From data and structure to models and controllers

van Waarde, Henk

DOI:
10.33612/diss.144254461

Publication date:
2020

From data and structure to models and controllers

Henk van Waarde

university of
groningen

The research described in this thesis has been carried out at the Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, and at the Engineering and Technology Institute Groningen, University of Groningen, The Netherlands.

The research has been financially supported by the Data Science and Systems Complexity Centre at the University of Groningen.

disc dutch institute
of systems
and control

The research reported in this dissertation is part of the research program of the Dutch Institute of Systems and Control (DISC). The author has successfully completed the educational program of the Graduate School DISC.

From data and structure to models and controllers
Henk van Waarde
PhD thesis University of Groningen

Printed by Ipskamp Printing

Cover painting by Tessa van Waarde

Thesis style based on André Miede's Classic Thesis

rijksuniversiteit
groningen

From data and structure to models and controllers

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

vrijdag 20 november 2020 om 12.45 uur

door

Hendrik Joannes van Waarde

geboren op 25 april 1993
te Groningen

Promotor

Prof. dr. M.K. Camlibel

Copromotor

Dr. P. Tesi

Beoordelingscommissie

Prof. dr. A.J. van der Schaft
Prof. dr. P.M.J. van den Hof
Prof. dr. R. Sepulchre

to my parents,
Aren and Maria

ACKNOWLEDGEMENTS

While I am writing these acknowledgements, the adagio from Mahler's 9th symphony is playing in the background. Mahler, with good sense of drama, writes "ersterbend" in the last bar. As the adagio dies out, I have come to the very final stage of my PhD. My somewhat gloomy state of mind can only mean that I have thoroughly enjoyed the journey that has led to this stage. Pursuing the PhD degree at the University of Groningen has been an amazing experience, and this success is largely due to a number of wonderful individuals. I would thus like to take a moment to thank all those who have educated, motivated and helped me along the way.

Firstly, I would like to express my gratitude towards my supervisors Pietro Tesi and Kanat Camlibel. Pietro, thank you for always being there for me. You moved to Italy during the first year of the project, but this did not stop you from regularly calling me and providing me with your detailed comments and ideas. You have a vast knowledge of the literature and a good nose for research topics. I very much appreciate your kindness towards me and your good sense of humor. I could not have wished for a better daily supervisor.

Kanat, thank you for our discussions and for all the times we were writing on your blackboard. You have a brilliant mind, and showed me the way even in situations where I thought progress was impossible. I appreciate the informal conversations with you about science, politics and virtually anything. "Kanat enters the room with a problem" has become a phenomenon in our office. Thank you also for inviting us to your house and for displaying your cooking skills.

I am grateful to Mehran Mesbahi for the opportunity to work in his research group at the University of Washington for half a year. Mehran, you are a true gentleman. Thank you for all of our discussions and for the nice dinners that we shared. I look forward to further collaborate with you in the future.

I would also like to thank Harry Trentelman, who I view as an "unofficial" supervisor and role model. Harry, thank you for everything that you taught me during my bachelor, master and PhD. Because of you I became interested in systems and control. Thanks to your humor (and occasional complaints) there is never a dull moment in the office.

My special thanks go to the members of the reading committee: Arjan van der Schaft, Paul van den Hof and Rodolphe Sepulchre, for reading the manuscript of this thesis and for providing their comments. I am grateful to the Data Science and Systems Complexity Centre at the University of Groningen for the funding that enabled this PhD project. I would like to thank all the people that were involved in organizing and teaching the DISC courses and making the course program a success every year. I am also thankful to Paul van den Hof, Sean Warnick and Ivan Markovsky for inviting me to give talks.

In addition, I would like to thank Jaap Eising and Mark Jeeninga for being my office mates during the last years and for agreeing to be my paranymphs. Office 389 of the Bernoulliborg has been a place where hard work and long scientific discussions could be paired with the occasional shenanigans like “bordbal" and brave attempts to start a grapefruit orchard. Jaap, thank you for our nice collaboration on data-driven control. Mark, thank you for creating the Bernoulli fish website that enhances our lunch once a week. I would like to thank Jiajia Jia and Brayan Shali for pleasant collaborations on strong structural controllability and pattern matrices. My thanks also go to all the other members of the SCAA group, in particular Bart Besselink, Anne-Men Huijzer, Junjie Jiao, Stephan Trenn and Paul Wijnbergen for many pleasant lunch breaks and game nights.

I am grateful to all of my colleagues from ENTEG. Specifically, I would like to thank Claudio De Persis for our discussions and the detailed comments he provided to the work in Chapter 2. I want to thank Jacquelien Scherpen for pleasant conversations and her encouragement, for example during the ENTEG days. My thanks go to Carlo Cenedese and Michele Cucuzzella for a very memorable trip to Miami and the Florida keys, where the number of Italians we met kept growing exponentially. I also thank Tobias van Damme for our joint supervision of bachelor projects, and Nima Monshizadeh for pleasant conversations at conferences.

During my visit in Seattle, I have gotten to know many great colleagues at the University of Washington. I would like to thank Siavash Alemzadeh, Jingjing Bu, Dillon Foight, Taylor Reynolds, Shahriar Talebi, Yue Yu and many others for interesting discussions about research, and for the enjoyable coffee hours, dinners and beers (I promise to always bring my passport next time!).

Finally, I am grateful to my friends and family. To Harmen Stoppels, for figuring out the bachelor courses together, and for saving numerous group projects with your programming skills. To all of my friends in Groningen, for many wonderful dinners, disc golf competitions, board games and whisky nights. To my parents, for everything they have taught me, for their encouragement and endless support throughout the years. And for the Sunday afternoons, playing long games of Agricola while listening to Diskotabel. To my “new" family, who have made me feel welcome from the start. Last but not least, to Tessa, thank you for your patience when I was distracted, and for your help and support during the last years. I am proud to be your husband, and look forward to what the future will bring us.

