mathfrak{B} \otimes F)$$

and

$$\widehat{H}_{\mathrm{AEP}}(s) = (\mathfrak{L}_{\mathrm{AEP}}\mathfrak{P} \otimes I_n)\left(sI - I_{\mathfrak{r}} \otimes A + \widehat{\mathfrak{L}}_{\mathrm{AEP}} \otimes B\right)^{-1}(\widehat{\mathfrak{B}} \otimes F).$$

As before, we assume that (4.5) is synchronized, so H is asymptotically stable. However, since the partition π is no longer assumed to be an AEP, the reduced transfer function $\widehat{H}$ need not be asymptotically stable anymore. Also, H_{AEP} and $\widehat{H}_{\mathrm{AEP}}$ need not be asymptotically stable. We will now first study under what conditions these are asymptotically stable. First note that, using Proposition 2.51, $\widehat{H}$ is asymptotically stable if and only if $A-\widehat{\lambda}B$ is Hurwitz for all nonzero eigenvalues $\widehat{\lambda}$ of $\widehat{\mathfrak{L}}$. Moreover, H_{AEP} and $\widehat{H}_{\mathrm{AEP}}$ are asymptotically stable if and only if $A-\lambda B$ is Hurwitz for all nonzero eigenvalues λ of $\mathfrak{L}_{\mathrm{AEP}}$. In the following, let $\lambda_{\min}(\mathfrak{L})$ and $\lambda_{\max}(\mathfrak{L})$ denote the smallest nonzero and largest eigenvalue of $\mathfrak{L}$, respectively. We have the following lemma about the location of the nonzero eigenvalues of $\widehat{\mathfrak{L}}$ and $\mathfrak{L}_{\mathrm{AEP}}$.

Lemma 4.17:
All nonzero eigenvalues of $\widehat{\mathfrak{L}}$ and of $\mathfrak{L}_{\mathrm{AEP}}$ lie in the closed interval $[\lambda_{\min}(\mathfrak{L}), \lambda_{\max}(\mathfrak{L})]$.$\Diamond$

Proof. The claim about the eigenvalues of $\widehat{\mathfrak{L}}$ follows from Theorem 2.6. Next, note that $\mathcal{P} = Q_1 Q_1^{\mathrm{T}}$, with $Q_1 = \mathfrak{P}\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-\frac{1}{2}}$. Since the columns of Q_1 are orthonormal, there exists a matrix $Q_2 \in \mathbb{R}^{\mathfrak{n}\times(\mathfrak{n}-\mathfrak{r})}$ such that $\begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$ is an orthogonal matrix. Then, we have $I_{\mathfrak{n}} - \mathcal{P} = Q_2 Q_2^{\mathrm{T}}$ and we find

$$\begin{aligned}
\mathfrak{L}_{\mathrm{AEP}} &= \mathcal{P}\mathfrak{L}\mathcal{P} + (I_{\mathfrak{n}} - \mathcal{P})\mathfrak{L}(I_{\mathfrak{n}} - \mathcal{P}) \\
&= Q_1 Q_1^{\mathrm{T}} \mathfrak{L} Q_1 Q_1^{\mathrm{T}} + Q_2 Q_2^{\mathrm{T}} \mathfrak{L} Q_2 Q_2^{\mathrm{T}} \\
&= \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} Q_1^{\mathrm{T}}\mathfrak{L}Q_1 & 0 \\ 0 & Q_2^{\mathrm{T}}\mathfrak{L}Q_2 \end{bmatrix} \begin{bmatrix} Q_1^{\mathrm{T}} \\ Q_2^{\mathrm{T}} \end{bmatrix}.
\end{aligned}$$

It follows that $\sigma(\mathfrak{L}_{\mathrm{AEP}}) = \sigma(Q_1^{\mathrm{T}}\mathfrak{L}Q_1) \cup \sigma(Q_2^{\mathrm{T}}\mathfrak{L}Q_2)$. By Theorem 2.6, both the eigenvalues of $Q_1^{\mathrm{T}}\mathfrak{L}Q_1$ and $Q_2^{\mathrm{T}}\mathfrak{L}Q_2$ are interlaced with the eigenvalues of $\mathfrak{L}$, so in particular we have that all eigenvalues λ of $\mathfrak{L}_{\mathrm{AEP}}$ satisfy $\lambda \leqslant \lambda_{\max}(\mathfrak{L})$. In order to prove the lower bound, note that $Q_1^{\mathrm{T}}\mathfrak{L}Q_1$ is similar to $\widehat{\mathfrak{L}}$, for which we know that its nonzero eigenvalues are between the nonzero eigenvalues of $\mathfrak{L}$. As for the eigenvalues of $Q_2^{\mathrm{T}}\mathfrak{L}Q_2$, note that $\|Q_2 x\|_2 = \|x\|_2$ for all $x \in \mathbb{R}^{\mathfrak{n}-\mathfrak{r}}$ and $\mathbb{1}^{\mathrm{T}} Q_2 = 0$ (since $Q_1\left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{\frac{1}{2}}\mathbb{1} = \mathbb{1}$). Thus, we find

$$\min_{\|x\|_2=1} x^{\mathrm{T}} Q_2^{\mathrm{T}} \mathfrak{L} Q_2 x \geqslant \min_{\substack{\mathbb{1}^{\mathrm{T}} y = 0 \\ \|y\|_2 = 1}} y^{\mathrm{T}} \mathfrak{L} y = \lambda_{\min}(\mathfrak{L}).$$

Therefore, the smallest eigenvalue of $Q_2^{\mathrm{T}}\mathfrak{L}Q_2$ is larger than the smallest positive eigenvalue of $\mathfrak{L}$. We conclude that indeed $\lambda \geqslant \lambda_{\min}(\mathfrak{L})$ for all nonzero eigenvalues λ of $\mathfrak{L}_{\mathrm{AEP}}$. $\square$

Using this lemma, we see that a sufficient condition for $\widehat{H}$, H_{AEP}, and $\widehat{H}_{\mathrm{AEP}}$ to be asymptotically stable is that for each $\lambda \in [\lambda_{\min}(\mathfrak{L}), \lambda_{\max}(\mathfrak{L})]$, the strict Lyapunov inequality

$$(A - \lambda B)X + X(A - \lambda B)^{\mathrm{T}} \prec 0$$

has a positive definite solution X. This sufficient condition can be checked by verifying solvability of a *single* linear matrix inequality, whose size does not depend on the number of agents, see [OP05]. After having checked this, it would then remain to establish upper bounds for the first and third term in (4.37). This can be done in an analogous way as in the previous subsection. Specifically, it can be shown that for $p = 2$ and $p = \infty$ we have

$$\begin{aligned}\|H - H_{\mathrm{AEP}}\|_{\mathcal{H}_p} \leqslant \Big(1 + \big\|(\mathfrak{L}_{\mathrm{AEP}} \otimes I_n)(sI - I_{\mathfrak{n}} \otimes A + \mathfrak{L}_{\mathrm{AEP}} \otimes B)^{-1}(I_{\mathfrak{n}} \otimes B)\big\|_{\mathcal{H}_\infty}\Big) \\ \cdot \big\|(\Delta\mathfrak{L} \otimes I_n)(sI - I_{\mathfrak{n}} \otimes A + \mathfrak{L} \otimes B)^{-1}(\mathfrak{B} \otimes F)\big\|_{\mathcal{H}_p}\end{aligned}$$

and

$$\|\widehat{H} - \widehat{H}_{\mathrm{AEP}}\|_{\mathcal{H}_p} = \left\|(\Delta\mathfrak{L}\mathfrak{P} \otimes I_n)\left(sI - I_{\mathfrak{r}} \otimes A + \widehat{\mathfrak{L}} \otimes B\right)^{-1}(\widehat{\mathfrak{B}} \otimes F)\right\|_{\mathcal{H}_p}.$$

4.2.8 Numerical examples

To illustrate the error bounds we have established in this section, consider the graph with 10 vertices taken from [MTC14], as shown in Figure 3.1. Its Laplacian matrix is

$$\mathfrak{L} = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 & -5 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & -3 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & -1 & -2 & -3 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 6 & -5 & 0 & 0 & 0 & 0 & 0 \\ 0 & -3 & -2 & -5 & 25 & -2 & -6 & -7 & 0 & 0 \\ -5 & -2 & -3 & 0 & -2 & 25 & -6 & -7 & 0 & 0 \\ 0 & 0 & 0 & 0 & -6 & -6 & 15 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & -7 & -7 & -1 & 15 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \end{bmatrix},$$

with spectrum (rounded to three significant digits)

$$\sigma(\mathfrak{L}) \approx \{0,\ 1,\ 1.08,\ 4.14,\ 5,\ 6.7,\ 8.36,\ 16.1,\ 28.2,\ 33.5\}.$$

First, we illustrate the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ error bounds from Theorems 4.5 and 4.14. We take $\pi = \{\{1, 2, 3, 4\}, \{5, 6\}, \{7\}, \{8\}, \{9, 10\}\}$ and

$$A = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix},\ B = F = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Note that, indeed, π is an AEP. Also, in order to satisfy the assumptions of Theorem 4.14, we have taken A and B symmetric. Note that $A - \lambda B$ is Hurwitz for all nonzero eigenvalues λ of the Laplacian matrix $\mathfrak{L}$. Therefore, the multi-agent system is synchronized. It remains to choose the set of leaders $\mathfrak{V}_{\mathrm{L}}$. For demonstration, we compute the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ upper bounds and the true errors for all possible choices of $\mathfrak{V}_{\mathrm{L}}$. Since the sets of leaders are nonempty subsets of $\mathfrak{V}$, it follows that there are $2^{10} - 1 = 1023$ possible sets of leaders. Figure 4.2 shows all the ratios of upper bounds and corresponding true errors, where we define $\frac{0}{0} := 1$. We see that in this example,

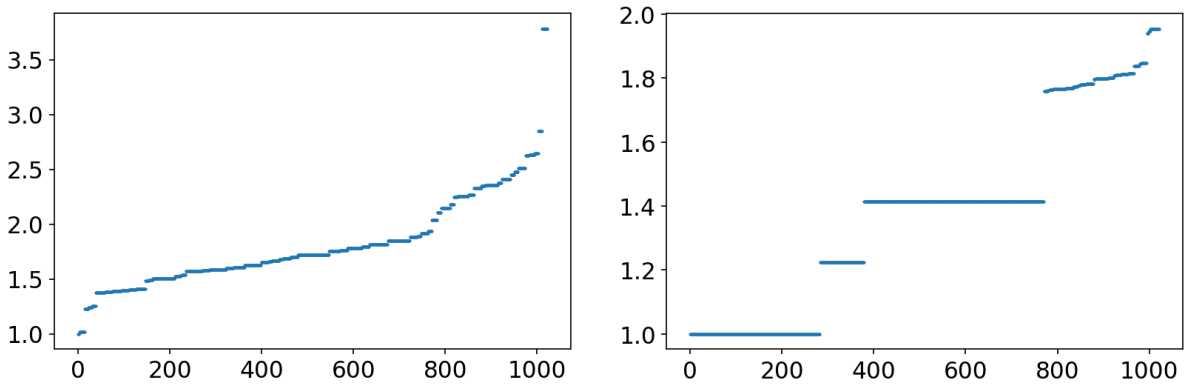

Figure 4.2: Ratios of $\mathcal{H}_2$ (left) and $\mathcal{H}_\infty$ (right) upper bounds and corresponding true errors, for a fixed almost equitable partition and all possible sets of leaders. In both figures, the sets of leaders are sorted such that the ratio is increasing (in particular, the ordering of the sets of leaders is not the same).

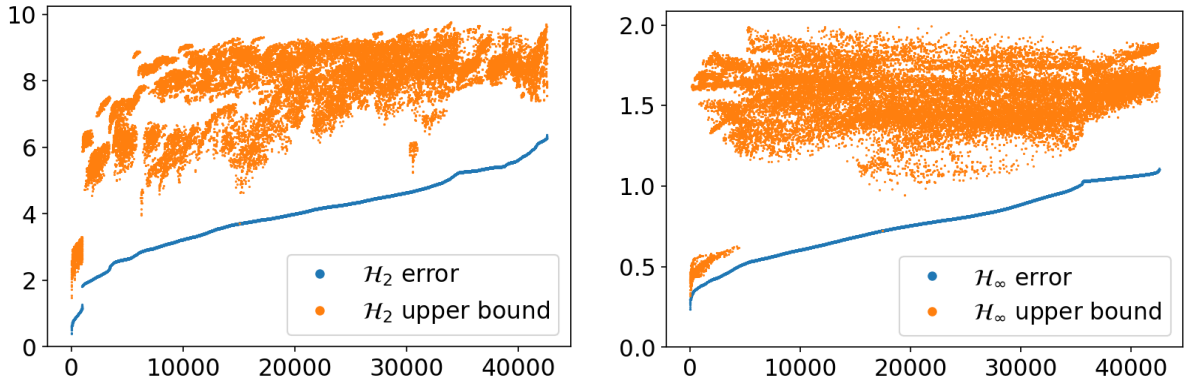

Figure 4.3: True $\mathcal{H}_2$ (left) and $\mathcal{H}_\infty$ (right) errors and upper bounds, for a fixed set of leaders and all partitions with five clusters. In each figure, partitions were sorted such that the true errors are increasing.

all true errors and upper bounds are within one order of magnitude, and that in most cases the ratio is below 2.

Next, we compare the true errors with the triangle inequality-based error bounds from (4.37) for a fixed set of leaders and all possible partitions consisting of five clusters. For the set of leaders, we take $\mathfrak{V}_\mathrm{L} = \{6, 7\}$, as was also used in [MTC14]. With this choice of leaders, the systems norms are $\|S\|_{\mathcal{H}_2} \approx 6.4$ and $\|S\|_{\mathcal{H}_\infty} \approx 1.03$ (rounded to three significant digits). Figure 4.3 shows true errors and upper bounds for all partitions of $\mathfrak{V}$ with five clusters (there are 42 525 such partitions). We observe that the upper bounds vary significantly as the true error increases, but the ratio is still less than one order of magnitude. Additionally, we notice that partitions giving small $\mathcal{H}_2$ errors give smaller upper bounds, as seen more clearly in the left subfigure of Figure 4.4. Furthermore, we observe a jump after the 966th partition. In fact, the 966

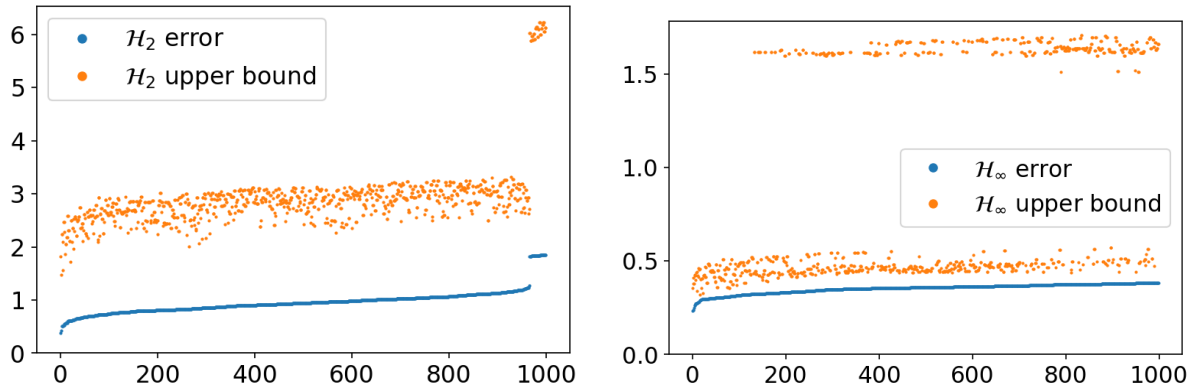

Figure 4.4: First 1000 true errors and upper bounds from Figure 4.3.

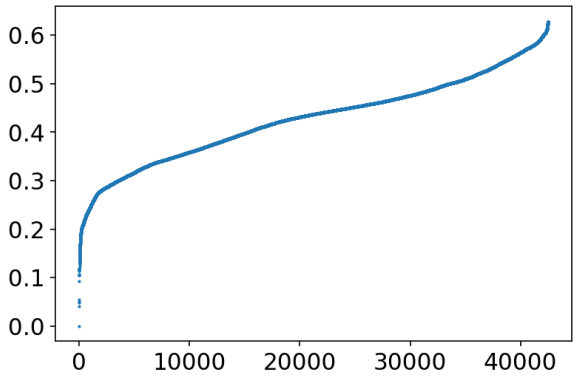

Figure 4.5: Relative error of $\mathfrak{L}$ by $\mathfrak{L}_{\text{AEP}}$ in Frobenius norm for all partitions with five clusters. The partitions are ordered such that the errors are increasing.

partitions giving the smallest $\mathcal{H}_2$ error are all those partitions where the leaders are the only members in their cluster. For the $\mathcal{H}_\infty$ error this is not the case, i.e., there are partitions with leaders sharing a cluster with more agents that give a smaller $\mathcal{H}_\infty$ error then a partition with leaders not sharing a cluster. On the other hand, partitions with the smallest $\mathcal{H}_2$ or $\mathcal{H}_\infty$ upper bound are close to the optimal true error.

In the following, we also compute the errors $\|\mathfrak{L} - \mathfrak{L}_{\text{AEP}}\|_{\text{F}}$ for all partitions with five clusters. Figure 4.5 shows the relative approximation errors $\frac{\|\mathfrak{L}-\mathfrak{L}_{\text{AEP}}\|_{\text{F}}}{\|\mathfrak{L}\|_{\text{F}}}$. We see that only a few (six, to be precise) partitions give a relative error less than 0.1. Irrespective of this, a small triangle inequality-based error bound (4.37) seems to indicate good partitions.

Finally, we compare the bound (4.37) with those from Ishizaki et al. [IKIA14, IKG+15, IKI16a]. There are also error bounds developed in [CKS16] and [BSJ16], but they depend on the proposed MOR methods and cannot be evaluated for an arbitrary partition. The $\mathcal{H}_2$ and $\mathcal{H}_\infty$ error bounds from Ishizaki et al. are based on the decomposition (see

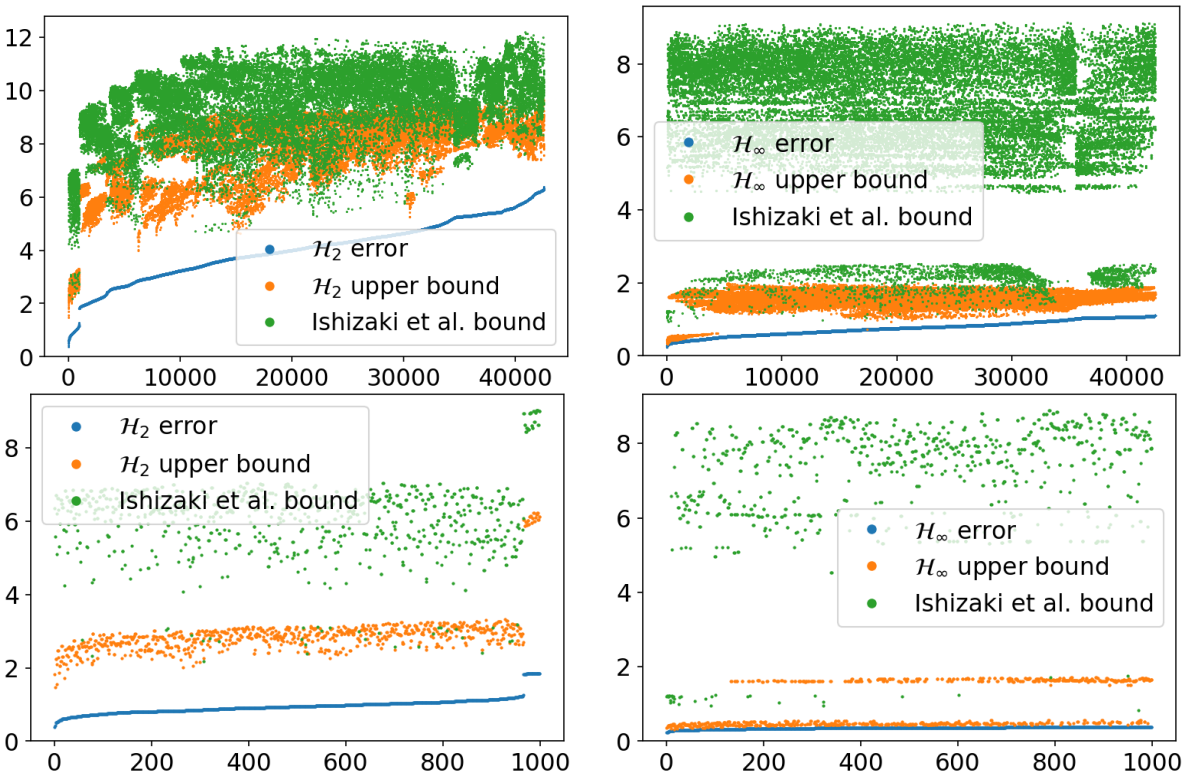

Figure 4.6: Comparison with error bounds from Ishizaki et al. [IKIA14, IKG+15, IKI16a]. The first column shows the $\mathcal{H}_2$ errors and bounds, the second column the $\mathcal{H}_\infty$ errors and bounds. The first row contains values for all partitions with five clusters, the second row only the first 1000 best ones.

equation (31) in [IKIA14], (20) in [II15], or (17) in [IKI16a])

$$H(s) - \widehat{H}(s) = \Xi(s) Q Q^{\mathrm{T}} X(s),$$

where

$$X(s) = \left(sI - I_{\mathfrak{n}} \otimes A + P^{\mathrm{T}} \mathfrak{L} P \otimes B\right)^{-1} (\mathfrak{B} \otimes F),$$
$$\Xi(s) = (\mathfrak{L} P \otimes I_n) \left(sI - I_{\mathfrak{n}} \otimes A + P^{\mathrm{T}} \mathfrak{L} P \otimes B\right)^{-1} \left(P^{\mathrm{T}} \otimes A - P^{\mathrm{T}} \mathfrak{L} \otimes B\right) + \mathfrak{L} \otimes I_n,$$

$P = \mathfrak{P} \left(\mathfrak{P}^{\mathrm{T}} \mathfrak{P}\right)^{-1}$, and Q is such that $\begin{bmatrix} P & Q \end{bmatrix}$ is orthogonal. The error bounds are then

$$\|H - \widehat{H}\|_{\mathcal{H}_p} \leqslant \|\Xi\|_{\mathcal{H}_\infty} \left\|Q Q^{\mathrm{T}} X\right\|_{\mathcal{H}_p} = \|\Xi\|_{\mathcal{H}_\infty} \left\|Q^{\mathrm{T}} X\right\|_{\mathcal{H}_p},$$

for $p = 2$ and $p = \infty$. Figure 4.6 shows the comparison between these bounds, the triangle inequality-based bound (4.37), and the true errors. In this example, our bounds are, for most partitions, lower than those from Ishizaki et al. Yet, they do share some qualitative properties: both vary significantly as the true error increases and those partitions with the small bounds are close to the optimal.

4.3 Exact clustering-based model order reduction for nonlinear power systems

4.3.1 Introduction

A power system is a network of electrical generators, loads, and their associated control elements. Each of these components may be thought of as vertex of a graph, while the transmission lines connecting them can be regarded as the edges of the graph. The vertices are modeled by physical laws that typically lead to a set of differential equations. These differential equations are coupled to each other across the edges. One question that has been of interest to power engineers over many years is how do the graph-theoretic properties of these types of electrical networks impact system-theoretic properties of the grid model [AA13].

Here, we study synchronization properties of power systems (see [DB14] for an overview) using graph-theoretic tools. Specifically, we show relations to graph symmetry and equitable partitions [RJME09], extending the work in [IKI16b] for linear systems to nonlinear power systems. Additionally, based on our results about synchronization, we propose a structure-preserving, clustering-based MOR framework for nonlinear power systems. Further, we show that for certain partitions this reduction is exact. In general, the dynamics of the reduced system can be used to approximate the dynamics of the original power system.

The motivation for clustering, in addition to reducing simulation time, is the possibility to simulate or control only a certain part of the grid, or a certain phenomenon that happens only over a certain time-scale. Some recent work on clustering of linear network systems can be found in [IKIA14, IKG+15, MGB15, CKS16, XC16, CKS17].

In Section 4.3.2, we describe the system we analyze. Next, we introduce synchronization for a pair of generators and prove necessary and sufficient conditions in Section 4.3.3. In Section 4.3.4, we continue in a similar way with two notions of synchronization with respect to a partition. We discuss clustering-based MOR in Section 4.3.5. Finally, we demonstrate our results in Section 4.3.6.

4.3.2 System description

We use the power system example in Figure 4.7 to introduce the type of system we analyze and to illustrate our results. As in the example in Figure 4.7, we consider power systems consisting of generators and buses, where each generator is connected to exactly one bus and buses can be classified into *generator buses* (those connected to one generator and some buses) and *non-generator buses* (those connected only to other buses). We follow the classical model of a synchronous generator [Kun94], which means that the generators' voltage amplitude is constant over time t.

Let $\mathcal{G} := \{1, 2, \dots, n\}$ and $\overline{\mathcal{G}} := \{n+1, n+2, \dots, n+\overline{n}\}$ denote the label sets of generator and non-generator buses. In the example in Figure 4.7, we have $n = 5$ and

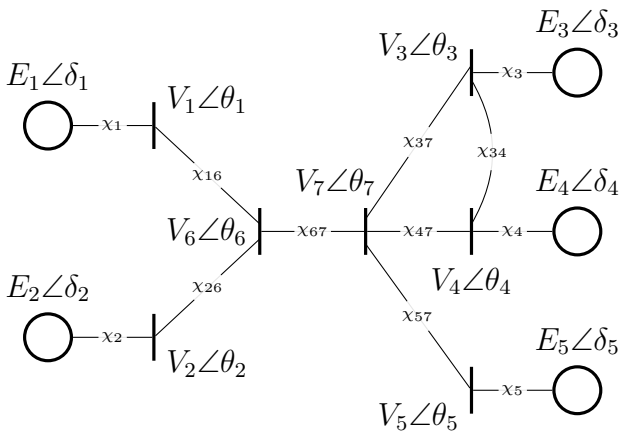

Figure 4.7: Power system consisting of generators (circles) and buses (vertical bars), where the ith generator is only connected to the ith bus. See Table 4.1 for the notation.

$\overline{n} = 2$. The vector of currents from generators to generator buses is given as

$$\boldsymbol{I}_{\mathcal{G}}(t) = \frac{1}{\boldsymbol{\imath}} L_{\mathrm{D}} \left(\boldsymbol{E}_{\mathcal{G}}(t) - \boldsymbol{V}_{\mathcal{G}}(t) \right), \tag{4.41}$$

where the vectors of voltages of generators and generator buses are denoted as

$$\begin{aligned} \boldsymbol{E}_{\mathcal{G}}(t) &\coloneqq [E_i(\cos\delta_i(t) + \boldsymbol{\imath}\sin\delta_i(t))]_{i\in\mathcal{G}} \in \mathbb{C}^n, \\ \boldsymbol{V}_{\mathcal{G}}(t) &\coloneqq [V_i(t)(\cos\theta_i(t) + \boldsymbol{\imath}\sin\theta_i(t))]_{i\in\mathcal{G}} \in \mathbb{C}^n, \end{aligned}$$

and L_{D} is a positive diagonal reactance matrix given as

$$L_{\mathrm{D}} \coloneqq \operatorname{diag}\Big(\left[\chi_i^{-1}\right]_{i\in\mathcal{G}} \Big),$$

where χ_i is the reactance between the ith generator and its bus (see Figure 4.7). We assume the generator voltage amplitudes E_i and reactances χ_i are given constants. Additionally, we assume the line resistances to be negligible.

The relation between the currents and voltages is given as

$$\begin{bmatrix} \boldsymbol{I}_{\mathcal{G}}(t) \\ 0 \end{bmatrix} = \frac{1}{\boldsymbol{\imath}} \begin{bmatrix} \mathfrak{L}_{11} & \mathfrak{L}_{12} \\ \mathfrak{L}_{12}^{\mathrm{T}} & \mathfrak{L}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_{\mathcal{G}}(t) \\ \boldsymbol{V}_{\overline{\mathcal{G}}}(t) \end{bmatrix}, \tag{4.42}$$

where the voltage vector of non-generator buses is denoted as

$$\boldsymbol{V}_{\overline{\mathcal{G}}}(t) \coloneqq [V_i(t)(\cos\theta_i(t) + \boldsymbol{\imath}\sin\theta_i(t))]_{i\in\overline{\mathcal{G}}} \in \mathbb{C}^{\overline{n}}$$

and $\mathfrak{L} = [\mathfrak{L}_{ij}] \in \mathbb{R}^{(n+\overline{n})\times(n+\overline{n})}$ denotes the weighted graph Laplacian of the reactance network. In particular, the (i,j)-th element of $\mathfrak{L}$ is $-\chi_{ij}^{-1}$ if the ith and jth buses are

Table 4.1: Notation

Symbol	Description
$[a_i]_{i\in S}$	vector $(a_{i_1}, a_{i_2}, \dots, a_{i_n})$, if $S = \{i_1, i_2, \dots, i_n\}$
sin, cos	functions applied element-wise to a vector or a matrix
$\mathcal{G}$	label set of generator buses
$\overline{\mathcal{G}}$	label set of non-generator buses
$\boldsymbol{E}_{\mathcal{G}}(t)$	voltages of the generators at time t
E_i	voltage amplitude of the ith generator
$\delta_i(t)$	voltage phase of the ith generator at time t
$\boldsymbol{V}_{\mathcal{G}}(t)$	voltages of the generator buses at time t
$\boldsymbol{V}_{\overline{\mathcal{G}}}(t)$	voltages of the non-generator buses at time t
$V_i(t)$	voltage amplitude of the ith bus at time t
$\theta_i(t)$	voltage phase of the ith bus at time t
$\boldsymbol{I}_{\mathcal{G}}(t)$	currents from generators to generator buses at time t
χ_i	reactance between the ith generator and its bus
χ_{ij}	reactance between the ith and jth bus
L_{D}	reactance matrix, $\operatorname{diag}\left([\chi_i^{-1}]_{i\in\mathcal{G}}\right)$
$\mathfrak{L}$	$[\mathfrak{L}_{ij}]_{i,j\in\{1,2\}}$, weighted graph Laplacian of the reactance network
$\delta(t)$	$[\delta_i(t)]_{i\in\mathcal{G}}$
M	diagonal matrix of inertias M_i of the generators
D	diagonal matrix of dissipativies D_i of the generators
f	vector of powers f_i to the generators
X	$(L_{\mathrm{D}} + \mathfrak{L}_{11} - \mathfrak{L}_{12}\mathfrak{L}_{22}^{-1}\mathfrak{L}_{12}^{\mathrm{T}})^{-1} L_{\mathrm{D}}$
Γ	$L_{\mathrm{D}}(L_{\mathrm{D}} + \mathfrak{L}_{11} - \mathfrak{L}_{12}\mathfrak{L}_{22}^{-1}\mathfrak{L}_{12}^{\mathrm{T}})^{-1} L_{\mathrm{D}}$
γ_{ij}	$[\Gamma]_{ij}^{-1}$
E	$[E_i]_{i\in\mathcal{G}}$
$V_{\mathcal{G}}(t)$	$[V_i(t)]_{i\in\mathcal{G}}$
$\theta_{\mathcal{G}}(t)$	$[\theta_i(t)]_{i\in\mathcal{G}}$
$\mathcal{X}_{ij}$	subspace of synchronism $\{x \in \mathbb{R}^n : x_i = x_j\}$
Π_{ij}	permutation matrix that swaps ith and jth components
$\mathcal{S}_{ij}$	set of symmetrical matrices $\{A \in \mathbb{R}^{n\times n} : A\Pi_{ij} = \Pi_{ij}A\}$
$\mathcal{X}_{\mathrm{cl}}$	$\bigcap_{\ell\in\widehat{\mathcal{G}}} \bigcap_{i,j\in\mathfrak{C}_\ell} \mathcal{X}_{ij}$
$\mathcal{S}_{\mathrm{cl}}$	$\bigcap_{\ell\in\widehat{\mathcal{G}}} \bigcap_{i,j\in\mathfrak{C}_\ell} \mathcal{S}_{ij}$

connected (see Figure 4.7) and the ith diagonal element is $\sum_{j\neq i} \chi_{ij}^{-1}$. In the following, we assume that the reactance network is connected, i.e., $\mathfrak{L}$ is irreducible. This assumption can be made without loss of generality because the same arguments can be applied to each connected component. For the example in Figure 4.7 with $\chi_{ij} = 1$ for all i, j, we have

$$\mathfrak{L} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 2 & -1 & 0 & 0 & -1 \\ 0 & 0 & -1 & 2 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ -1 & -1 & 0 & 0 & 0 & 3 & -1 \\ 0 & 0 & -1 & -1 & -1 & -1 & 4 \end{bmatrix}.$$

The dynamics of generators is given by

$$M\ddot{\delta}(t) + D\dot{\delta}(t) = f - \left[\frac{E_i V_i(t)}{\chi_i}\sin(\delta_i(t) - \theta_i(t))\right]_{i\in\mathcal{G}}, \tag{4.43a}$$

with voltage phases $\delta(t) \coloneqq [\delta_i(t)]_{i\in\mathcal{G}}$, inertia constants $M \coloneqq \operatorname{diag}\big([M_i]_{i\in\mathcal{G}}\big)$, $M_i > 0$, damping constants $D \coloneqq \operatorname{diag}\big([D_i]_{i\in\mathcal{G}}\big)$, $D_i \geqslant 0$, and input powers $f \in \mathbb{R}^n$ [Kun94]. Eliminating $\boldsymbol{I}_{\mathcal{G}}(t)$ from (4.41) and (4.42), we obtain

$$\begin{bmatrix} L_{\mathrm{D}}\left(\boldsymbol{E}_{\mathcal{G}}(t) - \boldsymbol{V}_{\mathcal{G}}(t)\right) \\ 0 \end{bmatrix} = \begin{bmatrix} \mathfrak{L}_{11} & \mathfrak{L}_{12} \\ \mathfrak{L}_{12}^{\mathrm{T}} & \mathfrak{L}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_{\mathcal{G}}(t) \\ \boldsymbol{V}_{\overline{\mathcal{G}}}(t) \end{bmatrix}, \tag{4.43b}$$

The set of equations (4.43) forms a differential-algebraic system. We can remove the algebraic constraints to find an equivalent set of differential equations using Kron reduction [Kro39]. First, from (4.43b), we find

$$\begin{aligned} \boldsymbol{V}_{\overline{\mathcal{G}}}(t) &= -\mathfrak{L}_{22}^{-1}\mathfrak{L}_{12}^{\mathrm{T}}\boldsymbol{V}_{\mathcal{G}}(t), \\ \boldsymbol{V}_{\mathcal{G}}(t) &= X\boldsymbol{E}_{\mathcal{G}}(t), \end{aligned} \tag{4.44}$$

where

$$X \coloneqq \left(L_{\mathrm{D}} + \mathfrak{L}_{11} - \mathfrak{L}_{12}\mathfrak{L}_{22}^{-1}\mathfrak{L}_{12}^{\mathrm{T}}\right)^{-1} L_{\mathrm{D}}. \tag{4.45}$$

It follows that

$$\Gamma \coloneqq L_{\mathrm{D}}\left(L_{\mathrm{D}} + \mathfrak{L}_{11} - \mathfrak{L}_{12}\mathfrak{L}_{22}^{-1}\mathfrak{L}_{12}^{\mathrm{T}}\right)^{-1} L_{\mathrm{D}} = L_{\mathrm{D}}X$$

is a positive definite matrix with positive elements, since $L_{\mathrm{D}} + \mathfrak{L}_{11} - \mathfrak{L}_{12}\mathfrak{L}_{22}^{-1}\mathfrak{L}_{12}^{\mathrm{T}}$ is positive definite and an irreducible M-matrix [BP94, pp. 141]. We denote its elements by $\gamma_{ij}^{-1} \coloneqq [\Gamma]_{ij}$. Then, multiplying (4.44) from the left by L_{D}, we find

$$\begin{aligned} \left[\frac{V_i(t)}{\chi_i}\cos\theta_i(t)\right]_{i\in\mathcal{G}} &= \Gamma\left[E_i\cos\delta_i(t)\right]_{i\in\mathcal{G}}, \\ \left[\frac{V_i(t)}{\chi_i}\sin\theta_i(t)\right]_{i\in\mathcal{G}} &= \Gamma\left[E_i\sin\delta_i(t)\right]_{i\in\mathcal{G}}, \end{aligned}$$

which, when inserted in (4.43a), gives us

$$\begin{aligned} M\ddot{\delta}(t) + D\dot{\delta}(t) &= f - \left[\frac{E_i V_i(t)}{\chi_i}\left(\sin\delta_i(t)\cos\theta_i(t) - \cos\delta_i(t)\sin\theta_i(t)\right)\right]_{i\in\mathcal{G}} \\ &= f - \left(\operatorname{diag}\big([E_i\sin\delta_i(t)]_{i\in\mathcal{G}}\big)\left[\frac{V_i(t)}{\chi_i}\cos\theta_i(t)\right]_{i\in\mathcal{G}}\right. \\ &\qquad \left. - \operatorname{diag}\big([E_i\cos\delta_i(t)]_{i\in\mathcal{G}}\big)\left[\frac{V_i(t)}{\chi_i}\sin\theta_i(t)\right]_{i\in\mathcal{G}}\right) \\ &= f - \big(\operatorname{diag}\big([E_i\sin\delta_i(t)]_{i\in\mathcal{G}}\big)\,\Gamma\,[E_i\cos\delta_i(t)]_{i\in\mathcal{G}} \\ &\qquad - \operatorname{diag}\big([E_i\cos\delta_i(t)]_{i\in\mathcal{G}}\big)\,\Gamma\,[E_i\sin\delta_i(t)]_{i\in\mathcal{G}}\big). \end{aligned}$$

Thus, now by using $\sin\delta_i(t)\cos\delta_j(t) - \cos\delta_i(t)\sin\delta_j(t) = \sin(\delta_i(t) - \delta_j(t))$, the Kron-reduced system of (4.43) is given as

$$M_i\ddot{\delta}_i(t) + D_i\dot{\delta}_i(t) = f_i - \sum_{k=1}^{n} \frac{E_iE_k}{\gamma_{ik}} \sin(\delta_i(t) - \delta_k(t)), \tag{4.46a}$$

with generator buses' voltages and phases satisfying

$$L_{\mathrm{D}}\boldsymbol{V}_{\mathcal{G}}(t) = \Gamma\boldsymbol{E}_{\mathcal{G}}(t). \tag{4.46b}$$

Denoting $E \coloneqq [E_i]_{i\in\mathcal{G}}$, $V_{\mathcal{G}}(t) \coloneqq [V_i(t)]_{i\in\mathcal{G}}$, and $\theta_{\mathcal{G}}(t) \coloneqq [\theta_i(t)]_{i\in\mathcal{G}}$, we can write (4.43a) and (4.46a) more compactly as

$$M\ddot{\delta}(t) + D\dot{\delta}(t) = f - L_{\mathrm{D}}\left(E \circ V_{\mathcal{G}}(t) \circ \sin(\delta(t) - \theta_{\mathcal{G}}(t))\right), \tag{4.47}$$

and

$$M\ddot{\delta}(t) + D\dot{\delta}(t) = f - \left(\Gamma \circ EE^{\mathrm{T}} \circ \sin\left(\delta(t)\mathbb{1}_n^{\mathrm{T}} - \mathbb{1}_n\delta(t)^{\mathrm{T}}\right)\right)\mathbb{1}_n.$$

4.3.3 Synchronization of generator pair

Let us denote the subspace of the synchronism between the ith and jth elements by

$$\mathcal{X}_{ij} \coloneqq \{x \in \mathbb{R}^n : x_i = x_j\}.$$

In this notation, we introduce the following notion of synchronism for the power system (4.43).