Henk van Waarde *Groningen, 15 October 2020*

CONTENTS

1 INTRODUCTION 13
- 1.1 From data to controllers 14
- 1.2 From data to network topology 16
- 1.3 From structure to identifiability 17
- 1.4 From structure to controllability 19
- 1.5 Outline and relations between chapters 21
- 1.6 Publications and origin of the chapters 22
- 1.7 General notation 24

2 WILLEMS' FUNDAMENTAL LEMMA FOR MULTIPLE DATASETS 25
- 2.1 Introduction 25
  - 2.1.1 Notation 26
- 2.2 Willems *et al.*'s fundamental lemma 27
- 2.3 Extension to multiple trajectories 30
- 2.4 Examples 33
  - 2.4.1 Identification with missing data samples 33
  - 2.4.2 Data-driven LQR of an unstable system 35
- 2.5 Conclusions 37

3 DATA INFORMATIVITY FOR ANALYSIS AND CONTROL 39
- 3.1 Introduction 39
- 3.2 Problem formulation 41
- 3.3 Data-driven analysis 44
- 3.4 Control using input and state data 48
  - 3.4.1 Stabilization by state feedback 49
  - 3.4.2 Informativity for linear quadratic regulation 54
  - 3.4.3 From data to LQ gain 59
- 3.5 Control using input and output data 61
  - 3.5.1 Stabilization using input, state and output data 62
  - 3.5.2 Stabilization using input and output data 66
- 3.6 Conclusions and future work 69

4 DATA-BASED PARAMETERIZATIONS OF SUBOPTIMAL CONTROLLERS 71
- 4.1 Introduction 71
- 4.2 Suboptimal control problems 72
  - 4.2.1 The suboptimal LQR problem 72
  - 4.2.2 The $\mathcal{H}_2$ suboptimal control problem 73
- 4.3 Problem formulation 74
- 4.4 Data-driven suboptimal LQR 76
- 4.5 Data-driven $\mathcal{H}_2$ suboptimal control 78
- 4.6 Illustrative example 81
- 4.7 Conclusions 82

5 CONTROL FROM NOISY DATA VIA THE MATRIX S-LEMMA 85

5.1 Introduction 85
5.2 Data-driven stabilization 88
5.2.1 Assumption on the noise 88
5.2.2 Problem formulation 90
5.2.3 Our approach 90
5.3 The matrix-valued S-lemma 92
5.3.1 Recap of the classical S-lemma 92
5.3.2 S-lemma with matrix variables 93
5.4 Data-driven stabilization revisited 98
5.5 Inclusion of performance specifications 101
5.5.1 $\mathcal{H}_2$ control 101
5.5.2 $\mathcal{H}_\infty$ control 104
5.6 Simulation examples 105
5.6.1 Stabilization using bounds on the noise samples 105
5.6.2 $\mathcal{H}_2$ control of a fighter aircraft 106
5.7 Discussion and conclusions 108

6 DATA INFORMATIVITY FOR DISSIPATIVITY 111
6.1 Introduction 111
6.2 Dissipativity of linear systems 112
6.3 Informativity: a vocabulary 113
6.4 Main results 115
6.4.1 A necessary condition for informativity 115
6.4.2 Informativity and noiseless data 117
6.4.3 Informativity and noisy data 118
6.5 Conclusions 121
6.6 Proofs of auxiliary results 121
6.6.1 Proof of Lemma 6.1 121
6.6.2 Proof of Lemma 6.2 122
6.6.3 Proof of Proposition 6.2 124

7 TOPOLOGY IDENTIFICATION OF HETEROGENEOUS NETWORKS 125
7.1 Introduction 125
7.2 Problem formulation 127
7.3 Conditions for topological identifiability 130
7.4 Topology identification 137
7.4.1 Identification of Markov parameters 138
7.4.2 Topology identification 139
7.4.3 Solving the generalized Sylvester equation 141
7.4.4 Robustness analysis 143
7.5 Conclusions 145

8 TOPOLOGY RECONSTRUCTION OF AUTONOMOUS NETWORKS 147
8.1 Introduction 147
8.2 Preliminaries 148
8.2.1 Systems theory 149
8.2.2 Graph theory 149

8.2.3 Consensus dynamics 150
8.3 Problem formulation 150
8.4 Solvability of the reconstruction problem 152
8.4.1 Solvability for general $\mathcal{K}$ 153
8.4.2 Solvability for $\mathcal{K} = \mathcal{Q}$ 155
8.4.3 Solvability for $\mathcal{K} = -\mathcal{L}$ and $\mathcal{K} = \mathcal{A}$ 155
8.5 The network reconstruction problem 157
8.5.1 Network reconstruction for general $\mathcal{K}$ 157
8.5.2 Network reconstruction for $\mathcal{K} = \mathcal{Q}$ 159
8.5.3 Network reconstruction for $\mathcal{K} = -\mathcal{L}$ and $\mathcal{K} = \mathcal{A}$ 161
8.6 Illustrative example 162
8.7 Conclusions 162
9 NETWORK IDENTIFIABILITY AND GRAPH SIMPLIFICATION 165
9.1 Introduction 165
9.2 Preliminaries 167
9.2.1 Rational functions and rational matrices 167
9.2.2 Graph theory 167
9.3 Problem statement and motivation 168
9.4 Rank conditions for identifiability 171
9.5 The graph simplification process 172
9.6 Identifiability and graph simplification 182
9.7 Constrained vertex-disjoint paths 183
9.8 Conclusions 186
9.9 Some proofs 187
9.9.1 Proof of Lemma 9.2 187
9.9.2 Proof of Theorem 9.5 187
10 NETWORK IDENTIFIABILITY OF UNDIRECTED NETWORKS 189
10.1 Introduction 189
10.2 Preliminaries 190
10.2.1 Graph theory 190
10.2.2 Zero forcing sets 191
10.2.3 Dynamical networks 191
10.2.4 Network identifiability 192
10.3 Problem statement 193
10.4 Main results 193
10.5 Higher-order node dynamics 198
10.6 Conclusions 199
11 A UNIFYING FRAMEWORK FOR STRUCTURAL CONTROLLABILITY 201
11.1 Introduction 201
11.2 Preliminaries 204
11.3 Problem formulation 204
11.4 Main results 205
11.5 Discussion of existing results 210
11.5.1 Graph-theoretic conditions 211

11.5.2 Algebraic conditions . . . 214

11.6 Proofs . . . 215

11.6.1 Proof of Theorem 11.1 . . . 215

11.6.2 Proof of Theorem 11.2 . . . 217

11.6.3 Proof of Theorem 11.3 . . . 218

11.6.4 Proof of Theorem 11.4 . . . 219

11.6.5 Proof of Lemma 11.1 . . . 219

11.7 Conclusions . . . 219

12 PROPERTIES OF PATTERN MATRICES 221

12.1 Introduction . . . 221

12.2 Pattern matrices . . . 222

12.3 Applications . . . 226

12.3.1 Controllability of linear DAE's . . . 227

12.3.2 Input-state observability . . . 229

12.3.3 Output controllability . . . 230

12.4 Conclusion . . . 231

13 A DISTANCE-BASED APPROACH TO TARGET CONTROLLABILITY 233

13.1 Introduction . . . 233

13.2 Preliminaries . . . 235

13.2.1 Qualitative class and pattern class . . . 235

13.2.2 Subclass of distance-information preserving matrices . . . 236

13.2.3 Zero forcing sets . . . 236

13.2.4 Output controllability of linear systems . . . 237

13.2.5 Targeted controllability of systems defined on graphs . . . 238

13.3 Problem statement . . . 239

13.4 Main results . . . 240

13.4.1 Sufficient condition for targeted controllability . . . 240

13.4.2 Sufficient richness of $Q_d(G)$ . . . 245

13.4.3 Necessary condition for targeted controllability . . . 246

13.4.4 Leader selection algorithm . . . 248

13.5 Illustrative example . . . 252

13.6 Conclusions . . . 255

14 CONCLUSIONS 257

14.1 Contributions . . . 257

14.2 Outlook . . . 259

BIBLIOGRAPHY 261

SUMMARY 281

SAMENVATTING 283

1 INTRODUCTION

Mathematical models are ubiquitous in science and engineering. They lie at the heart of most physical theories and play a fundamental role in systems biology and quantitative finance. A mathematical model is a simplified (quantitative) description of a real-world system or process. Such descriptions help to understand the most prominent features of the modeled phenomenon, and to make predictions about its future behavior. Although the ability of mathematics to describe reality may be puzzling to the philosophically inclined[1], many of us make use of models without much thought; for instance, when relying on the weather forecast or when browsing a playlist of recommended songs.

Clearly, the usefulness of a model is dependent on its ability to portray reality in a good way. The meaning of "good", in turn, depends on the intended use of the model. For example, crude mathematical models may not be able to make fully accurate predictions of a modeled system, but can still be useful for shaping the system's behavior by means of control [64].

In the field of systems and control, mathematical models are typically dynamical systems that are intended for analysis and control of the modeled process. Control theory has seen several remarkable developments, of which we mention the celebrated Kalman filter, Lyapunov stability analysis, and $\mathcal{H}_\infty$ optimal control. All of these techniques are based on mathematical models of the considered system. However, obtaining a suitable model is far from a simple task in practice. In fact, it is widely recognized that obtaining a process model is the single most time-consuming task in the application of model-based control [84, 156].

The difficulty of obtaining good models has multiple reasons. One of these is that modern control systems are becoming increasingly complex, which complicates (physical) modeling from first principles. In some cases, there is no clear physics underlying the system, for example, in the modeling of stock prices. Even in scenarios in which physical modeling is possible, the obtained model may be too complex for its intended purpose.

Systems complexity also manifests itself in the sheer size and interconnected nature of modern systems. There is a general trend in science and technology to study and design systems comprised of multiple interconnected subsystems. Such *networks* appear in an impressively wide variety of domains, ranging from social dynamics and neural networks to power grids and robotic systems. A trait of these systems is that their collective behavior is not only determined by the behavior of the individual subsystems, but is also influenced by the way these systems are coupled. Networks bring their own unique set of modeling challenges. One of such challenges stems from the fact that the network structure, or *topology*,

[1] See "The Unreasonable Effectiveness of Mathematics in the Natural Sciences" by E. P. Wigner [239].

is often unavailable; this is for instance the case in neural networks. Even in situations where we know the network topology, using this prior information effectively is a must, and a major challenge by itself.

In situations where a (precise) mathematical model is unavailable, this missing information about the system must be accounted for by something else. In this thesis, knowledge of mathematical models is substituted by two other ingredients, namely **data** and **structure**. By data, we mean measurements of a dynamical system, typically of its inputs and outputs. By structure, we generally refer to a zero/nonzero structure on the system parameters, for example induced by a network topology. The purpose of this thesis is to perform modeling, analysis and control of dynamical systems using data and structure. In particular, we will focus on four problems, namely data-driven control, topology identification, network identifiability analysis, and structural controllability analysis.

1.1 FROM DATA TO CONTROLLERS

Data-driven control refers to the problem of constructing control laws for an unknown dynamical system from data. The problem can be approached via different angles, for example using combined modeling (system identification) and model-based control, or by computing control laws from data without the intermediate modeling step. We will contribute to the second category of methods, aiming at analysis and control design directly from data.

The literature on data-driven control is expanding rapidly. We mention contributions to data-driven optimal control [1,4,10,50,56,62,71,150,162,193,197,222], PID control [59,99], predictive control [6,55,83,90,183], and nonlinear control [21,44,75,203,204]. Some of these techniques are iterative in nature: the controller is updated online when new data are presented. Examples of this include policy iteration methods [23] and iterative feedback tuning [85]. Other methods are one-shot in the sense that the controller is constructed offline from a batch of data. We mention, for instance, virtual reference feedback tuning [26] and methods based on Willems' fundamental lemma [241] (see also [220]). The latter line of work has been quite fruitful, with contributions ranging from output matching [125] and control by interconnection [132], to data-enabled predictive control [40] and data-based closed-loop system parameterizations [17,47] amenable for control. Additional recent research directions include data-driven control of networks [5,9] and the interplay between data-guided control and model reduction [140].

In addition to control problems, also *analysis* problems have been studied within a data-based framework. The authors of [164] analyze the stability of an input/output system using time series data. The papers [111,155,232,248] deal with data-based controllability and observability analysis. Moreover, the problem of verifying dissipativity on the basis of measured system trajectories has been studied in several recent contributions, see [16,100,101,131,178,179].

In this thesis, we will approach data-driven analysis and control from the angle of *data informativity*. Essentially, this means that we want to understand when the data contain sufficient information for the analysis of system properties and the design of controllers of the (unknown) system. Informative data are essential for control: without such data it is impossible to guarantee stability and performance of the system in interconnection with the data-driven controller. Although there are several methods for data-driven control with guaranteed stability and performance (c.f. [47, 125, 132]), a general definition of informative data and a characterization for different analysis and control problems is still largely missing. Therefore, in Chapter 3 we will define a general notion of data informativity for data-driven analysis and control. The basic idea is as follows. We will assume that the true data-generating system is contained in a known model class, for example a class of linear time-invariant systems. The measured data give rise to a subset of systems within the model class that all could have generated the data; a set that is reminiscent of the feasible systems set in set membership identification [138]. Roughly speaking, the data are then called *informative* if this set of systems explaining the data is “sufficiently small", so that we can analyze and control the true system using the given data.

In Chapters 3, 4, 5 and 6, we will put forward a fairly complete theory for data-guided analysis and control for model classes of linear time-invariant systems. We will consider both exact data (Chapters 3 and 4) and noisy data (Chapters 5 and 6). In Chapter 3 we focus on stability, stabilizability and controllability analysis, as well as stabilization and optimal control. For each of the problems, we provide necessary and sufficient conditions for data informativity, and for the control problems we also establish data-driven control design methods. In Chapter 4 we continue to study data-driven control, with a focus towards suboptimal linear quadratic regulation (LQR) and $\mathcal{H}_2$ control. Thereafter, we switch to a setting of noisy data in Chapter 5, where we study quadratic stabilization, $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control. Here, we will make the assumption that the noise has bounded energy on a finite time interval. Finally, in Chapter 6 we establish methods to determine dissipativity of linear systems from measured data, both in an exact and a noisy data setting.