Definition 4.18:
Consider the power system (4.43). The ith and jth generators are said to be *synchronized* if

$$\delta(t) \in \mathcal{X}_{ij} \text{ and } \boldsymbol{V}_{\mathcal{G}}(t) \in \mathcal{X}_{ij}, \text{ for all } t \geqslant 0$$

and for any initial condition $\delta(0), \dot{\delta}(0) \in \mathcal{X}_{ij}$. ◊

To characterize this generator synchronism in an algebraic manner, let us define a set of symmetrical matrices with respect to the permutation of the ith and jth columns and rows by

$$\mathcal{S}_{ij} \coloneqq \{A \in \mathbb{R}^{n\times n} : A\Pi_{ij} = \Pi_{ij}A\}, \tag{4.48}$$

where Π_{ij} denotes the permutation matrix associated with the ith and jth elements, i.e., all diagonal elements of Π_{ij} other than the ith and jth elements are 1, the (i,j)-th and (j,i)-th elements are 1, and the others are zero. Note that $\mathcal{S}_{ij}$ is not the set of usual symmetric (Hermitian) matrices; the condition in (4.48) represents the invariance with respect to the permutation of the ith and jth columns and rows, i.e., $\Pi_{ij}^{\mathrm{T}}A\Pi_{ij} = A$. The following lemma gives necessary and sufficient conditions for a symmetric matrix to be symmetrical.

Lemma 4.19:
Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix and $i, j \in \{1, 2, \ldots, n\}$ such that $i \neq j$. Then $A \in \mathcal{S}_{ij}$ if and only if $a_{ii} = a_{jj}$ and $a_{ik} = a_{jk}$ for all $k \neq i, j$. ◊

Proof. From the definition, it can be seen that $A \in \mathcal{S}_{ij}$ is equivalent to $a_{ii} = a_{jj}$, $a_{ij} = a_{ji}$, $a_{ik} = a_{jk}$, and $a_{ki} = a_{kj}$ for all $k \neq i, j$. Using that A is symmetric, the conditions of the lemma follow. □

Some basic properties of symmetrical matrices are given in the following lemma.

Lemma 4.20:
Let $A, B \in \mathcal{S}_{ij}$ for some $i, j \in \{1, 2, \ldots, n\}$ such that $i \neq j$ and $\alpha, \beta \in \mathbb{R}$. Then,

1. $\alpha A + \beta B \in \mathcal{S}_{ij}$,
2. $AB \in \mathcal{S}_{ij}$, and
3. if A is invertible, then $A^{-1} \in \mathcal{S}_{ij}$. ◊

Proof. Follows directly from the definition of $\mathcal{S}_{ij}$ in (4.48). □

We state the main result about synchronization of a pair of generators and prove it in the remainder of this section.

Theorem 4.21:
Consider the power system (4.43). The following two statements hold.

1. Let $n = 2$ and $M_1 = M_2$. Then the two generators are synchronized if and only if $D_1 = D_2$, $f_1 = f_2$, and $E_1 = E_2$.
2. Let $n \geqslant 3$ and $M \in \mathcal{S}_{ij}$. Then the ith and jth generators are synchronized if and only if $D \in \mathcal{S}_{ij}$, $f \in \mathcal{X}_{ij}$, $E \in \mathcal{X}_{ij}$, and $\Gamma \in \mathcal{S}_{ij}$. ◊

Remark 4.22:
Essentially, this result shows that the ith and jth generators are synchronized when the system equation are invariant under swapping the ith and jth label. ◊

We arrange the proof of Theorem 4.21 into a sequence of propositions in this section. We begin by analyzing the equations of the system (4.43) without assumptions on n and M.

Proposition 4.23:
The ith and jth generators are synchronized if and only if

$$\frac{D_i}{M_i} = \frac{D_j}{M_j}, \tag{4.49a}$$

$$\frac{f_i}{M_i} = \frac{f_j}{M_j}, \tag{4.49b}$$

$$\frac{E_i}{M_i\gamma_{ik}} = \frac{E_j}{M_j\gamma_{jk}}, \text{ for } k \neq i, j, \tag{4.49c}$$

$$\frac{\chi_i}{\gamma_{ik}} = \frac{\chi_j}{\gamma_{jk}}, \text{ for } k \neq i, j, \text{ and} \tag{4.49d}$$

$$\frac{\chi_i E_i}{\gamma_{ii}} + \frac{\chi_i E_j}{\gamma_{ij}} = \frac{\chi_j E_i}{\gamma_{ji}} + \frac{\chi_j E_j}{\gamma_{jj}}. \tag{4.49e}$$

◊

Proof. From (4.46a), we get

$$\ddot{\delta}_i - \ddot{\delta}_j = -\frac{D_i}{M_i}\dot{\delta}_i + \frac{D_j}{M_j}\dot{\delta}_j + \frac{f_i}{M_i} - \frac{f_j}{M_j} - \sum_{k=1}^{n}\left(\frac{E_i E_k}{M_i\gamma_{ik}}\sin(\delta_i - \delta_k) - \frac{E_j E_k}{M_j\gamma_{jk}}\sin(\delta_j - \delta_k)\right).$$

It is clear that, if (4.49a), (4.49b), and (4.49c) are true, then $\delta, \dot{\delta} \in \mathfrak{X}_{ij}$ implies $\ddot{\delta} \in \mathfrak{X}_{ij}$. For the other direction, let us assume that the ith and jth generators are synchronized. Then we necessarily have

$$-\left(\frac{D_i}{M_i} - \frac{D_j}{M_j}\right)\dot{\delta}_i + \left(\frac{f_i}{M_i} - \frac{f_j}{M_j}\right) - \sum_{k=1}^{n}\left(\left(\frac{E_i E_k}{M_i\gamma_{ik}} - \frac{E_j E_k}{M_j\gamma_{jk}}\right)\sin(\delta_i - \delta_k)\right) = 0,$$

for any δ_i, $\dot{\delta}_i$, and δ_k, $k \neq i, j$. Choosing $\dot{\delta}_i = 0$ and $\delta_k = \delta_i$, condition (4.49b) follows. Taking $\dot{\delta}_i = 1$ and $\delta_k = \delta_i$, we find condition (4.49a). Lastly, with $\delta_i - \delta_k = \frac{\pi}{2}$ for some $k \neq i, j$ and $\delta_i - \delta_\ell = 0$ for $\ell \neq i, j, k$, condition (4.49c) follows for the chosen k.

From (4.46b), we have

$$\begin{aligned} V_i\cos\theta_i - V_j\cos\theta_j = &\left(\frac{\chi_i E_i}{\gamma_{ii}} - \frac{\chi_j E_i}{\gamma_{ji}}\right)\cos\delta_i + \left(\frac{\chi_i E_j}{\gamma_{ij}} - \frac{\chi_j E_j}{\gamma_{jj}}\right)\cos\delta_j \\ &+ \sum_{\substack{k=1 \\ k\neq i,j}}^{n}\left(\frac{\chi_i}{\gamma_{ik}} - \frac{\chi_j}{\gamma_{jk}}\right)E_k\cos\delta_k, \\ V_i\sin\theta_i - V_j\sin\theta_j = &\left(\frac{\chi_i E_i}{\gamma_{ii}} - \frac{\chi_j E_i}{\gamma_{ji}}\right)\sin\delta_i + \left(\frac{\chi_i E_j}{\gamma_{ij}} - \frac{\chi_j E_j}{\gamma_{jj}}\right)\sin\delta_j \\ &+ \sum_{\substack{k=1 \\ k\neq i,j}}^{n}\left(\frac{\chi_i}{\gamma_{ik}} - \frac{\chi_j}{\gamma_{jk}}\right)E_k\sin\delta_k. \end{aligned}$$

Similarly, if we assume conditions (4.49d) and (4.49e) to be true, then $\delta_i = \delta_j$ implies $V_i\cos\theta_i = V_j\cos\theta_j$ and $V_i\sin\theta_i = V_j\sin\theta_j$, which in turn implies that $\boldsymbol{V}_\mathcal{G} \in \mathfrak{X}_{ij}$.

Conversely, we have

$$0 = \left(\frac{\chi_i E_i}{\gamma_{ii}} + \frac{\chi_i E_j}{\gamma_{ij}} - \frac{\chi_j E_i}{\gamma_{ji}} - \frac{\chi_j E_j}{\gamma_{jj}}\right)\cos\delta_i + \sum_{\substack{k=1\\k\neq i,j}}^{n}\left(\frac{\chi_i}{\gamma_{ik}} - \frac{\chi_j}{\gamma_{jk}}\right)E_k\cos\delta_k,$$

$$0 = \left(\frac{\chi_i E_i}{\gamma_{ii}} + \frac{\chi_i E_j}{\gamma_{ij}} - \frac{\chi_j E_i}{\gamma_{ji}} - \frac{\chi_j E_j}{\gamma_{jj}}\right)\sin\delta_i + \sum_{\substack{k=1\\k\neq i,j}}^{n}\left(\frac{\chi_i}{\gamma_{ik}} - \frac{\chi_j}{\gamma_{jk}}\right)E_k\sin\delta_k,$$

for arbitrary δ_i and δ_k for $k \neq i, j$. By appropriate choices of δ_i and δ_k, conditions (4.49d) and (4.49e) follow. □

The following lemma gives an important property of the matrix X.

Lemma 4.24:
For X as in (4.45), we have $X\mathbb{1} = \mathbb{1}$. ◊

Proof. Recalling the definition of X from (4.45), after some algebraic manipulation, it is clear that $X\mathbb{1} = \mathbb{1}$ is equivalent to

$$\left(\mathfrak{L}_{11} - \mathfrak{L}_{12}\mathfrak{L}_{22}^{-1}\mathfrak{L}_{12}^{\mathrm{T}}\right)\mathbb{1} = 0,$$

which follows from $\mathfrak{L}\mathbb{1} = 0$. □

Let us now assume that $E_i \neq E_j$ and find what follows from conditions of Proposition 4.23. From (4.49d) and Lemma 4.24, it follows that $\frac{\chi_i}{\gamma_{ii}} + \frac{\chi_i}{\gamma_{ij}} = \frac{\chi_j}{\gamma_{ji}} + \frac{\chi_j}{\gamma_{jj}}$. Then, by (4.49e) and $E_i \neq E_j$, it is necessary that $\frac{\chi_i}{\gamma_{ii}} = \frac{\chi_j}{\gamma_{ji}}$ and $\frac{\chi_i}{\gamma_{ij}} = \frac{\chi_j}{\gamma_{jj}}$. This, together with (4.49d), means that the ith and jth rows in X are equal, which is a contradiction with X being invertible. Therefore, for ith and jth generators to be synchronized, it is necessary that $E_i = E_j$. This allows us to simplify the statement of Proposition 4.23. We can simplify it further by assuming $M_i = M_j$, which gives us the following corollary.

Corollary 4.25:
Let $M_i = M_j$. Then the ith and jth generators are synchronized if and only if

$$\begin{aligned}
D_i &= D_j, \\
f_i &= f_j, \\
E_i &= E_j, \\
\gamma_{ik} &= \gamma_{jk}, \text{ for } k \neq i, j, && \text{(4.50a)}\\
\frac{\chi_i}{\gamma_{ik}} &= \frac{\chi_j}{\gamma_{jk}}, \text{ for } k \neq i, j, \text{ and} && \text{(4.50b)}\\
\frac{\chi_i}{\gamma_{ii}} + \frac{\chi_i}{\gamma_{ij}} &= \frac{\chi_j}{\gamma_{ji}} + \frac{\chi_j}{\gamma_{jj}}. && \text{(4.50c)}
\end{aligned}$$

◊

In the following, we separate the $n = 2$ and $n \geqslant 3$ cases. First, we use Corollary 4.25 to prove part 1 of Theorem 4.21.

Proof of Theorem 4.21, part 1. This is true since (4.50a) and (4.50b) are empty statements, while (4.50c) follows immediately from Lemma 4.24. □

Finally, to prove part 2 of Theorem 4.21, we simplify the statement of Corollary 4.25 for the case of $n \geqslant 3$. This gives us the following corollary.

Corollary 4.26:
Let $n \geqslant 3$ and $M_i = M_j$. Then the ith and jth generators are synchronized if and only if

$$\begin{aligned} D_i &= D_j, \\ f_i &= f_j, \\ E_i &= E_j, \\ \gamma_{ik} &= \gamma_{jk}, \text{ for } k \neq i, j, && \text{(4.51a)} \\ \chi_i &= \chi_j, \text{ and} && \text{(4.51b)} \\ \gamma_{ii} &= \gamma_{jj}. && \text{(4.51c)} \end{aligned}$$

◊

Proof. Condition (4.51b) follows from (4.50a) and (4.50b), using that there are at least three generators. Then (4.51c) follows from (4.50c), (4.51b), and symmetry $\gamma_{ij} = \gamma_{ji}$. □

Corollary 4.26, together with the following lemma allows us to complete the proof of Theorem 4.21.

Lemma 4.27:
Let $i, j \in \{1, 2, \ldots, n\}$ be such that $i \neq j$. We have $\Gamma \in \mathcal{S}_{ij}$ if and only if $L_\mathrm{D} \in \mathcal{S}_{ij}$ and $\mathfrak{L}_{11} - \mathfrak{L}_{12}\mathfrak{L}_{22}^{-1}\mathfrak{L}_{12}^\mathrm{T} \in \mathcal{S}_{ij}$. ◊

Proof. $\boxed{\Leftarrow}$ Follows from Lemma 4.20.

$\boxed{\Rightarrow}$ First we show that $L_\mathrm{D} \in \mathcal{S}_{ij}$. Using $\Gamma = L_\mathrm{D} X$, $\Pi_{ij}\mathbb{1} = \mathbb{1}$, and $X\mathbb{1} = \mathbb{1}$, from $\Gamma\Pi_{ij}\mathbb{1} = \Pi_{ij}\Gamma\mathbb{1}$ it follows that $L_\mathrm{D}\mathbb{1} = \Pi_{ij}L_\mathrm{D}\mathbb{1}$. Since L_D is a diagonal matrix, from this we see that $L_\mathrm{D} \in \mathcal{S}_{ij}$. Now $\mathfrak{L}_{11} - \mathfrak{L}_{12}\mathfrak{L}_{22}^{-1}\mathfrak{L}_{12}^\mathrm{T} \in \mathcal{S}_{ij}$ follows from Lemma 4.20. □

Now we can complete the proof of Theorem 4.21.

Proof of Theorem 4.21, part 2. Conditions (4.51a) and (4.51c), by Lemma 4.19, are equivalent to $\Gamma \in \mathcal{S}_{ij}$, which, by Lemma 4.27, is in turn equivalent to $L_\mathrm{D} \in \mathcal{S}_{ij}$ and $\mathfrak{L}_{11} - \mathfrak{L}_{12}\mathfrak{L}_{22}^{-1}\mathfrak{L}_{12}^\mathrm{T} \in \mathcal{S}_{ij}$. Therefore, (4.51a) and (4.51c) imply (4.51b). □

4.3.4 Synchronization of generator partition

Let $\pi = \{\mathfrak{C}_\ell\}_{\ell \in \widehat{\mathcal{G}}}$ be a partition of the set $\mathcal{G}$, where $\widehat{\mathcal{G}} = \{1, 2, \ldots, \widehat{n}\}$ and $\widehat{n} \leqslant n$. Let us denote

$$\mathcal{X}_\mathrm{cl} \coloneqq \bigcap_{\ell \in \widehat{\mathcal{G}}} \bigcap_{i,j \in \mathfrak{C}_\ell} \mathcal{X}_{ij}, \quad \mathcal{S}_\mathrm{cl} \coloneqq \bigcap_{\ell \in \widehat{\mathcal{G}}} \bigcap_{i,j \in \mathfrak{C}_\ell} \mathcal{S}_{ij}.$$

and $\mathfrak{P}$ the characteristic matrix of the partition π. Notice that $\mathcal{X}_{\mathrm{cl}} = \mathrm{im}(\mathfrak{P})$.

We define two notions generalizing the synchronization of two generators to a partition of generators.

Definition 4.28:
The system (4.43) is said to be *strongly synchronized with respect to partition* π if the ith and jth generators are synchronized for all $i, j \in \mathfrak{C}_\ell$ and all $\ell \in \widehat{\mathcal{G}}$, i.e., $\delta(t) \in \mathcal{X}_{ij}$ and $\boldsymbol{V}_{\mathcal{G}}(t) \in \mathcal{X}_{ij}$ for all $t \geqslant 0$ and for any $\delta(0), \dot{\delta}(0) \in \mathcal{X}_{ij}$, $i, j \in \mathfrak{C}_\ell$, and $\ell \in \widehat{\mathcal{G}}$.

The system (4.43) is said to be *weakly synchronized with respect to partition* π if, for arbitrary $\delta(0), \dot{\delta}(0) \in \mathcal{X}_{\mathrm{cl}}$, there exist functions $\widehat{\delta}\colon [0, \infty) \to \mathbb{R}^{\widehat{n}}$ and $\widehat{\boldsymbol{V}}_{\widehat{\mathcal{G}}}\colon [0, \infty) \to \mathbb{C}^{\widehat{n}}$ such that $\delta(t) = \mathfrak{P}\widehat{\delta}(t)$ and $\boldsymbol{V}_{\mathcal{G}}(t) = \mathfrak{P}\widehat{\boldsymbol{V}}_{\widehat{\mathcal{G}}}(t)$, i.e., $\delta(t) \in \mathcal{X}_{\mathrm{cl}}$ and $\boldsymbol{V}_{\mathcal{G}}(t) \in \mathcal{X}_{\mathrm{cl}}$ for all $t \geqslant 0$ and for any $\delta(0), \dot{\delta}(0) \in \mathcal{X}_{\mathrm{cl}}$. ◊

Remark 4.29:
Notice that strong synchronization is equivalent to $\mathcal{X}_{ij} \times \mathcal{X}_{ij} \times \mathcal{X}_{ij}$ being an invariant set for $(\delta, \dot{\delta}, \widehat{\boldsymbol{V}}_{\widehat{\mathcal{G}}})$ for any $i, j \in \mathfrak{C}_\ell$ and $\ell \in \widehat{\mathcal{G}}$, while weak synchronization is equivalent to an invariant set being $\mathcal{X}_{\mathrm{cl}} \times \mathcal{X}_{\mathrm{cl}} \times \mathcal{X}_{\mathrm{cl}}$. This means that, if the power system is strongly synchronized, when two generators and their buses in the same cluster have equal state, they will remain equal. If the power system is weakly synchronized, then when the states of every generator and its bus are equal to all others in the same cluster, they will stay equal. From this, we see that that if the system (4.43) is strongly synchronized with respect to π, then it is also weakly synchronized with respect to π, since $\mathcal{X}_{\mathrm{cl}} \times \mathcal{X}_{\mathrm{cl}} \times \mathcal{X}_{\mathrm{cl}} \subseteq \mathcal{X}_{ij} \times \mathcal{X}_{ij} \times \mathcal{X}_{ij}$, for all $i, j \in \mathfrak{C}_\ell$ and all $\ell \in \widehat{\mathcal{G}}$.

Further, the ith and jth generators are synchronized if and only if (4.43) is either strongly or weakly synchronized with respect to $\{\{i, j\}\} \cup \{\{k\} : k \neq i, j\}$.

Finally, notice that (4.43) is always both strongly and weakly synchronized with respect to $\{\{i\} : i \in \mathcal{G}\}$. ◊

In the following, we show necessary and sufficient conditions for the two synchronization notions. To start, in the next proposition, we present cases when the structure of Γ has no influence. It also illustrates the relation between strong and weak synchronization.

Proposition 4.30:
Let $\pi = \{\mathcal{G}\}$, $M, D \in \mathcal{S}_{\mathrm{cl}}$, and $f, E \in \mathcal{X}_{\mathrm{cl}}$. Then the system (4.43) is weakly synchronized with respect to $\{\mathcal{G}\}$. If additionally $n = 2$, then (4.43) is also strongly synchronized with respect to $\{\mathcal{G}\}$. ◊

Proof. From the assumptions, it follows that $M = \widehat{m}I$, $D = \widehat{d}I$, $f = \widehat{f}\mathbb{1}$, and $E = \widehat{E}\mathbb{1}$, for some $\widehat{m} > 0$, $\widehat{d} \geqslant 0$, and $\widehat{f}, \widehat{E} \in \mathbb{R}$. Notice that for $\pi = \{\mathcal{G}\}$, we have $\mathfrak{P} = \mathbb{1}$.

Let us assume that $\delta(0), \dot{\delta}(0) \in \mathrm{im}(\mathbb{1})$. To prove weak synchronization, we need to show that $\delta(t) \in \mathrm{im}(\mathbb{1})$ and $\boldsymbol{V}_{\mathcal{G}}(t) \in \mathrm{im}(\mathbb{1})$. For the former, it is enough to show that $\ddot{\delta}(t) \in \mathrm{im}(\mathbb{1})$ if $\delta(t), \dot{\delta}(t) \in \mathrm{im}(\mathbb{1})$, which is clear, since then $\ddot{\delta}(t) = -M^{-1}D\dot{\delta}(t) + M^{-1}f = -\frac{\widehat{d}}{\widehat{m}}\dot{\delta}(t) + \frac{\widehat{f}}{\widehat{m}}\mathbb{1}$. For the latter, we see that $\boldsymbol{V}_{\mathcal{G}} = L_{\mathrm{D}}^{-1}\Gamma\boldsymbol{E}_{\mathcal{G}} \in \mathrm{im}(\mathbb{1})$ whenever $\boldsymbol{E}_{\mathcal{G}} \in \mathrm{im}(\mathbb{1})$, which is equivalent to $\delta \in \mathrm{im}(\mathbb{1})$.

The second part follows from part 1 of Theorem 4.21. □

We continue with the first main result of this section—the necessary and sufficient conditions for strong synchronization. Here, symmetrical conditions for Γ are relevant.

Theorem 4.31:
Let $n \geqslant 3$, π arbitrary, and $M \in \mathcal{S}_{\mathrm{cl}}$. Then the system (4.43) is strongly synchronized with respect to π if and only if $D \in \mathcal{S}_{\mathrm{cl}}$, $f \in \mathcal{X}_{\mathrm{cl}}$, $E \in \mathcal{X}_{\mathrm{cl}}$, and $\Gamma \in \mathcal{S}_{\mathrm{cl}}$. ◊

Proof. Follows from applying part 2 of Theorem 4.21 for every ith and jth generator where $i, j \in \mathfrak{C}_\ell$ and $\ell \in \widehat{\mathcal{G}}$. □

We conclude this section with the second main result—the necessary and sufficient conditions for weak synchronization. Instead of symmetrical conditions, $\mathcal{X}_{\mathrm{cl}}$ being Γ-invariant is one of the conditions. Since $\mathcal{X}_{\mathrm{cl}} = \mathrm{im}(\mathfrak{P})$, this actually means that π is an equitable partition for a graph whose adjacency matrix is Γ (see Lemma 2.48).

Theorem 4.32:
Let $|\pi| \geqslant 2$, $M, D \in \mathcal{S}_{\mathrm{cl}}$, and $f, E \in \mathcal{X}_{\mathrm{cl}}$. Then the system (4.43) is weakly synchronized with respect to π if and only if

$$L_{\mathrm{D}} \in \mathcal{S}_{\mathrm{cl}} \quad \text{and} \quad \mathcal{X}_{\mathrm{cl}} \text{ is } \Gamma\text{-invariant.} \tag{4.52}$$

◊

Proof. From the definition, we see that (4.43) is weakly synchronized with respect to π if and only if

$$\left(\forall \delta, \dot{\delta} \in \mathcal{X}_{\mathrm{cl}}\right)\ M^{-1}\left(-D\dot{\delta} + f - \left(\Gamma \circ EE^{\mathrm{T}} \circ \sin\left(\delta \mathbb{1}_n^{\mathrm{T}} - \mathbb{1}_n \delta^{\mathrm{T}}\right)\right) \mathbb{1}_n\right) \in \mathcal{X}_{\mathrm{cl}} \tag{4.53}$$

and

$$\left(\forall \delta \in \mathcal{X}_{\mathrm{cl}}\right)\ L_{\mathrm{D}}^{-1} \Gamma \boldsymbol{E}_{\mathcal{G}} \in \mathcal{X}_{\mathrm{cl}}. \tag{4.54}$$

Since $M, D \in \mathcal{S}_{\mathrm{cl}}$ and $f \in \mathcal{X}_{\mathrm{cl}}$, condition (4.53) is equivalent to

$$\left(\forall \delta \in \mathcal{X}_{\mathrm{cl}}\right)\ \left(\Gamma \circ EE^{\mathrm{T}} \circ \sin\left(\delta \mathbb{1}_n^{\mathrm{T}} - \mathbb{1}_n \delta^{\mathrm{T}}\right)\right) \mathbb{1}_n \in \mathcal{X}_{\mathrm{cl}}.$$

Using $\delta = \mathfrak{P}\widehat{\delta}$, $E = \mathfrak{P}\widehat{E}$, $\mathbb{1}_n = \mathfrak{P}\mathbb{1}_{\widehat{n}}$, and that $v \in \mathcal{X}_{\mathrm{cl}}$ is equivalent to $\Pi_{ij} v = v$ for all $i, j \in \mathfrak{C}_\ell$ and $\ell \in \widehat{\mathcal{G}}$, we find that the above condition is equivalent to

$$\begin{aligned} &\left(\forall \widehat{\delta} \in \mathbb{R}^{\widehat{n}}\right)\left(\forall \ell \in \widehat{\mathcal{G}}\right)(\forall i, j \in \mathfrak{C}_\ell) \\ &\quad \left((\Gamma\mathfrak{P} - \Pi_{ij}\Gamma\mathfrak{P}) \circ \mathfrak{P}\widehat{E}\widehat{E}^{\mathrm{T}} \circ \mathfrak{P} \sin\left(\widehat{\delta}\mathbb{1}_{\widehat{n}}^{\mathrm{T}} - \mathbb{1}_{\widehat{n}}\widehat{\delta}^{\mathrm{T}}\right)\right) \mathbb{1}_{\widehat{n}} = 0. \end{aligned} \tag{4.55}$$

In a similar way, we find that the condition (4.54) is equivalent to

$$\left(\forall \widehat{\boldsymbol{E}}_{\widehat{\mathcal{G}}} \in \mathbb{C}^{\widehat{n}}\right)\left(\forall \ell \in \widehat{\mathcal{G}}\right)(\forall i, j \in \mathfrak{C}_\ell)\ L_{\mathrm{D}}^{-1}\Gamma\mathfrak{P}\widehat{\boldsymbol{E}}_{\widehat{\mathcal{G}}} = \Pi_{ij} L_{\mathrm{D}}^{-1}\Gamma\mathfrak{P}\widehat{\boldsymbol{E}}_{\widehat{\mathcal{G}}},$$

or, more simply,

$$\Big(\forall \ell \in \widehat{\mathcal{G}}\Big)(\forall i,j \in \mathfrak{C}_\ell)\ \ L_{\mathrm{D}}^{-1}\Gamma\mathfrak{P} = \Pi_{ij} L_{\mathrm{D}}^{-1}\Gamma\mathfrak{P}. \tag{4.56}$$

It is straightforward to check that (4.52) implies (4.55) and (4.56). For the other direction, choosing $\widehat{\delta} = e_{\ell_2}$ for $\ell_2 \neq \ell$ in (4.55), we find from the ith row that

$$\Big(\forall \ell, \ell_2 \in \widehat{\mathcal{G}}, \ell_2 \neq \ell\Big)(\forall i,j \in \mathfrak{C}_\ell)\ \sum_{k\in\mathfrak{C}_{\ell_2}} \frac{1}{\gamma_{ik}} = \sum_{k\in\mathfrak{C}_{\ell_2}} \frac{1}{\gamma_{jk}}. \tag{4.57}$$

The ith row and ℓ_2th column in condition (4.56) gives

$$\Big(\forall \ell, \ell_2 \in \widehat{\mathcal{G}}\Big)(\forall i,j \in \mathfrak{C}_\ell)\ \ \chi_i \sum_{k\in\mathfrak{C}_{\ell_2}} \frac{1}{\gamma_{ik}} = \chi_j \sum_{k\in\mathfrak{C}_{\ell_2}} \frac{1}{\gamma_{jk}}. \tag{4.58}$$

Since the assumption is that there are at least two clusters in π, from (4.57) and (4.58) we find that $\chi_i = \chi_j$, for all $i,j \in \mathfrak{C}_\ell$ and all $\ell \in \widehat{\mathcal{G}}$, i.e., $L_{\mathrm{D}} \in \mathcal{S}_{\mathrm{cl}}$. This, together with (4.58), gives

$$\Big(\forall \ell, \ell_2 \in \widehat{\mathcal{G}}\Big)(\forall i,j \in \mathfrak{C}_\ell)\ \sum_{k\in\mathfrak{C}_{\ell_2}} \frac{1}{\gamma_{ik}} = \sum_{k\in\mathfrak{C}_{\ell_2}} \frac{1}{\gamma_{jk}},$$

which is equivalent to $\operatorname{im}(\Gamma\mathfrak{P}) \subseteq \mathfrak{X}_{\mathrm{cl}}$, i.e., $\Gamma\mathfrak{X}_{\mathrm{cl}} \subseteq \mathfrak{X}_{\mathrm{cl}}$. □

4.3.5 Clustering of power systems

Let us assume that the system (4.43) is weakly synchronized with respect to a partition π. Let also the initial condition satisfy $\delta(0), \dot{\delta}(0) \in \mathfrak{X}_{\mathrm{cl}}$. Then there exist $\widehat{\delta}$ and $\widehat{\boldsymbol{V}}_{\widehat{\mathcal{G}}}$ such that $\delta(t) = \mathfrak{P}\widehat{\delta}(t)$ and $\boldsymbol{V}_{\mathcal{G}}(t) = \mathfrak{P}\widehat{\boldsymbol{V}}_{\widehat{\mathcal{G}}}(t)$, which also gives us $V_{\mathcal{G}}(t) = \mathfrak{P}\widehat{V}_{\widehat{\mathcal{G}}}(t)$ and $\theta_{\mathcal{G}}(t) = \mathfrak{P}\widehat{\theta}_{\widehat{\mathcal{G}}}(t)$. Inserting this into (4.43) with dynamics rewritten as in (4.47), we find

$$\begin{aligned} M\mathfrak{P}\ddot{\widehat{\delta}}(t) + D\mathfrak{P}\dot{\widehat{\delta}}(t) &= f - L_{\mathrm{D}}\left(E \circ \mathfrak{P}\widehat{V}_{\widehat{\mathcal{G}}}(t) \circ \sin\Big(\mathfrak{P}\widehat{\delta}(t) - \mathfrak{P}\widehat{\theta}_{\widehat{\mathcal{G}}}(t)\Big)\right), \\ \begin{bmatrix} L_{\mathrm{D}}\left(\boldsymbol{E}_{\mathcal{G}}(t) - \mathfrak{P}\widehat{\boldsymbol{V}}_{\widehat{\mathcal{G}}}(t)\right) \\ 0 \end{bmatrix} &= \begin{bmatrix} \mathfrak{L}_{11} & \mathfrak{L}_{12} \\ \mathfrak{L}_{12}^{\mathrm{T}} & \mathfrak{L}_{22} \end{bmatrix} \begin{bmatrix} \mathfrak{P}\widehat{\boldsymbol{V}}_{\widehat{\mathcal{G}}}(t) \\ \boldsymbol{V}_{\overline{\mathcal{G}}}(t) \end{bmatrix}. \end{aligned}$$

Assuming additionally that $E \in \mathfrak{X}_{\mathrm{cl}}$, i.e., $E = \mathfrak{P}\widehat{E}$ for some $\widehat{E} \in \mathbb{R}^{\widehat{n}}$, and premultiplying the above dynamics and first block-row of the constraint by $\mathfrak{P}^{\mathrm{T}}$, we obtain

$$\widehat{M}\ddot{\widehat{\delta}}(t) + \widehat{D}\dot{\widehat{\delta}}(t) = \widehat{f} - \widehat{L}_{\mathrm{D}}\left(\widehat{E} \circ \widehat{V}_{\widehat{\mathcal{G}}}(t) \circ \sin\Big(\widehat{\delta}(t) - \widehat{\theta}_{\widehat{\mathcal{G}}}(t)\Big)\right), \tag{4.59a}$$

$$\begin{bmatrix} \widehat{L}_{\mathrm{D}}\left(\widehat{\boldsymbol{E}}_{\widehat{\mathcal{G}}}(t) - \widehat{\boldsymbol{V}}_{\widehat{\mathcal{G}}}(t)\right) \\ 0 \end{bmatrix} = \begin{bmatrix} \widehat{\mathfrak{L}}_{11} & \widehat{\mathfrak{L}}_{12} \\ \widehat{\mathfrak{L}}_{12}^{\mathrm{T}} & \mathfrak{L}_{22} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{V}}_{\widehat{\mathcal{G}}}(t) \\ \boldsymbol{V}_{\overline{\mathcal{G}}}(t) \end{bmatrix}, \tag{4.59b}$$

where $\widehat{M} = \mathfrak{P}^{\mathrm{T}} M \mathfrak{P}$, $\widehat{D} = \mathfrak{P}^{\mathrm{T}} D \mathfrak{P}$, $\widehat{f} = \mathfrak{P}^{\mathrm{T}} f$, $\widehat{L}_{\mathrm{D}} = \mathfrak{P}^{\mathrm{T}} L_{\mathrm{D}} \mathfrak{P}$, $\widehat{\mathfrak{L}}_{11} = \mathfrak{P}^{\mathrm{T}} \mathfrak{L}_{11} \mathfrak{P}$, $\widehat{\mathfrak{L}}_{12} = \mathfrak{P}^{\mathrm{T}} \mathfrak{L}_{12}$. Moreover, from $\delta(t) = \mathfrak{P}\widehat{\delta}(t)$ follows that $\widehat{\delta}(0) = \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1} \mathfrak{P}^{\mathrm{T}} \delta(0)$ and $\dot{\widehat{\delta}}(0) = \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1} \mathfrak{P}^{\mathrm{T}} \dot{\delta}(0)$.

Notice that the reduced model (4.59) is again a power system of the same form as (4.43). In particular, we have that $\widehat{M}$, $\widehat{D}$, and $\widehat{L}_{\mathrm{D}}$ are positive definite diagonal matrices and that $\widehat{\mathfrak{L}}$ is a Laplacian matrix. Additionally, note that this projection-based MOR can be done for arbitrary power system and arbitrary partition. In general, we can take (4.59) with $\widehat{\delta}(0) = \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1} \mathfrak{P}^{\mathrm{T}} \delta(0)$, $\dot{\widehat{\delta}}(0) = \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1} \mathfrak{P}^{\mathrm{T}} \dot{\delta}(0)$, and $\widehat{E} = \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1} \mathfrak{P}^{\mathrm{T}} E$. We can also apply Kron reduction to this reduced model.

4.3.6 Illustrative example

For the example in Figure 4.7, let $\chi_i = 1$ and $\chi_{ij} = 1$ for all i, j. Then we have

$$\Gamma = \frac{1}{32} \begin{bmatrix} 21 & 5 & 2 & 2 & 2 \\ 5 & 21 & 2 & 2 & 2 \\ 2 & 2 & 16 & 8 & 4 \\ 2 & 2 & 8 & 16 & 4 \\ 2 & 2 & 4 & 4 & 20 \end{bmatrix}.$$

Additionally, let $M = D = I_5$, $f = 0$, and $E = \mathbb{1}_5$. Then, using Theorem 4.21, we see that the first and second generators are synchronized, and that the same is true for the third and fourth. By definition, this implies that the system is strongly synchronized with respect to $\{\{1,2\},\{3,4\},\{5\}\}$. On the other hand, from Theorem 4.32 and

$$\Gamma \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 13 & 3 \\ 13 & 3 \\ 2 & 14 \\ 2 & 14 \\ 2 & 14 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \left(\frac{1}{16} \begin{bmatrix} 13 & 3 \\ 2 & 14 \end{bmatrix} \right),$$

we see that the system is weakly synchronized with respect to $\{\{1,2\},\{3,4,5\}\}$, but not strongly. Using the partition $\pi = \{\{1,2\},\{3,4,5\}\}$ for clustering, we find the following reduced quantities: $\widehat{M} = \widehat{D} = \left[\begin{smallmatrix} 2 & 0 \\ 0 & 3 \end{smallmatrix}\right]$, $\widehat{f} = 0$, $\widehat{E} = \mathbb{1}_2$, $\widehat{L}_{\mathrm{D}} = \widehat{\mathfrak{L}}_{11} = \left[\begin{smallmatrix} 2 & 0 \\ 0 & 3 \end{smallmatrix}\right]$, $\widehat{\mathfrak{L}}_{12} = \left[\begin{smallmatrix} -2 & 0 \\ 0 & -3 \end{smallmatrix}\right]$, $\widehat{\Gamma} = \frac{1}{8}\left[\begin{smallmatrix} 13 & 3 \\ 3 & 21 \end{smallmatrix}\right]$. The Figure 4.8 shows the partition and Figure 4.9 the associated reduced power system. From the definition of weak synchronization, we know that this reduced power system exactly reproduces the initial value response of the original system for any initial condition $\delta(0), \dot{\delta}(0) \in \mathcal{X}_{\mathrm{cl}}$, taking the initial condition of the reduced model to be $\widehat{\delta}(0) = \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1} \mathfrak{P}^{\mathrm{T}} \delta(0)$ and $\dot{\widehat{\delta}}(0) = \left(\mathfrak{P}^{\mathrm{T}}\mathfrak{P}\right)^{-1} \mathfrak{P}^{\mathrm{T}} \dot{\delta}(0)$.