Our results lead to multiple interesting conclusions. In the noiseless setup of Chapters 3 and 4 we conclude that the data informativity conditions for stabilization and suboptimal control are generally *weaker* than those for system identification. The interpretation is that it is generally easier to learn a stabilizing controller than it is to learn a system model from data. However, for the LQR problem, we show that the data informativity conditions are practically the same as for identification. The conditions for data informativity are thus *dependent* on the control problem. In the noisy setup of Chapters 5 and 6 our analysis also leads to new types of robust control results that are interesting in their own right. For example, in Chapter 5 we derive a generalization of the classical S-lemma [243] to matrix variables. In addition, in Chapter 6 we provide a variant of the dualization lemma to prove the equivalence of different noise models and to establish data-driven tests for dissipativity.

Some parts of this thesis (namely, the motivation of Chapter 3 and the identification approach of Chapter 7) are based on the notion of *persistency of excitation*, which is discussed in detail in Chapter 2. Chapter 2 studies Willems *et al.*'s fundamental lemma [241]. This result asserts that under controllability and persistency of excitation conditions, all trajectories of a linear system are parameterized in terms of a single given one. As we will see, the result has implications for both system identification and data-driven control. In fact, the conditions of Willems' lemma can be interpreted as *experiment design* conditions, enabling the generation of informative data for modeling and control in a noiseless data setup. The main purpose of Chapter 2 is to establish a generalization of Willems' lemma to the situation in which *multiple* trajectories are given instead of a single one. This result aids the identification from data sets with missing samples, and is also shown to be beneficial in the data-guided control of unstable systems.

1.2 FROM DATA TO NETWORK TOPOLOGY

Topology identification entails the problem of identifying the structure of a networked system from data. The problem is not only important in the systems and control community, but has also received attention in physics [206] and biology [213]. The interest in topology identification is motivated by the fact that many real-world networks have a network topology that is either completely unavailable or uncertain. Some examples include neural networks [213], genetic networks [97], and networks of interconnected stock prices [129]. In the case that one is only interested in *control* of the networked system, we envision the possibility of applying direct methods in the spirit of Chapters 3, 4, 5 and 6. However, there are many situations in which one is interested in the network topology *an sich*, rather than control of the networked system. For instance, in the examples mentioned above, the main problem is to understand the different interactions between subsystems. The network topology also plays a fundamental role in the success of distributed algorithms [158], and can be used to make predictions about their rate of convergence. Therefore, we consider the problem of topology identification in Chapters 7 and 8.

There are several existing methods for topology reconstruction from data. The paper [70] studies dynamical structure function reconstruction, see also [246]. A node-knockout scheme for topology identification was introduced in [153] and further investigated in [202]. Moreover, the paper [184] studies topology identification using compressed sensing, while [130] considers network reconstruction using Wiener filtering. A distributed algorithm for network reconstruction has also been studied [145]. The paper [190] studies topology identification using power spectral analysis. A Bayesian approach to the network identification problem was investigated in [32]. The network topology was inferred from multiple independent observations of consensus dynamics in [189]. The paper [41] studies topology identification via subspace methods. There are also several results for topology reconstruction of nonlinear systems, see e.g., [192, 206, 231] albeit in this case few guarantees on the accuracy of identification can be given.

Most existing work on topology identification emphasizes the role of the network topology by considering relatively simple node dynamics. For example, networks of single integrators have been studied in [79, 145, 153, 226]. In addition, the papers [202] and [190] consider homogeneous networks comprised of identical single-input single-output systems.

The goal of Chapter 7 is to provide a comprehensive treatment of topology identification for linear multi-input multi-output (MIMO) heterogeneous networks. We will consider both the problem of *identifiability*, as well as *reconstruction* of the network topology. The study of identifiability of the network topology deals with the question whether there exists a data set from which the topology can be uniquely identified. Identifiability of the topology is hence a property of the node systems and the network graph, and is *independent* of any data. Topological identifiability is an important property. Indeed, if it is not satisfied, then it is impossible to uniquely identify the network topology, regardless of the amount and richness of the data. After studying topological identifiability, we will turn our attention towards reconstruction, which involves the development of algorithms that identify the network graph from data.

Our identifiability results recover and generalize a result for the special case of networks of single integrators [163, 226]. We will also see that homogeneous networks of single-input single-output systems have quite special identifiability properties that do not extend to the general case of heterogeneous networks. Our topology identification scheme makes use of Willems' lemma (Chapter 2). Willems' lemma can be leveraged to identify the network's Markov parameters. Then, the idea is to reconstruct the network interconnection matrix by solving a *generalized Sylvester equation* involving the Markov parameters. We prove that the network topology can be uniquely reconstructed in this way, under the assumptions of topological identifiability and persistently exciting inputs.

In Chapter 8 we investigate a more specific network setup, where the dynamics of each node is a single integrator, and the network is *autonomous*. In this case, excitation has to be secured through the initial conditions of the network. The more specialized setting of Chapter 8 allows us to come up with more specific reconstruction methods, in terms of Lyapunov equations.

1.3 FROM STRUCTURE TO IDENTIFIABILITY

As we have explained in the previous section, there are several examples of networks with unknown topology. Nevertheless, the assumption that the network topology is *known* is reasonable for other systems, in particular for engineering systems such as water distribution networks.

In the case that the network structure is known, we need new techniques to exploit this prior information. In Chapters 9 and 10 we will utilize the known graph structure of the network in order to characterize a notion of global identifiability. In the setting of these chapters, identifiability is a property of the network model set that guarantees that a unique model can be identified, given informative data and prior knowledge of the network topology. There are multiple reasons why understanding identifiability from a graph-theoretic perspective is interesting. First, conditions based on the network topology are desirable since they give insight on the types of network structures that allow identification. Secondly, as demonstrated in [31], graph-theoretic conditions may aid in the *selection* of excited and measured nodes guaranteeing identifiability.

Identifiability of dynamical networks is an active research area, see e.g., [3, 80, 81, 152, 223, 224, 233–235] and the references therein. However, considerably less results are available on the connections between identifiability and graph structure [81, 152, 233]. In [152], sufficient graph-theoretic conditions for identifiability have been presented for a class of consensus networks. In [81], graph-theoretic conditions have been established for *generic* identifiability. That is, conditions were given under which transfer functions in the network can be identified for *"almost all"* network matrices associated with the graph. The authors of [81] show that generic identifiability is equivalent to the existence of certain *vertex-disjoint paths*, which yields elegant conditions for generic identifiability. Similar results were also presented in a "dual" setup in [233].