To demonstrate the possibility to cluster using any partition, including those with respect to which the power system is not weakly synchronized, and any initial condition, we show simulation result for partition $\{\{1,2,3\},\{4,5\}\}$ in Figure 4.10. We see that, in this case, the reduced model matches the steady state and approximates the transient behavior. Finding sufficient conditions for matching the steady state and deriving error bounds is a possible topic of future research.

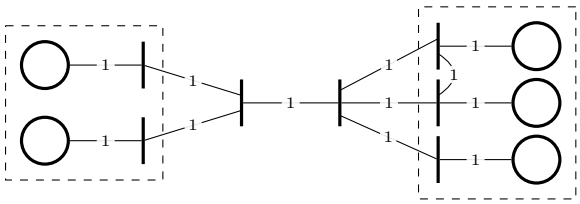

Figure 4.8: Partition $\{\{1,2\},\{3,4,5\}\}$ applied to the original power system in Figure 4.7 with $\chi_i = \chi_{ij} = 1$ for all i, j.

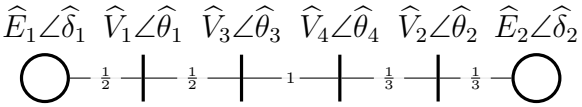

Figure 4.9: Reduced power system obtained by clustering the system in Figure 4.8 with $M = D = I_5$, $f = 0$, and $E = \mathbb{1}_5$.

4.4 Conclusion

In Section 4.2, we have extended results on MOR of leader-follower networks with single integrator agent dynamics from [MTC14] to leader-follower networks with arbitrary linear multivariable agent dynamics. We have also extended these results to the case that the approximation error is measured in the $\mathcal{H}_\infty$-norm. The proposed MOR technique reduces the complexity of the network topology by clustering the agents. We have shown that clustering amounts to applying a specific Petrov-Galerkin projection associated with the graph partition. The resulting reduced order model can be interpreted as a networked multi-agent system with a weighted, directed network graph. If the original network is clustered using an almost equitable graph partition, then its consensus properties are preserved. We have provided a priori upper bounds on the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ model reduction errors in this case. These error bounds depend on an auxiliary system related to the agent dynamics, the eigenvalues of the Laplacian matrices of the original and the reduced network, and on the number of cellmates of the leaders in the network. Finally, we have provided some insight into the general case of clustering according to arbitrary, not necessarily almost equitable, partitions. Here, direct computation of a priori upper bounds on the error is not as straightforward as in the case of almost equitable partitions. We have shown that in this more general case, one can bound the model reduction errors by first optimally approximating the original network by a new network for which the chosen partition is almost equitable, and then bounding the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ errors using the triangle inequality.

In Section 4.3, we analyzed power systems consisting of generators and buses. We introduced a notion of synchronization for a pair of generators and two for a partition of the set of generators. We proved equivalent conditions depending on the Kron-reduced system being symmetrical or equitable. This additionally gives a relation between symmetrical matrices and equitable partitions. We showed how a synchronized power

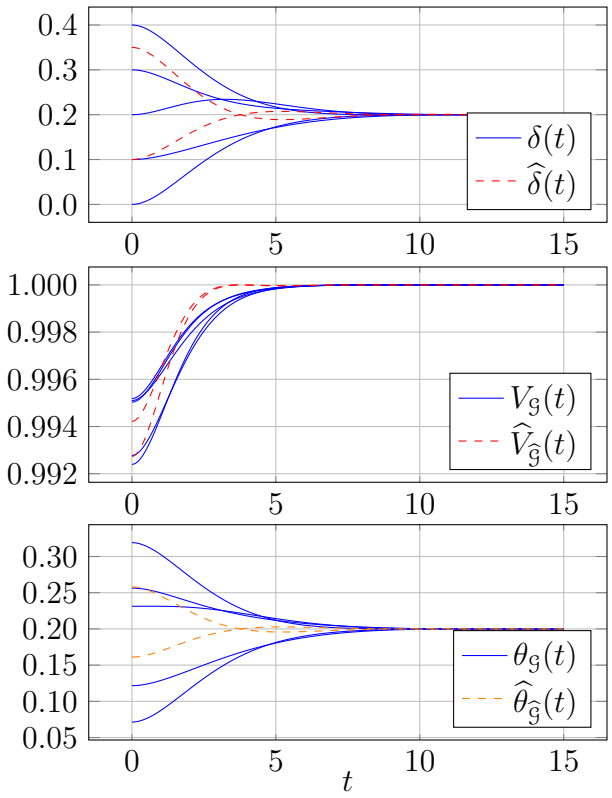

Figure 4.10: Initial value response of the original power system from Figure 4.7 and a reduced system obtained by clustering with partition $\{\{1,2,3\},\{4,5\}\}$. Original system's parameters are $\chi_i = \chi_{ij} = 1$ for all i,j, $M = D = I_5$, $f = 0$, and $E = \mathbb{1}_5$. The initial value is $\delta(0) = (0, 0.1, 0.2, 0.3, 0.4)$ and $\dot{\delta}(0) = 0$.

system can be exactly approximated with a reduced system by clustering generators and their buses. Furthermore, this provides a clustering-based MOR method for arbitrary power systems, although finding bounds for the approximation error remains an open problem.

These results give motivation for further research into efficient clustering methods for linear and nonlinear multi-agent systems.

CHAPTER 5

SUBSYSTEM REDUCTION FOR INTERCONNECTED SYSTEMS

Contents

5.1 Introduction

Here, we consider MOR of interconnected systems, in particular of LTI systems linearly interconnected through their inputs and outputs. Moreover, we want to preserve the interconnection structure by only reducing the subsystems, while also preserving the stability of the coupled system. This is a complementary approach to clustering-based methods discussed in previous chapters.

Several approaches were investigated in the literature. Reis and Stykel [RS07, RS08b] proposed using BT for each subsystem and proved a sufficient condition for stability of the interconnected system. Furthermore, they developed an a priori error bound of the network system based on the error bounds for the individual subsystems. Vandendorpe and Van Dooren [VVD08] showed sufficient conditions for transfer function interpolation of the interconnected system while preserving the interconnection structure. Sandberg and Murray [SM09] used block-diagonal generalized Gramians to extend BT

for unstructured LTI systems to interconnected systems. Monshizadeh et al. [MTC13] proposed methods preserving stability or synchronization for multi-agent systems.

We consider balancing-based and $\mathcal{H}_2$-optimal MOR of interconnected systems in Section 5.2 and Section 5.3, respectively.

5.2 Stability-preserving balancing-based model order reduction

5.2.1 Preliminaries

We extend the stability-preserving MOR method, based on bounded real balanced truncation (BRBT), for multi-agent systems from [MTC13] to coupled systems considered in [RS07]. More information about BRBT is in the following section.

Reis and Stykel [RS07] study systems of $\mathfrak{n}$ coupled LTI subsystems

$$\begin{aligned} E_i\dot{x}_i(t) &= A_i x_i(t) + B_i u_i(t), \\ y_i(t) &= C_i x_i(t), \end{aligned} \tag{5.1a}$$

with interconnections and external input

$$u_i(t) = K_{i1}y_1(t) + \cdots + K_{i\mathfrak{n}}y_{\mathfrak{n}}(t) + F_i u(t), \tag{5.1b}$$

and external output

$$y(t) = G_1 y_1(t) + \cdots + G_{\mathfrak{n}} y_{\mathfrak{n}}(t), \tag{5.1c}$$

where $i \in \{1, 2, \ldots, \mathfrak{n}\}$, $E_i, A_i \in \mathbb{R}^{n_i \times n_i}$, $B_i \in \mathbb{R}^{n_i \times m_i}$, $C_i \in \mathbb{R}^{p_i \times n_i}$, $x_i(t) \in \mathbb{R}^{n_i}$ is the state, $u_i(t) \in \mathbb{R}^{m_i}$ is the internal input, $y_i(t) \in \mathbb{R}^{p_i}$ is the internal output, $K_{ij} \in \mathbb{R}^{m_i \times p_j}$, $F_i \in \mathbb{R}^{m_i \times m}$, $u(t) \in \mathbb{R}^m$ is the external input, $G_i \in \mathbb{R}^{p \times p_i}$, and $y(t) \in \mathbb{R}^p$ is the external output. Here, we assume the matrices E_i are invertible and that the matrix pencils $A_i - \lambda E_i$ are asymptotically stable. The LTI system (5.1) can be written (see [RS07]) as

$$\begin{aligned} \mathcal{E}\dot{x}(t) &= \mathcal{A}x(t) + \mathcal{B}u(t), \\ y(t) &= \mathcal{C}x(t), \end{aligned} \tag{5.2}$$

with $\mathcal{E}, \mathcal{A} \in \mathbb{R}^{n \times n}$, $\mathcal{B} \in \mathbb{R}^{n \times m}$, $\mathcal{C} \in \mathbb{R}^{p \times n}$, and $x(t) = \operatorname{col}(x_1(t), x_2(t), \ldots, x_{\mathfrak{n}}(t)) \in \mathbb{R}^n$, where

$$\begin{aligned} &\mathcal{E} = E_{\mathrm{D}}, \quad \mathcal{A} = A_{\mathrm{D}} + B_{\mathrm{D}} K C_{\mathrm{D}}, \quad \mathcal{B} = B_{\mathrm{D}} F, \quad \mathcal{C} = G C_{\mathrm{D}}, \\ &E_{\mathrm{D}} = \operatorname{diag}(E_1, E_2, \ldots, E_{\mathfrak{n}}), \quad A_{\mathrm{D}} = \operatorname{diag}(A_1, A_2, \ldots, A_{\mathfrak{n}}), \\ &B_{\mathrm{D}} = \operatorname{diag}(B_1, B_2, \ldots, B_{\mathfrak{n}}), \quad C_{\mathrm{D}} = \operatorname{diag}(C_1, C_2, \ldots, C_{\mathfrak{n}}), \\ &K = \begin{bmatrix} K_{11} & K_{12} & \cdots & K_{1\mathfrak{n}} \\ K_{21} & K_{22} & \cdots & K_{2\mathfrak{n}} \\ \vdots & \vdots & \ddots & \vdots \\ K_{\mathfrak{n}1} & K_{\mathfrak{n}2} & \cdots & K_{\mathfrak{n}\mathfrak{n}} \end{bmatrix}, \quad F = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_{\mathfrak{n}} \end{bmatrix}, \quad G = \begin{bmatrix} G_1 & G_2 & \cdots & G_{\mathfrak{n}} \end{bmatrix}. \end{aligned} \tag{5.3}$$

Algorithm 5.1: Bounded real balanced truncation [OJ88]

Input: Asymptotically stable system $(E; A, B, C)$ and $\gamma > 0$ such that $\|H\|_{\mathcal{H}_\infty} < \gamma$.

Output: ROM $(\widehat{E}; \widehat{A}, \widehat{B}, \widehat{C})$.

1 Compute the maximal solutions P and Q of

$$APE^{\mathrm{T}} + EPA^{\mathrm{T}} + BB^{\mathrm{T}} + \frac{1}{\gamma^2} EPC^{\mathrm{T}}CPE^{\mathrm{T}} = 0,$$

$$A^{\mathrm{T}}QE + E^{\mathrm{T}}QA + C^{\mathrm{T}}C + \frac{1}{\gamma^2} E^{\mathrm{T}}QBB^{\mathrm{T}}QE = 0.$$

2 Proceed as in Algorithm 2.1 using the obtained P and Q.

Denote $H_i(s) = C_i(sE_i - A_i)^{-1}B_i$ and $\mathcal{H}(s) = \mathcal{C}(s\mathcal{E} - \mathcal{A})^{-1}\mathcal{B}$ the transfer functions of the subsystems and the interconnected system respectively. Furthermore, let $H_{\mathrm{D}}(s) = C_{\mathrm{D}}(sE_{\mathrm{D}} - A_{\mathrm{D}})^{-1}B_{\mathrm{D}} = \mathrm{diag}(H_1(s), H_2(s), \ldots, H_{\mathfrak{n}}(s))$. Then, we have

$$\mathcal{H}(s) = G(I - H_{\mathrm{D}}(s)K)^{-1}H_{\mathrm{D}}(s)F = GH_{\mathrm{D}}(s)(I - KH_{\mathrm{D}}(s))^{-1}F.$$

5.2.2 Bounded real balanced truncation

BRBT is a modification of the standard BT method, originally developed to preserve strict bounded realness, or equivalently the bound $\|H\|_{\mathcal{H}_\infty} < 1$, of the original system [OJ88]. Algorithm 5.1 describes a straightforward extension of BRBT which preserves the bound $\|H\|_{\mathcal{H}_\infty} < \gamma$, for any given $\gamma > 0$. The following theorem presents some properties of BRBT and an $\mathcal{H}_\infty$-error bound, and is also a straightforward extension of results in [OJ88].

Theorem 5.1:
Let $H(s) = C(sE - A)^{-1}B$ be an LTI system of order n such that $\|H\|_{\mathcal{H}_\infty} < \gamma$ and $\widehat{H}(s) = \widehat{C}(s\widehat{E} - \widehat{A})^{-1}\widehat{B}$ its ROM of order $r < n$ obtained by BRBT. Then $\widehat{H}$ is asymptotically stable, $\|\widehat{H}\|_{\mathcal{H}_\infty} < \gamma$, and

$$\|H - \widehat{H}\|_{\mathcal{H}_\infty} \leqslant 2 \sum_{i=r+1}^{n} \xi_i,$$

where ξ_i are the *bounded real characteristic values* (square roots of eigenvalues of $PE^{\mathrm{T}}QE$ in Algorithm 5.1). ◊

5.2.3 Stability-preserving model order reduction

In the above setting, the multi-agent systems studied by Monshizadeh et al. [MTC13] can be represented with

$$\begin{gathered} E_i = I_n, \quad A_i = A_1, \quad B_i = B_1, \quad C_i = C_1, \quad \text{for } i = 1, 2, \ldots, \mathfrak{n}, \\ K = -\mathfrak{L} \otimes I_{m_1}, \quad m_1 = p_1, \end{gathered} \tag{5.4}$$

where $\mathfrak{L}$ is the Laplacian matrix of an undirected interconnection graph. For these systems, the following sufficient condition for asymptotic stability was proven.

Lemma 5.2 ([MTC13, Lemma 3.1]):
If $\|\mathfrak{L}\|_2 \, \|H_1\|_{\mathcal{H}_\infty} < 1$, then the multi-agent system (5.2) with (5.4) is asymptotically stable. ◊

Based on this, they propose using BRBT to reduce the subsystem (A_1, B_1, C_1) to $(\widehat{A}_1, \widehat{B}_1, \widehat{C}_1)$ with transfer function $\widehat{H}_1$ such that $\|\widehat{H}_1\|_{\mathcal{H}_\infty} < \frac{1}{\|\mathfrak{L}\|_2}$. Since the interconnection structure is preserved in the reduced network system, i.e., the Laplacian matrix $\mathfrak{L}$ remains the same for the reduced multi-agent system, asymptotic stability is also preserved.

On the other hand, Reis and Stykel [RS07] prove the following sufficient condition for asymptotic stability of the interconnected system (5.2).

Theorem 5.3 ([RS07, Corollary 2.3]):
Let $\Phi_2 \in \mathbb{R}^{\mathfrak{n} \times \mathfrak{n}}$ be given by

$$\Phi_2 = \begin{bmatrix} \|K_{11}\|_2 & \|K_{12}\|_2 & \cdots & \|K_{1\mathfrak{n}}\|_2 \\ \|K_{21}\|_2 & \|K_{22}\|_2 & \cdots & \|K_{2\mathfrak{n}}\|_2 \\ \vdots & \vdots & \ddots & \vdots \\ \|K_{\mathfrak{n}1}\|_2 & \|K_{\mathfrak{n}2}\|_2 & \cdots & \|K_{\mathfrak{n}\mathfrak{n}}\|_2 \end{bmatrix}$$

and $\Psi \in \mathbb{R}^{\mathfrak{n} \times \mathfrak{n}}$ by

$$\Psi = \Phi_2 \operatorname{diag}\left(\|H_1\|_{\mathcal{H}_\infty}, \|H_2\|_{\mathcal{H}_\infty}, \ldots, \|H_{\mathfrak{n}}\|_{\mathcal{H}_\infty}\right).$$

If $\rho(\Psi) < 1$, then the system (5.2) is asymptotically stable. ◊

The similarity between the conditions in Lemma 5.2 and Theorem 5.3 motivates us to extend the use of BRBT from multi-agent systems to general interconnected systems. Notice that in the case of multi-agent systems, we have $\Phi_2 = |\mathfrak{L}|$, $\Psi = |\mathfrak{L}| \, \|H_1\|_{\mathcal{H}_\infty}$, and $\rho(\Psi) = \rho(|\mathfrak{L}|) \, \|H_1\|_{\mathcal{H}_\infty}$. Next, from [HJ85, Theorem 8.1.18], we have $\rho(|\mathfrak{L}|) \geqslant \rho(\mathfrak{L})$. Since $\mathfrak{L}$ is positive semi-definite, we have $\rho(\mathfrak{L}) = \|\mathfrak{L}\|_2$ and thus $\rho(|\mathfrak{L}|) \geqslant \|\mathfrak{L}\|_2$. Therefore, $\rho(\Psi) \geqslant \|\mathfrak{L}\|_2 \, \|H_1\|_{\mathcal{H}_\infty}$.

The following result allows us to extend the idea from [MTC13] for preserving the sufficient condition for asymptotic stability by applying BRBT to subsystems of the interconnected system (5.1).

Proposition 5.4:
Let $\rho(\Psi) < 1$ and

$$\|\widehat{H}_i\|_{\mathcal{H}_\infty} < \frac{1}{\rho(\Psi)} \|H_i\|_{\mathcal{H}_\infty}, \text{ for } i = 1, 2, \ldots, \mathfrak{n}.$$

Then $\rho(\widehat{\Psi}) < 1$. ◊

Proof. We see that there must exist $\alpha \in (0, 1)$ such that

$$\|\widehat{H}_i\|_{\mathcal{H}_\infty} \leqslant \alpha \frac{1}{\rho(\Psi)} \|H_i\|_{\mathcal{H}_\infty}, \text{ for } i = 1, 2, \ldots, \mathfrak{n}.$$

It follows that

$$\begin{aligned}
\rho(\widehat{\Psi}) &= \rho\big(\Phi_2 \operatorname{diag}\big(\|\widehat{H}_1\|_{\mathcal{H}_\infty}, \|\widehat{H}_2\|_{\mathcal{H}_\infty}, \ldots, \|\widehat{H}_{\mathfrak{n}}\|_{\mathcal{H}_\infty}\big)\big) \\
&\leqslant \rho\left(\Phi_2 \operatorname{diag}\left(\frac{\alpha}{\rho(\Psi)} \|H_1\|_{\mathcal{H}_\infty}, \frac{\alpha}{\rho(\Psi)} \|H_2\|_{\mathcal{H}_\infty}, \ldots, \frac{\alpha}{\rho(\Psi)} \|H_{\mathfrak{n}}\|_{\mathcal{H}_\infty}\right)\right) \\
&= \frac{\alpha}{\rho(\Psi)} \rho(\Phi_2 \operatorname{diag}\big(\|H_1\|_{\mathcal{H}_\infty}, \|H_2\|_{\mathcal{H}_\infty}, \ldots, \|H_{\mathfrak{n}}\|_{\mathcal{H}_\infty}\big)) \\
&< 1,
\end{aligned}$$

where we used that ρ is increasing on nonnegative matrices ([HJ85, Corollary 8.1.19]) and homogeneous. □

Clearly, $\|H_i\|_{\mathcal{H}_\infty} < \frac{1}{\rho(\Psi)} \|H_i\|_{\mathcal{H}_\infty}$ when $\rho(\Psi) < 1$. Thus, the idea is to apply BRBT to H_i with $\gamma_i = \frac{1}{\rho(\Psi)} \|H_i\|_{\mathcal{H}_\infty}$. Using Theorem 5.1, we find that

$$\|H_i - \widehat{H}_i\|_{\mathcal{H}_\infty} \leqslant 2 \sum_{j=r_i+1}^{n_i} \xi_j^{(i)}(\gamma_i), \tag{5.5}$$

where r_i is the order of $\widehat{H}_i$ and $\xi_1^{(i)}(\gamma_i), \xi_2^{(i)}(\gamma_i), \ldots, \xi_{n_i}^{(i)}(\gamma_i)$ are the bounded real characteristic values of H_i. From Corollary 4.2 in [RS07], we find the upper bound for the $\mathcal{H}_\infty$-error from reducing subsystems by BRBT in the interconnected system (5.1).

Theorem 5.5:
Denote

$$\delta := \max_{i=1,2,\ldots,\mathfrak{n}} 2 \sum_{j=r_i+1}^{n_i} \xi_j^{(i)}(\gamma_i),$$

the maximum of the bounds from (5.5). Furthermore, let

$$\begin{gathered}
g = \left\|K(I - HK)^{-1}\right\|_{\mathcal{H}_\infty}, \; g_1 = \left\|G(I - HK)^{-1}\right\|_{\mathcal{H}_\infty}, \; g_2 = \left\|(I - KH)^{-1}F\right\|_{\mathcal{H}_\infty}, \\
c_1 = g_1(\|F\|_2 + g \|HF\|_{\mathcal{H}_\infty}), \text{ and } c_2 = g_2(\|R\|_2 + g \|GH\|_{\mathcal{H}_\infty}).
\end{gathered} \tag{5.6}$$

If $g\delta < 1$, then

$$\|\mathcal{H} - \widehat{\mathcal{H}}\|_{\mathcal{H}_\infty} \leqslant \min\{c_1, c_2\} \frac{\delta}{1 - g\delta}. \tag{5.7}$$

◊

Algorithm 5.2: Balancing method for network systems preserving stability and structure

Input: Network system $(\mathcal{E}; \mathcal{A}, \mathcal{B}, \mathcal{C})$ from (5.2) with $\rho(\Psi) < 1$, tolerance $\varepsilon > 0$.

Output: Asymptotically stable reduced network system $(\widehat{\mathcal{E}}; \widehat{\mathcal{A}}, \widehat{\mathcal{B}}, \widehat{\mathcal{C}})$ with $\|\mathcal{H} - \widehat{\mathcal{H}}\|_{\mathcal{H}_\infty} \leqslant \varepsilon$.

1 Compute $\gamma_i = \frac{1}{\rho(\Psi)} \|H_i\|_{\mathcal{H}_\infty}$ for $i = 1, 2, \ldots, \mathfrak{n}$.

2 Compute bounded real characteristic values $\xi_j^{(i)}(\gamma_i)$ for $i = 1, 2, \ldots, \mathfrak{n}$, $j = 1, 2, \ldots, n_i$.

3 Compute g, c_1, and c_2 from (5.6).

4 Find minimal r_i such that

$$2 \sum_{j=r_i+1}^{n_i} \xi_j^{(i)}(\gamma_i) \leqslant \frac{\varepsilon}{\min\{c_1, c_2\} + g\varepsilon}$$

for $i = 1, 2, \ldots, \mathfrak{n}$.

5 Reduce the ith subsystem to order r_i using BRBT with $\gamma = \gamma_i$ (see Algorithm 5.1), for $i = 1, 2, \ldots, \mathfrak{n}$.

Theorem 5.5 enables us to adaptively choose reduced orders $r_1, r_2, \ldots, r_\mathfrak{n}$ of the subsystems. To see this, let $\varepsilon > 0$ be a tolerance for the error $\|\mathcal{H} - \widehat{\mathcal{H}}\|_{\mathcal{H}_\infty}$. From (5.7), we see that the bound $\|\mathcal{H} - \widehat{\mathcal{H}}\|_{\mathcal{H}_\infty} \leqslant \varepsilon$ can be achieved by enforcing

$$\min\{c_1, c_2\} \frac{\delta}{1 - g\delta} \leqslant \varepsilon,$$

which is equivalent to

$$\delta \leqslant \frac{\varepsilon}{\min\{c_1, c_2\} + g\varepsilon}.$$

Finding minimal r_i such that the right hand side in (5.5) is less than or equal to $\frac{\varepsilon}{\min\{c_1, c_2\} + g\varepsilon}$ guarantees that $\|\mathcal{H} - \widehat{\mathcal{H}}\|_{\mathcal{H}_\infty} \leqslant \varepsilon$. Note that $g \cdot \frac{\varepsilon}{\min\{c_1, c_2\} + g\varepsilon} < 1$, which is the assumption in Theorem 5.5.

5.2.4 Numerical example

We use the interconnected string-beam example from [RS07, Section 6], illustrated in Figure 5.1. After discretizing the associated partial differential equation by finite differences, the string subsystem has $n_1 = 1006$ states, $m_1 = 3$ inputs, and $p_1 = 2$ outputs, while the beam subsystem has $n_2 = 1006$ states, $m_2 = 2$ inputs, and $p_2 = 2$ outputs. The interconnected system has $n = n_1 + n_2 = 2012$ states, $m = 1$ input, and $p = 2$ outputs. The left figure in Figure 5.2 shows magnitude plots of the subsystems and the interconnected system.

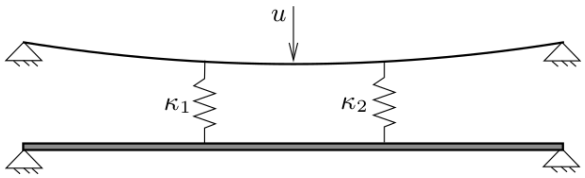

Figure 5.1: Interconnected string-beam example from [RS07]

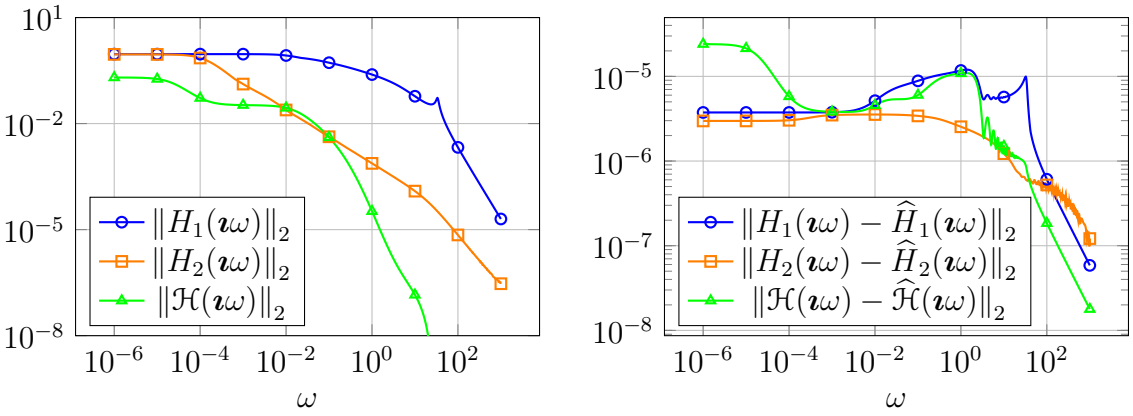

Figure 5.2: Magnitude plot of the full-order and error systems

For this system, we have $\rho(\Psi) = 0.9157 < 1$, which guarantees the interconnected system is asymptotically stable by Theorem 5.3. Furthermore, this allows us to use the method described in the previous section.

Let us set the reduction error tolerance for the coupled system to $\varepsilon = 10^{-3}$. From this, we find that the error tolerance for each of the subsystems is $\frac{\varepsilon}{\min\{c_1,c_2\}+g\varepsilon} = 8.5821 \cdot 10^{-5}$. Applying BRBT from Algorithm 5.1, with $\gamma_i = \frac{\|H_i\|_{\mathcal{H}_\infty}}{\rho(\Psi)}$ and adaptive reduced order decision, we determine reduced orders to be $r_1 = 30$ and $r_2 = 14$ for the string and beam subsystems respectively. We find that $\|H_1 - \widehat{H}_1\|_{\mathcal{H}_\infty} = 1.2001 \cdot 10^{-5}$, $\|H_2 - \widehat{H}_2\|_{\mathcal{H}_\infty} = 3.544 \cdot 10^{-6}$, and $\|\mathcal{H} - \widehat{\mathcal{H}}\|_{\mathcal{H}_\infty} = 2.3893 \cdot 10^{-5} \leqslant 10^{-3}$. The right figure in Figure 5.2 shows magnitude plots of error systems.

5.3 $\mathcal{H}_2$-optimal subsystem reduction

Here, we are interested in subsystem reduction for interconnected systems (5.1) which minimizes the $\mathcal{H}_2$-error of the overall network system. We also consider multi-agent systems, where the additional constraint is that all subsystems are equal.

Notice that one special case of subsystem reduction for interconnected systems is weighted MOR [VVD08, Section 4], for which there are interpolatory $\mathcal{H}_2$-optimality conditions [ABGA13, BBG15]. We will focus on the general case.

5.3.1 Interconnected systems

We use the same notation as in the previous section. We want to find a reduced model with reduced subsystems

$$\begin{aligned}\widehat{E}_i\dot{\widehat{x}}_i(t) &= \widehat{A}_i\widehat{x}_i(t) + \widehat{B}_i\widehat{u}_i(t),\\ \widehat{y}_i(t) &= \widehat{C}_i\widehat{x}_i(t),\end{aligned} \tag{5.8a}$$

but preserved interconnections and external input

$$\widehat{u}_i(t) = K_{i1}\widehat{y}_1(t) + \cdots + K_{i\mathfrak{n}}\widehat{y}_{\mathfrak{n}}(t) + F_i u(t), \tag{5.8b}$$

and external output

$$\widehat{y}(t) = G_1\widehat{y}_1(t) + \cdots + G_{\mathfrak{n}}\widehat{y}_{\mathfrak{n}}(t), \tag{5.8c}$$

where $\widehat{E}_i, \widehat{A}_i \in \mathbb{R}^{r_i\times r_i}$, $\widehat{B} \in \mathbb{R}^{r_i\times m_i}$, $\widehat{C} \in \mathbb{R}^{p_i\times r_i}$, for some $r_i < n_i$, $i = 1, 2, \ldots, \mathfrak{n}$. Similarly as for the original system, the reduced system (5.8) can be represented by

$$\begin{aligned}\widehat{\mathcal{E}}\dot{\widehat{x}}(t) &= \widehat{\mathcal{A}}\widehat{x}(t) + \widehat{\mathcal{B}}u(t),\\ \widehat{y}(t) &= \widehat{\mathcal{C}}\widehat{x}(t),\end{aligned} \tag{5.9}$$

with $\widehat{x}(t) = \operatorname{col}(\widehat{x}_1(t), \widehat{x}_2(t), \ldots, \widehat{x}_{\mathfrak{n}}(t))$ and

$$\begin{aligned}\widehat{\mathcal{E}} = \widehat{E}, \quad \widehat{\mathcal{A}} = \widehat{A} + \widehat{B}K\widehat{C}, \quad \widehat{\mathcal{B}} = \widehat{B}F, \quad \widehat{\mathcal{C}} = G\widehat{C},\\ \widehat{E} = \operatorname{diag}\left(\widehat{E}_1, \widehat{E}_2, \ldots, \widehat{E}_{\mathfrak{n}}\right), \quad \widehat{A} = \operatorname{diag}\left(\widehat{A}_1, \widehat{A}_2, \ldots, \widehat{A}_{\mathfrak{n}}\right),\\ \widehat{B} = \operatorname{diag}\left(\widehat{B}_1, \widehat{B}_2, \ldots, \widehat{B}_{\mathfrak{n}}\right), \quad \widehat{C} = \operatorname{diag}\left(\widehat{C}_1, \widehat{C}_2, \ldots, \widehat{C}_{\mathfrak{n}}\right).\end{aligned} \tag{5.10}$$

To find Gramian-based optimality conditions, we need to find gradients of the squared $\mathcal{H}_2$-error

$$\mathcal{J}(\widehat{E}_i, \widehat{A}_i, \widehat{B}_i, \widehat{C}_i) = \|H - H\|^2_{\mathcal{H}_2}. \tag{5.11}$$

With the Gramian-based formulation of the $\mathcal{H}_2$-norm, we can think of $\mathcal{J}$ as a function $f\colon X \to \mathbb{R}$ given implicitly by $f(x) = g(x, y)$, where $h(x, y) = 0$ and $\mathrm{D}_y h(x, y)$ is bijective for all x and y. Here, x would represent the matrices of the ROM and y the Gramians. The following lemma states that we can use Lagrange multiplier method to compute $\mathrm{D}f(x)$.

Lemma 5.6:
Let X, Y, Z be Banach spaces, U an open subset of X, x_0 an element of X, and $f\colon U \to Z$, $g\colon U \times Y \to Z$, $h\colon U \times Y \to Y$ functions. Furthermore, suppose there exists $y_0 \in Y$ such that $h(x_0, y_0) = 0$, g is Fréchet differentiable at (x_0, y_0), $\mathrm{D}_y h$ exists on $U \times Y$, $\mathrm{D}_y h(x_0, y_0)$ is bijective, and h is of class C^1 in a neighborhood of (x_0, y_0). Additionally, let $\mathcal{L}\colon U \times Y \times B(Y, Z) \to Z$ be such that $\mathcal{L}(x, y, \lambda) = g(x, y) - \lambda(h(x, y))$.

Then f is Fréchet differentiable at x_0 and

$$\mathrm{D}f(x_0) = \mathrm{D}_x\mathcal{L}(x_0, y_0, \lambda_0)$$

where λ_0 is such that $\mathrm{D}_y\mathcal{L}(x_0, y_0, \lambda_0) = 0$. ◊

Proof. Using Theorem 2.19 on h, we get that there exists an open subset U_{x_0} of U and a function $k\colon U_{x_0} \to Y$ such that $k(x_0) = y_0$, $h(x, k(x)) = 0$ for $x \in U_{x_0}$, and $\mathrm{D}k(x_0) = -\mathrm{D}_y h(x_0, y_0)^{-1}\,\mathrm{D}_x h(x_0, y_0)$. Then

$$\begin{aligned}\mathrm{D}f(x_0) &= \mathrm{D}_x g(x_0, y_0) + \mathrm{D}_y g(x_0, y_0)\,\mathrm{D}k(x_0)\\ &= \mathrm{D}_x g(x_0, y_0) - \mathrm{D}_y g(x_0, y_0)\mathrm{D}_y h(x_0, y_0)^{-1}\,\mathrm{D}_x h(x_0, y_0).\end{aligned}$$

On the other hand, we have

$$\mathrm{D}_x \mathcal{L}(x_0, y_0, \lambda_0) = \mathrm{D}_x g(x_0, y_0) - \lambda_0\,\mathrm{D}_x h(x_0, y_0), \tag{5.12a}$$
$$\mathrm{D}_y \mathcal{L}(x_0, y_0, \lambda_0) = \mathrm{D}_y g(x_0, y_0) - \lambda_0\,\mathrm{D}_y h(x_0, y_0). \tag{5.12b}$$

Therefore, from (5.12b) we get $\lambda_0 = \mathrm{D}_y g(x_0, y_0)\mathrm{D}_y h(x_0, y_0)^{-1}$ and inserting it into (5.12a), we find

$$\mathrm{D}_x \mathcal{L}(x_0, y_0, \lambda_0) = \mathrm{D}_x g(x_0, y_0) - \mathrm{D}_y g(x_0, y_0)\mathrm{D}_y h(x_0, y_0)^{-1}\,\mathrm{D}_x h(x_0, y_0),$$

which completes the proof. □

The following theorem gives the gradients of the squared $\mathcal{H}_2$-error (5.11).