Inspired by the work on generic identifiability [81, 233], in Chapter 9 we are interested in graph-theoretic conditions for a stronger notion, namely identifiability *for all* network matrices associated with the graph. This notion is referred to as *global identifiability* of the model set. We will study a network model, introduced in [214], where relations between nodes are modeled by proper transfer functions.

Our goal of studying global identifiability is motivated by the fact that generic identifiability provides an *indication* of identifiability, rather than *guarantees*. Indeed, although generic identifiability guarantees identifiability for almost all network matrices, there are meaningful examples of network matrices that are not contained in this set of almost all systems. As a consequence, a situation may arise in which the unknown system under consideration is not identifiable, even though the conditions for generic identifiability are satisfied. For an example of such a situation, we refer to Section 9.3. On the other hand, if the conditions derived in Chapter 9 are satisfied, then it is guaranteed that the network is identifiable *for all* network matrices associated with the graph. It turns out that in order to characterize global identifiability, we need a new graph-theoretic concept called the *graph simplification process*. We will provide necessary and sufficient conditions for identifiability in terms of this concept in Chapter 9. An interesting outcome of our results is that identifiability can often be achieved with relatively few measured nodes. This shows the effectiveness of using prior knowledge of the topology. In comparison, in case that the topology is unknown it can be shown that all network nodes (except for one) need to be measured.

In Chapter 10 we are interested in a similar notion of global identifiability, but for a different class of undirected networks described by state-space systems. In this case we provide sufficient conditions for identifiability in terms of so-called *zero forcing sets*. Zero forcing sets have been studied in relation to structural controllability [142], but the connection to identifiability has not been studied before. The results of Chapter 10 reveal that in the more specialized (undirected) setup, identifiability can be achieved not only with a limited number of measured nodes, but also with a limited number of *excited* nodes.

1.4 FROM STRUCTURE TO CONTROLLABILITY

Global identifiability, as studied in Chapters 9 and 10, is a *structural property*: it can be completely characterized in terms of the graph structure and locations of excited/measured nodes. Another example of a structural property (and arguably, the prime example) is *structural controllability*. Structural controllability has a rich history that started with the classical paper by Lin [110]. The concept involves a pair of matrices (A, B), where each entry of these matrices is either a fixed zero or a free parameter. (Weak) structural controllability then requires almost all realizations of (A, B) to be controllable. That is, for almost all parameter settings of the free entries of A and B, the resulting numerical pair of matrices is controllable in the classical sense by Kalman. Lin provided a graph-theoretic condition under which (A, B) is weakly structurally controllable in the single-input case. The extension to multiple inputs was also studied in [67] and [194].

Later on, Mayeda and Yamada introduced the notion of *strong* structural controllability [133]. They considered a *zero/nonzero* structure on the matrices A and B, meaning that each entry of these matrices is either a fixed zero or a *nonzero* free parameter. Strong structural controllability then requires *all* numerical realizations of (A, B) to be controllable.

There has been a renewed surge of interest in structural controllability that was initiated by the publication of the Nature paper [113] studying structural controllability of networks. Several contributions followed, both for the weak [42, 134, 147] and strong [29, 142, 207] variants of controllability. In the context of networks, the structure of the matrix A results from a given graph structure. In addition, the matrix B often has a specific structure reflecting the fact that each input of the network directly affects only one network node (called a leader). Structural controllability of networks is therefore not essentially different from “classical" structural controllability studied by Lin; the differences mostly lie in the interpretation of A and the special structure of B. The contributions to controllability of networks therefore mainly involved new graph-theoretic characterizations of controllability in terms of maximal matchings [113], constrained matchings [29] and zero forcing sets [142]. These new graph-theoretic conditions were also shown to be amenable from a design point of view, in the sense that they enable the selection of a set of leaders guaranteeing (strong) structural controllability.

Nowadays, structural controllability is still an active research area. Some new research lines involve structural output controllability [39, 63, 141, 219], structural controllability with dependencies amongst entries in A and B [94, 112, 134] and the study of controllability of networks with higher order node dynamics [35, 36].

We recall that Mayeda and Yamada [133] studied strong structural controllability for (A, B) pairs having *zero/nonzero* structure. This basic assumption is predominant in the literature. Nonetheless, as we demonstrate in Chapter 11, this assumption is not always realistic in the sense that there are many examples in which we do not know whether an entry of the system matrices is zero or nonzero. Therefore, in Chapter 11 we extend the zero/nonzero structure to a more general zero/nonzero/arbitrary structure, and we study strong structural controllability in this framework. We will provide both algebraic and graph-theoretic conditions for strong structural controllability. We also find that seemingly incomparable results of [207] and [142] follow from our main results, which reveals an overarching theory. For this reason, Chapter 11 can be seen as a unifying approach to strong structural controllability of linear time-invariant systems.

We continue our study of the zero/nonzero/arbitrary structure in Chapter 12. In this chapter, we take a closer look at properties of so-called pattern matrices. A pattern matrix is an array of symbols, where each symbol captures some structural information. The pattern matrices that we consider thus contain three different symbols: 0 (zero), $*$ (nonzero) and $?$ (arbitrary). We will define notions of addition and multiplication of such pattern matrices, and study the properties of pattern matrices that are either the sum or the product of two pattern matrices. Subsequently, we will apply these results to assess strong structural input-state observability, output controllability, and controllability of linear differential algebraic equations.

We follow up in Chapter 13 by studying strong structural output controllability in a network setting. Here, the output of the network consists of the states of a subset of network nodes, called target nodes. The goal is to understand under which conditions the target nodes can be controlled by applying inputs to the leader nodes. Due to the specific form of the network output, output controllability is often referred to as *targeted controllability* in this context. Strong structural targeted controllability has been considered before in the paper [141]. We will follow up on this work by studying targeted controllability for a subclass of A-matrices, called distance-information preserving matrices. For this subclass, we are able to come up with more powerful sufficient conditions for strong structural targeted controllability. We also provide necessary conditions for targeted controllability, as well as a strategy for the selection of leader nodes.

1.5 OUTLINE AND RELATIONS BETWEEN CHAPTERS

To summarize, Chapter 2 studies Willems' fundamental lemma, and Chapters 3, 4, 5 and 6 treat data-driven analysis and control. Topology identification is the main topic of Chapters 7 and 8, while Chapters 9 and 10 consider network identifiability from a graph-theoretic perspective. We study strong structural controllability, input-state observability and output controllability in Chapters 11, 12 and 13. Finally, our conclusions are provided in Chapter 14. A graph of relations between the chapters of this thesis is displayed in Figure 1.1.