Theorem 5.7:
Let (5.2) and (5.9) be an asymptotically stable systems. Then

$$\begin{aligned}\nabla_{\widehat{E}_i} \mathcal{J} &= 2\sum_{j=1}^{\mathfrak{n}} \widetilde{Q}_{ji}^{\mathrm{T}} A_j \widetilde{P}_{ji} + 2\sum_{j,k=1}^{\mathfrak{n}} \widetilde{Q}_{ji}^{\mathrm{T}} B_j K_{jk} C_k \widetilde{P}_{ki}\\ &\quad + 2\sum_{j=1}^{\mathfrak{n}} \widehat{Q}_{ij} \widehat{A}_j \widehat{P}_{ji} + 2\sum_{j,k=1}^{\mathfrak{n}} \widehat{Q}_{ij} \widehat{B}_j K_{jk} \widehat{C}_k \widehat{P}_{ki},\\ \nabla_{\widehat{A}_i} \mathcal{J} &= 2\sum_{j=1}^{\mathfrak{n}} \widetilde{Q}_{ji}^{\mathrm{T}} E_j \widetilde{P}_{ji} + 2\sum_{j=1}^{\mathfrak{n}} \widehat{Q}_{ij} \widehat{E}_j \widehat{P}_{ji},\\ \nabla_{\widehat{B}_i} \mathcal{J} &= 2\sum_{j,k=1}^{\mathfrak{n}} \widetilde{Q}_{ji}^{\mathrm{T}} E_j \widetilde{P}_{jk} \widehat{C}_k^{\mathrm{T}} K_{ik}^{\mathrm{T}} + 2\sum_{j=1}^{\mathfrak{n}} \widetilde{Q}_{ji}^{\mathrm{T}} B_j F_j F_i^{\mathrm{T}}\\ &\quad + 2\sum_{j,k=1}^{\mathfrak{n}} \widehat{Q}_{ij} \widehat{E}_j \widehat{P}_{jk} \widehat{C}_k^{\mathrm{T}} K_{ik}^{\mathrm{T}} + 2\sum_{j=1}^{\mathfrak{n}} \widehat{Q}_{ji}^{\mathrm{T}} \widehat{B}_j F_j F_i^{\mathrm{T}},\\ \nabla_{\widehat{C}_i} \mathcal{J} &= -2\sum_{j=1}^{\mathfrak{n}} G_i^{\mathrm{T}} G_j C_j \widetilde{P}_{ji} + 2\sum_{j=1}^{\mathfrak{n}} G_i^{\mathrm{T}} G_j \widehat{C}_j \widehat{P}_{ji}\\ &\quad + 2\sum_{j,k=1}^{\mathfrak{n}} K_{ji}^{\mathrm{T}} \widehat{B}_j^{\mathrm{T}} \widetilde{Q}_{kj}^{\mathrm{T}} E_k \widetilde{P}_{ki} + 2\sum_{j,k=1}^{\mathfrak{n}} K_{ji}^{\mathrm{T}} \widehat{B}_j^{\mathrm{T}} \widehat{Q}_{kj}^{\mathrm{T}} \widehat{E}_k \widehat{P}_{ki}.\end{aligned}$$

◊

Proof. We proceed as in the proof of Theorem 2.44. We find similarly that the Lagrange function is

$$\begin{aligned}\mathcal{L} &= \operatorname{tr}\Big(\mathcal{C}\mathcal{P}\mathcal{C}^{\mathrm{T}} - 2\mathcal{C}\widetilde{\mathcal{P}}\widehat{\mathcal{C}}^{\mathrm{T}} + \widehat{\mathcal{C}}\widehat{\mathcal{P}}\widehat{\mathcal{C}}^{\mathrm{T}}\Big) \\ &\quad + \operatorname{tr}\Big(2\widetilde{\mathcal{Q}}^{\mathrm{T}}\mathcal{A}\widetilde{\mathcal{P}}\widehat{\mathcal{E}}^{\mathrm{T}} + 2\widetilde{\mathcal{Q}}^{\mathrm{T}}\mathcal{E}\widetilde{\mathcal{P}}\widehat{\mathcal{A}}^{\mathrm{T}} + 2\widetilde{\mathcal{Q}}^{\mathrm{T}}\mathcal{B}\widehat{\mathcal{B}}^{\mathrm{T}}\Big) \\ &\quad + \operatorname{tr}\Big(2\widehat{\mathcal{Q}}\widehat{\mathcal{A}}\widehat{\mathcal{P}}\widehat{\mathcal{E}}^{\mathrm{T}} + \widehat{\mathcal{Q}}\widehat{\mathcal{B}}\widehat{\mathcal{B}}^{\mathrm{T}}\Big),\end{aligned}$$

and after inserting (5.3), (5.10),

$$\begin{aligned}\mathcal{L} &= \operatorname{tr}\Big(GC\mathcal{P}C^{\mathrm{T}}G^{\mathrm{T}} - 2GC\widetilde{\mathcal{P}}\widehat{C}^{\mathrm{T}}G^{\mathrm{T}} + G\widehat{C}\widehat{\mathcal{P}}\widehat{C}^{\mathrm{T}}G^{\mathrm{T}}\Big) \\ &\quad + \operatorname{tr}\left(2\widetilde{\mathcal{Q}}^{\mathrm{T}}(A+BKC)\widetilde{\mathcal{P}}\widehat{E}^{\mathrm{T}} + 2\widetilde{\mathcal{Q}}^{\mathrm{T}}E\widetilde{\mathcal{P}}\left(\widehat{A}+\widehat{B}K\widehat{C}\right)^{\mathrm{T}} + 2\widetilde{\mathcal{Q}}^{\mathrm{T}}BFF^{\mathrm{T}}\widehat{B}^{\mathrm{T}}\right) \\ &\quad + \operatorname{tr}\Big(2\widehat{\mathcal{Q}}\left(\widehat{A}+\widehat{B}K\widehat{C}\right)\widehat{\mathcal{P}}\widehat{E}^{\mathrm{T}} + \widehat{\mathcal{Q}}\widehat{B}FF^{\mathrm{T}}\widehat{B}^{\mathrm{T}}\Big).\end{aligned}$$

Now we find the gradients

$$\begin{aligned}\nabla_{\widehat{E}_i}\mathcal{L} &= 2\sum_{j=1}^{n}\widetilde{Q}_{ji}^{\mathrm{T}}A_j\widetilde{P}_{ji} + 2\sum_{j,k=1}^{n}\widetilde{Q}_{ji}^{\mathrm{T}}B_jK_{jk}C_k\widetilde{P}_{ki} \\ &\quad + 2\sum_{j=1}^{n}\widehat{Q}_{ij}\widehat{A}_j\widehat{P}_{ji} + 2\sum_{j,k=1}^{n}\widehat{Q}_{ij}\widehat{B}_jK_{jk}\widehat{C}_k\widehat{P}_{ki}, \\ \nabla_{\widehat{A}_i}\mathcal{L} &= 2\sum_{j=1}^{n}\widetilde{Q}_{ji}^{\mathrm{T}}E_j\widetilde{P}_{ji} + 2\sum_{j=1}^{n}\widehat{Q}_{ij}\widehat{E}_j\widehat{P}_{ji}, \\ \nabla_{\widehat{B}_i}\mathcal{L} &= 2\sum_{j,k=1}^{n}\widetilde{Q}_{ji}^{\mathrm{T}}E_j\widetilde{P}_{jk}\widehat{C}_k^{\mathrm{T}}K_{ik}^{\mathrm{T}} + 2\sum_{j=1}^{n}\widetilde{Q}_{ji}^{\mathrm{T}}B_jF_jF_i^{\mathrm{T}} \\ &\quad + 2\sum_{j,k=1}^{n}\widehat{Q}_{ij}\widehat{E}_j\widehat{P}_{jk}\widehat{C}_k^{\mathrm{T}}K_{ik}^{\mathrm{T}} + 2\sum_{j=1}^{n}\widehat{Q}_{ji}^{\mathrm{T}}\widehat{B}_jF_jF_i^{\mathrm{T}}, \\ \nabla_{\widehat{C}_i}\mathcal{L} &= -2\sum_{j=1}^{n}G_i^{\mathrm{T}}G_jC_j\widetilde{P}_{ji} + 2\sum_{j=1}^{n}G_i^{\mathrm{T}}G_j\widehat{C}_j\widehat{P}_{ji} \\ &\quad + 2\sum_{j,k=1}^{n}K_{ji}^{\mathrm{T}}\widehat{B}_j^{\mathrm{T}}\widetilde{Q}_{kj}^{\mathrm{T}}E_k\widetilde{P}_{ki} + 2\sum_{j,k=1}^{n}K_{ji}^{\mathrm{T}}\widehat{B}_j^{\mathrm{T}}\widehat{Q}_{kj}^{\mathrm{T}}\widehat{E}_k\widehat{P}_{ki}.\end{aligned}$$

Using Lemma 5.6 completes the proof. □

The following corollary gives Wilson-type necessary optimality conditions for $\mathcal{H}_2$-optimal subsystem reduction for interconnected systems.

Corollary 5.8:
Let (5.9) be an $\mathcal{H}_2$-optimal reduced network systems for (5.2). Then

$$\begin{aligned}
0 &= \sum_{j=1}^{\mathfrak{n}} \widetilde{Q}_{ji}^{\mathrm{T}} E_j \widetilde{P}_{ji} + \sum_{j=1}^{\mathfrak{n}} \widehat{Q}_{ij} \widehat{E}_j \widehat{P}_{ji}, \\
0 &= \sum_{j=1}^{\mathfrak{n}} \widetilde{Q}_{ji}^{\mathrm{T}} A_j \widetilde{P}_{ji} + \sum_{j=1}^{\mathfrak{n}} \widehat{Q}_{ij} \widehat{A}_j \widehat{P}_{ji} \\
&\quad + \sum_{j,k=1}^{\mathfrak{n}} \widetilde{Q}_{ji}^{\mathrm{T}} B_j K_{jk} C_k \widetilde{P}_{ki} + \sum_{j,k=1}^{\mathfrak{n}} \widehat{Q}_{ij} \widehat{B}_j K_{jk} \widehat{C}_k \widehat{P}_{ki}, \\
0 &= \sum_{j=1}^{\mathfrak{n}} \widetilde{Q}_{ji}^{\mathrm{T}} B_j F_j F_i^{\mathrm{T}} + \sum_{j=1}^{\mathfrak{n}} \widehat{Q}_{ji}^{\mathrm{T}} \widehat{B}_j F_j F_i^{\mathrm{T}} \\
&\quad + \sum_{j,k=1}^{\mathfrak{n}} \widetilde{Q}_{ji}^{\mathrm{T}} E_j \widetilde{P}_{jk} \widehat{C}_k^{\mathrm{T}} K_{ik}^{\mathrm{T}} + \sum_{j,k=1}^{\mathfrak{n}} \widehat{Q}_{ij} \widehat{E}_j \widehat{P}_{jk} \widehat{C}_k^{\mathrm{T}} K_{ik}^{\mathrm{T}}, \\
0 &= -\sum_{j=1}^{\mathfrak{n}} G_i^{\mathrm{T}} G_j C_j \widetilde{P}_{ji} + \sum_{j=1}^{\mathfrak{n}} G_i^{\mathrm{T}} G_j \widehat{C}_j \widehat{P}_{ji} \\
&\quad + \sum_{j,k=1}^{\mathfrak{n}} K_{ji}^{\mathrm{T}} \widehat{B}_j^{\mathrm{T}} \widetilde{Q}_{kj}^{\mathrm{T}} E_k \widetilde{P}_{ki} + \sum_{j,k=1}^{\mathfrak{n}} K_{ji}^{\mathrm{T}} \widehat{B}_j^{\mathrm{T}} \widehat{Q}_{kj}^{\mathrm{T}} \widehat{E}_k \widehat{P}_{ki}. \qquad \lozenge
\end{aligned}$$

5.3.2 Multi-agent systems

Here, we consider multi-agent systems of the form

$$\begin{aligned}
\mathcal{E}\dot{x}(t) &= \mathcal{A}x(t) + \mathcal{B}u(t), \\
y(t) &= \mathcal{C}x(t),
\end{aligned} \tag{5.13a}$$

with

$$\mathcal{E} = \mathfrak{M} \otimes E, \quad \mathcal{A} = \mathfrak{M} \otimes A - \mathfrak{L} \otimes BKC, \quad \mathcal{B} = \mathfrak{B} \otimes B, \quad \mathcal{C} = \mathfrak{C} \otimes C, \tag{5.13b}$$

where $E, A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$ and $K \in \mathbb{R}^{m\times p}$. We want to find a ROM with reduced agents

$$\begin{aligned}
\widehat{\mathcal{E}}\dot{\widehat{x}}(t) &= \widehat{\mathcal{A}}\widehat{x}(t) + \widehat{\mathcal{B}}u(t), \\
\widehat{y}(t) &= \widehat{\mathcal{C}}\widehat{x}(t),
\end{aligned} \tag{5.14a}$$

with

$$\widehat{\mathcal{E}} = \mathfrak{M} \otimes \widehat{E}, \quad \widehat{\mathcal{A}} = \mathfrak{M} \otimes \widehat{A} - \mathfrak{L} \otimes \widehat{B}K\widehat{C}, \quad \widehat{\mathcal{B}} = \mathfrak{B} \otimes \widehat{B}, \quad \widehat{\mathcal{C}} = \mathfrak{C} \otimes \widehat{C}, \tag{5.14b}$$

where $\widehat{E}, \widehat{A} \in \mathbb{R}^{r\times r}$, $\widehat{B} \in \mathbb{R}^{r\times m}$, and $\widehat{C} \in \mathbb{R}^{p\times r}$ for some $r < n$.

We proceed similarly to the previous section.

Theorem 5.9:
Let (5.13) and (5.14) be asymptotically stable systems. Then for the squared $\mathcal{H}_2$-error, we have

$$\begin{aligned}
\nabla_{\widehat{E}}\mathcal{J} &= 2\sum_{i,j=1}^{\mathfrak{n}} \mathfrak{m}_i\mathfrak{m}_j\widetilde{Q}_{ji}^{\mathrm{T}}A\widetilde{P}_{ji} + 2\sum_{i,j=1}^{\mathfrak{n}} \mathfrak{m}_i\mathfrak{m}_j\widehat{Q}_{ij}\widehat{A}\widehat{P}_{ji} \\
&\quad - 2\sum_{i,j,k=1}^{\mathfrak{n}} \mathfrak{m}_i[\mathfrak{L}]_{jk}\widetilde{Q}_{ji}^{\mathrm{T}}BKC\widetilde{P}_{ki} - 2\sum_{i,j,k=1}^{\mathfrak{n}} \mathfrak{m}_i[\mathfrak{L}]_{jk}\widehat{Q}_{ij}\widehat{B}K\widehat{C}\widehat{P}_{ki}, \\
\nabla_{\widehat{A}}\mathcal{J} &= 2\sum_{i,j=1}^{\mathfrak{n}} \mathfrak{m}_i\mathfrak{m}_j\widetilde{Q}_{ji}^{\mathrm{T}}E\widetilde{P}_{ji} + 2\sum_{i,j=1}^{\mathfrak{n}} \mathfrak{m}_i\mathfrak{m}_j\widehat{Q}_{ij}\widehat{E}\widehat{P}_{ji}, \\
\nabla_{\widehat{B}}\mathcal{J} &= -2\sum_{i,j,k=1}^{\mathfrak{n}} \mathfrak{m}_j[\mathfrak{L}]_{ik}\widetilde{Q}_{ji}^{\mathrm{T}}E\widetilde{P}_{jk}\widehat{C}^{\mathrm{T}}K^{\mathrm{T}} - 2\sum_{i,j,k=1}^{\mathfrak{n}} \mathfrak{m}_j[\mathfrak{L}]_{ik}\widehat{Q}_{ij}\widehat{E}\widehat{P}_{jk}\widehat{C}^{\mathrm{T}}K^{\mathrm{T}} \\
&\quad + 2\sum_{i,j=1}^{\mathfrak{n}} [\mathfrak{B}\mathfrak{B}^{\mathrm{T}}]_{ji}\widetilde{Q}_{ji}^{\mathrm{T}}B + 2\sum_{i,j=1}^{\mathfrak{n}} [\mathfrak{B}\mathfrak{B}^{\mathrm{T}}]_{ji}\widehat{Q}_{ij}\widehat{B}, \\
\nabla_{\widehat{C}}\mathcal{J} &= -2K\widehat{B}^{\mathrm{T}}\sum_{i,j,k=1}^{\mathfrak{n}} \mathfrak{m}_j[\mathfrak{L}]_{ik}\widetilde{Q}_{ji}^{\mathrm{T}}E\widetilde{P}_{jk} - 2K\widehat{B}^{\mathrm{T}}\sum_{i,j,k=1}^{\mathfrak{n}} \mathfrak{m}_j[\mathfrak{L}]_{ik}\widehat{Q}_{ij}\widehat{E}\widehat{P}_{jk} \\
&\quad - 2C\sum_{j,k=1}^{\mathfrak{n}} [\mathfrak{C}^{\mathrm{T}}\mathfrak{C}]_{jk}\widetilde{P}_{jk} + 2\widehat{C}\sum_{j,k=1}^{\mathfrak{n}} [\mathfrak{C}^{\mathrm{T}}\mathfrak{C}]_{jk}\widehat{P}_{jk}. \qquad \Diamond
\end{aligned}$$

Proof. As in the proof of 5.8, the Lagrange function is

$$\begin{aligned}
\mathcal{L} &= \operatorname{tr}\Big((\mathfrak{C}\otimes C)P\left(\mathfrak{C}^{\mathrm{T}}\otimes C^{\mathrm{T}}\right) - 2(\mathfrak{C}\otimes C)\widetilde{P}\left(\mathfrak{C}^{\mathrm{T}}\otimes \widehat{C}^{\mathrm{T}}\right) + \left(\mathfrak{C}\otimes \widehat{C}\right)\widehat{P}\left(\mathfrak{C}^{\mathrm{T}}\otimes \widehat{C}^{\mathrm{T}}\right)\Big) \\
&\quad + \operatorname{tr}\Big(2\widetilde{Q}^{\mathrm{T}}(\mathfrak{M}\otimes A - \mathfrak{L}\otimes BKC)\widetilde{P}\left(\mathfrak{M}\otimes \widehat{E}^{\mathrm{T}}\right)\Big) \\
&\quad + \operatorname{tr}\Big(2\widetilde{Q}^{\mathrm{T}}(\mathfrak{M}\otimes E)\widetilde{P}\left(\mathfrak{M}\otimes \widehat{A}^{\mathrm{T}} - \mathfrak{L}^{\mathrm{T}}\otimes \widehat{C}^{\mathrm{T}}K^{\mathrm{T}}\widehat{B}^{\mathrm{T}}\right)\Big) \\
&\quad + \operatorname{tr}\Big(2\widetilde{Q}^{\mathrm{T}}(\mathfrak{B}\otimes B)\left(\mathfrak{B}^{\mathrm{T}}\otimes \widehat{B}^{\mathrm{T}}\right)\Big) \\
&\quad + \operatorname{tr}\Big(2\widehat{Q}\left(\mathfrak{M}\otimes \widehat{A} - \mathfrak{L}\otimes \widehat{B}K\widehat{C}\right)\widehat{P}\left(\mathfrak{M}\otimes \widehat{E}^{\mathrm{T}}\right) + \widehat{Q}\left(\mathfrak{B}\otimes \widehat{B}\right)\left(\mathfrak{B}^{\mathrm{T}}\otimes \widehat{B}^{\mathrm{T}}\right)\Big).
\end{aligned}$$

The gradients are

$$\begin{aligned}
\nabla_{\widehat{E}}\mathcal{L} &= 2\sum_{i,j=1}^{\mathfrak{n}} \mathfrak{m}_i\mathfrak{m}_j\widetilde{Q}_{ji}^{\mathrm{T}}A\widetilde{P}_{ji} + 2\sum_{i,j=1}^{\mathfrak{n}} \mathfrak{m}_i\mathfrak{m}_j\widehat{Q}_{ij}\widehat{A}\widehat{P}_{ji} \\
&\quad - 2\sum_{i,j,k=1}^{\mathfrak{n}} \mathfrak{m}_i[\mathfrak{L}]_{jk}\widetilde{Q}_{ji}^{\mathrm{T}}BKC\widetilde{P}_{ki} - 2\sum_{i,j,k=1}^{\mathfrak{n}} \mathfrak{m}_i[\mathfrak{L}]_{jk}\widehat{Q}_{ij}\widehat{B}K\widehat{C}\widehat{P}_{ki},
\end{aligned}$$

$$\begin{aligned}
\nabla_{\widehat{A}}\mathcal{L} &= 2\sum_{i,j=1}^{n}\mathfrak{m}_i\mathfrak{m}_j\widetilde{Q}_{ji}^{\mathrm{T}}E\widetilde{P}_{ji} + 2\sum_{i,j=1}^{n}\mathfrak{m}_i\mathfrak{m}_j\widehat{Q}_{ij}\widehat{E}\widehat{P}_{ji},\\
\nabla_{\widehat{B}}\mathcal{L} &= -2\sum_{i,j,k=1}^{n}\mathfrak{m}_j[\mathfrak{L}]_{ik}\widetilde{Q}_{ji}^{\mathrm{T}}E\widetilde{P}_{jk}\widehat{C}^{\mathrm{T}}K^{\mathrm{T}} - 2\sum_{i,j,k=1}^{n}\mathfrak{m}_j[\mathfrak{L}]_{ik}\widehat{Q}_{ij}\widehat{E}\widehat{P}_{jk}\widehat{C}^{\mathrm{T}}K^{\mathrm{T}}\\
&\quad + 2\sum_{i,j=1}^{n}[\mathfrak{B}\mathfrak{B}^{\mathrm{T}}]_{ji}\widetilde{Q}_{ji}^{\mathrm{T}}B + 2\sum_{i,j=1}^{n}[\mathfrak{B}\mathfrak{B}^{\mathrm{T}}]_{ji}\widehat{Q}_{ij}\widehat{B},\\
\nabla_{\widehat{C}}\mathcal{L} &= -2K\widehat{B}^{\mathrm{T}}\sum_{i,j,k=1}^{n}\mathfrak{m}_j[\mathfrak{L}]_{ik}\widetilde{Q}_{ji}^{\mathrm{T}}E\widetilde{P}_{jk} - 2K\widehat{B}^{\mathrm{T}}\sum_{i,j,k=1}^{n}\mathfrak{m}_j[\mathfrak{L}]_{ik}\widehat{Q}_{ij}\widehat{E}\widehat{P}_{jk}\\
&\quad - 2C\sum_{j,k=1}^{n}[\mathfrak{C}^{\mathrm{T}}\mathfrak{C}]_{jk}\widetilde{P}_{jk} + 2\widehat{C}\sum_{j,k=1}^{n}[\mathfrak{C}^{\mathrm{T}}\mathfrak{C}]_{jk}\widehat{P}_{jk}.
\end{aligned}$$

□

As a direct consequence of the previous theorem, we get the Wilson-type necessary optimality conditions for $\mathcal{H}_2$-optimal MOR of multi-agent systems.

Corollary 5.10:
Let (5.14) be an $\mathcal{H}_2$-optimal reduced multi-agent system for (5.13). Then

$$\begin{aligned}
0 &= \sum_{i,j=1}^{n}\mathfrak{m}_i\mathfrak{m}_j\widetilde{Q}_{ji}^{\mathrm{T}}A\widetilde{P}_{ji} + \sum_{i,j=1}^{n}\mathfrak{m}_i\mathfrak{m}_j\widehat{Q}_{ij}\widehat{A}\widehat{P}_{ji}\\
&\quad - \sum_{i,j,k=1}^{n}\mathfrak{m}_i[\mathfrak{L}]_{jk}\widetilde{Q}_{ji}^{\mathrm{T}}BKC\widetilde{P}_{ki} - \sum_{i,j,k=1}^{n}\mathfrak{m}_i[\mathfrak{L}]_{jk}\widehat{Q}_{ij}\widehat{B}K\widehat{C}\widehat{P}_{ki},\\
0 &= \sum_{i,j=1}^{n}\mathfrak{m}_i\mathfrak{m}_j\widetilde{Q}_{ji}^{\mathrm{T}}E\widetilde{P}_{ji} + \sum_{i,j=1}^{n}\mathfrak{m}_i\mathfrak{m}_j\widehat{Q}_{ij}\widehat{E}\widehat{P}_{ji},\\
0 &= -\sum_{i,j,k=1}^{n}\mathfrak{m}_j[\mathfrak{L}]_{ik}\widetilde{Q}_{ji}^{\mathrm{T}}E\widetilde{P}_{jk}\widehat{C}^{\mathrm{T}}K^{\mathrm{T}} - \sum_{i,j,k=1}^{n}\mathfrak{m}_j[\mathfrak{L}]_{ik}\widehat{Q}_{ij}\widehat{E}\widehat{P}_{jk}\widehat{C}^{\mathrm{T}}K^{\mathrm{T}}\\
&\quad + \sum_{i,j=1}^{n}[\mathfrak{B}\mathfrak{B}^{\mathrm{T}}]_{ji}\widetilde{Q}_{ji}^{\mathrm{T}}B + \sum_{i,j=1}^{n}[\mathfrak{B}\mathfrak{B}^{\mathrm{T}}]_{ji}\widehat{Q}_{ij}\widehat{B},\\
0 &= -K\widehat{B}^{\mathrm{T}}\sum_{i,j,k=1}^{n}\mathfrak{m}_j[\mathfrak{L}]_{ik}\widetilde{Q}_{ji}^{\mathrm{T}}E\widetilde{P}_{jk} - K\widehat{B}^{\mathrm{T}}\sum_{i,j,k=1}^{n}\mathfrak{m}_j[\mathfrak{L}]_{ik}\widehat{Q}_{ij}\widehat{E}\widehat{P}_{jk}\\
&\quad - C\sum_{j,k=1}^{n}[\mathfrak{C}^{\mathrm{T}}\mathfrak{C}]_{jk}\widetilde{P}_{jk} + \widehat{C}\sum_{j,k=1}^{n}[\mathfrak{C}^{\mathrm{T}}\mathfrak{C}]_{jk}\widehat{P}_{jk}.
\end{aligned}$$

◊

5.4 Conclusion

In Section 5.2, we developed a stability-preserving method consisting of applying BRBT to subsystems. The preservation of asymptotic stability is based on a sufficient condition. As such, the method can only be applied to asymptotically stable interconnected systems which additionally satisfy this sufficient condition. On the other hand, it could be possible in practical applications to design the interconnected systems to satisfy this condition. This would not only allow the use of the developed method, but also give an a priori guarantee for the asymptotic stability of the original interconnected system.

In Section 5.3, we derived Wilson-type necessary optimality conditions for $\mathcal{H}_2$-optimal subsystem reduction of interconnected and multi-agent systems. These conditions can be used with a gradient-based optimization, with backtracking line search to ensure asymptotic stability of the ROM, to find locally $\mathcal{H}_2$-optimal ROMs.

CHAPTER 6

$\mathcal{H}_2$-OPTIMAL MODEL ORDER REDUCTION OF FURTHER STRUCTURED SYSTEMS

Contents

6.1 Introduction

In this chapter, we discuss $\mathcal{H}_2$-optimal MOR of structured systems which are not represented using networks as in the previous chapters. Just as before, we are interested in structure-preserving MOR, which can be beneficial for preserving the physical interpretation of the system. Additionally, simulation and optimization methods tailored for specific system structures can be used for the ROMs.

Second-order systems are considered in Section 6.2, port-Hamiltonian systems in Section 6.3, parametric systems in Section 6.4, and time-delay systems in Section 6.5.

6.2 Second-order systems

Here, we consider LTI second-order systems

$$\begin{aligned} M\ddot{x}(t) + E\dot{x}(t) + Kx(t) &= Bu(t), \\ y(t) &= C_p x(t) + C_v \dot{x}(t), \end{aligned} \tag{6.1}$$

where $M, E, K \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C_p, C_v \in \mathbb{R}^{p\times n}$, $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, and $y(t) \in \mathbb{R}^p$. We assume that M is invertible and that the matrix pencil $\lambda^2 M + \lambda E + K$ is asymptotically stable. These systems appear, e.g., when analyzing mechanical or electrical systems. We want to find a ROM of the same structure

$$\begin{aligned} \widehat{M}\ddot{\widehat{x}}(t) + \widehat{E}\dot{\widehat{x}}(t) + \widehat{K}\widehat{x}(t) &= \widehat{B}u(t), \\ \widehat{y}(t) &= \widehat{C}_p \widehat{x}(t) + \widehat{C}_v \dot{\widehat{x}}(t), \end{aligned} \tag{6.2}$$

where $\widehat{M}, \widehat{E}, \widehat{K} \in \mathbb{R}^{r\times r}$, $\widehat{B} \in \mathbb{R}^{r\times m}$, and $\widehat{C}_p, \widehat{C}_v \in \mathbb{R}^{p\times r}$, with $r \ll n$, such that $\|H - \widehat{H}\|_{\mathcal{H}_2}$ is minimized. In particular, the matrix pencil $\lambda^2\widehat{M} + \lambda\widehat{E} + \widehat{K}$ should be asymptotically stable.

There is some work towards $\mathcal{H}_2$-optimal MOR of second-order systems. Beattie and Gugercin [BG09] showed it is possible to find a structured ROM which interpolates the original model. Wyatt [Wya12] proposed several iterative methods based on IRKA using the result from [BG09], but $\mathcal{H}_2$-optimality was not investigated. Beattie and Benner [BB14] derived interpolation-based necessary $\mathcal{H}_2$-optimality conditions for second-order systems, but the ROM was restricted to be modally-damped, i.e., such that $\widehat{M}^{-1}\widehat{E}$ and $\widehat{M}^{-1}\widehat{K}$ are simultaneously diagonalizable. Additionally, finding an algorithm which would achieve the interpolation conditions remains an open problem. Sato [Sat17] developed a method based on Riemannian optimization for $\mathcal{H}_2$-optimal MOR of second-order systems with symmetric positive definite mass, damping, and stiffness matrices. We consider a more general setting here.

There are several balancing-based methods which preserve the second-order structure [MS96, CLVVD06, RS08a]. They can also find ROMs with small $\mathcal{H}_2$-errors, but there is no a priori error bound as for the standard BT.

We derive Wilson-type necessary $\mathcal{H}_2$-optimality conditions for second-order systems in Section 6.2.1 and corresponding interpolatory conditions in Section 6.2.2.

6.2.1 Wilson-type conditions

In [BB14, Section 5], the authors focus on interpolation-based necessary optimality conditions for $\mathcal{H}_2$-optimal MOR of second-order systems. Here, we use the Gramian-based approach to derive $\mathcal{H}_2$-optimality conditions similar to Wilson conditions for first-order systems. In the next section, we derive interpolatory conditions and compare them to the results from [BB14, Section 5].

As is well known, the second-order system (6.1) has an equivalent first-order representation

$$\underbrace{\begin{bmatrix} I & 0 \\ 0 & M \end{bmatrix}}_{\mathcal{E}} \dot{z}(t) = \underbrace{\begin{bmatrix} 0 & I \\ -K & -E \end{bmatrix}}_{\mathcal{A}} z(t) + \underbrace{\begin{bmatrix} 0 \\ B \end{bmatrix}}_{\mathcal{B}} u(t),$$
$$y(t) = \underbrace{\begin{bmatrix} C_p & C_v \end{bmatrix}}_{\mathcal{C}} z(t),$$

where $z(t) = \operatorname{col}(x(t), \dot{x}(t))$, $\mathcal{E}, \mathcal{A} \in \mathbb{R}^{2n \times 2n}$, $\mathcal{B} \in \mathbb{R}^{2n \times m}$, and $\mathcal{C} \in \mathbb{R}^{p \times 2n}$. Let $\mathcal{P} \in \mathbb{R}^{2n \times 2n}$ and $\mathcal{Q} \in \mathbb{R}^{2n \times 2n}$ be the controllability and observability Gramians of this system, i.e., solutions to the following Lyapunov equations:

$$\mathcal{A}\mathcal{P}\mathcal{E}^{\mathrm{T}} + \mathcal{E}\mathcal{P}\mathcal{A}^{\mathrm{T}} + \mathcal{B}\mathcal{B}^{\mathrm{T}} = 0,$$
$$\mathcal{A}^{\mathrm{T}}\mathcal{Q}\mathcal{E} + \mathcal{E}^{\mathrm{T}}\mathcal{Q}\mathcal{A} + \mathcal{C}^{\mathrm{T}}\mathcal{C} = 0.$$

An equivalent first-order representation of (6.2) is

$$\underbrace{\begin{bmatrix} I & 0 \\ 0 & \widehat{M} \end{bmatrix}}_{\widehat{\mathcal{E}}} \dot{\widehat{z}}(t) = \underbrace{\begin{bmatrix} 0 & I \\ -\widehat{K} & -\widehat{E} \end{bmatrix}}_{\widehat{\mathcal{A}}} \widehat{z}(t) + \underbrace{\begin{bmatrix} 0 \\ \widehat{B} \end{bmatrix}}_{\widehat{\mathcal{B}}} u(t),$$
$$\widehat{y}(t) = \underbrace{\begin{bmatrix} \widehat{C}_p & \widehat{C}_v \end{bmatrix}}_{\widehat{\mathcal{C}}} \widehat{z}(t),$$

with $\widehat{\mathcal{E}}, \widehat{\mathcal{A}} \in \mathbb{R}^{2r \times 2r}$, $\widehat{\mathcal{B}} \in \mathbb{R}^{2r \times m}$, $\widehat{\mathcal{C}} \in \mathbb{R}^{p \times 2r}$, and Gramians $\widehat{\mathcal{P}}, \widehat{\mathcal{Q}} \in \mathbb{R}^{2r \times 2r}$ satisfying

$$\widehat{\mathcal{A}}\widehat{\mathcal{P}}\widehat{\mathcal{E}}^{\mathrm{T}} + \widehat{\mathcal{E}}\widehat{\mathcal{P}}\widehat{\mathcal{A}}^{\mathrm{T}} + \widehat{\mathcal{B}}\widehat{\mathcal{B}}^{\mathrm{T}} = 0, \tag{6.3a}$$
$$\widehat{\mathcal{A}}^{\mathrm{T}}\widehat{\mathcal{Q}}\widehat{\mathcal{E}} + \widehat{\mathcal{E}}^{\mathrm{T}}\widehat{\mathcal{Q}}\widehat{\mathcal{A}} + \widehat{\mathcal{C}}^{\mathrm{T}}\widehat{\mathcal{C}} = 0. \tag{6.3b}$$

The error system is

$$\underbrace{\begin{bmatrix} \mathcal{E} & 0 \\ 0 & \widehat{\mathcal{E}} \end{bmatrix}}_{\mathcal{E}_{\mathrm{err}}} \begin{bmatrix} \dot{z}(t) \\ \dot{\widehat{z}}(t) \end{bmatrix} = \underbrace{\begin{bmatrix} \mathcal{A} & 0 \\ 0 & \widehat{\mathcal{A}} \end{bmatrix}}_{\mathcal{A}_{\mathrm{err}}} \begin{bmatrix} z(t) \\ \widehat{z}(t) \end{bmatrix} + \underbrace{\begin{bmatrix} \mathcal{B} \\ \widehat{\mathcal{B}} \end{bmatrix}}_{\mathcal{B}_{\mathrm{err}}} u(t)$$
$$y(t) - \widehat{y}(t) = \underbrace{\begin{bmatrix} \mathcal{C} & -\widehat{\mathcal{C}} \end{bmatrix}}_{\mathcal{C}_{\mathrm{err}}} \begin{bmatrix} z(t) \\ \widehat{z}(t) \end{bmatrix},$$

with Gramians

$$\mathcal{P}_{\mathrm{err}} = \begin{bmatrix} \mathcal{P} & \widetilde{\mathcal{P}} \\ \widetilde{\mathcal{P}}^{\mathrm{T}} & \widehat{\mathcal{P}} \end{bmatrix}, \quad \mathcal{Q}_{\mathrm{err}} = \begin{bmatrix} \mathcal{Q} & \widetilde{\mathcal{Q}} \\ \widetilde{\mathcal{Q}}^{\mathrm{T}} & \widehat{\mathcal{Q}} \end{bmatrix},$$

where $\widetilde{\mathcal{P}} \in \mathbb{R}^{2n\times 2r}$ and $\widetilde{\mathcal{Q}} \in \mathbb{R}^{2n\times 2r}$ satisfy Sylvester equations

$$\mathcal{A}\widetilde{\mathcal{P}}\widehat{\mathcal{E}}^{\mathrm{T}} + \mathcal{E}\widetilde{\mathcal{P}}\widehat{\mathcal{A}}^{\mathrm{T}} + \mathcal{B}\widehat{\mathcal{B}}^{\mathrm{T}} = 0, \tag{6.4a}$$
$$\mathcal{A}^{\mathrm{T}}\widetilde{\mathcal{Q}}\widehat{\mathcal{E}} + \mathcal{E}^{\mathrm{T}}\widetilde{\mathcal{Q}}\widehat{\mathcal{A}} - \mathcal{C}^{\mathrm{T}}\widehat{\mathcal{C}} = 0. \tag{6.4b}$$

We denote

$$\widetilde{\mathcal{P}} = \begin{bmatrix} \widetilde{P}_{pp} & \widetilde{P}_{pv} \\ \widetilde{P}_{vp} & \widetilde{P}_{vv} \end{bmatrix}, \quad \widehat{\mathcal{P}} = \begin{bmatrix} \widehat{P}_{pp} & \widehat{P}_{pv} \\ \widehat{P}_{vp} & \widehat{P}_{vv} \end{bmatrix}, \quad \widetilde{\mathcal{Q}} = \begin{bmatrix} \widetilde{Q}_{pp} & \widetilde{Q}_{pv} \\ \widetilde{Q}_{vp} & \widetilde{Q}_{vv} \end{bmatrix}, \quad \widehat{\mathcal{Q}} = \begin{bmatrix} \widehat{Q}_{pp} & \widehat{Q}_{pv} \\ \widehat{Q}_{vp} & \widehat{Q}_{vv} \end{bmatrix},$$

where $\widetilde{P}_{ij}, \widetilde{Q}_{ij} \in \mathbb{R}^{n\times r}$ and $\widehat{P}_{ij}, \widehat{Q}_{ij} \in \mathbb{R}^{r\times r}$ for $i, j \in \{p, v\}$.

We find the squared $\mathcal{H}_2$-error as in Theorem 5.7.

Theorem 6.1:
Consider the second-order system (6.1). Let (6.2) be asymptotically stable. Then for the squared $\mathcal{H}_2$-error $\mathcal{J}$, we have

$$\begin{aligned}
\nabla_{\widehat{M}}\mathcal{J} &= -2\widetilde{Q}_{vv}^{\mathrm{T}}K\widetilde{P}_{pv} + 2\widetilde{Q}_{pv}^{\mathrm{T}}\widetilde{P}_{vv} - 2\widetilde{Q}_{vv}^{\mathrm{T}}E\widetilde{P}_{vv} - 2\widehat{Q}_{vv}\widehat{K}\widehat{P}_{pv} + 2\widehat{Q}_{pv}^{\mathrm{T}}\widehat{P}_{vv} - 2\widehat{Q}_{vv}\widehat{E}\widehat{P}_{vv},\\
\nabla_{\widehat{E}}\mathcal{J} &= -2\widetilde{Q}_{pv}^{\mathrm{T}}\widetilde{P}_{pv} - 2\widetilde{Q}_{vv}^{\mathrm{T}}M\widetilde{P}_{vv} - 2\widehat{Q}_{pv}^{\mathrm{T}}\widehat{P}_{pv} - 2\widehat{Q}_{vv}\widehat{M}\widehat{P}_{vv},\\
\nabla_{\widehat{K}}\mathcal{J} &= -2\widetilde{Q}_{pv}^{\mathrm{T}}\widetilde{P}_{pp} - 2\widetilde{Q}_{vv}^{\mathrm{T}}M\widetilde{P}_{vp} - 2\widehat{Q}_{pv}^{\mathrm{T}}\widehat{P}_{pp} - 2\widehat{Q}_{vv}\widehat{M}\widehat{P}_{vp},\\
\nabla_{\widehat{B}}\mathcal{J} &= 2\widetilde{Q}_{vv}^{\mathrm{T}}B + 2\widehat{Q}_{vv}\widehat{B},\\
\nabla_{\widehat{C}_p}\mathcal{J} &= -2C_p\widetilde{P}_{pp} - 2C_v\widetilde{P}_{vp} + 2\widehat{C}_p\widehat{P}_{pp} + 2\widehat{C}_v\widehat{P}_{vp},\\
\nabla_{\widehat{C}_v}\mathcal{J} &= -2C_p\widetilde{P}_{pv} - 2C_v\widetilde{P}_{vv} + 2\widehat{C}_p\widehat{P}_{pv} + 2\widehat{C}_v\widehat{P}_{vv}.
\end{aligned}$$ ◊

Proof. Similar to the proof of Theorem 5.7, the Lagrange function is

$$\begin{aligned}
\mathcal{L} = {} & \operatorname{tr}\left(\mathcal{C}\mathcal{P}\mathcal{C}^{\mathrm{T}} - 2\mathcal{C}\widetilde{\mathcal{P}}\widehat{\mathcal{C}}^{\mathrm{T}} + \widehat{\mathcal{C}}\widehat{\mathcal{P}}\widehat{\mathcal{C}}^{\mathrm{T}}\right)\\
& + \operatorname{tr}\left(\widetilde{\Lambda}^{\mathrm{T}}\mathcal{A}\widetilde{\mathcal{P}}\widehat{\mathcal{E}}^{\mathrm{T}} + \widetilde{\Lambda}^{\mathrm{T}}\mathcal{E}\widetilde{\mathcal{P}}\widehat{\mathcal{A}}^{\mathrm{T}} + \widetilde{\Lambda}^{\mathrm{T}}\mathcal{B}\widehat{\mathcal{B}}^{\mathrm{T}}\right)\\
& + \operatorname{tr}\left(\widehat{\Lambda}^{\mathrm{T}}\widehat{\mathcal{A}}\widehat{\mathcal{P}}\widehat{\mathcal{E}}^{\mathrm{T}} + \widehat{\Lambda}^{\mathrm{T}}\widehat{\mathcal{E}}\widehat{\mathcal{P}}\widehat{\mathcal{A}}^{\mathrm{T}} + \widehat{\Lambda}^{\mathrm{T}}\widehat{\mathcal{B}}\widehat{\mathcal{B}}^{\mathrm{T}}\right),
\end{aligned}$$

where $\widetilde{\Lambda} \in \mathbb{R}^{2n\times 2r}$ and $\widehat{\Lambda} \in \mathbb{R}^{2r\times 2r}$ are the Lagrange multipliers. From (6.4b), (6.3b), and

$$\begin{aligned}
\nabla_{\widetilde{\mathcal{P}}}\mathcal{L} &= -2\mathcal{C}^{\mathrm{T}}\widehat{\mathcal{C}} + \mathcal{A}^{\mathrm{T}}\widetilde{\Lambda}\widehat{\mathcal{E}} + \mathcal{E}^{\mathrm{T}}\widetilde{\Lambda}\widehat{\mathcal{A}},\\
\nabla_{\widehat{\mathcal{P}}}\mathcal{L} &= \widehat{\mathcal{C}}^{\mathrm{T}}\widehat{\mathcal{C}} + \widehat{\mathcal{A}}^{\mathrm{T}}\widehat{\Lambda}\widehat{\mathcal{E}} + \widehat{\mathcal{E}}^{\mathrm{T}}\widehat{\Lambda}\widehat{\mathcal{A}},
\end{aligned}$$

it follows that $\widetilde{\Lambda} = 2\widetilde{\Omega}$ and $\widehat{\Lambda} = \widehat{\Omega}$. Inserting this into the Lagrange function, we find

$$\begin{aligned}
\mathcal{L} &= \operatorname{tr}\left(\mathcal{C}\mathcal{P}\mathcal{C}^{\mathrm{T}} - 2\begin{bmatrix} C_p & C_v \end{bmatrix}\widetilde{\mathcal{P}}\begin{bmatrix} \widehat{C}_p & \widehat{C}_v \end{bmatrix}^{\mathrm{T}} + \begin{bmatrix} \widehat{C}_p & \widehat{C}_v \end{bmatrix}\widehat{\mathcal{P}}\begin{bmatrix} \widehat{C}_p & \widehat{C}_v \end{bmatrix}^{\mathrm{T}}\right) \\
&\quad + \operatorname{tr}\left(2\widetilde{\Omega}^{\mathrm{T}}\begin{bmatrix} 0 & I \\ -K & -E \end{bmatrix}\widetilde{\mathcal{P}}\begin{bmatrix} I & 0 \\ 0 & \widehat{M} \end{bmatrix}^{\mathrm{T}} + 2\widetilde{\Omega}^{\mathrm{T}}\begin{bmatrix} I & 0 \\ 0 & M \end{bmatrix}\widetilde{\mathcal{P}}\begin{bmatrix} 0 & I \\ -\widehat{K} & -\widehat{E} \end{bmatrix}^{\mathrm{T}}\right) \\
&\quad + \operatorname{tr}\left(2\widetilde{\Omega}^{\mathrm{T}}\begin{bmatrix} 0 \\ B \end{bmatrix}\begin{bmatrix} 0 \\ \widehat{B} \end{bmatrix}^{\mathrm{T}}\right) \\
&\quad + \operatorname{tr}\left(2\widehat{\Omega}\begin{bmatrix} 0 & I \\ -\widehat{K} & -\widehat{E} \end{bmatrix}\widehat{\mathcal{P}}\begin{bmatrix} I & 0 \\ 0 & \widehat{M} \end{bmatrix}^{\mathrm{T}} + \widehat{\Omega}\begin{bmatrix} 0 \\ \widehat{B} \end{bmatrix}\begin{bmatrix} 0 \\ \widehat{B} \end{bmatrix}^{\mathrm{T}}\right) \\
&= \operatorname{tr}\left(\mathcal{C}\mathcal{P}\mathcal{C}^{\mathrm{T}} - 2C_p\widetilde{P}_{pp}\widehat{C}_p^{\mathrm{T}} - 2C_p\widetilde{P}_{pv}\widehat{C}_v^{\mathrm{T}} - 2C_v\widetilde{P}_{vp}\widehat{C}_p^{\mathrm{T}} - 2C_v\widetilde{P}_{vv}\widehat{C}_v^{\mathrm{T}}\right) \\
&\quad + \operatorname{tr}\left(\widehat{C}_p\widehat{P}_{pp}\widehat{C}_p^{\mathrm{T}} + 2\widehat{C}_p\widehat{P}_{pv}\widehat{C}_v^{\mathrm{T}} + \widehat{C}_v\widehat{P}_{vv}\widehat{C}_v^{\mathrm{T}}\right) \\
&\quad + \operatorname{tr}\left(-2\widetilde{Q}_{vp}^{\mathrm{T}}K\widetilde{P}_{pp} + 2\widetilde{Q}_{pp}^{\mathrm{T}}\widetilde{P}_{vp} - 2\widetilde{Q}_{vp}^{\mathrm{T}}E\widetilde{P}_{vp}\right) \\
&\quad + \operatorname{tr}\left(-2\widetilde{Q}_{vv}^{\mathrm{T}}K\widetilde{P}_{pv}\widehat{M}^{\mathrm{T}} + 2\widetilde{Q}_{pv}^{\mathrm{T}}\widetilde{P}_{vv}\widehat{M}^{\mathrm{T}} - 2\widetilde{Q}_{vv}^{\mathrm{T}}E\widetilde{P}_{vv}\widehat{M}^{\mathrm{T}}\right) \\
&\quad + \operatorname{tr}\left(2\widetilde{Q}_{pp}^{\mathrm{T}}\widetilde{P}_{pv} + 2\widetilde{Q}_{vp}^{\mathrm{T}}M\widetilde{P}_{vv}\right) \\
&\quad + \operatorname{tr}\left(-2\widetilde{Q}_{pv}^{\mathrm{T}}\widetilde{P}_{pp}\widehat{K}^{\mathrm{T}} - 2\widetilde{Q}_{pv}^{\mathrm{T}}\widetilde{P}_{pv}\widehat{E}^{\mathrm{T}} - 2\widetilde{Q}_{vv}^{\mathrm{T}}M\widetilde{P}_{vp}\widehat{K}^{\mathrm{T}} - 2\widetilde{Q}_{vv}^{\mathrm{T}}M\widetilde{P}_{vv}\widehat{E}^{\mathrm{T}}\right) \\
&\quad + \operatorname{tr}\left(2\widetilde{Q}_{vv}^{\mathrm{T}}B\widehat{B}^{\mathrm{T}}\right) \\
&\quad + \operatorname{tr}\left(-2\widehat{Q}_{vp}^{\mathrm{T}}\widehat{K}\widehat{P}_{pp} + 2\widehat{Q}_{pp}^{\mathrm{T}}\widehat{P}_{vp} - 2\widehat{Q}_{vp}^{\mathrm{T}}\widehat{E}\widehat{P}_{vp}\right) \\
&\quad + \operatorname{tr}\left(-2\widehat{Q}_{vv}^{\mathrm{T}}\widehat{K}\widehat{P}_{pv}\widehat{M}^{\mathrm{T}} + 2\widehat{Q}_{pv}^{\mathrm{T}}\widehat{P}_{vv}\widehat{M}^{\mathrm{T}} - 2\widehat{Q}_{vv}^{\mathrm{T}}\widehat{E}\widehat{P}_{vv}\widehat{M}^{\mathrm{T}}\right) \\
&\quad + \operatorname{tr}\left(\widehat{Q}_{vv}^{\mathrm{T}}\widehat{B}\widehat{B}^{\mathrm{T}}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\nabla_{\widehat{M}}\mathcal{L} &= -2\widetilde{Q}_{vv}^{\mathrm{T}}K\widetilde{P}_{pv} + 2\widetilde{Q}_{pv}^{\mathrm{T}}\widetilde{P}_{vv} - 2\widetilde{Q}_{vv}^{\mathrm{T}}E\widetilde{P}_{vv} - 2\widehat{Q}_{vv}\widehat{K}\widehat{P}_{pv} + 2\widehat{Q}_{pv}^{\mathrm{T}}\widehat{P}_{vv} - 2\widehat{Q}_{vv}\widehat{E}\widehat{P}_{vv}, \\
\nabla_{\widehat{E}}\mathcal{L} &= -2\widetilde{Q}_{pv}^{\mathrm{T}}\widetilde{P}_{pv} - 2\widetilde{Q}_{vv}^{\mathrm{T}}M\widetilde{P}_{vv} - 2\widehat{Q}_{pv}^{\mathrm{T}}\widehat{P}_{pv} - 2\widehat{Q}_{vv}\widehat{M}\widehat{P}_{vv}, \\
\nabla_{\widehat{K}}\mathcal{L} &= -2\widetilde{Q}_{pv}^{\mathrm{T}}\widetilde{P}_{pp} - 2\widetilde{Q}_{vv}^{\mathrm{T}}M\widetilde{P}_{vp} - 2\widehat{Q}_{pv}^{\mathrm{T}}\widehat{P}_{pp} - 2\widehat{Q}_{vv}\widehat{M}\widehat{P}_{vp}, \\
\nabla_{\widehat{B}}\mathcal{L} &= 2\widetilde{Q}_{vv}^{\mathrm{T}}B + 2\widehat{Q}_{vv}\widehat{B}, \\
\nabla_{\widehat{C}_p}\mathcal{L} &= -2C_p\widetilde{P}_{pp} - 2C_v\widetilde{P}_{vp} + 2\widehat{C}_p\widehat{P}_{pp} + 2\widehat{C}_v\widehat{P}_{vp}, \\
\nabla_{\widehat{C}_v}\mathcal{L} &= -2C_p\widetilde{P}_{pv} - 2C_v\widetilde{P}_{vv} + 2\widehat{C}_p\widehat{P}_{pv} + 2\widehat{C}_v\widehat{P}_{vv}.
\end{aligned}$$

The statement follows from Lemma 5.6. □

As a direct consequence, we obtain the Wilson-type necessary optimality conditions.

Corollary 6.2:
Consider the second-order system (6.1). Let (6.2) be its locally $\mathcal{H}_2$-optimal ROM. Then

$$\begin{aligned}
\begin{bmatrix} \widetilde{Q}_{pv} \\ \widetilde{Q}_{vv} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} I & 0 \\ 0 & M \end{bmatrix} \widetilde{\mathcal{P}} + \begin{bmatrix} \widehat{Q}_{pv} \\ \widehat{Q}_{vv} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} I & 0 \\ 0 & \widehat{M} \end{bmatrix} \widehat{\mathcal{P}} &= 0, \\
\begin{bmatrix} \widetilde{Q}_{pv} \\ \widetilde{Q}_{vv} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 0 & I \\ -K & -E \end{bmatrix} \begin{bmatrix} \widetilde{P}_{pv} \\ \widetilde{P}_{vv} \end{bmatrix} + \begin{bmatrix} \widehat{Q}_{pv} \\ \widehat{Q}_{vv} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 0 & I \\ -\widehat{K} & -\widehat{E} \end{bmatrix} \begin{bmatrix} \widehat{P}_{pv} \\ \widehat{P}_{vv} \end{bmatrix} &= 0, \\
\begin{bmatrix} \widetilde{Q}_{pv} \\ \widetilde{Q}_{vv} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 0 \\ B \end{bmatrix} + \begin{bmatrix} \widehat{Q}_{pv} \\ \widehat{Q}_{vv} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 0 \\ \widehat{B} \end{bmatrix} &= 0, \\
\begin{bmatrix} C_p & C_v \end{bmatrix} \widetilde{\mathcal{P}} - \begin{bmatrix} \widehat{C}_p & \widehat{C}_v \end{bmatrix} \widehat{\mathcal{P}} &= 0. \qquad \lozenge
\end{aligned}$$

We can also consider second-order systems (6.1) with $C_v = 0$, where one would then want enforce $\widehat{C}_v = 0$ in (6.2).

Corollary 6.3:
Consider the second-order system (6.1) with $\widehat{C}_v = 0$. Let (6.2) be its locally $\mathcal{H}_2$-optimal ROM with $\widehat{C}_v = 0$. Then

$$\begin{aligned}
\begin{bmatrix} \widetilde{Q}_{pv} \\ \widetilde{Q}_{vv} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} I & 0 \\ 0 & M \end{bmatrix} \widetilde{\mathcal{P}} + \begin{bmatrix} \widehat{Q}_{pv} \\ \widehat{Q}_{vv} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} I & 0 \\ 0 & \widehat{M} \end{bmatrix} \widehat{\mathcal{P}} &= 0, \\
\begin{bmatrix} \widetilde{Q}_{pv} \\ \widetilde{Q}_{vv} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 0 & I \\ -K & -E \end{bmatrix} \begin{bmatrix} \widetilde{P}_{pv} \\ \widetilde{P}_{vv} \end{bmatrix} + \begin{bmatrix} \widehat{Q}_{pv} \\ \widehat{Q}_{vv} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 0 & I \\ -\widehat{K} & -\widehat{E} \end{bmatrix} \begin{bmatrix} \widehat{P}_{pv} \\ \widehat{P}_{vv} \end{bmatrix} &= 0, \\
\widetilde{Q}_{vv}^{\mathrm{T}} B + \widehat{Q}_{vv}^{\mathrm{T}} \widehat{B} &= 0, \\
C_p \widetilde{P}_{pp} - \widehat{C}_p \widehat{P}_{pp} &= 0. \qquad \lozenge
\end{aligned}$$

Proof. Following the proof of Theorem 6.1, we see that it is enough to ignore $\nabla_{\widehat{C}_v} \mathcal{L}$ and replace C_v and $\widehat{C}_v$ with zero. □

Similarly, we can consider the case of $C_p = 0$ and $\widehat{C}_p = 0$ and the results can be obtained as in Corollary 6.3.

6.2.2 Interpolatory conditions

The following theorem shows how to derive interpolatory conditions using Wilson-type conditions, similar to Theorem 2.45. As in [BB14, Section 5], we assume that the ROM is such that $\widehat{M}^{-1}\widehat{E}$ and $\widehat{M}^{-1}\widehat{K}$ are simultaneously diagonalizable. This is the case for second-order, modally damped dynamical systems.

Theorem 6.4:
Consider the second-order system (6.1). Let (6.2) be a locally $\mathcal{H}_2$-optimal ROM, such that $\widehat{M}^{-1}\widehat{E}$ and $\widehat{M}^{-1}\widehat{K}$ are simultaneously diagonalizable, i.e., there exist invertible S and T such that $S^{\mathrm{T}}\widehat{M}T = I_r$, $S^{\mathrm{T}}\widehat{E}T = -(\Lambda_1 + \Lambda_2)$, $S^{\mathrm{T}}\widehat{K}T = \Lambda_1\Lambda_2$, where $\Lambda_1 = \operatorname{diag}(\lambda_{1,i})$ and $\Lambda_2 = \operatorname{diag}(\lambda_{2,i})$. Additionally, let all $\lambda_{1,i}$ and $\lambda_{2,j}$ be pairwise distinct. Denote $t_i = Te_i$, $c_{p,i} = \widehat{C}_p t_i$, $c_{v,i} = \widehat{C}_v t_i$, $s_i = Se_i$, and $b_i^{\mathrm{T}} = s_i^{\mathrm{T}}\widehat{B}$, which gives us

$$\widehat{H}(s) = \sum_{i=1}^{r} \frac{(c_{p,i} + sc_{v,i})b_i^{\mathrm{T}}}{(s-\lambda_{1,i})(s-\lambda_{2,i})}.$$

Then,

$$\begin{aligned}
(H(-\lambda_{1,i}) - H(-\lambda_{2,i}))\, b_i &= \left(\widehat{H}(-\lambda_{1,i}) - \widehat{H}(-\lambda_{2,i})\right) b_i,\\
(\lambda_{1,i}H(-\lambda_{1,i}) - \lambda_{2,i}H(-\lambda_{2,i}))\, b_i &= \left(\lambda_{1,i}\widehat{H}(-\lambda_{1,i}) - \lambda_{2,i}\widehat{H}(-\lambda_{2,i})\right) b_i,\\
(c_{p,i} + \lambda_{1,i}c_{v,i})^{\mathrm{T}}\,(H(-\lambda_{1,i}) - H(-\lambda_{2,i})) &= (c_{p,i} + \lambda_{1,i}c_{v,i})^{\mathrm{T}}\left(\widehat{H}(-\lambda_{1,i}) - \widehat{H}(-\lambda_{2,i})\right),\\
(c_{p,i} + \lambda_{1,i}c_{v,i})^{\mathrm{T}}H'(-\lambda_{1,i})b_i &= (c_{p,i} + \lambda_{1,i}c_{v,i})^{\mathrm{T}}\widehat{H}'(-\lambda_{1,i})b_i,\\
(c_{p,i} + \lambda_{2,i}c_{v,i})^{\mathrm{T}}H'(-\lambda_{2,i})b_i &= (c_{p,i} + \lambda_{2,i}c_{v,i})^{\mathrm{T}}\widehat{H}'(-\lambda_{2,i})b_i,
\end{aligned}$$

for $i = 1, 2, \ldots, r$. ◊

Proof. From the assumptions, we have $\widehat{E}t_i = -(\lambda_{1,i} + \lambda_{2,i})\widehat{M}t_i$, $s_i^{\mathrm{T}}\widehat{E} = -(\lambda_{1,i} + \lambda_{2,i})s_i^{\mathrm{T}}\widehat{M}$, $\widehat{K}t_i = \lambda_{1,i}\lambda_{2,i}\widehat{M}t_i$, and $s_i^{\mathrm{T}}\widehat{K} = \lambda_{1,i}\lambda_{2,i}s_i^{\mathrm{T}}\widehat{M}$. Postmultiplying (6.4a) by

$$\begin{bmatrix} \widehat{M}^{\mathrm{T}}s_i & 0 \\ 0 & s_i \end{bmatrix}\begin{bmatrix} -\lambda_{2,i} & -\lambda_{1,i} \\ 1 & 1 \end{bmatrix}$$

and using

$$\begin{bmatrix} 0 & -\lambda_{1,i}\lambda_{2,i} \\ 1 & \lambda_{1,i} + \lambda_{2,i} \end{bmatrix}\begin{bmatrix} -\lambda_{2,i} & -\lambda_{1,i} \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} -\lambda_{2,i} & -\lambda_{1,i} \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \lambda_{1,i} & 0 \\ 0 & \lambda_{2,i} \end{bmatrix},$$

we find

$$\begin{aligned}
&\mathcal{A}\widetilde{\mathcal{P}}\begin{bmatrix} -\lambda_{2,i}\widehat{M}^{\mathrm{T}}s_i & -\lambda_{1,i}\widehat{M}^{\mathrm{T}}s_i \\ \widehat{M}^{\mathrm{T}}s_i & \widehat{M}^{\mathrm{T}}s_i \end{bmatrix} + \mathcal{E}\widetilde{\mathcal{P}}\begin{bmatrix} -\lambda_{2,i}\widehat{M}^{\mathrm{T}}s_i & -\lambda_{1,i}\widehat{M}^{\mathrm{T}}s_i \\ \widehat{M}^{\mathrm{T}}s_i & \widehat{M}^{\mathrm{T}}s_i \end{bmatrix}\begin{bmatrix} \lambda_{1,i} & 0 \\ 0 & \lambda_{2,i} \end{bmatrix}\\
&\quad + \mathcal{B}\begin{bmatrix} b_i & b_i \end{bmatrix} = 0.
\end{aligned}$$

This gives us

$$\begin{aligned}
-\lambda_{2,i}\begin{bmatrix} \widetilde{P}_{pp} \\ \widetilde{P}_{vp} \end{bmatrix}\widehat{M}^{\mathrm{T}}s_i + \begin{bmatrix} \widetilde{P}_{pv} \\ \widetilde{P}_{vv} \end{bmatrix}\widehat{M}^{\mathrm{T}}s_i &= (-\lambda_{1,i}\mathcal{E} - \mathcal{A})^{-1}\mathcal{B}b_i,\\
-\lambda_{1,i}\begin{bmatrix} \widetilde{P}_{pp} \\ \widetilde{P}_{vp} \end{bmatrix}\widehat{M}^{\mathrm{T}}s_i + \begin{bmatrix} \widetilde{P}_{pv} \\ \widetilde{P}_{vv} \end{bmatrix}\widehat{M}^{\mathrm{T}}s_i &= (-\lambda_{2,i}\mathcal{E} - \mathcal{A})^{-1}\mathcal{B}b_i,
\end{aligned}$$

which implies

$$\begin{aligned}(\lambda_{1,i}-\lambda_{2,i})\begin{bmatrix}\widetilde{P}_{pp}\\ \widetilde{P}_{vp}\end{bmatrix}\widehat{M}^{\mathrm{T}}s_i &= (-\lambda_{1,i}\mathcal{E}-\mathcal{A})^{-1}\mathcal{B}b_i-(-\lambda_{2,i}\mathcal{E}-\mathcal{A})^{-1}\mathcal{B}b_i,\\ (\lambda_{1,i}-\lambda_{2,i})\begin{bmatrix}\widetilde{P}_{pv}\\ \widetilde{P}_{vv}\end{bmatrix}\widehat{M}^{\mathrm{T}}s_i &= \lambda_{1,i}(-\lambda_{1,i}\mathcal{E}-\mathcal{A})^{-1}\mathcal{B}b_i-\lambda_{2,i}(-\lambda_{2,i}\mathcal{E}-\mathcal{A})^{-1}\mathcal{B}b_i.\end{aligned} \tag{6.5}$$

Similarly, we find

$$\begin{aligned}(\lambda_{1,i}-\lambda_{2,i})\begin{bmatrix}\widehat{P}_{pp}\\ \widehat{P}_{vp}\end{bmatrix}\widehat{M}^{\mathrm{T}}s_i &= \left(-\lambda_{1,i}\widehat{\mathcal{E}}-\widehat{\mathcal{A}}\right)^{-1}\widehat{\mathcal{B}}b_i-\left(-\lambda_{2,i}\widehat{\mathcal{E}}-\widehat{\mathcal{A}}\right)^{-1}\widehat{\mathcal{B}}b_i,\\ (\lambda_{1,i}-\lambda_{2,i})\begin{bmatrix}\widehat{P}_{pv}\\ \widehat{P}_{vv}\end{bmatrix}\widehat{M}^{\mathrm{T}}s_i &= \lambda_{1,i}\left(-\lambda_{1,i}\widehat{\mathcal{E}}-\widehat{\mathcal{A}}\right)^{-1}\widehat{\mathcal{B}}b_i-\lambda_{2,i}\left(-\lambda_{2,i}\widehat{\mathcal{E}}-\widehat{\mathcal{A}}\right)^{-1}\widehat{\mathcal{B}}b_i.\end{aligned} \tag{6.6}$$

Analogously, postmultiplying (6.4b) by

$$\begin{bmatrix}t_i & 0\\ 0 & t_i\end{bmatrix}\begin{bmatrix}1 & 1\\ \lambda_{1,i} & \lambda_{2,i}\end{bmatrix}$$

and using

$$\begin{bmatrix}0 & 1\\ -\lambda_{1,i}\lambda_{2,i} & \lambda_{1,i}+\lambda_{2,i}\end{bmatrix}\begin{bmatrix}1 & 1\\ \lambda_{1,i} & \lambda_{2,i}\end{bmatrix}=\begin{bmatrix}1 & 1\\ \lambda_{1,i} & \lambda_{2,i}\end{bmatrix}\begin{bmatrix}\lambda_{1,i} & 0\\ 0 & \lambda_{2,i}\end{bmatrix},$$

we find

$$\begin{aligned}&\mathcal{A}^{\mathrm{T}}\widetilde{\mathcal{Q}}\begin{bmatrix}t_i & t_i\\ \lambda_{1,i}\widehat{M}t_i & \lambda_{2,i}\widehat{M}t_i\end{bmatrix}\begin{bmatrix}t_i & 0\\ 0 & \widehat{M}t_i\end{bmatrix}+\mathcal{E}^{\mathrm{T}}\widetilde{\mathcal{Q}}\begin{bmatrix}t_i & t_i\\ \lambda_{1,i}\widehat{M}t_i & \lambda_{2,i}\widehat{M}t_i\end{bmatrix}\begin{bmatrix}\lambda_{1,i} & 0\\ 0 & \lambda_{2,i}\end{bmatrix}\\ &\quad-\mathcal{C}^{\mathrm{T}}\begin{bmatrix}c_{p,i}+\lambda_{1,i}c_{v,i} & c_{p,i}+\lambda_{2,i}c_{v,i}\end{bmatrix}=0.\end{aligned}$$

This in turn gives us

$$\begin{aligned}\begin{bmatrix}\widetilde{Q}_{pp}\\ \widetilde{Q}_{vp}\end{bmatrix}t_i+\lambda_{1,i}\begin{bmatrix}\widetilde{Q}_{pv}\\ \widetilde{Q}_{vv}\end{bmatrix}\widehat{M}t_i &= -(-\lambda_{1,i}\mathcal{E}-\mathcal{A})^{-\mathrm{T}}\mathcal{C}^{\mathrm{T}}(c_{p,i}+\lambda_{1,i}c_{v,i}),\\ \begin{bmatrix}\widetilde{Q}_{pp}\\ \widetilde{Q}_{vp}\end{bmatrix}t_i+\lambda_{2,i}\begin{bmatrix}\widetilde{Q}_{pv}\\ \widetilde{Q}_{vv}\end{bmatrix}\widehat{M}t_i &= -(-\lambda_{2,i}\mathcal{E}-\mathcal{A})^{-\mathrm{T}}\mathcal{C}^{\mathrm{T}}(c_{p,i}+\lambda_{2,i}c_{v,i}),\end{aligned}$$

which implies

$$\begin{aligned}(\lambda_{1,i}-\lambda_{2,i})\begin{bmatrix}\widetilde{Q}_{pv}\\ \widetilde{Q}_{vv}\end{bmatrix}\widehat{M}t_i = &-(-\lambda_{1,i}\mathcal{E}-\mathcal{A})^{-\mathrm{T}}\mathcal{C}^{\mathrm{T}}(c_{p,i}+\lambda_{1,i}c_{v,i})\\ &+(-\lambda_{2,i}\mathcal{E}-\mathcal{A})^{-\mathrm{T}}\mathcal{C}^{\mathrm{T}}(c_{p,i}+\lambda_{2,i}c_{v,i}).\end{aligned} \tag{6.7}$$

We similarly find

$$\begin{aligned}(\lambda_{1,i}-\lambda_{2,i})\begin{bmatrix}\widehat{Q}_{pv}\\ \widehat{Q}_{vv}\end{bmatrix}\widehat{M}t_i = &\left(-\lambda_{1,i}\widehat{\mathcal{E}}-\widehat{\mathcal{A}}\right)^{-\mathrm{T}}\widehat{\mathcal{C}}^{\mathrm{T}}(c_{p,i}+\lambda_{1,i}c_{v,i})\\ &-\left(-\lambda_{2,i}\widehat{\mathcal{E}}-\widehat{\mathcal{A}}\right)^{-\mathrm{T}}\widehat{\mathcal{C}}^{\mathrm{T}}(c_{p,i}+\lambda_{2,i}c_{v,i}).\end{aligned} \tag{6.8}$$

Now we use the Wilson-type conditions from Corollary 6.2 to derive the interpolatory conditions. Postmultiplying

$$\begin{bmatrix}C_p & C_v\end{bmatrix}\widetilde{\mathcal{P}}-\begin{bmatrix}\widehat{C}_p & \widehat{C}_v\end{bmatrix}\widehat{\mathcal{P}}=0,$$

by $\operatorname{diag}\left(\widehat{M}^{\mathrm{T}}s_i,\widehat{M}^{\mathrm{T}}s_i\right)$ and using (6.5) and (6.6), we get

$$\begin{aligned}&\mathcal{C}\left(-\lambda_{1,i}\mathcal{E}-\mathcal{A}\right)^{-1}\mathcal{B}b_i-\mathcal{C}\left(-\lambda_{2,i}\mathcal{E}-\mathcal{A}\right)^{-1}\mathcal{B}b_i\\ &\quad=\widehat{\mathcal{C}}\left(-\lambda_{1,i}\widehat{\mathcal{E}}-\widehat{\mathcal{A}}\right)^{-1}\widehat{\mathcal{B}}b_i-\widehat{\mathcal{C}}\left(-\lambda_{2,i}\widehat{\mathcal{E}}-\widehat{\mathcal{A}}\right)^{-1}\widehat{\mathcal{B}}b_i,\\ &\lambda_{1,i}\mathcal{C}\left(-\lambda_{1,i}\mathcal{E}-\mathcal{A}\right)^{-1}\mathcal{B}b_i-\lambda_{2,i}\mathcal{C}\left(-\lambda_{2,i}\mathcal{E}-\mathcal{A}\right)^{-1}\mathcal{B}b_i\\ &\quad=\lambda_{1,i}\widehat{\mathcal{C}}\left(-\lambda_{1,i}\widehat{\mathcal{E}}-\widehat{\mathcal{A}}\right)^{-1}\widehat{\mathcal{B}}b_i-\lambda_{2,i}\widehat{\mathcal{C}}\left(-\lambda_{2,i}\widehat{\mathcal{E}}-\widehat{\mathcal{A}}\right)^{-1}\widehat{\mathcal{B}}b_i,\end{aligned}$$

i.e.,

$$\begin{aligned}\left(H(-\lambda_{1,i})-H(-\lambda_{2,i})\right)b_i&=\left(\widehat{H}(-\lambda_{1,i})-\widehat{H}(-\lambda_{2,i})\right)b_i,\\ \left(\lambda_{1,i}H(-\lambda_{1,i})-\lambda_{2,i}H(-\lambda_{2,i})\right)b_i&=\left(\lambda_{1,i}\widehat{H}(-\lambda_{1,i})-\lambda_{2,i}\widehat{H}(-\lambda_{2,i})\right)b_i,\end{aligned}$$

which are exactly the first two optimality conditions. Premultiplying

$$\begin{bmatrix}\widetilde{Q}_{pv}\\ \widetilde{Q}_{vv}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}0\\ B\end{bmatrix}+\begin{bmatrix}\widehat{Q}_{pv}\\ \widehat{Q}_{vv}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}0\\ \widehat{B}\end{bmatrix}=0,$$

by $t_i^{\mathrm{T}}\widehat{M}^{\mathrm{T}}$ and using (6.7) and (6.8), we get

$$\begin{aligned}&(c_{p,i}+\lambda_{1,i}c_{v,i})^{\mathrm{T}}\mathcal{C}\left(-\lambda_{1,i}\mathcal{E}-\mathcal{A}\right)^{-1}\mathcal{B}-(c_{p,i}+\lambda_{2,i}c_{v,i})^{\mathrm{T}}\mathcal{C}\left(-\lambda_{2,i}\mathcal{E}-\mathcal{A}\right)^{-1}\mathcal{B}\\ &\quad=(c_{p,i}+\lambda_{1,i}c_{v,i})^{\mathrm{T}}\widehat{\mathcal{C}}\left(-\lambda_{1,i}\widehat{\mathcal{E}}-\widehat{\mathcal{A}}\right)^{-1}\widehat{\mathcal{B}}-(c_{p,i}+\lambda_{2,i}c_{v,i})^{\mathrm{T}}\widehat{\mathcal{C}}\left(-\lambda_{2,i}\widehat{\mathcal{E}}-\widehat{\mathcal{A}}\right)^{-1}\widehat{\mathcal{B}},\end{aligned}$$

i.e.,

$$(c_{p,i}+\lambda_{1,i}c_{v,i})^{\mathrm{T}}\left(H(-\lambda_{1,i})-H(-\lambda_{2,i})\right)=(c_{p,i}+\lambda_{1,i}c_{v,i})^{\mathrm{T}}\left(\widehat{H}(-\lambda_{1,i})-\widehat{H}(-\lambda_{2,i})\right),$$

which is the third optimality condition. From

$$\begin{bmatrix} \widetilde{Q}_{pv} \\ \widetilde{Q}_{vv} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} I & 0 \\ 0 & M \end{bmatrix} \widetilde{\mathcal{P}} + \begin{bmatrix} \widehat{Q}_{pv} \\ \widehat{Q}_{vv} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} I & 0 \\ 0 & \widehat{M} \end{bmatrix} \widehat{\mathcal{P}} = 0,$$

we get

$$\begin{aligned}
&\left((c_{p,i} + \lambda_{1,i} c_{v,i})^{\mathrm{T}} \mathcal{C} \left(-\lambda_{1,i}\mathcal{E} - \mathcal{A}\right)^{-1} - (c_{p,i} + \lambda_{2,i} c_{v,i})^{\mathrm{T}} \mathcal{C} \left(-\lambda_{2,i}\mathcal{E} - \mathcal{A}\right)^{-1}\right) \\
&\quad \cdot \mathcal{E} \left(\left(-\lambda_{1,i}\mathcal{E} - \mathcal{A}\right)^{-1} \mathcal{B} b_i - \left(-\lambda_{2,i}\mathcal{E} - \mathcal{A}\right)^{-1} \mathcal{B} b_i\right) \\
&\qquad = \left((c_{p,i} + \lambda_{1,i} c_{v,i})^{\mathrm{T}} \widehat{\mathcal{C}} \left(-\lambda_{1,i}\widehat{\mathcal{E}} - \widehat{\mathcal{A}}\right)^{-1} - (c_{p,i} + \lambda_{2,i} c_{v,i})^{\mathrm{T}} \widehat{\mathcal{C}} \left(-\lambda_{2,i}\widehat{\mathcal{E}} - \widehat{\mathcal{A}}\right)^{-1}\right) \\
&\qquad\quad \cdot \widehat{\mathcal{E}} \left(\left(-\lambda_{1,i}\widehat{\mathcal{E}} - \widehat{\mathcal{A}}\right)^{-1} \widehat{\mathcal{B}} b_i - \left(-\lambda_{2,i}\widehat{\mathcal{E}} - \widehat{\mathcal{A}}\right)^{-1} \widehat{\mathcal{B}} b_i\right), \\
&\left((c_{p,i} + \lambda_{1,i} c_{v,i})^{\mathrm{T}} \mathcal{C} \left(-\lambda_{1,i}\mathcal{E} - \mathcal{A}\right)^{-1} - (c_{p,i} + \lambda_{2,i} c_{v,i})^{\mathrm{T}} \mathcal{C} \left(-\lambda_{2,i}\mathcal{E} - \mathcal{A}\right)^{-1}\right) \\
&\quad \cdot \mathcal{E} \left(\lambda_{1,i} \left(-\lambda_{1,i}\mathcal{E} - \mathcal{A}\right)^{-1} \mathcal{B} b_i - \lambda_{2,i} \left(-\lambda_{2,i}\mathcal{E} - \mathcal{A}\right)^{-1} \mathcal{B} b_i\right) \\
&\qquad = \left((c_{p,i} + \lambda_{1,i} c_{v,i})^{\mathrm{T}} \widehat{\mathcal{C}} \left(-\lambda_{1,i}\widehat{\mathcal{E}} - \widehat{\mathcal{A}}\right)^{-1} - (c_{p,i} + \lambda_{2,i} c_{v,i})^{\mathrm{T}} \widehat{\mathcal{C}} \left(-\lambda_{2,i}\widehat{\mathcal{E}} - \widehat{\mathcal{A}}\right)^{-1}\right) \\
&\qquad\quad \cdot \widehat{\mathcal{E}} \left(\lambda_{1,i} \left(-\lambda_{1,i}\widehat{\mathcal{E}} - \widehat{\mathcal{A}}\right)^{-1} \widehat{\mathcal{B}} b_i - \lambda_{2,i} \left(-\lambda_{2,i}\widehat{\mathcal{E}} - \widehat{\mathcal{A}}\right)^{-1} \widehat{\mathcal{B}} b_i\right),
\end{aligned}$$

thus

$$\begin{aligned}
&(c_{p,i} + \lambda_{1,i} c_{v,i})^{\mathrm{T}} H'(-\lambda_{1,i}) b_i + (c_{p,i} + \lambda_{2,i} c_{v,i})^{\mathrm{T}} H'(-\lambda_{2,i}) b_i \\
&\qquad = (c_{p,i} + \lambda_{1,i} c_{v,i})^{\mathrm{T}} \widehat{H}'(-\lambda_{1,i}) b_i + (c_{p,i} + \lambda_{2,i} c_{v,i})^{\mathrm{T}} \widehat{H}'(-\lambda_{2,i}) b_i, \\
&\lambda_{1,i} (c_{p,i} + \lambda_{1,i} c_{v,i})^{\mathrm{T}} H'(-\lambda_{1,i}) b_i + \lambda_{2,i} (c_{p,i} + \lambda_{2,i} c_{v,i})^{\mathrm{T}} H'(-\lambda_{2,i}) b_i \\
&\qquad = \lambda_{1,i} (c_{p,i} + \lambda_{1,i} c_{v,i})^{\mathrm{T}} \widehat{H}'(-\lambda_{1,i}) b_i + \lambda_{2,i} (c_{p,i} + \lambda_{2,i} c_{v,i})^{\mathrm{T}} \widehat{H}'(-\lambda_{2,i}) b_i,
\end{aligned}$$

i.e., since $\lambda_{1,i} \neq \lambda_{2,i}$,

$$\begin{aligned}
(c_{p,i} + \lambda_{1,i} c_{v,i})^{\mathrm{T}} H'(-\lambda_{1,i}) b_i &= (c_{p,i} + \lambda_{1,i} c_{v,i})^{\mathrm{T}} \widehat{H}'(-\lambda_{1,i}) b_i, \\
(c_{p,i} + \lambda_{2,i} c_{v,i})^{\mathrm{T}} H'(-\lambda_{2,i}) b_i &= (c_{p,i} + \lambda_{2,i} c_{v,i})^{\mathrm{T}} \widehat{H}'(-\lambda_{2,i}) b_i,
\end{aligned}$$

which are the final two interpolatory optimality conditions. □

Here are the interpolatory conditions with additional assumptions that $C_v = 0$ and $\widehat{C}_v = 0$, as derived in [BB14, Section 5] using a different approach.

Corollary 6.5:
Let the assumptions in Theorem 6.4 hold. Additionally, let $C_v = 0$ and $\widehat{C}_v = 0$. Then,

$$\begin{aligned}
\left(H(-\lambda_{1,i}) - H(-\lambda_{2,i})\right) b_i &= \left(\widehat{H}(-\lambda_{1,i}) - \widehat{H}(-\lambda_{2,i})\right) b_i, \\
c_{p,i}^{\mathrm{T}} \left(H(-\lambda_{1,i}) - H(-\lambda_{2,i})\right) &= c_{p,i}^{\mathrm{T}} \left(\widehat{H}(-\lambda_{1,i}) - \widehat{H}(-\lambda_{2,i})\right), \\
c_{p,i}^{\mathrm{T}} H'(-\lambda_{1,i}) b_i &= c_{p,i}^{\mathrm{T}} \widehat{H}'(-\lambda_{1,i}) b_i, \\
c_{p,i}^{\mathrm{T}} H'(-\lambda_{2,i}) b_i &= c_{p,i}^{\mathrm{T}} \widehat{H}'(-\lambda_{2,i}) b_i,
\end{aligned}$$

for $i = 1, 2, \ldots, r$. ◊

Proof. Proceed as the proof of Theorem 6.4, using the Wilson-type conditions from Corollary 6.3. □

6.3 Port-Hamiltonian systems

We consider LTI, input-state-output port-Hamiltonian systems (see [vdSJ14]) without algebraic constraints of the form

$$\begin{aligned} \dot{x}(t) &= (J - R)Sx(t) + Bu(t), \\ y(t) &= B^{\mathrm{T}}Sx(t), \end{aligned} \tag{6.9}$$

where, $J, R, S \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $J^{\mathrm{T}} = -J$, $R \succcurlyeq 0$, and $S \succ 0$. Matrix S is called the *energy matrix* and R the *dissipation matrix*.