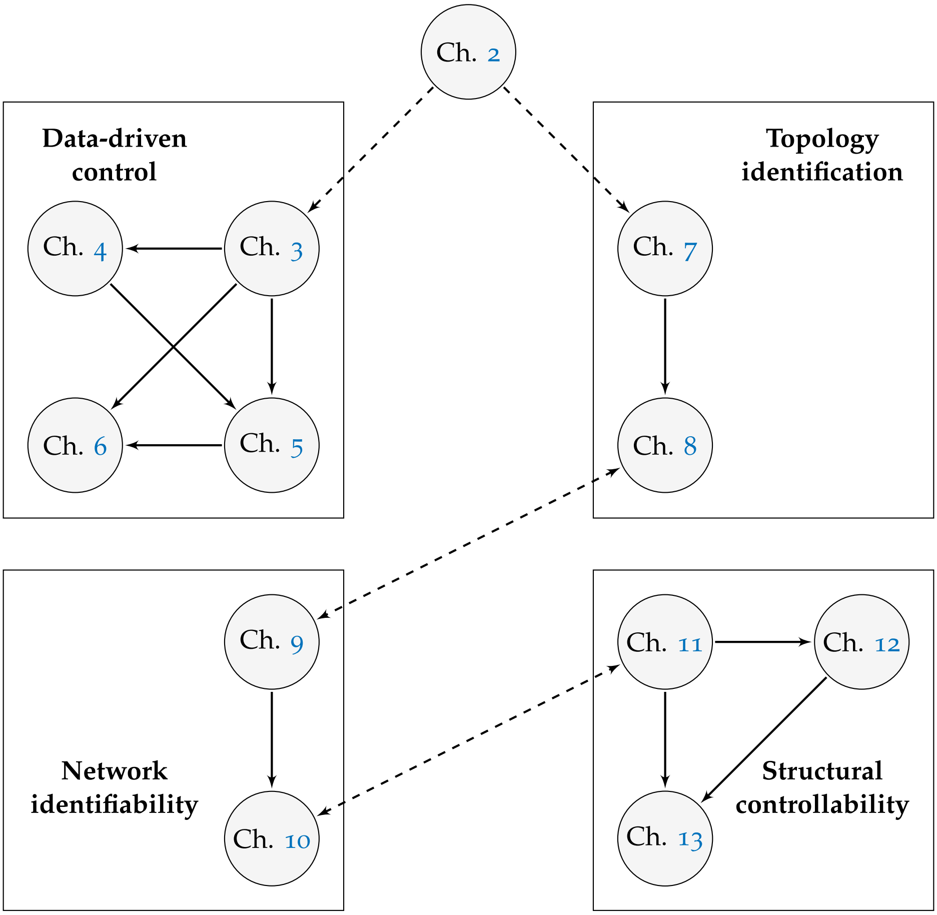

Figure 1.1: Graph of relations between chapters. Solid links represent strong relations, e.g., Chapter 5 directly extends results from Chapters 3 and 4 to noisy data. Dashed links indicate weaker relations, e.g., Willems' lemma (Chapter 2) is applied in the topology identification approach of Chapter 7.

1.6 PUBLICATIONS AND ORIGIN OF THE CHAPTERS

Journal publications

1. H.J. van Waarde, C. De Persis, M.K. Camlibel, and P. Tesi, “Willems’ Fundamental Lemma for State-space Systems and its Extension to Multiple Datasets", *in IEEE Control Systems Letters*, vol. 4, no. 3, pp. 602–607, July 2020 (**Ch. 2**).

2. H.J. van Waarde, J. Eising, H.L. Trentelman, and M.K. Camlibel, “Data informativity: a new perspective on data-driven analysis and control", *to appear in IEEE Transactions on Automatic Control*, 2020 (**Ch. 3**).

3. H.J. van Waarde, P. Tesi, and M.K. Camlibel, “Topology Identification of Heterogeneous Networks: Identifiability and Reconstruction", *accepted for publication in Automatica*, 2020 (**Ch. 7**).

4. H.J. van Waarde, P. Tesi, and M.K. Camlibel, “Topology Reconstruction of Dynamical Networks via Constrained Lyapunov Equations" *in IEEE Transactions on Automatic Control*, vol. 64, no. 10, pp. 4300–4306, Oct. 2019 (**Ch. 8**).

5. H.J. van Waarde, P. Tesi, and M.K. Camlibel, “Necessary and Sufficient Topological Conditions for Identifiability of Dynamical Networks", *to appear in IEEE Transactions on Automatic Control*, 2020 (**Ch. 9**).

6. H.J. van Waarde, P. Tesi, and M.K. Camlibel, “Identifiability of Undirected Dynamical Networks: A Graph-Theoretic Approach", *in IEEE Control Systems Letters*, vol. 2, no. 4, pp. 683–688, Oct. 2018 (**Ch. 10**).

7. J. Jia, H.J. van Waarde, H.L. Trentelman, and M.K. Camlibel, “A Unifying Framework for Strong Structural Controllability", *to appear in IEEE Transactions on Automatic Control*, 2020 (**Ch. 11**).

8. H.J. van Waarde, M.K. Camlibel, and H.L. Trentelman, “A Distance-based Approach to Strong Target Control of Dynamical Networks", *in IEEE Transactions on Automatic Control*, vol. 62, no. 12, pp. 6266–6277, Dec. 2017 (**Ch. 13**).

9. H.J. van Waarde, M.K. Camlibel, and H.L. Trentelman, “Comments on "On the Necessity of Diffusive Couplings in Linear Synchronization Problems With Quadratic Cost"", *in IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 3099–3101, June 2017.

Conference publications

1. H.J. van Waarde, M. Mesbahi, “Data-driven parameterizations of suboptimal LQR and H2 controllers", *accepted for publication in Proceedings of the IFAC World Congress*, Berlin, Germany, 2020 (**Ch.** 4).
2. H.J. van Waarde, P. Tesi, and M.K. Camlibel, “Topology Identification of Heterogeneous Networks of Linear Systems", *in Proceedings of the IEEE Conference on Decision and Control*, Nice, France, pp. 5513–5518, 2019 (**Ch.** 7).
3. H.J. van Waarde, P. Tesi, and M.K. Camlibel, “Topological Conditions for Identifiability of Dynamical Networks with Partial Node Measurements", *in Proceedings of the IFAC Workshop on Distributed Estimation and Control in Networked Systems*, Groningen, The Netherlands, pp. 319–324, 2018 (**Ch.** 9).
4. F. Veldman-de Roo, A. Tejada, H.J. van Waarde, and H.L. Trentelman, “Towards observer-based fault detection and isolation for branched water distribution networks without cycles", *in Proceedings of the European Control Conference*, Linz, Austria, pp. 3280–3285, 2015.