Remark 6.6:
The standard notation in literature for the energy matrix is Q. Since we use Q for the observability Gramian, we decided to use S to denote the energy matrix. ◊

The *Hamiltonian* $\frac{1}{2}x(t)^{\mathrm{T}}Sx(t)$ defines the total energy of the system. The state $x(t) \in \mathbb{R}^n$ are called the *energy variables*, while $u(t), y(t) \in \mathbb{R}^m$ the *power variables*. The inner product of the power variables $u(t)^{\mathrm{T}}y(t)$ is the power supplied to the system. The system is passive, i.e., the rate of increase of total energy is bounded by the power provided to the system, as seen from

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}x(t)^{\mathrm{T}}Sx(t)\right) = u(t)^{\mathrm{T}}y(t) - x(t)^{\mathrm{T}}SRSx(t) \leqslant u(t)^{\mathrm{T}}y(t).$$

We are looking for a ROM

$$\begin{aligned} \dot{\widehat{x}}(t) &= \left(\widehat{J} - \widehat{R}\right)\widehat{S}\widehat{x}(t) + \widehat{B}u(t), \\ \widehat{y}(t) &= \widehat{B}^{\mathrm{T}}\widehat{S}\widehat{x}(t), \end{aligned} \tag{6.10}$$

which preserves the port-Hamiltonian structure in (6.9) (in particular, such that $\widehat{J}^{\mathrm{T}} = -\widehat{J}$, $\widehat{R} \succcurlyeq 0$, $\widehat{S} \succ 0$) and minimizes the $\mathcal{H}_2$-error $\|H - \widehat{H}\|_{\mathcal{H}_2}$, where H and $\widehat{H}$ are respectively the transfer function of (6.9) and (6.10). Additionally, we assume $(J-R)S$ is Hurwitz, which is guaranteed when $R \succ 0$.

An equivalent form of the port-Hamiltonian system (6.9) is using the *co-energy variables* $e(t) = Sx(t)$, given by

$$\begin{aligned} S^{-1}\dot{e}(t) &= (J - R)e(t) + Bu(t), \\ y(t) &= B^{\mathrm{T}}e(t). \end{aligned} \tag{6.11}$$

From this, we see that a system is port-Hamiltonian if it has an equivalent representation $(\widetilde{E}; \widetilde{A}, \widetilde{B}, \widetilde{B}^{\mathrm{T}})$ with $\widetilde{E} \succ 0$ and $\widetilde{A} + \widetilde{A}^{\mathrm{T}} \preccurlyeq 0$. Clearly, applying a Galerkin projection to (6.11) with some full-rank matrix $V \in \mathbb{R}^{n \times r}$ will preserve the structure, since

$$\left(V^{\mathrm{T}} J V\right)^{\mathrm{T}} = -V^{\mathrm{T}} J V, \ V^{\mathrm{T}} R V \succcurlyeq 0, \text{ and } \left(V^{\mathrm{T}} S^{-1} V\right)^{-1} \succ 0.$$

In [GPBvdS12], the authors propose projecting the matrices by

$$\widehat{J} = W^{\mathrm{T}} J W, \ \widehat{R} = W^{\mathrm{T}} R W, \ \widehat{S} = V^{\mathrm{T}} S V, \text{ and } \widehat{B} = W^{\mathrm{T}} B,$$

where $W = SV(V^{\mathrm{T}} S V)^{-1}$, giving the ROM

$$\left(\left(W^{\mathrm{T}} J W - W^{\mathrm{T}} R W\right) V^{\mathrm{T}} S V, \ W^{\mathrm{T}} B, \ B^{\mathrm{T}} W V^{\mathrm{T}} S V\right). \tag{6.12}$$

It can be seen that system (6.12) is equivalent to projecting (6.11) by Galerkin projection $V_{\mathrm{G}} = SV$ and to projecting (6.9) by Petrov-Galerkin projection

$$\begin{aligned} V_{\mathrm{PG}} &= V = V \left(V^{\mathrm{T}} S V\right)^{-1} V^{\mathrm{T}} S V, \\ W_{\mathrm{PG}} &= S V \left(V^{\mathrm{T}} S V\right)^{-1}, \end{aligned}$$

also showed in the proof of [GPBvdS12, Theorem 7].

If V is chosen such that

$$\operatorname{im}(V) = \operatorname{im}\left(\left[(\sigma_1 I - (J-R)S)^{-1} B b_1 \quad \cdots \quad (\sigma_r I - (J-R)S)^{-1} B b_r\right]\right),$$

then, according to Theorem 2.41, we have

$$H(\sigma_i) b_i = \widehat{H}(\sigma_i) b_i, \quad i = 1, 2, \ldots, r.$$

Based on this, [GPBvdS12] proposes an iterative algorithm similar to IRKA, called IRKA-PH, which finds a ROM with a transfer function

$$\widehat{H}(s) = \sum_{i=1}^{r} \frac{c_i b_i^{\mathrm{T}}}{s - \lambda_i}$$

which, upon convergence, satisfies

$$H(-\lambda_i) b_i = \widehat{H}(-\lambda_i) b_i, \quad i = 1, 2, \ldots, r.$$

This is one of the interpolatory necessary $\mathcal{H}_2$-optimality conditions for unstructured first-order systems (see Theorem 2.42). The idea is that IRKA-PH should give a ROM which is close to $\mathcal{H}_2$-optimal. But, it is not clear a priori how close it will be.

Theorem 1 in [BB14] shows the following interpolatory necessary $\mathcal{H}_2$-optimality conditions.

Theorem 6.7 ([BB14, Theorem 1]):
Suppose that $\widehat{H}$ is an $\mathcal{H}_2$-optimal ROM with $\widehat{R} \succ 0$, has r distinct poles, and is represented as $\widehat{H}(s) = \sum_{i=1}^{r} \frac{c_i b_i^{\mathrm{T}}}{s-\lambda_i}$. Then

$$c_i^{\mathrm{T}}(H(-\lambda_i) - H(-\lambda_j))b_j = c_i^{\mathrm{T}}\left(\widehat{H}(-\lambda_i) - \widehat{H}(-\lambda_j)\right)b_j,$$
$$c_i^{\mathrm{T}} H'(-\lambda_i)b_i = c_i^{\mathrm{T}}\widehat{H}'(-\lambda_i)b_i,$$

for $i, j = 1, 2, \ldots, r$. ◊

Similar to the interpolatory necessary optimality conditions for second-order systems, an algorithm which would satisfy these conditions is not known.

In the following sections, we will find Wilson-type optimality conditions and derive interpolatory conditions as in Theorem 6.7.

6.3.1 Wilson-type conditions

Similar to Section 6.2.1, we first derive gradients of the squared $\mathcal{H}_2$-error.

Theorem 6.8:
Let (6.10) be asymptotically stable, with $\widehat{S} = I$, $\widehat{J} = \widehat{J}_2 - \widehat{J}_2^{\mathrm{T}}$, and $\widehat{R} = \widehat{R}_2\widehat{R}_2^{\mathrm{T}}$. Then for the squared $\mathcal{H}_2$-error $\mathcal{J}$, we have

$$\nabla_{\widehat{J}_2}\mathcal{J} = 2\widetilde{Q}^{\mathrm{T}}\widetilde{P} - 2\widetilde{P}^{\mathrm{T}}\widetilde{Q} + 2\widehat{Q}\widehat{P} - 2\widehat{P}\widehat{Q},$$
$$\nabla_{\widehat{R}_2}\mathcal{J} = -2\widetilde{Q}^{\mathrm{T}}\widetilde{P}\widehat{R}_2 - 2\widetilde{P}^{\mathrm{T}}\widetilde{Q}\widehat{R}_2 - 2\widehat{Q}\widehat{P}\widehat{R}_2 - 2\widehat{P}\widehat{Q}\widehat{R}_2,$$
$$\nabla_{\widehat{B}}\mathcal{J} = -2\widetilde{P}^{\mathrm{T}}SB + 2\widehat{P}\widehat{B} + 2\widetilde{Q}^{\mathrm{T}}B + 2\widehat{Q}\widehat{B}.$$ ◊

Proof. As in the proof of Theorem 2.44, we find the Lagrange function is

$$\begin{aligned}\mathcal{L} = \operatorname{tr}&\left(B^{\mathrm{T}}SPSB - 2B^{\mathrm{T}}S\widetilde{P}\widehat{B} + \widehat{B}^{\mathrm{T}}\widehat{P}\widehat{B}\right)\\ &+ \operatorname{tr}\left(2\widetilde{Q}^{\mathrm{T}}(J-R)S\widetilde{P} + 2\widetilde{Q}^{\mathrm{T}}\widetilde{P}\left(\widehat{J}_2^{\mathrm{T}} - \widehat{J}_2 - \widehat{R}_2\widehat{R}_2^{\mathrm{T}}\right) + 2\widetilde{Q}^{\mathrm{T}}B\widehat{B}^{\mathrm{T}}\right)\\ &+ \operatorname{tr}\left(2\widehat{Q}\left(\widehat{J}_2 - \widehat{J}_2^{\mathrm{T}} - \widehat{R}_2\widehat{R}_2^{\mathrm{T}}\right)\widehat{P} + \widehat{Q}\widehat{B}\widehat{B}^{\mathrm{T}}\right).\end{aligned}$$

Gradients with respect to the reduced matrices are

$$\nabla_{\widehat{J}_2}\mathcal{L} = 2\widetilde{Q}^{\mathrm{T}}\widetilde{P} - 2\widetilde{P}^{\mathrm{T}}\widetilde{Q} + 2\widehat{Q}\widehat{P} - 2\widehat{P}\widehat{Q},$$
$$\nabla_{\widehat{R}_2}\mathcal{L} = -2\widetilde{Q}^{\mathrm{T}}\widetilde{P}\widehat{R}_2 - 2\widetilde{P}^{\mathrm{T}}\widetilde{Q}\widehat{R}_2 - 2\widehat{Q}\widehat{P}\widehat{R}_2 - 2\widehat{P}\widehat{Q}\widehat{R}_2,$$
$$\nabla_{\widehat{B}}\mathcal{L} = -2\widetilde{P}^{\mathrm{T}}SB + 2\widehat{P}\widehat{B} + 2\widetilde{Q}^{\mathrm{T}}B + 2\widehat{Q}\widehat{B}.$$

The result follows from Lemma 5.6. □

As a consequence, we get the Wilson-type necessary optimality conditions for $\mathcal{H}_2$-optimal structure-preserving MOR of port-Hamiltonian systems.

Theorem 6.9:
Let (6.10), with $\widehat{S} = I$, be an $\mathcal{H}_2$-optimal ROM for (6.9) with $\widehat{R} \succ 0$. Then

$$\begin{aligned} \widetilde{Q}^{\mathrm{T}}\widetilde{P} + \widehat{Q}\widehat{P} &= 0, \\ \operatorname{sym}\left(\widetilde{Q}^{\mathrm{T}}(J-R)S\widetilde{P} + \widehat{Q}^{\mathrm{T}}\left(\widehat{J}-\widehat{R}\right)\widehat{P}\right) &= 0, \\ \left(\widetilde{Q}^{\mathrm{T}} - \widetilde{P}^{\mathrm{T}}S\right)B + \left(\widehat{Q}+\widehat{P}\right)\widehat{B} &= 0. \end{aligned}$$ ◊

Proof. From Theorem 6.8, we have

$$\begin{aligned} 0 &= \widetilde{Q}^{\mathrm{T}}\widetilde{P} - \widetilde{P}^{\mathrm{T}}\widetilde{Q} + \widehat{Q}\widehat{P} - \widehat{P}\widehat{Q}, \\ 0 &= \widetilde{Q}^{\mathrm{T}}\widetilde{P}\widehat{R}_2 + \widetilde{P}^{\mathrm{T}}\widetilde{Q}\widehat{R}_2 + \widehat{Q}\widehat{P}\widehat{R}_2 + \widehat{P}\widehat{Q}\widehat{R}_2, \\ 0 &= -\widetilde{P}^{\mathrm{T}}SB + \widehat{P}\widehat{B} + \widetilde{Q}^{\mathrm{T}}B + \widehat{Q}\widehat{B}. \end{aligned}$$

Using that $\widehat{R}_2$ is invertible, it follows that

$$\widetilde{Q}^{\mathrm{T}}\widetilde{P} + \widehat{Q}\widehat{P} = 0, \tag{6.13}$$

$$\left(\widetilde{Q}^{\mathrm{T}} - \widetilde{P}^{\mathrm{T}}S\right)B + \left(\widehat{Q}+\widehat{P}\right)\widehat{B} = 0. \tag{6.14}$$

Summing

$$\begin{aligned} \widetilde{Q}^{\mathrm{T}}\left((J-R)S\widetilde{P} + \widetilde{P}\left(\widehat{J}_2^{\mathrm{T}} - \widehat{J}_2 - \widehat{R}_2\widehat{R}_2^{\mathrm{T}}\right) + B\widehat{B}^{\mathrm{T}}\right) &= 0, \\ \widehat{Q}\left(\left(\widehat{J}_2 - \widehat{J}_2^{\mathrm{T}} - \widehat{R}_2\widehat{R}_2^{\mathrm{T}}\right)\widehat{P} + \widehat{P}\left(\widehat{J}_2^{\mathrm{T}} - \widehat{J}_2 - \widehat{R}_2\widehat{R}_2^{\mathrm{T}}\right) + \widehat{B}\widehat{B}^{\mathrm{T}}\right) &= 0, \\ \widetilde{P}^{\mathrm{T}}\left(S(-J-R)\widetilde{Q} + \widetilde{Q}\left(\widehat{J}_2 - \widehat{J}_2^{\mathrm{T}} - \widehat{R}_2\widehat{R}_2^{\mathrm{T}}\right) - SB\widehat{B}^{\mathrm{T}}\right) &= 0, \\ \widehat{P}\left(\left(\widehat{J}_2^{\mathrm{T}} - \widehat{J}_2 - \widehat{R}_2\widehat{R}_2^{\mathrm{T}}\right)\widehat{Q} + \widehat{Q}\left(\widehat{J}_2 - \widehat{J}_2^{\mathrm{T}} - \widehat{R}_2\widehat{R}_2^{\mathrm{T}}\right) + \widehat{B}\widehat{B}^{\mathrm{T}}\right) &= 0, \end{aligned}$$

and using (6.13) and (6.14) gives

$$\operatorname{sym}\left(\widetilde{Q}^{\mathrm{T}}(J-R)S\widetilde{P} + \widehat{Q}\left(\widehat{J}_2 - \widehat{J}_2^{\mathrm{T}} - \widehat{R}_2\widehat{R}_2^{\mathrm{T}}\right)\widehat{P}\right) = 0.$$ □

6.3.2 Interpolatory conditions

Here, we derive interpolatory conditions using Theorem 6.9. Assuming diagonalizability of $\widehat{J}-\widehat{R}$, we obtain results about bitangential interpolation from [BB14, Theorem 1] (see Theorem 6.7). Under an additional assumption, we derive a tangential interpolation condition.

Theorem 6.10:
Let the assumptions in Theorem 6.9 hold. Furthermore, let T be an invertible matrix such that $T^{-1}(\widehat{J}-\widehat{R})T = \Lambda = \operatorname{diag}(\lambda_i)$, where λ_i-s are pairwise distinct. Denote $t_i = Te_i$, $s_i^{\mathrm{T}} = e_i^{\mathrm{T}}T^{-1}$, $c_i = \widehat{B}^{\mathrm{T}}t_i$, and $b_i^{\mathrm{T}} = s_i^{\mathrm{T}}\widehat{B}$ so that $\widehat{H}(s) = \sum_{i=1}^{r} \frac{c_i b_i^{\mathrm{T}}}{s-\lambda_i}$. Then

$$c_i^{\mathrm{T}}\left(H(-\lambda_i) - H(-\lambda_j)\right)b_j = c_i^{\mathrm{T}}\left(\widehat{H}(-\lambda_i) - \widehat{H}(-\lambda_j)\right)b_j,$$
$$c_i^{\mathrm{T}}H'(-\lambda_i)b_i = c_i^{\mathrm{T}}H'(-\lambda_i)b_i,$$

for $i,j = 1,2,\ldots,r$.

Additionally, if T can be chosen to be unitary, which is equivalent to $\widehat{J}-\widehat{R}$ being normal and to $\widehat{J}$ and $\widehat{R}$ commuting, then $\widehat{H}(s) = \sum_{i=1}^{r} \frac{\overline{b_i}b_i^{\mathrm{T}}}{s-\lambda_i}$ and also

$$\left(H(-\lambda_i) + H(-\lambda_i)^*\right)b_i = \left(\widehat{H}(-\lambda_i) + \widehat{H}(-\lambda_i)^*\right)b_i,$$

for $i = 1,2,\ldots,r$. ◊

Proof. From the assumptions, we have $(\widehat{J}-\widehat{R})t_i = \lambda_i t_i$ and $s_i^{\mathrm{T}}(\widehat{J}-\widehat{R}) = \lambda_i s_i^{\mathrm{T}}$. From

$$(J-R)S\widetilde{P}s_i + \widetilde{P}\left(-\widehat{J}-\widehat{R}\right)s_i + B\widehat{B}^{\mathrm{T}}s_i = 0,$$

we get

$$(J-R)S\widetilde{P}s_i + \lambda_i\widetilde{P}s_i + B\widehat{B}^{\mathrm{T}}s_i = 0,$$

and then

$$\begin{aligned}
\widetilde{P}s_i &= (-\lambda_i I - (J-R)S)^{-1}Bb_i,\\
\widehat{P}s_i &= \left(-\lambda_i I - \left(\widehat{J}-\widehat{R}\right)\right)^{-1}\widehat{B}b_i,\\
\widetilde{Q}t_i &= -(-\lambda_i I - (J-R)S)^{-\mathrm{T}}SBc_i,\\
\widehat{Q}t_i &= \left(-\lambda_i I - \left(\widehat{J}-\widehat{R}\right)\right)^{-\mathrm{T}}\widehat{B}c_i.
\end{aligned}$$

From this and proceeding as in the proof of Theorem 2.45, we find

$$\begin{aligned}
t_i^{\mathrm{T}}\left(\widetilde{Q}^{\mathrm{T}}\widetilde{P} + \widehat{Q}\widehat{P}\right)s_i &= c_i^{\mathrm{T}}(H'(-\lambda_i) - H'(-\lambda_i))b_i,\\
t_i^{\mathrm{T}}\left(\widetilde{Q}^{\mathrm{T}}\widetilde{P} + \widehat{Q}\widehat{P}\right)s_j &= c_i^{\mathrm{T}}\left(\frac{H(-\lambda_i) - H(-\lambda_j)}{(-\lambda_j)-(-\lambda_j)} - \frac{\widehat{H}(-\lambda_i) - \widehat{H}(-\lambda_j)}{(-\lambda_j)-(-\lambda_j)}\right)b_j.
\end{aligned}$$

Let additionally $T^{-1} = T^*$, which is possible if and only if $\widehat{J}-\widehat{R}$ is normal. We can see that $(\widehat{J}-\widehat{R})(\widehat{J}-\widehat{R})^{\mathrm{T}} = (\widehat{J}-\widehat{R})^{\mathrm{T}}(\widehat{J}-\widehat{R})$ is equivalent to $\widehat{J}\widehat{R} = \widehat{R}\widehat{J}$, i.e., $\widehat{J}$ and $\widehat{R}$

commute. From $T^{-1} = T^*$, we have $s_i = \overline{t_i}$ and $c_i = \overline{b_i}$. Then

$$\begin{aligned}
&t_i^{\mathrm{T}} \left(\widetilde{Q}^{\mathrm{T}} - \widetilde{P}^{\mathrm{T}} S\right) B + t_i^{\mathrm{T}} \left(\widehat{Q} + \widehat{P}\right) \widehat{B} \\
&= t_i^{\mathrm{T}} \widetilde{Q}^{\mathrm{T}} B - s_i^* \widetilde{P}^{\mathrm{T}} S B + t_i^{\mathrm{T}} \widehat{Q} \widehat{B} + s_i^* \widehat{P} \widehat{B} \\
&= -c_i^{\mathrm{T}} B^{\mathrm{T}} S (-\lambda_i I - (J-R)S)^{-1} B - b_i^* B^{\mathrm{T}} \left(-\overline{\lambda_i} I - (J-R)S\right)^{-\mathrm{T}} S B \\
&\quad + c_i^{\mathrm{T}} \widehat{B}^{\mathrm{T}} \left(-\lambda_i I - \left(\widehat{J} - \widehat{R}\right)\right)^{-1} \widehat{B} + b_i^* \widehat{B}^{\mathrm{T}} \left(-\overline{\lambda_i} I - \left(\widehat{J} - \widehat{R}\right)\right)^{-\mathrm{T}} \widehat{B} \\
&= -b_i^* \left(H(-\lambda_i) + H(-\lambda_i)^*\right) + b_i^* \left(\widehat{H}(-\lambda_i) + \widehat{H}(-\lambda_i)^*\right). \qquad \square
\end{aligned}$$

6.4 Linear parametric systems

Consider a parameterized LTI (PLTI) system

$$\begin{aligned}
E(\mathsf{p})\dot{x}(t,\mathsf{p}) &= A(\mathsf{p})x(t,\mathsf{p}) + B(\mathsf{p})u(t), \\
y(t,\mathsf{p}) &= C(\mathsf{p})x(t,\mathsf{p}),
\end{aligned} \tag{6.15}$$

where $\mathsf{p} \in \mathsf{P}$ is the parameter, $\mathsf{P} \subset \mathbb{R}^d$ is a compact set, and $E(\mathsf{p}), A(\mathsf{p}) \in \mathbb{R}^{n\times n}$, $B(\mathsf{p}) \in \mathbb{R}^{n\times m}$, $C(\mathsf{p}) \in \mathbb{R}^{p\times n}$ are continuous matrix-valued functions. We assume $E(\mathsf{p})$ is invertible and $\sigma(A(\mathsf{p}), E(\mathsf{p})) \subset \mathbb{C}_-$, for all $\mathsf{p} \in \mathsf{P}$.

6.4.1 $\mathcal{H}_2 \otimes \mathcal{L}_2$-optimal model order reduction

Following [BBBG11], we define the $\mathcal{H}_2 \otimes \mathcal{L}_2$-norm of (6.15) with

$$\|H\|^2_{\mathcal{H}_2\otimes\mathcal{L}_2(\mathcal{P})} := \frac{1}{2\pi} \int_{\mathsf{P}} \int_{-\infty}^{\infty} \|H(\imath\omega,\mathsf{p})\|_{\mathrm{F}}^2 \,\mathrm{d}\omega \,\mathrm{d}\mathsf{p} = \int_{\mathsf{P}} \|H(\cdot,\mathsf{p})\|^2_{\mathcal{H}_2} \,\mathrm{d}\mathsf{p},$$

where H is the parameterized transfer function of (6.15):

$$H(s,\mathsf{p}) = C(\mathsf{p})(sE(\mathsf{p}) - A(\mathsf{p}))^{-1} B(\mathsf{p}).$$

The norm is well-defined since $\|H(\cdot,\mathsf{p})\|^2_{\mathcal{H}_2}$ is continuous with respect to p and P is compact.

We are interested in finding an $\mathcal{H}_2 \otimes \mathcal{L}_2$-optimal ROM

$$\begin{aligned}
\widehat{E}(\mathsf{p})\dot{\widehat{x}}(t,\mathsf{p}) &= \widehat{A}(\mathsf{p})\widehat{x}(t,\mathsf{p}) + \widehat{B}(\mathsf{p})u(t), \\
\widehat{y}(t,\mathsf{p}) &= \widehat{C}(\mathsf{p})\widehat{x}(t,\mathsf{p}),
\end{aligned} \tag{6.16}$$

with a parameterized transfer function $\widehat{H}$, where $\widehat{E}(\mathsf{p}), \widehat{A}(\mathsf{p}) \in \mathbb{R}^{r\times r}$, $\widehat{B}(\mathsf{p}) \in \mathbb{R}^{r\times m}$, $\widehat{C}(\mathsf{p}) \in \mathbb{R}^{p\times r}$. But, clearly, $\|H - \widehat{H}\|_{\mathcal{H}_2\otimes\mathcal{L}_2(\mathcal{P})}$ is minimized if and only if $\|H(\cdot,\mathsf{p}) - \widehat{H}(\cdot,\mathsf{p})\|_{\mathcal{H}_2}$ is minimized for all $\mathsf{p} \in \mathsf{P}$. This would mean that, in this setting, $\mathcal{H}_2 \otimes \mathcal{L}_2$-optimal

MOR would consist of performing $\mathcal{H}_2$-optimal MOR for every parameter value $\mathsf{p} \in \mathsf{P}$, which is a bottleneck if the ROM (6.16) needs to be computed for many parameter values.

To overcome this, we need to restrict the structure of the ROM (6.16). One possible approach is to take that the parametric matrices are separable (see [Haa17, Definition 2.6])

$$\begin{aligned} \widehat{E}(\mathsf{p}) &= \sum_{i=1}^{q_{\widehat{E}}} \theta_i^{\widehat{E}}(\mathsf{p})\widehat{E}_i, \quad \widehat{A}(\mathsf{p}) = \sum_{i=1}^{q_{\widehat{A}}} \theta_i^{\widehat{A}}(\mathsf{p})\widehat{A}_i, \\ \widehat{B}(\mathsf{p}) &= \sum_{i=1}^{q_{\widehat{B}}} \theta_i^{\widehat{B}}(\mathsf{p})\widehat{B}_i, \quad \widehat{C}(\mathsf{p}) = \sum_{i=1}^{q_{\widehat{C}}} \theta_i^{\widehat{C}}(\mathsf{p})\widehat{C}_i. \end{aligned} \tag{6.17}$$

where $\theta_i^{\widehat{E}}, \theta_i^{\widehat{A}}, \theta_i^{\widehat{B}}, \theta_i^{\widehat{C}} \colon \mathsf{P} \to \mathbb{R}$ are given continuous functions, and then to optimize the non-parametric matrices $\widehat{E}_i$, $\widehat{A}_i$, $\widehat{B}_i$, $\widehat{C}_i$. Clearly, with the form (6.17), if additionally $q_{\widehat{E}}$, $q_{\widehat{A}}$, $q_{\widehat{B}}$, $q_{\widehat{C}}$ are small and functions $\theta_i^{\widehat{E}}$, $\theta_i^{\widehat{A}}$, $\theta_i^{\widehat{B}}$, $\theta_i^{\widehat{C}}$ are easy to compute, the ROM (6.16) can be computed for many parameter values in an efficient manner.

Parameter separability is a beneficial property for full-order matrices in projection-based methods. To see this, let $A(\mathsf{p}) = \sum_{i=1}^{q_A} \theta_i^A(\mathsf{p})A_i$. Then, when using a Petrov-Galerkin projection to find reduced matrices, we have $W^{\mathrm{T}}A(\mathsf{p})V = \sum_{i=1}^{q_A} \theta_i^{\widehat{A}}(\mathsf{p})W^{\mathrm{T}}A_iV$. Therefore, after precomputing matrices $W^{\mathrm{T}}A_iV$, the matrix $W^{\mathrm{T}}A(\mathsf{p})V$ can be assembled for any parameter value $\mathsf{p} \in \mathsf{P}$ with time complexity which is independent of the order of the original model. If the assumption of parameter separability is not satisfied, an approximation is done, e.g., using the empirical interpolation method [BMNP04] (see also [Haa17, Section 2.3.7]). Here, we will only assume that the ROM has separable parametric matrices. This enables structure preservation when full-order matrices are separable and otherwise avoids an intermediate approximation step.

The following theorem gives the gradients of the squared $\mathcal{H}_2 \otimes \mathcal{L}_2$-error.

Theorem 6.11:
Let (6.16) be of the form (6.17) and asymptotically stable for all $\mathsf{p} \in \mathsf{P}$. Then for the the squared $\mathcal{H}_2 \otimes \mathcal{L}_2$-error, we have

$$\begin{aligned} \nabla_{\widehat{E}_i}\mathcal{J} &= 2\int_{\mathsf{P}} \theta_i^{\widehat{E}}(\mathsf{p}) \left(\widetilde{Q}(\mathsf{p})^{\mathrm{T}}A(\mathsf{p})\widetilde{P}(\mathsf{p}) + \widehat{Q}(\mathsf{p})\widehat{A}(\mathsf{p})\widehat{P}(\mathsf{p})\right) \mathrm{d}\mathsf{p}, \\ \nabla_{\widehat{A}_i}\mathcal{J} &= 2\int_{\mathsf{P}} \theta_i^{\widehat{A}}(\mathsf{p}) \left(\widetilde{Q}(\mathsf{p})^{\mathrm{T}}E(\mathsf{p})\widetilde{P}(\mathsf{p}) + \widehat{Q}(\mathsf{p})\widehat{E}(\mathsf{p})\widehat{P}(\mathsf{p})\right) \mathrm{d}\mathsf{p}, \\ \nabla_{\widehat{B}_i}\mathcal{J} &= 2\int_{\mathsf{P}} \theta_i^{\widehat{B}}(\mathsf{p}) \left(\widetilde{Q}(\mathsf{p})^{\mathrm{T}}B(\mathsf{p}) + \widehat{Q}(\mathsf{p})\widehat{B}(\mathsf{p})\right) \mathrm{d}\mathsf{p}, \\ \nabla_{\widehat{C}_i}\mathcal{J} &= 2\int_{\mathsf{P}} \theta_i^{\widehat{C}}(\mathsf{p}) \left(-C(\mathsf{p})\widetilde{P}(\mathsf{p}) + \widehat{C}(\mathsf{p})\widehat{P}(\mathsf{p})\right) \mathrm{d}\mathsf{p}. \end{aligned}$$

◊

Proof. Since $\widetilde{P}$ and $\widehat{P}$ are continuous functions over a compact set P, we have $\widetilde{P} \in \mathcal{L}_2(\mathsf{P};\mathbb{R}^{n\times r})$ and $\widehat{P} \in \mathcal{L}_2(\mathsf{P};\mathbb{R}^{r\times r})$. Similar to Theorem 2.44, we find the Lagrange function is

$$\begin{aligned}\mathcal{L} = &\int_{\mathsf{P}} \operatorname{tr}\Big(C(\mathsf{p})P(\mathsf{p})C(\mathsf{p})^{\mathrm{T}} - 2C(\mathsf{p})\widetilde{P}(\mathsf{p})\widehat{C}(\mathsf{p})^{\mathrm{T}} + \widehat{C}(\mathsf{p})\widehat{P}(\mathsf{p})\widehat{C}(\mathsf{p})^{\mathrm{T}}\Big)\,\mathrm{d}\mathsf{p}\\ &+ \int_{\mathsf{P}} \operatorname{tr}\Big(\widetilde{\Lambda}(\mathsf{p})^{\mathrm{T}}A(\mathsf{p})\widetilde{P}(\mathsf{p})\widehat{E}(\mathsf{p})^{\mathrm{T}} + \widetilde{\Lambda}(\mathsf{p})^{\mathrm{T}}E(\mathsf{p})\widetilde{P}(\mathsf{p})\widehat{A}(\mathsf{p})^{\mathrm{T}} + \widetilde{\Lambda}(\mathsf{p})^{\mathrm{T}}B(\mathsf{p})\widehat{B}(\mathsf{p})^{\mathrm{T}}\Big)\,\mathrm{d}\mathsf{p}\\ &+ \int_{\mathsf{P}} \operatorname{tr}\Big(\widehat{\Lambda}(\mathsf{p})^{\mathrm{T}}\widehat{A}(\mathsf{p})\widehat{P}(\mathsf{p})\widehat{E}(\mathsf{p})^{\mathrm{T}} + \widehat{\Lambda}(\mathsf{p})^{\mathrm{T}}\widehat{E}(\mathsf{p})\widehat{P}(\mathsf{p})\widehat{A}(\mathsf{p})^{\mathrm{T}} + \widehat{\Lambda}(\mathsf{p})^{\mathrm{T}}\widehat{B}(\mathsf{p})\widehat{B}(\mathsf{p})^{\mathrm{T}}\Big)\,\mathrm{d}\mathsf{p}.\end{aligned}$$

We see that the gradients with respect to $\widetilde{P}$ and $\widehat{P}$ are

$$\begin{aligned}\nabla_{\widetilde{P}}\mathcal{L} &= -2C^{\mathrm{T}}\widehat{C} + A^{\mathrm{T}}\widetilde{\Lambda}\widehat{E} + E^{\mathrm{T}}\widetilde{\Lambda}\widehat{A} \in \mathcal{L}_2(\mathsf{P};\mathbb{R}^{n\times r}),\\ \nabla_{\widehat{P}}\mathcal{L} &= \widehat{C}^{\mathrm{T}}\widehat{C} + \widehat{A}^{\mathrm{T}}\widehat{\Lambda}\widehat{E} + \widehat{E}^{\mathrm{T}}\widehat{\Lambda}\widehat{A} \in \mathcal{L}_2(\mathsf{P};\mathbb{R}^{r\times r}).\end{aligned}$$

Equating with zero, it follows that $\widetilde{\Lambda} = 2\widetilde{Q}$ and $\widehat{\Lambda} = \widehat{Q}$. Therefore, the Lagrange function simplifies to

$$\begin{aligned}\mathcal{L} = &\\ &\int_{\mathsf{P}} \operatorname{tr}\Big(C(\mathsf{p})P(\mathsf{p})C(\mathsf{p})^{\mathrm{T}} - 2C(\mathsf{p})\widetilde{P}(\mathsf{p})\widehat{C}(\mathsf{p})^{\mathrm{T}} + \widehat{C}(\mathsf{p})\widehat{P}(\mathsf{p})\widehat{C}(\mathsf{p})^{\mathrm{T}}\Big)\,\mathrm{d}\mathsf{p}\\ &+ \int_{\mathsf{P}} \operatorname{tr}\Big(2\widetilde{Q}(\mathsf{p})^{\mathrm{T}}A(\mathsf{p})\widetilde{P}(\mathsf{p})\widehat{E}(\mathsf{p})^{\mathrm{T}} + 2\widetilde{Q}(\mathsf{p})^{\mathrm{T}}E(\mathsf{p})\widetilde{P}(\mathsf{p})\widehat{A}(\mathsf{p})^{\mathrm{T}} + 2\widetilde{Q}(\mathsf{p})^{\mathrm{T}}B(\mathsf{p})\widehat{B}(\mathsf{p})^{\mathrm{T}}\Big)\,\mathrm{d}\mathsf{p}\\ &+ \int_{\mathsf{P}} \operatorname{tr}\Big(2\widehat{Q}(\mathsf{p})\widehat{A}(\mathsf{p})\widehat{P}(\mathsf{p})\widehat{E}(\mathsf{p})^{\mathrm{T}} + \widehat{Q}(\mathsf{p})\widehat{B}(\mathsf{p})\widehat{B}(\mathsf{p})^{\mathrm{T}}\Big)\,\mathrm{d}\mathsf{p}.\end{aligned}$$

Finally, the gradients with respect to reduced matrices are

$$\begin{aligned}\nabla_{\widehat{E}_i}\mathcal{L} &= 2\int_{\mathsf{P}} \theta_i^{\widehat{E}}(\mathsf{p})\left(\widetilde{Q}(\mathsf{p})^{\mathrm{T}}A(\mathsf{p})\widetilde{P}(\mathsf{p}) + \widehat{Q}(\mathsf{p})\widehat{A}(\mathsf{p})\widehat{P}(\mathsf{p})\right)\mathrm{d}\mathsf{p},\\ \nabla_{\widehat{A}_i}\mathcal{L} &= 2\int_{\mathsf{P}} \theta_i^{\widehat{A}}(\mathsf{p})\left(\widetilde{Q}(\mathsf{p})^{\mathrm{T}}E(\mathsf{p})\widetilde{P}(\mathsf{p}) + \widehat{Q}(\mathsf{p})\widehat{E}(\mathsf{p})\widehat{P}(\mathsf{p})\right)\mathrm{d}\mathsf{p},\\ \nabla_{\widehat{B}_i}\mathcal{L} &= 2\int_{\mathsf{P}} \theta_i^{\widehat{B}}(\mathsf{p})\left(\widetilde{Q}(\mathsf{p})^{\mathrm{T}}B(\mathsf{p}) + \widehat{Q}(\mathsf{p})\widehat{B}(\mathsf{p})\right)\mathrm{d}\mathsf{p},\\ \nabla_{\widehat{C}_i}\mathcal{L} &= 2\int_{\mathsf{P}} \theta_i^{\widehat{C}}(\mathsf{p})\left(-C(\mathsf{p})\widetilde{P}(\mathsf{p}) + \widehat{C}(\mathsf{p})\widehat{P}(\mathsf{p})\right)\mathrm{d}\mathsf{p}.\end{aligned}$$

The claim follows from Lemma 5.6. □

Directly, we obtain the Wilson-type necessary optimality conditions for $\mathcal{H}_2\otimes\mathcal{L}_2$-optimal MOR.

Corollary 6.12:
Let (6.16) be an $\mathcal{H}_2 \otimes \mathcal{L}_2$-optimal ROM with the form (6.17) for (6.15). Then

$$\begin{aligned}
\int_{\mathsf{P}} \theta_i^{\widehat{A}}(\mathsf{p}) \left(\widetilde{Q}(\mathsf{p})^{\mathrm{T}} E(\mathsf{p}) \widetilde{P}(\mathsf{p}) + \widehat{Q}(\mathsf{p}) \widehat{E}(\mathsf{p}) \widehat{P}(\mathsf{p})\right) \mathrm{d}\mathsf{p} = 0, \qquad & i = 1, 2, \ldots, q_{\widehat{A}}, \\
\int_{\mathsf{P}} \theta_i^{\widehat{E}}(\mathsf{p}) \left(\widetilde{Q}(\mathsf{p})^{\mathrm{T}} A(\mathsf{p}) \widetilde{P}(\mathsf{p}) + \widehat{Q}(\mathsf{p}) \widehat{A}(\mathsf{p}) \widehat{P}(\mathsf{p})\right) \mathrm{d}\mathsf{p} = 0, \qquad & i = 1, 2, \ldots, q_{\widehat{E}}, \\
\int_{\mathsf{P}} \theta_i^{\widehat{B}}(\mathsf{p}) \left(\widetilde{Q}(\mathsf{p})^{\mathrm{T}} B(\mathsf{p}) + \widehat{Q}(\mathsf{p}) \widehat{B}(\mathsf{p})\right) \mathrm{d}\mathsf{p} = 0, \qquad & i = 1, 2, \ldots, q_{\widehat{B}}, \\
\int_{\mathsf{P}} \theta_i^{\widehat{C}}(\mathsf{p}) \left(C(\mathsf{p}) \widetilde{P}(\mathsf{p}) - \widehat{C}(\mathsf{p}) \widehat{P}(\mathsf{p})\right) \mathrm{d}\mathsf{p} = 0, \qquad & i = 1, 2, \ldots, q_{\widehat{C}}. \qquad \lozenge
\end{aligned}$$

6.4.2 Interpolatory conditions

Using Corollary 6.12, we derive interpolatory necessary optimality conditions. To have the pole-residue form of the ROM, we assume $\widehat{A}(\mathsf{p})$ and $\widehat{E}(\mathsf{p})$ can be diagonalized using parameter-independent transformation matrices. Note that we also assumed diagonalizability for second-order systems in Theorem 6.4 and port-Hamiltonian systems in Theorem 6.10.

Theorem 6.13:
Let (6.16) be an $\mathcal{H}_2 \otimes \mathcal{L}_2$-optimal ROM with the form (6.17) for (6.15). Assume that $\{\widehat{E}_1^{-1}\widehat{E}_2, \ldots, \widehat{E}_1^{-1}\widehat{E}_{q_{\widehat{E}}}, \widehat{E}_1^{-1}\widehat{A}_1, \ldots, \widehat{E}_1^{-1}\widehat{A}_{q_{\widehat{A}}}\}$ is a simultaneously diagonalizable family of matrices. Let S and T be invertible matrices such that $S^{\mathrm{T}}\widehat{E}_i T = \Lambda^{\widehat{E}_i} = \operatorname{diag}\left(\lambda_j^{\widehat{E}_i}\right)$ and $S^{\mathrm{T}}\widehat{A}_i T = \Lambda^{\widehat{A}_i} = \operatorname{diag}\left(\lambda_j^{\widehat{A}_i}\right)$, with $\Lambda^{\widehat{E}_i} = I_r$. Furthermore, let

$$\begin{gathered}
t_j = T e_j, \quad s_j^{\mathrm{T}} = e_j^{\mathrm{T}} S, \\
\lambda_j^{\widehat{A}}(\mathsf{p}) = \sum_{i=1}^{q_{\widehat{A}}} \theta_i^{\widehat{A}}(\mathsf{p}) \lambda_j^{\widehat{A}_i}, \quad \lambda_j^{\widehat{E}}(\mathsf{p}) = \sum_{i=1}^{q_{\widehat{E}}} \theta_i^{\widehat{E}}(\mathsf{p}) \lambda_j^{\widehat{E}_i}, \quad \lambda_j(\mathsf{p}) = \frac{\lambda_j^{\widehat{A}}(\mathsf{p})}{\lambda_j^{\widehat{E}}(\mathsf{p})}, \\
b_j(\mathsf{p}) = \frac{1}{\lambda_j^{\widehat{E}}(\mathsf{p})} \widehat{B}(\mathsf{p})^{\mathrm{T}} s_j, \text{ and } c_j(\mathsf{p}) = \frac{1}{\lambda_j^{\widehat{E}}(\mathsf{p})} \widehat{C}(\mathsf{p}) t_j,
\end{gathered}$$

where $\lambda_j(\mathsf{p})$ are pairwise distinct for almost all $\mathsf{p} \in \mathsf{P}$. Then $\widehat{H}(s, \mathsf{p}) = \sum_{j=1}^{r} \lambda_j^{\widehat{E}}(\mathsf{p}) \frac{c_j(\mathsf{p}) b_j(\mathsf{p})^{\mathrm{T}}}{s - \lambda_j(\mathsf{p})}$ and

$$\begin{aligned}
\int_{\mathsf{P}} \theta_i^{\widehat{C}}(\mathsf{p}) \cdot H(-\lambda_j(\mathsf{p}), \mathsf{p}) b_j(\mathsf{p}) \, \mathrm{d}\mathsf{p} &= \int_{\mathsf{P}} \theta_i^{\widehat{C}}(\mathsf{p}) \cdot \widehat{H}(-\lambda_j(\mathsf{p}), \mathsf{p}) b_j(\mathsf{p}) \, \mathrm{d}\mathsf{p}, && i \in [q_{\widehat{C}}], \\
\int_{\mathsf{P}} \theta_i^{\widehat{B}}(\mathsf{p}) \cdot c_j(\mathsf{p})^{\mathrm{T}} H(-\lambda_j(\mathsf{p}), \mathsf{p}) \, \mathrm{d}\mathsf{p} &= \int_{\mathsf{P}} \theta_i^{\widehat{B}}(\mathsf{p}) \cdot c_j(\mathsf{p})^{\mathrm{T}} \widehat{H}(-\lambda_j(\mathsf{p}), \mathsf{p}) \, \mathrm{d}\mathsf{p}, && i \in [q_{\widehat{B}}], \\
\int_{\mathsf{P}} \theta_i^{\widehat{A}}(\mathsf{p}) \cdot c_j(\mathsf{p})^{\mathrm{T}} H'(-\lambda_j(\mathsf{p}), \mathsf{p}) b_j(\mathsf{p}) \, \mathrm{d}\mathsf{p} &= \int_{\mathsf{P}} \theta_i^{\widehat{A}}(\mathsf{p}) \cdot c_j(\mathsf{p})^{\mathrm{T}} \widehat{H}'(-\lambda_j(\mathsf{p}), \mathsf{p}) b_j(\mathsf{p}) \, \mathrm{d}\mathsf{p}, && i \in [q_{\widehat{A}}],
\end{aligned}$$

for $j = 1, 2, \ldots, r$, where $[k] := \{1, 2, \ldots, k\}$. $\lozenge$

Proof. We have $\widehat{E}_i t_j = \lambda_j^{\widehat{E}_i} \widehat{E}_1 t_j$, $\widehat{A}_i t_j = \lambda_j^{\widehat{A}_i} \widehat{E}_1 t_j$, $s_j^{\mathrm{T}} \widehat{E}_i = \lambda_j^{\widehat{E}_i} s_j^{\mathrm{T}} \widehat{E}_1$, $s_j^{\mathrm{T}} \widehat{A}_i = \lambda_j^{\widehat{A}_i} s_j^{\mathrm{T}} \widehat{E}_1$. Therefore, we also have $\widehat{E}(\mathsf{p}) t_j = \lambda_j^{\widehat{E}}(\mathsf{p}) \widehat{E}_1 t_j$, $\widehat{A}(\mathsf{p}) t_j = \lambda_j^{\widehat{A}_i}(\mathsf{p}) \widehat{E}_1 t_j$, $s_j^{\mathrm{T}} \widehat{E}_i(\mathsf{p}) = \lambda_j^{\widehat{E}_i}(\mathsf{p}) s_j^{\mathrm{T}} \widehat{E}_1$, $s_j^{\mathrm{T}} \widehat{A}_i(\mathsf{p}) = \lambda_j^{\widehat{A}_i}(\mathsf{p}) s_j^{\mathrm{T}} \widehat{E}_1$. From

$$A(\mathsf{p}) \widetilde{P}(\mathsf{p}) \widehat{E}(\mathsf{p})^{\mathrm{T}} s_j + E(\mathsf{p}) \widetilde{P}(\mathsf{p}) \widehat{A}(\mathsf{p})^{\mathrm{T}} s_j + B(\mathsf{p}) \widehat{B}(\mathsf{p})^{\mathrm{T}} s_j = 0,$$

it follows that

$$\left(\lambda_j^{\widehat{E}}(\mathsf{p}) A(\mathsf{p}) + \lambda_j^{\widehat{A}}(\mathsf{p}) E(\mathsf{p}) \right) \widetilde{P}(\mathsf{p}) \widehat{E}_1^{\mathrm{T}} s_j + B(\mathsf{p}) \widehat{B}(\mathsf{p})^{\mathrm{T}} s_j = 0.$$

Then

$$\widetilde{P}(\mathsf{p}) \widehat{E}_1^{\mathrm{T}} s_j = \left(-\lambda_j(\mathsf{p}) E(\mathsf{p}) - A(\mathsf{p}) \right)^{-1} B(\mathsf{p}) b_j(\mathsf{p}),$$

and similarly

$$\begin{aligned}
\widehat{P}(\mathsf{p}) \widehat{E}_1^{\mathrm{T}} s_j &= \left(-\lambda_j(\mathsf{p}) \widehat{E}(\mathsf{p}) - \widehat{A}(\mathsf{p}) \right)^{-1} \widehat{B}(\mathsf{p}) b_j(\mathsf{p}), \\
\widetilde{Q}(\mathsf{p}) \widehat{E}_1 t_j &= \left(-\lambda_j(\mathsf{p}) E(\mathsf{p}) - A(\mathsf{p}) \right)^{-\mathrm{T}} C(\mathsf{p})^{\mathrm{T}} c_j(\mathsf{p}), \\
\widehat{Q}(\mathsf{p}) \widehat{E}_1 t_j &= \left(-\lambda_j(\mathsf{p}) \widehat{E}(\mathsf{p}) - \widehat{A}(\mathsf{p}) \right)^{-\mathrm{T}} \widehat{C}(\mathsf{p})^{\mathrm{T}} c_j(\mathsf{p}).
\end{aligned}$$

From

$$\begin{aligned}
\int_{\mathsf{P}} \theta_i^{\widehat{C}}(\mathsf{p}) \left(C(\mathsf{p}) \widetilde{P}(\mathsf{p}) - \widehat{C}(\mathsf{p}) \widehat{P}(\mathsf{p}) \right) \widehat{E}_1^{\mathrm{T}} s_j \, \mathrm{d}\mathsf{p} = 0, \\
\int_{\mathsf{P}} \theta_i^{\widehat{B}}(\mathsf{p}) t_j^{\mathrm{T}} \widehat{E}_1^{\mathrm{T}} \left(\widetilde{Q}(\mathsf{p})^{\mathrm{T}} B(\mathsf{p}) + \widehat{Q}(\mathsf{p}) \widehat{B}(\mathsf{p}) \right) \mathrm{d}\mathsf{p} = 0, \\
\int_{\mathsf{P}} \theta_i^{\widehat{A}}(\mathsf{p}) t_j^{\mathrm{T}} \widehat{E}_1^{\mathrm{T}} \left(\widetilde{Q}(\mathsf{p})^{\mathrm{T}} E(\mathsf{p}) \widetilde{P}(\mathsf{p}) + \widehat{Q}(\mathsf{p}) \widehat{E}(\mathsf{p}) \widehat{P}(\mathsf{p}) \right) \widehat{E}_1^{\mathrm{T}} s_j \, \mathrm{d}\mathsf{p} = 0,
\end{aligned}$$

we find the statement of the theorem. □

6.5 Linear time-delay systems

We consider a linear time-delay (LTD) system with a single delay

$$\begin{aligned}
E\dot{x}(t) &= A_0 x(t) + A_\tau x(t-\tau) + Bu(t), \\
y(t) &= Cx(t),
\end{aligned} \tag{6.18}$$

where $E, A_0, A_\tau \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, and $C \in \mathbb{R}^{p \times n}$, with E invertible. Additionally, we assume the system to be exponentially stable.

The transfer function of this system is

$$H(s) = C \left(sE - A_0 - e^{-\tau s} A_\tau \right)^{-1} B.$$

As shown in [JVM11], the $\mathcal{H}_2$-norm $\|H\|_{\mathcal{H}_2}$ can be computed using the Gramians similarly as for LTI systems without delay. The Gramians of the system (6.18) are solutions to a boundary value problem involving a delay differential Lyapunov equation:

$$\dot{P}(t)E^{\mathrm{T}} = P(t)A_0^{\mathrm{T}} + P(t-\tau)A_\tau^{\mathrm{T}},\ t \geqslant 0, \tag{6.19a}$$
$$P(-t) = P(t)^{\mathrm{T}}, \tag{6.19b}$$
$$-BB^{\mathrm{T}} = A_0P(0)E^{\mathrm{T}} + EP(0)A_0^{\mathrm{T}} + A_\tau P(\tau)E^{\mathrm{T}} + EP(-\tau)A_\tau^{\mathrm{T}}, \tag{6.19c}$$

and

$$\dot{Q}(t)E = Q(t)A_0 + Q(t-\tau)A_\tau,\ t \geqslant 0, \tag{6.20a}$$
$$Q(-t) = Q(t)^{\mathrm{T}}, \tag{6.20b}$$
$$-C^{\mathrm{T}}C = A_0^{\mathrm{T}}Q(0)E + E^{\mathrm{T}}Q(0)A_0 + A_\tau^{\mathrm{T}}Q(\tau)E + E^{\mathrm{T}}Q(-\tau)A_\tau. \tag{6.20c}$$

The $\mathcal{H}_2$-norm can then be computed as (see [JVM11, Theorem 1])

$$\|H\|_{\mathcal{H}_2}^2 = \operatorname{tr}\big(CP(0)C^{\mathrm{T}}\big) = \operatorname{tr}\big(B^{\mathrm{T}}Q(0)B\big).$$

The system (6.19) can be solved analytically, as done in [JVM11, Section III.A]. First, notice that differentiating (6.19b) gives us $-\dot{P}(-t) = \dot{P}(t)^{\mathrm{T}}$. Now, transposing (6.19a), substituting t with $-t+\tau$ and using the previous expression, we find

$$E\dot{P}(t-\tau) = -A_0P(t-\tau) - A_\tau P(t),\ t \leqslant \tau.$$

Defining $z(t) = \operatorname{col}(\operatorname{vec}(P(t)), \operatorname{vec}(P(t-\tau)))$ gives us a system of ODEs with a boundary value condition

$$\begin{bmatrix} E\otimes I & 0 \\ 0 & I\otimes E \end{bmatrix}\dot{z}(t) = \begin{bmatrix} A_0\otimes I & A_\tau\otimes I \\ -I\otimes A_\tau & -I\otimes A_0 \end{bmatrix} z(t), \quad t\in[0,\tau],$$
$$\begin{bmatrix} -\operatorname{vec}\big(BB^{\mathrm{T}}\big) \\ 0 \end{bmatrix} = \begin{bmatrix} A_0\otimes E & A_\tau\otimes E \\ I & 0 \end{bmatrix} z(0) + \begin{bmatrix} E\otimes A_\tau & E\otimes A_0 \\ 0 & -I \end{bmatrix} z(\tau),$$

which can be used to find the analytic solution. Therefore, an equivalent form of (6.19) is

$$\dot{P}(t)E^{\mathrm{T}} = P(t)A_0^{\mathrm{T}} + P(t-\tau)A_\tau^{\mathrm{T}},\ t\in[0,\tau], \tag{6.21a}$$
$$E\dot{P}(t-\tau) = -A_0P(t-\tau) - A_\tau P(t),\ t\in[0,\tau], \tag{6.21b}$$
$$-BB^{\mathrm{T}} = A_0P(0)E^{\mathrm{T}} + EP(0)A_0^{\mathrm{T}} + A_\tau P(\tau)E^{\mathrm{T}} + EP(-\tau)A_\tau^{\mathrm{T}}, \tag{6.21c}$$

and similarly, system (6.20) is equivalent to

$$\dot{Q}(t)E = Q(t)A_0 + Q(t-\tau)A_\tau,\ t\in[0,\tau], \tag{6.22a}$$
$$E^{\mathrm{T}}\dot{Q}(t-\tau) = -A_0^{\mathrm{T}}Q(t-\tau) - A_\tau^{\mathrm{T}}Q(t),\ t\in[0,\tau], \tag{6.22b}$$
$$-C^{\mathrm{T}}C = A_0^{\mathrm{T}}Q(0)E + E^{\mathrm{T}}Q(0)A_0 + A_\tau^{\mathrm{T}}Q(\tau)E + E^{\mathrm{T}}Q(-\tau)A_\tau. \tag{6.22c}$$

We will prefer this form in the following.

6.5.1 Wilson-type conditions

We want to find $\mathcal{H}_2$-optimality conditions for the ROM

$$\begin{aligned} \widehat{E}\dot{\widehat{x}}(t) &= \widehat{A}_0\widehat{x}(t) + \widehat{A}_\tau\widehat{x}(t-\tau) + \widehat{B}u(t), \\ \widehat{y}(t) &= \widehat{C}\widehat{x}(t). \end{aligned} \tag{6.23}$$

with $\widehat{E}, \widehat{A}_0, \widehat{A}_\tau \in \mathbb{R}^{r\times r}$, $\widehat{B} \in \mathbb{R}^{r\times m}$, and $\widehat{C} \in \mathbb{R}^{p\times r}$, where $\widehat{E}$ is invertible. Just as for the full-order model (FOM), the Gramians of the ROM (6.23) satisfy

$$\begin{aligned} \dot{\widehat{P}}(t)\widehat{E}^{\mathrm{T}} &= \widehat{P}(t)\widehat{A}_0^{\mathrm{T}} + \widehat{P}(t-\tau)\widehat{A}_\tau^{\mathrm{T}},\ t\in[0,\tau], \\ \widehat{E}\dot{\widehat{P}}(t-\tau) &= -\widehat{A}_0\widehat{P}(t-\tau) - \widehat{A}_\tau\widehat{P}(t),\ t\in[0,\tau], \\ -\widehat{B}\widehat{B}^{\mathrm{T}} &= \widehat{A}_0\widehat{P}(0)\widehat{E}^{\mathrm{T}} + \widehat{E}\widehat{P}(0)\widehat{A}_0^{\mathrm{T}} + \widehat{A}_\tau\widehat{P}(\tau)\widehat{E}^{\mathrm{T}} + \widehat{E}\widehat{P}(-\tau)\widehat{A}_\tau^{\mathrm{T}}, \end{aligned}$$

and

$$\begin{aligned} \dot{\widehat{Q}}(t)\widehat{E} &= \widehat{Q}(t)\widehat{A}_0 + \widehat{Q}(t-\tau)\widehat{A}_\tau,\ t\in[0,\tau], \\ \widehat{E}^{\mathrm{T}}\dot{\widehat{Q}}(t-\tau) &= -\widehat{A}_0^{\mathrm{T}}\widehat{Q}(t-\tau) - \widehat{A}_\tau^{\mathrm{T}}\widehat{Q}(t),\ t\in[0,\tau], \\ -\widehat{C}^{\mathrm{T}}\widehat{C} &= \widehat{A}_0^{\mathrm{T}}\widehat{Q}(0)\widehat{E} + \widehat{E}^{\mathrm{T}}\widehat{Q}(0)\widehat{A}_0 + \widehat{A}_\tau^{\mathrm{T}}\widehat{Q}(\tau)\widehat{E} + \widehat{E}^{\mathrm{T}}\widehat{Q}(-\tau)\widehat{A}_\tau. \end{aligned}$$

We can define the error system as

$$\begin{aligned} \begin{bmatrix} E & 0 \\ 0 & \widehat{E} \end{bmatrix} \begin{bmatrix} \dot{x}(t) \\ \dot{\widehat{x}}(t) \end{bmatrix} &= \begin{bmatrix} A_0 & 0 \\ 0 & \widehat{A}_0 \end{bmatrix} \begin{bmatrix} x(t) \\ \widehat{x}(t) \end{bmatrix} + \begin{bmatrix} A_\tau & 0 \\ 0 & \widehat{A}_\tau \end{bmatrix} \begin{bmatrix} x(t-\tau) \\ \widehat{x}(t-\tau) \end{bmatrix} + \begin{bmatrix} B \\ \widehat{B} \end{bmatrix} u(t), \\ y(t) - \widehat{y}(t) &= \begin{bmatrix} C & -\widehat{C} \end{bmatrix} \begin{bmatrix} x(t) \\ \widehat{x}(t) \end{bmatrix}. \end{aligned}$$

We find the Gramians of the error system have the form

$$\begin{bmatrix} P(t) & \widetilde{P}(t) \\ \widetilde{P}(-t)^{\mathrm{T}} & \widehat{P}(t) \end{bmatrix}, \quad \begin{bmatrix} Q(t) & \widetilde{Q}(t) \\ \widetilde{Q}(-t)^{\mathrm{T}} & \widehat{Q}(t) \end{bmatrix},$$

where

$$\begin{aligned} \dot{\widetilde{P}}(t)\widehat{E}^{\mathrm{T}} &= \widetilde{P}(t)\widehat{A}_0^{\mathrm{T}} + \widetilde{P}(t-\tau)\widehat{A}_\tau^{\mathrm{T}}, \\ E\dot{\widetilde{P}}(t-\tau) &= -A_0\widetilde{P}(t-\tau) - A_\tau\widetilde{P}(t), \\ -B\widehat{B}^{\mathrm{T}} &= A_0\widetilde{P}(0)\widehat{E}^{\mathrm{T}} + E\widetilde{P}(0)\widehat{A}_0^{\mathrm{T}} + A_\tau\widetilde{P}(\tau)\widehat{E}^{\mathrm{T}} + E\widetilde{P}(-\tau)\widehat{A}_\tau^{\mathrm{T}}, \end{aligned}$$

and

$$\begin{aligned} \dot{\widetilde{Q}}(t)\widehat{E} &= \widetilde{Q}(t)\widehat{A}_0 + \widetilde{Q}(t-\tau)\widehat{A}_\tau, \\ E^{\mathrm{T}}\dot{\widetilde{Q}}(t-\tau) &= -A_0^{\mathrm{T}}\widetilde{Q}(t-\tau) - A_\tau^{\mathrm{T}}\widetilde{Q}(t), \\ C^{\mathrm{T}}\widehat{C} &= A_0^{\mathrm{T}}\widetilde{Q}(0)\widehat{E} + E^{\mathrm{T}}\widetilde{Q}(0)\widehat{A}_0 + A_\tau^{\mathrm{T}}\widetilde{Q}(\tau)\widehat{E} + E^{\mathrm{T}}\widetilde{Q}(-\tau)\widehat{A}_\tau. \end{aligned}$$

Theorem 6.14:
Let (6.23) be asymptotically stable. Then for the squared $\mathcal{H}_2$-error $\mathcal{J}$, we have

$$\begin{aligned}
\nabla_{\widehat{E}}\mathcal{J} &= -2\int_0^\tau \widetilde{Q}(\tau-t)^{\mathrm{T}} A_\tau \dot{\widetilde{P}}(t)\,\mathrm{d}t + 2\widetilde{Q}(0)^{\mathrm{T}} A_0 \widetilde{P}(0) + 2\widetilde{Q}(0)^{\mathrm{T}} A_\tau \widetilde{P}(\tau) \\
&\quad - 2\int_0^\tau \widehat{Q}(\tau-t)^{\mathrm{T}} \widehat{A}_\tau \dot{\widehat{P}}(t)\,\mathrm{d}t + 2\widehat{Q}(0)\widehat{A}_0\widehat{P}(0) + 2\widehat{Q}(0)\widehat{A}_\tau\widehat{P}(\tau), \\
\nabla_{\widehat{A}_0}\mathcal{J} &= 2\int_0^\tau \widetilde{Q}(\tau-t)^{\mathrm{T}} A_\tau \widetilde{P}(t)\,\mathrm{d}t + 2\widetilde{Q}(0)^{\mathrm{T}} E \widetilde{P}(0) \\
&\quad + 2\int_0^\tau \widehat{Q}(\tau-t)^{\mathrm{T}} \widehat{A}_\tau \widehat{P}(t)\,\mathrm{d}t + 2\widehat{Q}(0)\widehat{E}\widehat{P}(0), \\
\nabla_{\widehat{A}_\tau}\mathcal{J} &= 2\int_0^\tau \widetilde{Q}(\tau-t)^{\mathrm{T}} A_\tau \widetilde{P}(t-\tau)\,\mathrm{d}t + 2\widetilde{Q}(0)^{\mathrm{T}} E \widetilde{P}(-\tau) \\
&\quad + 2\int_0^\tau \widehat{Q}(\tau-t)^{\mathrm{T}} \widehat{A}_\tau \widehat{P}(t-\tau)\,\mathrm{d}t + 2\widehat{Q}(0)\widehat{E}\widehat{P}(-\tau), \\
\nabla_{\widehat{B}}\mathcal{J} &= 2\widetilde{Q}(0)^{\mathrm{T}} B + 2\widehat{Q}(0)\widehat{B}, \\
\nabla_{\widehat{C}}\mathcal{J} &= -2C\widetilde{P}(0) + 2\widehat{C}\widehat{P}(0).
\end{aligned}$$

◊

Proof. We proceed similar to the proof of Theorem 5.7. The $\mathcal{H}_2$-error is given by

$$\begin{aligned}
\|H - \widehat{H}\|_{\mathcal{H}_2}^2 &= \operatorname{tr}\left(\begin{bmatrix} C & -\widehat{C} \end{bmatrix} \begin{bmatrix} P(0) & \widetilde{P}(0) \\ \widetilde{P}(0)^{\mathrm{T}} & \widehat{P}(0) \end{bmatrix} \begin{bmatrix} C^{\mathrm{T}} \\ -\widehat{C}^{\mathrm{T}} \end{bmatrix}\right) \\
&= \operatorname{tr}\left(CP(0)C^{\mathrm{T}} - 2C\widetilde{P}(0)\widehat{C}^{\mathrm{T}} + \widehat{C}\widehat{P}(0)\widehat{C}^{\mathrm{T}}\right).
\end{aligned}$$

Therefore, we consider the optimization problem

$$\begin{aligned}
\underset{\widehat{E},\widehat{A}_0,\widehat{A}_\tau,\widehat{B},\widehat{C},\widetilde{P},\widehat{P}}{\text{minimize}} \quad & \operatorname{tr}\left(CP(0)C^{\mathrm{T}} - 2C\widetilde{P}(0)\widehat{C}^{\mathrm{T}} + \widehat{C}\widehat{P}(0)\widehat{C}^{\mathrm{T}}\right), \\
\text{subject to} \quad & \dot{\widetilde{P}}(t)\widehat{E}^{\mathrm{T}} = \widetilde{P}(t)\widehat{A}_0^{\mathrm{T}} + \widetilde{P}(t-\tau)\widehat{A}_\tau^{\mathrm{T}}, \quad t \in [0,\tau], \\
& E\dot{\widetilde{P}}(t-\tau) = -A_0\widetilde{P}(t-\tau) - A_\tau\widetilde{P}(t), \quad t \in [0,\tau], \\
& -B\widehat{B}^{\mathrm{T}} = A_0\widetilde{P}(0)\widehat{E}^{\mathrm{T}} + E\widetilde{P}(0)\widehat{A}_0^{\mathrm{T}} + A_\tau\widetilde{P}(\tau)\widehat{E}^{\mathrm{T}} + E\widetilde{P}(-\tau)\widehat{A}_\tau^{\mathrm{T}}, \\
& \dot{\widehat{P}}(t)\widehat{E}^{\mathrm{T}} = \widehat{P}(t)\widehat{A}_0^{\mathrm{T}} + \widehat{P}(t-\tau)\widehat{A}_\tau^{\mathrm{T}}, \quad t \in [0,\tau], \\
& \widehat{E}\dot{\widehat{P}}(t-\tau) = -\widehat{A}_0\widehat{P}(t-\tau) - \widehat{A}_\tau\widehat{P}(t), \quad t \in [0,\tau], \\
& -\widehat{B}\widehat{B}^{\mathrm{T}} = \widehat{A}_0\widehat{P}(0)\widehat{E}^{\mathrm{T}} + \widehat{E}\widehat{P}(0)\widehat{A}_0^{\mathrm{T}} + \widehat{A}_\tau\widehat{P}(\tau)\widehat{E}^{\mathrm{T}} + \widehat{E}\widehat{P}(-\tau)\widehat{A}_\tau^{\mathrm{T}},
\end{aligned}$$

where we have $\widetilde{P} \in H^1([-\tau,\tau];\mathbb{R}^{n\times r})$ and $\widehat{P} \in H^1([-\tau,\tau];\mathbb{R}^{r\times r})$. The Lagrange func-

tion is

$$\begin{aligned}
\mathcal{L} = {} & \operatorname{tr}\Big(CP(0)C^{\mathrm{T}} - 2C\widetilde{P}(0)\widehat{C}^{\mathrm{T}} + \widehat{C}\widehat{P}(0)\widehat{C}^{\mathrm{T}}\Big) \\
& + \int_0^\tau \operatorname{tr}\Big(\widetilde{\Lambda}_1(t)^{\mathrm{T}}\left(\dot{\widetilde{P}}(t)\widehat{E}^{\mathrm{T}} - \widetilde{P}(t)\widehat{A}_0^{\mathrm{T}} - \widetilde{P}(t-\tau)\widehat{A}_\tau^{\mathrm{T}}\right)\Big)\,\mathrm{d}t \\
& + \int_0^\tau \operatorname{tr}\Big(\widetilde{\Lambda}_2(t)^{\mathrm{T}}\left(E\dot{\widetilde{P}}(t-\tau) + A_0\widetilde{P}(t-\tau) + A_\tau\widetilde{P}(t)\right)\Big)\,\mathrm{d}t \\
& + \operatorname{tr}\Big(\widetilde{\Lambda}_3^{\mathrm{T}}\left(A_0\widetilde{P}(0)\widehat{E}^{\mathrm{T}} + E\widetilde{P}(0)\widehat{A}_0^{\mathrm{T}} + A_\tau\widetilde{P}(\tau)\widehat{E}^{\mathrm{T}} + E\widetilde{P}(-\tau)\widehat{A}_\tau^{\mathrm{T}} + B\widehat{B}^{\mathrm{T}}\right)\Big) \\
& + \int_0^\tau \operatorname{tr}\Big(\widehat{\Lambda}_1(t)^{\mathrm{T}}\left(\dot{\widehat{P}}(t)\widehat{E}^{\mathrm{T}} - \widehat{P}(t)\widehat{A}_0^{\mathrm{T}} - \widehat{P}(t-\tau)\widehat{A}_\tau^{\mathrm{T}}\right)\Big)\,\mathrm{d}t \\
& + \int_0^\tau \operatorname{tr}\Big(\widehat{\Lambda}_2(t)^{\mathrm{T}}\left(\widehat{E}\dot{\widehat{P}}(t-\tau) + \widehat{A}_0\widehat{P}(t-\tau) + \widehat{A}_\tau\widehat{P}(t)\right)\Big)\,\mathrm{d}t \\
& + \operatorname{tr}\Big(\widehat{\Lambda}_3^{\mathrm{T}}\left(\widehat{A}_0\widehat{P}(0)\widehat{E}^{\mathrm{T}} + \widehat{E}\widehat{P}(0)\widehat{A}_0^{\mathrm{T}} + \widehat{A}_\tau\widehat{P}(\tau)\widehat{E}^{\mathrm{T}} + \widehat{E}\widehat{P}(-\tau)\widehat{A}_\tau^{\mathrm{T}} + \widehat{B}\widehat{B}^{\mathrm{T}}\right)\Big),
\end{aligned}$$

where $\widetilde{\Lambda}_1, \widetilde{\Lambda}_2 \in H^1([0,\tau];\mathbb{R}^{n\times r})$, $\widetilde{\Lambda}_3 \in \mathbb{R}^{n\times r}$, $\widehat{\Lambda}_1, \widehat{\Lambda}_2 \in H^1([0,\tau];\mathbb{R}^{r\times r})$, and $\widehat{\Lambda}_3 \in \mathbb{R}^{r\times r}$ are Lagrange multipliers.

Directional derivatives with respect to $\widetilde{P}$ and $\widehat{P}$, in directions $\widetilde{D}$ and $\widehat{D}$, are

$$\begin{aligned}
\mathrm{d}_{\widetilde{P}}\mathcal{L}(\widetilde{D}) = {} & \int_0^\tau \operatorname{tr}\Big(\left(-\widehat{E}^{\mathrm{T}}\dot{\widetilde{\Lambda}}_1(t)^{\mathrm{T}} - \widehat{A}_0^{\mathrm{T}}\widetilde{\Lambda}_1(t)^{\mathrm{T}} + \widetilde{\Lambda}_2(t)^{\mathrm{T}}A_\tau\right)\widetilde{D}(t)\Big)\,\mathrm{d}t \\
& + \int_0^\tau \operatorname{tr}\Big(\left(-\widehat{A}_\tau^{\mathrm{T}}\widetilde{\Lambda}_1(t)^{\mathrm{T}} - \dot{\widetilde{\Lambda}}_2(t)^{\mathrm{T}}E + \widetilde{\Lambda}_2(t)^{\mathrm{T}}A_0\right)\widetilde{D}(t-\tau)\Big)\,\mathrm{d}t, \\
& + \operatorname{tr}\Big(\left(-2\widehat{C}^{\mathrm{T}}C - \widehat{E}^{\mathrm{T}}\widetilde{\Lambda}_1(0)^{\mathrm{T}} + \widetilde{\Lambda}_2(\tau)^{\mathrm{T}}E + \widehat{E}^{\mathrm{T}}\widetilde{\Lambda}_3^{\mathrm{T}}A_0 + \widehat{A}_0^{\mathrm{T}}\widetilde{\Lambda}_3^{\mathrm{T}}E\right)\widetilde{D}(0)\Big) \\
& + \operatorname{tr}\Big(\left(\widehat{E}^{\mathrm{T}}\widetilde{\Lambda}_1(\tau)^{\mathrm{T}} + \widehat{E}^{\mathrm{T}}\widetilde{\Lambda}_3^{\mathrm{T}}A_\tau\right)\widetilde{D}(\tau)\Big) \\
& + \operatorname{tr}\Big(\left(-\widetilde{\Lambda}_2(0)^{\mathrm{T}}E + \widehat{A}_\tau^{\mathrm{T}}\widetilde{\Lambda}_3^{\mathrm{T}}E\right)\widetilde{D}(-\tau)\Big) \\
\mathrm{d}_{\widehat{P}}\mathcal{L}(\widehat{D}) = {} & \int_0^\tau \operatorname{tr}\Big(\left(-\widehat{E}^{\mathrm{T}}\dot{\widehat{\Lambda}}_1(t)^{\mathrm{T}} - \widehat{A}_0^{\mathrm{T}}\widehat{\Lambda}_1(t)^{\mathrm{T}} + \widehat{\Lambda}_2(t)^{\mathrm{T}}\widehat{A}_\tau\right)\widehat{D}(t)\Big)\,\mathrm{d}t \\
& + \int_0^\tau \operatorname{tr}\Big(\left(-\widehat{A}_\tau^{\mathrm{T}}\widehat{\Lambda}_1(t)^{\mathrm{T}} - \dot{\widehat{\Lambda}}_2(t)^{\mathrm{T}}\widehat{E} + \widehat{\Lambda}_2(t)^{\mathrm{T}}\widehat{A}_0\right)\widehat{D}(t-\tau)\Big)\,\mathrm{d}t \\
& + \operatorname{tr}\Big(\left(\widehat{C}^{\mathrm{T}}\widehat{C} - \widehat{E}^{\mathrm{T}}\widehat{\Lambda}_1(0)^{\mathrm{T}} + \widehat{\Lambda}_2(\tau)^{\mathrm{T}}\widehat{E} + \widehat{E}^{\mathrm{T}}\widehat{\Lambda}_3^{\mathrm{T}}\widehat{A}_0 + \widehat{A}_0^{\mathrm{T}}\widehat{\Lambda}_3^{\mathrm{T}}\widehat{E}\right)\widehat{D}(0)\Big) \\
& + \operatorname{tr}\Big(\left(\widehat{E}^{\mathrm{T}}\widehat{\Lambda}_1(\tau)^{\mathrm{T}} + \widehat{E}^{\mathrm{T}}\widehat{\Lambda}_3^{\mathrm{T}}\widehat{A}_\tau\right)\widehat{D}(\tau)\Big) \\
& + \operatorname{tr}\Big(\left(-\widehat{\Lambda}_2(0)^{\mathrm{T}}\widehat{E} + \widehat{A}_\tau^{\mathrm{T}}\widehat{\Lambda}_3^{\mathrm{T}}\widehat{E}\right)\widehat{D}(-\tau)\Big).
\end{aligned}$$

Equating with zero, it follows that

$$\begin{aligned}
\dot{\widetilde{\Lambda}}_1(t)\widehat{E} &= -\widetilde{\Lambda}_1(t)\widehat{A}_0 + A_\tau^{\mathrm{T}}\widetilde{\Lambda}_2(t), \quad \text{for a.e. } t \in [0,\tau],\\
E^{\mathrm{T}}\dot{\widetilde{\Lambda}}_2(t) &= A_0^{\mathrm{T}}\widetilde{\Lambda}_2(t) - \widetilde{\Lambda}_1(t)\widehat{A}_\tau, \quad \text{for a.e. } t \in [0,\tau],\\
2C^{\mathrm{T}}\widehat{C} &= A_0^{\mathrm{T}}\widetilde{\Lambda}_3\widehat{E} + E^{\mathrm{T}}\widetilde{\Lambda}_3\widehat{A}_0 - \widetilde{\Lambda}_1(0)\widehat{E} + E^{\mathrm{T}}\widetilde{\Lambda}_2(\tau),\\
\widetilde{\Lambda}_1(\tau) &= -A_\tau^{\mathrm{T}}\widetilde{\Lambda}_3,\\
\widetilde{\Lambda}_2(0) &= \widetilde{\Lambda}_3\widehat{A}_\tau,
\end{aligned}$$

and

$$\begin{aligned}
\dot{\widehat{\Lambda}}_1(t)\widehat{E} &= -\widehat{\Lambda}_1(t)\widehat{A}_0 + \widehat{A}_\tau^{\mathrm{T}}\widehat{\Lambda}_2(t), \quad \text{for a.e. } t \in [0,\tau],\\
\widehat{E}^{\mathrm{T}}\dot{\widehat{\Lambda}}_2(t) &= \widehat{A}_0^{\mathrm{T}}\widehat{\Lambda}_2(t) - \widehat{\Lambda}_1(t)\widehat{A}_\tau, \quad \text{for a.e. } t \in [0,\tau],\\
-\widehat{C}^{\mathrm{T}}\widehat{C} &= \widehat{A}_0^{\mathrm{T}}\widehat{\Lambda}_3\widehat{E} + \widehat{E}^{\mathrm{T}}\widehat{\Lambda}_3\widehat{A}_0 - \widehat{\Lambda}_1(0)\widehat{E} + \widehat{E}^{\mathrm{T}}\widehat{\Lambda}_2(\tau),\\
\widehat{\Lambda}_1(\tau) &= -\widehat{A}_\tau^{\mathrm{T}}\widehat{\Lambda}_3,\\
\widehat{\Lambda}_2(0) &= \widehat{\Lambda}_3\widehat{A}_\tau.
\end{aligned}$$