Peer-reviewed extended abstracts

1. H.J. van Waarde, J. Eising, H.L. Trentelman, and M.K. Camlibel, “An exciting, but not persistently exciting perspective on data-driven analysis and control", *accepted for publication in International Symposium on Mathematical Theory of Networks and Systems*, Cambridge, United Kingdom, 2021 (**Ch.** 3).
2. H.J. van Waarde, P. Tesi, and M.K. Camlibel, “Topology Reconstruction of Dynamical Networks via Constrained Lyapunov Equations", *in Proceedings of the International Conference on Complex Networks and Their Applications*, Lyon, France, pp. 187–189, 2017 (**Ch.** 8).

Book chapters

1. H.J. van Waarde, N. Monshizadeh, H.L. Trentelman, and M.K. Camlibel, “Strong structural controllability and zero forcing", *in Structural Methods in the Study of Complex Systems*, ser. Lecture Notes in Control and Information Sciences, E. Zattoni, A. Perdon, and G. Conte, Eds., Springer, 2019 (**Ch.** 13).

Preprints

1. H.J. van Waarde, M.K. Camlibel, and M. Mesbahi, “From noisy data to feedback controllers: non-conservative design via a matrix S-lemma", *submitted to IEEE Transactions on Automatic Control*, 2020 (**Ch.** 5).
2. M.K. Camlibel, P. Rapisarda, H.J. van Waarde, and H.L. Trentelman, “Informativity for data-driven dissipativity", *under preparation*, 2020 (**Ch.** 6).

3. B. Shali, H.J. van Waarde, H.L. Trentelman, and M.K. Camlibel, “Properties of pattern matrices with applications to structured systems", *under preparation*, 2020 (**Ch. 12**).

1.7 GENERAL NOTATION

In this section we define some notation that we will use throughout the thesis. More specific notation that is used in one or only a few chapters will be defined within the chapters themselves.

Sets

We denote the set of natural, real, and complex numbers by $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{C}$ respectively. Let $\mathbb{R}^n$ ($\mathbb{C}^n$) denote the linear space of vectors with n real (complex) components. Moreover, the set of real (complex) $m \times n$ matrices is denoted by $\mathbb{R}^{m\times n}$ ($\mathbb{C}^{m\times n}$).

The image of a matrix $A \in \mathbb{R}^{n\times m}$ is denoted by $\operatorname{im} A$ and defined as

$$\operatorname{im} A := \{Av \in \mathbb{R}^n \mid v \in \mathbb{R}^m\}.$$

The kernel of A is denoted by $\ker A$ and defined as

$$\ker A := \{v \in \mathbb{R}^m \mid Av = 0\}.$$

The left kernel of A is defined as $\ker A^\top$, where $A^\top$ denotes the transpose of A. The spectrum of a square matrix $A \in \mathbb{C}^{n\times n}$ is the set of its eigenvalues and denoted by $\sigma(A)$. The cardinality of a set S is denoted by $|S|$.

Matrices and vectors

The $n \times n$ identity matrix is denoted by I_n. The zero vector of dimension n is denoted by 0_n, and the zero matrix of dimension $n \times m$ is denoted by $0_{n\times m}$. We denote the n-dimensional vector of ones by $\mathbb{1}_n$. If the dimensions of I_n, 0_n, $0_{n\times m}$ and $\mathbb{1}_n$ are clear from the context, we simply write I, 0 and $\mathbb{1}$.

The real and imaginary parts of a vector $v \in \mathbb{C}^n$ are denoted by $\operatorname{Re} v$ and $\operatorname{Im} v$, respectively. Its conjugate transpose is denoted by v^*. A matrix $A \in \mathbb{R}^{n\times n}$ is called positive definite, denoted by $A > 0$, if $v^\top Av > 0$ for all nonzero $v \in \mathbb{R}^n$. It is called positive semidefinite, denoted by $A \geqslant 0$, if $v^\top Av \geqslant 0$ for all $v \in \mathbb{R}^n$. Negative definite and negative semidefinite matrices are defined analogously, and denoted by $A < 0$ and $A \leqslant 0$, respectively. The trace $\operatorname{tr} A$ of a square matrix A is the sum of its diagonal entries. We denote the Kronecker product of $A \in \mathbb{C}^{n\times m}$ and $B \in \mathbb{C}^{p\times q}$ by $A \otimes B \in \mathbb{C}^{np\times mq}$. Finally, the concatenation of matrices $A_1, A_2, \dots, A_k$ of compatible dimensions is defined as

$$\operatorname{col}(A_1, A_2, \dots, A_k) := \begin{pmatrix} A_1^\top & A_2^\top & \cdots & A_k^\top \end{pmatrix}^\top.$$

2 WILLEMS' FUNDAMENTAL LEMMA FOR MULTIPLE DATASETS

In this chapter we revisit a result by Willems, Rapisarda, Markovsky and De Moor. The result is often referred to as the fundamental lemma. Essentially, this lemma gives a condition under which a measured trajectory of a linear system can be used to parameterize *all* trajectories that the system can produce. The measured trajectory thereby implicitly serves as a non-parametric system model. Here, we provide an alternative proof of the fundamental lemma. We will also prove an extension of the lemma that applies to the scenario in which *multiple* system trajectories are measured.

2.1 INTRODUCTION

In the seminal work by Willems and coauthors [241], it was shown that a single, sufficiently exciting trajectory of a linear system can be used to parameterize *all* trajectories that the system can produce. This result has later been named the *fundamental lemma* [125, 128], and plays an important role in the learning and control of dynamical systems on the basis of measured data.

An immediate consequence of the fundamental lemma is that a sufficiently long, persistently exciting trajectory captures the entire behavior of the data-generating system, thus allowing successful identification of a system model using, e.g., subspace methods [227]. The lemma also enables data-driven simulation [125], which involves the computation of the system's response to a given reference input. In addition, Willems' lemma is instrumental in the design of controllers from data. The result has been applied to tackle several control problems, ranging from output matching [125] to control by interconnection [132], predictive control [18,40,90], optimal and robust control [47], linear quadratic regulation [47,125,181] as well as set-invariance control [20].