The solution is

$$\begin{aligned}
\widetilde{\Lambda}_1(t) &= -2A_\tau^{\mathrm{T}}\widetilde{Q}(\tau - t), \quad \widetilde{\Lambda}_2(t) = 2\widetilde{Q}(-t)\widehat{A}_\tau, \quad \widetilde{\Lambda}_3 = 2\widetilde{Q}(0),\\
\widehat{\Lambda}_1(t) &= -\widehat{A}_\tau^{\mathrm{T}}\widehat{Q}(\tau - t), \quad \widehat{\Lambda}_2(t) = \widehat{Q}(-t)\widehat{A}_\tau, \quad \widehat{\Lambda}_3 = \widehat{Q}(0).
\end{aligned}$$

Gradients with respect to reduced matrices are

$$\begin{aligned}
\nabla_{\widehat{E}}\mathcal{L} &= \int_0^\tau \widetilde{\Lambda}_1(t)^{\mathrm{T}}\dot{\widetilde{P}}(t)\,\mathrm{d}t + \widetilde{\Lambda}_3^{\mathrm{T}}\left(A_0\widetilde{P}(0) + A_\tau\widetilde{P}(\tau)\right)\\
&\quad + \int_0^\tau \widehat{\Lambda}_1(t)^{\mathrm{T}}\dot{\widehat{P}}(t)\,\mathrm{d}t + \int_0^\tau \widehat{\Lambda}_2(t)\dot{\widehat{P}}(t-\tau)^{\mathrm{T}}\,\mathrm{d}t\\
&\quad + \widehat{\Lambda}_3^{\mathrm{T}}\widehat{A}_0\widehat{P}(0) + \widehat{\Lambda}_3\widehat{A}_0\widehat{P}(0)^{\mathrm{T}} + \widehat{\Lambda}_3^{\mathrm{T}}\widehat{A}_\tau\widehat{P}(\tau) + \widehat{\Lambda}_3\widehat{A}_\tau\widehat{P}(-\tau)^{\mathrm{T}}\\
&= -2\int_0^\tau \widetilde{Q}(\tau-t)^{\mathrm{T}}A_\tau\dot{\widetilde{P}}(t)\,\mathrm{d}t + 2\widetilde{Q}(0)^{\mathrm{T}}A_0\widetilde{P}(0) + 2\widetilde{Q}(0)^{\mathrm{T}}A_\tau\widetilde{P}(\tau)\\
&\quad - 2\int_0^\tau \widehat{Q}(\tau-t)^{\mathrm{T}}\widehat{A}_\tau\dot{\widehat{P}}(t)\,\mathrm{d}t + 2\widehat{Q}(0)\widehat{A}_0\widehat{P}(0) + 2\widehat{Q}(0)\widehat{A}_\tau\widehat{P}(\tau),\\
\nabla_{\widehat{A}_0}\mathcal{L} &= -\int_0^\tau \widetilde{\Lambda}_1(t)^{\mathrm{T}}\widetilde{P}(t)\,\mathrm{d}t + \widetilde{\Lambda}_3^{\mathrm{T}}E\widetilde{P}(0)\\
&\quad - \int_0^\tau \widehat{\Lambda}_1(t)^{\mathrm{T}}\widehat{P}(t)\,\mathrm{d}t + \int_0^\tau \widehat{\Lambda}_2(t)\widehat{P}(t-\tau)^{\mathrm{T}}\,\mathrm{d}t\\
&\quad + \widehat{\Lambda}_3\widehat{E}\widehat{P}(0)^{\mathrm{T}} + \widehat{\Lambda}_3^{\mathrm{T}}\widehat{E}\widehat{P}(0)\\
&= 2\int_0^\tau \widetilde{Q}(\tau-t)^{\mathrm{T}}A_\tau\widetilde{P}(t)\,\mathrm{d}t + 2\widetilde{Q}(0)^{\mathrm{T}}E\widetilde{P}(0)\\
&\quad + 2\int_0^\tau \widehat{Q}(\tau-t)^{\mathrm{T}}\widehat{A}_\tau\widehat{P}(t)\,\mathrm{d}t + 2\widehat{Q}(0)\widehat{E}\widehat{P}(0),
\end{aligned}$$

$$
\begin{aligned}
\nabla_{\widehat{A}_\tau}\mathcal{L} &= -\int_0^\tau \widetilde{\Lambda}_1(t)^{\mathrm{T}}\widetilde{P}(t-\tau)\,\mathrm{d}t + \widetilde{\Lambda}_3^{\mathrm{T}}E\widetilde{P}(-\tau) \\
&\quad - \int_0^\tau \widehat{\Lambda}_1(t)^{\mathrm{T}}\widehat{P}(t-\tau)\,\mathrm{d}t + \int_0^\tau \widehat{\Lambda}_2(t)\widehat{P}(t)^{\mathrm{T}}\,\mathrm{d}t \\
&\quad + \widehat{\Lambda}_3\widehat{E}\widehat{P}(\tau)^{\mathrm{T}} + \widehat{\Lambda}_3^{\mathrm{T}}\widehat{E}\widehat{P}(-\tau) \\
&= 2\int_0^\tau \widetilde{Q}(\tau-t)^{\mathrm{T}}A_\tau\widetilde{P}(t-\tau)\,\mathrm{d}t + 2\widetilde{Q}(0)^{\mathrm{T}}E\widetilde{P}(-\tau) \\
&\quad + 2\int_0^\tau \widehat{Q}(\tau-t)^{\mathrm{T}}\widehat{A}_\tau\widehat{P}(t-\tau)\,\mathrm{d}t + 2\widehat{Q}(0)\widehat{E}\widehat{P}(-\tau), \\
\nabla_{\widehat{B}}\mathcal{L} &= \widetilde{\Lambda}_3^{\mathrm{T}}B + \widehat{\Lambda}_3\widehat{B} + \widehat{\Lambda}_3^{\mathrm{T}}\widehat{B} \\
&= 2\widetilde{Q}(0)^{\mathrm{T}}B + 2\widehat{Q}(0)\widehat{B}, \\
\nabla_{\widehat{C}}\mathcal{L} &= -2C\widetilde{P}(0) + 2\widehat{C}\widehat{P}(0).
\end{aligned}
$$

□

From the previous theorem, we directly find the Wilson-type necessary optimality conditions for $\mathcal{H}_2$-optimal MOR of LTD systems.

Corollary 6.15:
Let (6.23) be an $\mathcal{H}_2$-optimal ROM for (6.18). Then

$$
\begin{aligned}
0 &= \widetilde{Q}(0)^{\mathrm{T}}E\widetilde{P}(0) + \int_0^\tau \widetilde{Q}(\tau-t)^{\mathrm{T}}A_\tau\widetilde{P}(t)\,\mathrm{d}t \\
&\quad + \widehat{Q}(0)\widehat{E}\widehat{P}(0) + \int_0^\tau \widehat{Q}(\tau-t)^{\mathrm{T}}\widehat{A}_\tau\widehat{P}(t)\,\mathrm{d}t, \\
0 &= \widetilde{Q}(0)^{\mathrm{T}}E\widetilde{P}(-\tau) + \int_0^\tau \widetilde{Q}(\tau-t)^{\mathrm{T}}A_\tau\widetilde{P}(t-\tau)\,\mathrm{d}t \\
&\quad + \widehat{Q}(0)\widehat{E}\widehat{P}(-\tau) + \int_0^\tau \widehat{Q}(\tau-t)^{\mathrm{T}}\widehat{A}_\tau\widehat{P}(t-\tau)\,\mathrm{d}t, \\
0 &= \widetilde{Q}(0)^{\mathrm{T}}A_0\widetilde{P}(0) + \widetilde{Q}(0)^{\mathrm{T}}A_\tau\widetilde{P}(\tau) - \int_0^\tau \widetilde{Q}(\tau-t)^{\mathrm{T}}A_\tau\dot{\widetilde{P}}(t)\,\mathrm{d}t \\
&\quad + \widehat{Q}(0)\widehat{A}_0\widehat{P}(0) + \widehat{Q}(0)\widehat{A}_\tau\widehat{P}(\tau) - \int_0^\tau \widehat{Q}(\tau-t)^{\mathrm{T}}\widehat{A}_\tau\dot{\widehat{P}}(t)\,\mathrm{d}t, \\
0 &= \widetilde{Q}(0)^{\mathrm{T}}B + \widehat{Q}(0)\widehat{B}, \\
0 &= C\widetilde{P}(0) - \widehat{C}\widehat{P}(0).
\end{aligned}
$$

◊

6.6 Conclusion

We derived Wilson-type necessary optimality conditions for $\mathcal{H}_2$-optimal MOR of second-order systems, port-Hamiltonian systems, LTD systems with a single delay, and $\mathcal{H}_2\otimes\mathcal{L}_2$-optimal MOR of PLTI systems. Similar approach can be taken for, e.g., $\mathcal{H}_2$-optimal MOR of linear time-varying (LTV) systems and $\mathcal{L}_2$-optimal MOR of stationary parametric systems.

CHAPTER 7

CONCLUSIONS AND OUTLOOK

Contents

7.1 Summary

In this thesis, we have investigated structure-preserving MOR problems for different types of structured systems, mainly network systems.

In Chapter 3, we studied clustering-based MOR of multi-agent systems. In the first part, we have focused on linear multi-agent systems and considered the problem of $\mathcal{H}_2$-optimal clustering-based MOR. Based on the relaxation of the discrete optimization problem, we have proposed combining IRKA with a clustering algorithm. Next, we have generalized this to a framework of combining a projection-based method with a clustering algorithm and applied it to nonlinear multi-agent systems.

In Chapter 4, we have considered more theoretically the error due to clustering. First, we have looked at linear multi-agent systems and derived $\mathcal{H}_2$ and $\mathcal{H}_\infty$ error bounds when using an AEP. We have also proposed an extension to arbitrary partitions using the distance to a graph for which the partition becomes almost equitable. Next, we have considered nonlinear power systems, a type of nonlinear multi-agent systems. There, we derive equivalence conditions for clustering with zero error based on graph symmetries and equitable partitions.

In Chapter 5, we have studied the problem of subsystem reduction for linear network systems. In the first part, we extend a balancing-based MOR method which preserves stability for network systems satisfying a certain small-gain condition. Using the known a priori $\mathcal{H}_\infty$ error bound, it allows automatic choice of the order of the reduced subsystems. In the second part, we considered $\mathcal{H}_2$-optimal subsystem reduction. Using the Gramian-based formulation of the $\mathcal{H}_2$-error, we have derived gradients with respect to matrices defining the ROM. Thereby, we have also obtained Wilson-type necessary

optimality conditions.

In Chapter 6, we have used the ideas from Chapter 5 to other structure-preserving $\mathcal{H}_2$-optimal MOR problems. In particular, we have considered structure-preserving MOR for second-order systems, port-Hamiltonian systems, and time-delay systems. Additionally, we have also considered $\mathcal{H}_2 \otimes \mathcal{L}_2$-optimal MOR for parametric systems. We have derived Wilson-type necessary optimality conditions and for some systems also the interpolatory optimality conditions.

7.2 Future research perspectives

We considered different aspects of structure-preserving MOR, which motivate future research in new methods for network systems and other structured systems.

In Chapter 3, we have found on a small-scale example that combining a projection-based method and a clustering algorithms gives a partition close to the optimal one. However, it is not clear whether this is true in general. Deriving an error bound for this approach would be an interesting goal for future research. Furthermore, comparing different clustering algorithms and with a theoretical analysis would be a worthwhile. Additionally, extending the framework to multi-agent system evolving over directed graphs should be possible.

In Chapter 4, we have derived error bounds for clustering-based MOR of certain linear multi-agent systems. Extending the results to more general multi-agent systems, e.g., with different output functions or with directed underlying graph, would be an interesting problem. Finding easy to compute error bounds for non-almost equitable partitions remains an open problem. For power systems, we have focused on reduction with zero error. Deriving general error bounds remains an open problem.

In the first part of Chapter 5, we have considered subsystem reduction for network systems satisfying a particular small-gain condition. Relaxing this condition, e.g., by dissipativity, would extend the applicability of this approach. In the second part, we have derived Wilson-type necessary $\mathcal{H}_2$-optimality conditions. Finding interpolatory conditions and developing an interpolatory method could be an interesting problem.

In Chapter 6, as in the previous, finding interpolatory conditions for $\mathcal{H}_2$-optimal MOR of time-delay systems would be an interesting research question. Additionally, efficient implementation for large-scale systems is an open problem.

BIBLIOGRAPHY

[AA13] A. M. Annaswamy and M. Amin. *IEEE Vision for Smart Grid Controls: 2030 and Beyond*. IEEE, June 2013. doi:10.1109/IEEESTD.2013.6577608. 84

[ABG10] A. C. Antoulas, C. A. Beattie, and S. Gugercin. Interpolatory model reduction of large-scale dynamical systems. In Javad Mohammadpour and Karolos M. Grigoriadis, editors, *Efficient Modeling and Control of Large-Scale Systems*, pages 3–58. Springer US, 2010. doi:10.1007/978-1-4419-5757-3_1. 19, 20, 22, 23

[ABGA13] B. Anić, C. Beattie, S. Gugercin, and A. C. Antoulas. Interpolatory weighted-$\mathcal{H}_2$ model reduction. *Automatica J. IFAC*, 49(5):1275–1280, 2013. doi:10.1016/j.automatica.2013.01.040. 105

[Ant05] A. C. Antoulas. *Approximation of Large-Scale Dynamical Systems*, volume 6 of *Adv. Des. Control*. SIAM Publications, Philadelphia, PA, 2005. doi:10.1137/1.9780898718713. 1, 13, 14, 15, 16, 17, 18, 19, 20

[BB14] C. A. Beattie and P. Benner. $\mathcal{H}_2$-optimality conditions for structured dynamical systems. Preprint MPIMD/14-18, Max Planck Institute Magdeburg, 2014. Available from http://www.mpi-magdeburg.mpg.de/preprints/. 114, 118, 122, 124, 125, 126

[BBBG11] U. Baur, C. A. Beattie, P. Benner, and S. Gugercin. Interpolatory projection methods for parameterized model reduction. *SIAM J. Sci. Comput.*, 33(5):2489–2518, 2011. doi:10.1137/090776925. 128

[BBG15] T. Breiten, C. Beattie, and S. Gugercin. Near-optimal frequency-weighted interpolatory model reduction. *Syst. Control Lett.*, 78:8–18, 2015. doi:10.1016/j.sysconle.2015.01.005. 105

[BBG19] T. Breiten, C. A. Beattie, and S. Gugercin. $\mathcal{H}_2$-gap model reduction for stabilizable and detectable systems. e-print 1909.13764, arXiv, 2019. math.NA. URL: https://arxiv.org/abs/1909.13764. 30

[BFF+14] P. Benner, R. Findeisen, D. Flockerzi, U. Reichl, and K. Sundmacher, editors. *Large-Scale Networks in Engineering and Life Sciences*. Modeling and Simulation in Science, Engineering and Technology. Birkhäuser, Basel, CH, 2014. doi:10.1007/978-3-319-08437-4. 1

[BG09] C. A. Beattie and S. Gugercin. Interpolatory projection methods for structure-preserving model reduction. *Syst. Control Lett.*, 58(3):225–232, 2009. doi:10.1016/j.sysconle.2008.10.016. 114

[BG12] C. A. Beattie and S. Gugercin. Realization-independent $\mathcal{H}_2$-approximation. In *51st IEEE Conference on Decision and Control (CDC)*, pages 4953–4958, 2012. doi:10.1109/CDC.2012.6426344. 29

[BGM17] C. Beattie, S. Gugercin, and V. Mehrmann. Model reduction for systems with inhomogeneous initial conditions. *Syst. Control Lett.*, 99:99–106, 2017. doi:10.1016/j.sysconle.2016.11.007. 19

[BGW12] C. A. Beattie, S. Gugercin, and S. Wyatt. Inexact solves in interpolatory model reduction. *Linear Algebra Appl.*, 436(8):2916–2943, 2012. Special Issue dedicated to Danny Sorensen's 65th birthday. doi:10.1016/j.laa.2011.07.015. 43

[BKS11] P. Benner, M. Köhler, and J. Saak. Sparse-dense Sylvester equations in $\mathcal{H}_2$-model order reduction. Preprint MPIMD/11-11, Max Planck Institute Magdeburg, December 2011. Available from http://www.mpi-magdeburg.mpg.de/preprints/. 22, 29

[BMNP04] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera. An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations. *C. R. Math. Acad. Sci. Paris*, 339(9):667–672, 2004. doi:10.1016/j.crma.2004.08.006. 129

[BMS05] P. Benner, V. Mehrmann, and D. C. Sorensen. *Dimension Reduction of Large-Scale Systems*, volume 45 of *Lect. Notes Comput. Sci. Eng.* Springer-Verlag, Berlin/Heidelberg, Germany, 2005. doi:10.1007/3-540-27909-1. 1

[BNBG10] C. Boess, N. K. Nichols, and A. Bunse-Gerstner. Model order reduction for discrete unstable control systems using a balanced truncation approach. Preprint MPS_2010_06, University of Reading, 2010. Available from https://www.reading.ac.uk/maths-and-stats/research/maths-preprints.aspx. 29

[BOCW17] P. Benner, M. Ohlberger, A. Cohen, and K. Willcox. *Model Reduction and Approximation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2017. doi:10.1137/1.9781611974829. 1

[BP94] A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Society for Industrial and Applied Mathematics, 1994. doi:10.1137/1.9781611971262. 87

[BS13] P. Benner and J. Saak. Numerical solution of large and sparse continuous time algebraic matrix Riccati and Lyapunov equations: a state of the art survey. *GAMM-Mitteilungen*, 36(1):32–52, 2013. doi:10.1002/gamm.201310003. 18

[BSJ16] B. Besselink, H. Sandberg, and K. H. Johansson. Clustering-based model reduction of networked passive systems. *IEEE Trans. Autom. Control*, 61(10):2958–2973, 2016. doi:10.1109/TAC.2015.2505418. 2, 34, 40, 82

[CDR07] D. M. Cardoso, C. Delorme, and P. Rama. Laplacian eigenvectors and eigenvalues and almost equitable partitions. *European J. Combin.*, 28(3):665–673, 2007. doi:10.1016/j.ejc.2005.03.006. 34

[Chu87] E. K. Chu. The solution of the matrix equations $AXB-CXD = E$ and $(YA - DZ, YC - BZ) = (E, F)$. *Linear Algebra Appl.*, 93(0):93–105, 1987. doi:10.1016/S0024-3795(87)90314-4. 8

[CKS16] X. Cheng, Y. Kawano, and J. M. A. Scherpen. Graph structure-preserving model reduction of linear network systems. In *European Control Conference (ECC)*, pages 1970–1975, 2016. doi:10.1109/ECC.2016.7810580. 1, 34, 36, 82, 84

[CKS17] X. Cheng, Y. Kawano, and J. M. A. Scherpen. Reduction of second-order network systems with structure preservation. *IEEE Trans. Autom. Control*, 62(10):5026–5038, 2017. doi:10.1109/TAC.2017.2679479. 1, 84

[CKS18] X. Cheng, Y. Kawano, and J. M. A. Scherpen. Model reduction of multi-agent systems using dissimilarity-based clustering. *IEEE Trans. Autom. Control*, 2018. doi:10.1109/TAC.2018.2853578. 34

[CLVVD06] Y. Chahlaoui, D. Lemonnier, A. Vandendorpe, and P. Van Dooren. Second-order balanced truncation. *Linear Algebra Appl.*, 415(2–3):373–384, 2006. doi:10.1016/j.laa.2004.03.032. 114

[CM11] A. Chapman and M. Mesbahi. UAV flocking with wind gusts: Adaptive topology and model reduction. In *Proceedings of the American Control Conference*, pages 1045–1050, June 2011. doi:10.1109/ACC.2011.5990799. 39

[Col12] R. Coleman. *Calculus on Normed Vector Spaces*. Universitext. Springer New York, 2012. doi:10.1007/978-1-4614-3894-6. 8, 9, 10, 11, 12

[DB14] F. Dörfler and F. Bullo. Synchronization in complex networks of phase oscillators: A survey. *Automatica J. IFAC*, 50(6):1539–1564, 2014. doi:10.1016/j.automatica.2014.04.012. 84

[DP00] G. E. Dullerud and F. Paganini. *A Course in Robust Control Theory*. Springer New York, 2000. doi:10.1007/978-1-4757-3290-0. 15, 16, 17

[EFHO10] E. Estrada, M. Fox, D. J. Higham, and G.-L. Oppo, editors. *Network Science: Complexity in Nature and Technology*. Springer-Verlag, London, UK, 2010. doi:10.1007/978-1-84996-396-1. 1

[Enn85] D. F. Enns. *Model Reduction for Control System Design*. PhD thesis, Stanford Univ., March 1985. URL: https://ntrs.nasa.gov/search.jsp?R=19850014087. 29

[GAB08] S. Gugercin, A. C. Antoulas, and C. Beattie. $\mathcal{H}_2$ model reduction for large-scale linear dynamical systems. *SIAM J. Matrix Anal. Appl.*, 30(2):609–638, 2008. doi:10.1137/060666123. 22, 23

[GPBvdS12] S. Gugercin, R. V. Polyuga, C. Beattie, and A. van der Schaft. Structure-preserving tangential interpolation for model reduction of port-Hamiltonian systems. *Automatica*, 48(9):1963–1974, 2012. doi:10.1016/j.automatica.2012.05.052. 124

[GR01] C. Godsil and G. Royle. *Algebraic graph theory*, volume 207 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2001. doi:10.1007/978-1-4613-0163-9. 30, 31

[GV13] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, fourth edition, 2013. 6, 7

[Haa17] B. Haasdonk. Reduced basis methods for parametrized PDEs—a tutorial introduction for stationary and instationary problems. In P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, editors, *Model Reduction and Approximation: Theory and Algorithms*, pages 65–136. SIAM, 2017. doi:10.1137/1.9781611974829.ch2. 129

[HJ85] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1985. 6, 7, 102, 103

[HRA11] M. Heinkenschloss, T. Reis, and A. C. Antoulas. Balanced truncation model reduction for systems with inhomogeneous initial conditions. *Automatica J. IFAC*, 47(3):559–564, 2011. doi:10.1016/j.automatica.2010.12.002. 19

[HV05] M. Hinze and S. Volkwein. Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control. In P. Benner, V. Mehrmann, and D.C. Sorensen, editors,

Dimension Reduction of Large-Scale Systems, volume 45 of *Lect. Notes Comput. Sci. Eng.*, pages 261–306. Springer-Verlag, Berlin/Heidelberg, Germany, 2005. 51

[HW79] J. A. Hartigan and M. A. Wong. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979. doi:10.2307/2346830. 42

[II15] T. Ishizaki and J. Imura. Clustered model reduction of interconnected second-order systems. *Nonlinear Theory and Its Applications, IEICE*, 6(1):26–37, 2015. doi:10.1587/nolta.6.26. 1, 83

[IKG+12] T. Ishizaki, K. Kashima, A. Girard, J. Imura, L. Chen, and K. Aihara. Clustering-based $\mathcal{H}_2$-state aggregation of positive networks and its application to reduction of chemical master equations. In *51st IEEE Conference on Decision and Control (CDC)*, pages 4175–4180, December 2012. doi:10.1109/CDC.2012.6426793. 40

[IKG+15] T. Ishizaki, K. Kashima, A. Girard, J. Imura, L. Chen, and K. Aihara. Clustered model reduction of positive directed networks. *Automatica J. IFAC*, 59:238–247, 2015. doi:10.1016/j.automatica.2015.06.027. xi, 1, 34, 36, 82, 83, 84

[IKI16a] T. Ishizaki, R. Ku, and J. Imura. Clustered model reduction of networked dissipative systems. In *American Control Conference (ACC)*, pages 3662–3667, 2016. doi:10.1109/ACC.2016.7525482. xi, 2, 34, 35, 82, 83

[IKI16b] T. Ishizaki, R. Ku, and J. Imura. Eigenstructure analysis from symmetrical graph motives with application to aggregated controller design. In *55th IEEE Conference on Decision and Control (CDC)*, pages 5744–5749, 2016. doi:10.1109/CDC.2016.7799152. 84

[IKIA14] T. Ishizaki, K. Kashima, J. Imura, and K. Aihara. Model reduction and clusterization of large-scale bidirectional networks. *IEEE Trans. Autom. Control*, 59(1):48–63, January 2014. doi:10.1109/TAC.2013.2275891. xi, 1, 34, 36, 39, 40, 56, 82, 83, 84

[JVM11] E. Jarlebring, J. Vanbiervliet, and W. Michiels. Characterizing and computing the $\mathcal{H}_2$ norm of time-delay systems by solving the delay Lyapunov equation. *IEEE Trans. Autom. Control*, 56(4):814–825, April 2011. doi:10.1109/TAC.2010.2067510. 133

[Kro39] G. Kron. *Tensor analysis of networks*. Wiley, New York, 1939. 87

[Kun94] P Kundur. *Power System Stability and Control*. McGraw-Hill, 1994. 84, 87

[LDCH10] Z. Li, Z. Duan, G. Chen, and L. Huang. Consensus of multiagent systems and synchronization of complex networks: A unified viewpoint. *IEEE Trans. Circuits Syst. I, Regular Papers*, 57(1):213–224, January 2010. doi:10.1109/TCSI.2009.2023937. 1, 36

[MBG10] C. Magruder, C. Beattie, and S. Gugercin. Rational krylov methods for optimal $\mathcal{L}_2$ model reduction. In *49th IEEE Conference on Decision and Control (CDC)*, pages 6797–6802, December 2010. doi:10.1109/CDC.2010.5717454. 30

[ME10] M. Mesbahi and M. Egerstedt. *Graph Theoretic Methods in Multiagent Networks*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2010. doi:10.1515/9781400835355. 30, 31, 33, 34, 59

[MEB08] S. Martini, M. Egerstedt, and A. Bicchi. Controllability decompositions of networked systems through quotient graphs. In *47th IEEE Conference on Decision and Control (CDC)*, pages 5244–5249, December 2008. doi:10.1109/CDC.2008.4739213. 39

[MEB10] S. Martini, M. Egerstedt, and A. Bicchi. Controllability analysis of multi-agent systems using relaxed equitable partitions. *Int. J. Syst., Control Commun.*, 2(1/2/3):100–121, January 2010. doi:10.1504/IJSCC.2010.031160. 39

[MGB15] P. Mlinarić, S. Grundel, and P. Benner. Efficient model order reduction for multi-agent systems using QR decomposition-based clustering. In *54th IEEE Conference on Decision and Control (CDC)*, pages 4794–4799, December 2015. doi:10.1109/CDC.2015.7402967. 84

[ML67] L. Meier and D. G. Luenberger. Approximation of linear constant systems. *IEEE Trans. Autom. Control*, 12(5):585–588, 1967. doi:10.1109/TAC.1967.1098680. 22

[MN99] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Probability and Statistics. Wiley, 1999. 7

[Mor05] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Trans. Autom. Control*, 50(2):169–182, February 2005. doi:10.1109/TAC.2004.841888. 1

[MS96] D. G. Meyer and S. Srinivasan. Balancing and model reduction for second-order form linear systems. *IEEE Trans. Autom. Control*, 41(11):1632–1644, 1996. doi:10.1109/9.544000. 114

[MS05] V. Mehrmann and T. Stykel. Balanced truncation model reduction for large-scale systems in descriptor form. In P. Benner, V. Mehrmann, and D. C. Sorensen, editors, *Dimension Reduction of Large-Scale Systems*, volume 45 of *Lect. Notes Comput. Sci. Eng.*, pages 83–115. Springer-Verlag, Berlin/Heidelberg, Germany, 2005. doi:10.1007/3-540-27909-1_3. 17

[MTC13] N. Monshizadeh, H. L. Trentelman, and M. K. Camlibel. Stability and synchronization preserving model reduction of multi-agent systems. *Syst. Control Lett.*, 62(1):1–10, 2013. doi:10.1016/j.sysconle.2012.10.011. 2, 34, 36, 100, 102

[MTC14] N. Monshizadeh, H. L. Trentelman, and M. K. Camlibel. Projection-based model reduction of multi-agent systems using graph partitions. *IEEE Trans. Control Netw. Syst.*, 1(2):145–154, June 2014. doi:10.1109/TCNS.2014.2311883. xi, 31, 34, 36, 37, 40, 44, 45, 46, 47, 54, 58, 61, 68, 80, 81, 97, 151

[MZ10] C.-Q. Ma and J.-F. Zhang. Necessary and sufficient conditions for consensusability of linear multi-agent systems. *IEEE Trans. Autom. Control*, 55(5):1263–1268, May 2010. doi:10.1109/TAC.2010.2042764. 1

[New10] M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA, March 2010. doi:10.1093/acprof:oso/9780199206650.001.0001. 1

[OJ88] P. C. Opdenacker and E. A. Jonckheere. A contraction mapping preserving balanced reduction scheme and its infinity norm error bounds. *IEEE Trans. Circuits Syst.*, 35(2):184–189, 1988. doi:10.1109/31.1720. xv, 101

[OP05] R.C.L.F. Oliveira and P.L.D. Peres. Stability of polytopes of matrices via affine parameter-dependent Lyapunov functions: Asymptotically exact LMI conditions. *Linear Algebra Appl.*, 405:209–228, 2005. doi:10.1016/j.laa.2005.03.019. 80

[ORW13] M. Opmeer, T. Reis, and W. Wollner. Finite-rank ADI iteration for operator Lyapunov equations. *SIAM J. Control Optim.*, 51(5):4084–4117, 2013. doi:10.1137/120885310. 29

[OSM03] R. Olfati-Saber and R. M. Murray. Consensus protocols for networks of dynamic agents. In *Proceedings of the American Controls Conference*,

volume 2, pages 951–956, June 2003. doi:10.1109/ACC.2003.1239709. 1

[Ran96] A. Rantzer. On the Kalman-Yakubovich-Popov lemma. *Syst. Cont. Lett.*, 28(1):7–10, 1996. doi:10.1016/0167-6911(95)00063-1. 56

[RJME09] A. Rahmani, M. Ji, M. Mesbahi, and M. Egerstedt. Controllability of multi-agent systems from a graph-theoretic perspective. *SIAM J. Control Optim.*, 48(1):162–186, 2009. doi:10.1137/060674909. 39, 54, 84

[RS07] T. Reis and T. Stykel. Stability analysis and model order reduction of coupled systems. *Math. Comput. Model. Dyn. Syst.*, 13(5):413–436, 2007. doi:10.1080/13873950701189071. xii, 2, 99, 100, 102, 103, 104, 105

[RS08a] T. Reis and T. Stykel. Balanced truncation model reduction of second-order systems. *Math. Comput. Model. Dyn. Syst.*, 14(5):391–406, 2008. doi:10.1080/13873950701844170. 114

[RS08b] T. Reis and T. Stykel. A survey on model reduction of coupled systems. In W. H. A. Schilders, H. A. van der Vorst, and J. Rommes, editors, *Model Order Reduction: Theory, Research Aspects and Applications*, pages 133–155. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. doi:10.1007/978-3-540-78841-6_7. 2, 99

[Sat17] K. Sato. Riemannian optimal model reduction of linear second-order systems. *IEEE Contr. Syst. Lett.*, 1(1):2–7, July 2017. doi:10.1109/LCSYS.2017.2698178. 114

[Sch07] S. E. Schaeffer. Graph clustering. *Comput. Sci. Rev.*, 1(1):27–64, 2007. doi:10.1016/j.cosrev.2007.05.001. 39, 41

[Sim16] V. Simoncini. Computational methods for linear matrix equations. *SIAM Review*, 58(3):377–441, 2016. doi:10.1137/130912839. 18

[SM09] H. Sandberg and R. M. Murray. Model reduction of interconnected linear systems. *Optim. Control Appl. Methods*, 30(3):225–245, 2009. doi:10.1002/oca.854. 2, 99

[TTM13] H. L. Trentelman, K. Takaba, and N. Monshizadeh. Robust synchronization of uncertain linear multi-agent systems. *IEEE Trans. Autom. Control*, 58(6):1511–1523, 2013. doi:10.1109/TAC.2013.2239011. 59

[VDGA08] P. Van Dooren, K. Gallivan, and P.-A. Absil. $\mathcal{H}_2$-optimal model reduction of MIMO systems. *Appl. Math. Lett.*, 21:1267–1273, 2008. doi:10.1016/j.aml.2007.09.015. 22, 24, 26

[VDGA10] P. Van Dooren, K. A. Gallivan, and P.-A. Absil. $\mathcal{H}_2$-optimal model reduction with higher-order poles. *SIAM J. Matrix Anal. Appl.*, 31(5):2738–2753, 2010. doi:10.1137/080731591. 22, 29

[vdSJ14] A. van der Schaft and D. Jeltsema. *Port-Hamiltonian Systems Theory: An Introductory Overview*. now, 2014. doi:10.1561/2600000002. 123

[VVD08] A. Vandendorpe and P. Van Dooren. Model reduction of interconnected systems. In W. H. A. Schilders, H. A. van der Vorst, and J. Rommes, editors, *Model Order Reduction: Theory, Research Aspects and Applications*, volume 13 of *Mathematics in Industry*, pages 305–321. Springer, Berlin, Heidelberg, 2008. doi:10.1007/978-3-540-78841-6_14. 2, 99, 105

[Wil70] D. A. Wilson. Optimum solution of model-reduction problem. *Proceedings of the Institution of Electrical Engineers*, 117(6):1161–1165, 1970. doi:10.1049/piee.1970.0227. 22, 24

[Wya12] S. Wyatt. *Issues in Interpolatory Model Reduction: Inexact Solves, Second-order Systems and DAEs*. PhD thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA, May 2012. URL: http://hdl.handle.net/10919/27668. 114

[XC16] N. Xue and A. Chakrabortty. Optimal control of large-scale networks using clustering based projections. e-print 1609.05265, arXiv, 2016. cs.SY. URL: https://arxiv.org/abs/1609.05265. 84

[XZ11] Y. Xu and T. Zeng. Optimal $\mathcal{H}_2$ model reduction for large scale MIMO systems via tangential interpolation. *Int. J. Numer. Anal. Model.*, 8(1):174–188, 2011. URL: http://www.math.ualberta.ca/ijnam/Volume-8-2011/No-1-11/2011-01-10.pdf. 22, 26

[YCDAGX93] J. Yang, C. S. Chen, J. A. De Abreu-Garcia, and Y. Xu. Model reduction of unstable systems. *International Journal of Systems Science*, 24(12):2407–2414, 1993. doi:10.1080/00207729308949638. 29

[ZCC14] S. Zhang, M. Cao, and M. K. Camlibel. Upper and lower bounds for controllable subspaces of networks of diffusively coupled agents. *IEEE Trans. Autom. Control*, 59(3):745–750, March 2014. doi:10.1109/TAC.2013.2275666. 34

[ZDG96] K. Zhou, J. C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice-Hall, Upper Saddle River, NJ, 1996. 18

[Zei85a] E. Zeidler. *Nonlinear Functional Analysis and its Applications I: Fixed Point Theorems*. Springer-Verlag, 1985. 8, 9, 10, 11, 12

[Zei85b] E. Zeidler. *Nonlinear Functional Analysis and its Applications III: Variational Methods and Optimization*. Springer-Verlag, 1985. doi:10.1007/978-1-4612-5020-3. 12

[ZHD+01] H. Zha, X. He, C. Ding, H. Simon, and M. Gu. Spectral relaxation for k-means clustering. In *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, pages 1057–1064, 2001. URL: https://papers.nips.cc/paper/1992-spectral-relaxation-for-k-means-clustering.pdf. xv, 41, 42

[Zil91] A. Zilouchian. Balanced structures and model reduction of unstable systems. In *IEEE Proceedings of the SOUTHEASTCON '91*, pages 1198–1201 vol.2, April 1991. doi:10.1109/SECON.1991.147956. 29

[ZSW99] K. Zhou, G. Salomon, and E. Wu. Balanced realization and model reduction for unstable systems. *Internat. J. Robust Nonlinear Control*, 9(3):183–198, 1999. doi:10.1002/(SICI)1099-1239(199903)9:3<183::AID-RNC399>3.0.CO;2-E. 30

STATEMENT OF SCIENTIFIC COOPERATIONS

This work is based on articles and reports (published and unpublished) that have been obtained in cooperation with various coauthors. To guarantee a fair assessment of this thesis, this statement clarifies the contributions that each individual coauthor has made. The following people contributed to the content of this work:

- Peter Benner (PB), Max Planck Institute for Dynamics of Complex Technical Systems;
- Aranya Chakrabortty (AC), North Carolina State University;
- Sara Grundel (SG), Max Planck Institute for Dynamics of Complex Technical Systems;
- Takayuki Ishizaki (TI), Tokyo Institute of Technology;
- Hidde-Jan Jongsma (HJ), Netherlands Organisation for Applied Scientific Research;
- Harry L. Trentelman (HT), University of Groningen.

Chapter 3

The work on this chapter started from PB pointing me to [MTC14]. All theoretical and computational results were obtained by myself, but proofread and improved by PB and SG.

Chapter 4

HJ and HT extended the results about $\mathcal{H}_2$-error expressions for multi-agent systems with single-integrator agents from [MTC14] to upper bounds for $\mathcal{H}_2$-error for more general multi-agent systems (Theorem 4.5). I derived the expression for $\mathcal{H}_\infty$-error for special LTI systems in Lemma 4.2. Jointly with HJ and HT, we derived similar upper bounds for $\mathcal{H}_\infty$-error for symmetric multi-agent systems (Theorem 4.12 and Theorem 4.14). Based on an idea from HJ and HT, I derived an upper bound for non-almost equitable partitions in Section 4.2.7.

Section 4.3 is work I extended with help from TI. The problem setting was motivated by AC.

Chapter 5

PB had the idea to extend to results from multi-agent systems to general network systems. All theoretical and computational results were obtained by myself, but proofread and improved by PB and SG.

Chapter 6

The idea and results of this chapter were completely obtained by me.

EHRENERKLÄRUNG

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; verwendete fremde und eigene Quellen sind als solche kenntlich gemacht.

Ich habe insbesondere nicht wissentlich:

- Ergebnisse erfunden oder widersprüchliche Ergebnisse verschwiegen,
- statistische Verfahren absichtlich missbraucht, um Daten in ungerechtfertigter Weise zu interpretieren,
- fremde Ergebnisse oder Veröffentlichungen plagiiert oder verzerrt wiedergegeben.

Mir ist bekannt, dass Verstöße gegen das Urheberrecht Unterlassungs- und Schadenersatzansprüche des Urhebers sowie eine strafrechtliche Ahndung durch die Strafverfolgungsbehörden begründen kann.

Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form als Dissertation eingereicht und ist als Ganzes auch noch nicht veröffentlicht.

Magdeburg, 30.10.2019

Petar Mlinarić