All of the above examples show the value of the fundamental lemma in modeling, simulation and control using a *single* measured system trajectory. Nonetheless, there are many scenarios in which *multiple* system trajectories are measured instead of a single one. For example, performing multiple short experiments becomes desirable when the data-generating system has unstable dynamics. Also, as pointed out in [88], a single system trajectory collected during normal operations may be too poorly excited to reveal the system dynamics. In contrast, multiple archival data may *collectively* provide a well-excited experiment. Another situation is when a single trajectory is measured but some of the samples are corrupted or missing. In this case, we have access to multiple system trajectories consisting of the remaining, uncorrupted, data samples. System identification from multiple experiments [120, 123] and from data with missing samples [121, 122, 126] has been studied. However, a proof of Willems' lemma for multiple trajectories is still missing. Therefore, in this chapter we aim at extending Willems' fundamental lemma to the case where multiple trajectories, possibly of different lengths, are given instead of a single one.

Originally, the fundamental lemma was formulated and proven in a behavioral context. The starting point in this chapter, however, is a reformulation of the lemma in terms of state-space systems. Such a version of Willems' fundamental lemma has appeared before in [47, Lem. 2] and [16, Thm. 3] but no proof of the statement was given in this context. Our first contribution is to provide a complete and self-contained proof of the lemma for state-space systems. Strictly speaking, such an alternative proof is not necessary since the original proof of [241] applies to state-space systems as a special case. Nonetheless, we believe that our proof can be of interest to researchers who want to apply Willems' lemma to state-space systems. In fact, the proof is *elementary* in the sense that it only makes use of basic concepts such as the Cayley-Hamilton theorem and Kalman controllability test. The proof is also direct, and in contrast to [241] does not rely on a contradiction argument.

Our second contribution involves the extension of the fundamental lemma to the case of multiple trajectories. To this end, we first introduce a notion of *collective* persistency of excitation. Then, analogous to Willems' lemma, we show that a finite number of given trajectories can be used to parameterize all trajectories of the system, assuming that collective persistency of excitation holds. We will illustrate this result by two examples. First, we will show that the extended fundamental lemma enables the identification of linear systems from data sets with missing samples. Next, we will show how the result can be used to compute controllers of unstable systems from multiple short system trajectories, even when this is problematic from a single long trajectory.

The chapter is organized as follows: in Section 2.2 we formulate and prove Willems' fundamental lemma. Section 2.3 extends the lemma to multiple trajectories. In Section 2.4 we provide applications of this result. Finally, Section 2.5 contains our conclusions.

2.1.1 Notation

Consider a signal $f : \mathbb{Z} \to \mathbb{R}^{\bullet}$ and let $i, j \in \mathbb{Z}$ be integers such that $i \leqslant j$. We denote by $f_{[i,j]}$ the restriction of f to the interval $[i, j]$, that is,

$$f_{[i,j]} := \begin{bmatrix} f(i)^\top & f(i+1)^\top & \cdots & f(j)^\top \end{bmatrix}^\top.$$

With slight abuse of notation, we will also use the notation $f_{[i,j]}$ to refer to the sequence $f(i), f(i+1), \ldots, f(j)$. Let k be a positive integer such that $k \leqslant j-i+1$ and define the *Hankel matrix* of depth k, associated with $f_{[i,j]}$, as

$$\mathcal{H}_k(f_{[i,j]}) := \begin{bmatrix} f(i) & f(i+1) & \cdots & f(j-k+1) \\ f(i+1) & f(i+2) & \cdots & f(j-k+2) \\ \vdots & \vdots & & \vdots \\ f(i+k-1) & f(i+k) & \cdots & f(j) \end{bmatrix}.$$

Note that the subscript k refers to the number of block rows of the Hankel matrix.

Definition 2.1. The sequence $f_{[i,j]}$ is said to be *persistently exciting of order k* if $\mathcal{H}_k(f_{[i,j]})$ has full row rank.

2.2 WILLEMS *ET AL.*'S FUNDAMENTAL LEMMA

In this section we explain the fundamental lemma [241] in a state-space setting. Our goal is to provide a simple and self-contained proof of the result within this context. Consider the linear time-invariant (LTI) system

$$\begin{aligned} \mathbf{x}(t+1) &= A\mathbf{x}(t) + B\mathbf{u}(t) && \text{(2.1a)} \\ \mathbf{y}(t) &= C\mathbf{x}(t) + D\mathbf{u}(t), && \text{(2.1b)} \end{aligned}$$

where $\mathbf{x} \in \mathbb{R}^n$ denotes the state, $\mathbf{u} \in \mathbb{R}^m$ is the input and $\mathbf{y} \in \mathbb{R}^p$ is the output. Let $(u_{[0,T-1]}, y_{[0,T-1]})$ be a given input/output trajectory[1] of (2.1). We consider the Hankel matrices of these inputs and outputs, given by:

$$\begin{bmatrix} \mathcal{H}_L(u_{[0,T-1]}) \\ \mathcal{H}_L(y_{[0,T-1]}) \end{bmatrix} = \begin{bmatrix} u(0) & u(1) & \cdots & u(T-L) \\ \vdots & \vdots & & \vdots \\ u(L-1) & u(L) & \cdots & u(T-1) \\ y(0) & y(1) & \cdots & y(T-L) \\ \vdots & \vdots & & \vdots \\ y(L-1) & y(L) & \cdots & y(T-1) \end{bmatrix}, \tag{2.2}$$

where $L \geqslant 1$. Clearly, each column of (2.2) contains a length L input/output trajectory of (2.1). By linearity of the system, every linear combination of the columns of (2.2) is also a trajectory of (2.1). In other words,

$$\begin{bmatrix} \bar{u}_{[0,L-1]} \\ \bar{y}_{[0,L-1]} \end{bmatrix} := \begin{bmatrix} \mathcal{H}_L(u_{[0,T-1]}) \\ \mathcal{H}_L(y_{[0,T-1]}) \end{bmatrix} g \tag{2.3}$$

[1] Throughout this chapter, we denote variables such as $\mathbf{u}$ and $\mathbf{y}$ by bold font characters, and specific instances of such variables in normal font, e.g., $u(0), u(1), \ldots$ and $y(0), y(1), \ldots$.

is an input/output trajectory of (2.1) for any real vector g.

The powerful crux of Willems *et al.*'s fundamental lemma is that *every* length L input/output trajectory of (2.1) can be expressed in terms of $(u_{[0,T-1]}, y_{[0,T-1]})$ as in (2.3), assuming that $u_{[0,T-1]}$ is persistently exciting. The result has appeared first in a behavioral context in [241, Thm. 1]. In Theorem 2.1, we will formulate the fundamental lemma for systems of the form (2.1). The theorem consists of two statements. First, under controllability and excitation assumptions, a rank condition on the state and input Hankel matrices (2.4) is satisfied. Second, under the same conditions, all length L input/output trajectories of (2.1) can be written as a linear combination of the columns of the matrix (2.2).

Theorem 2.1. Consider the system (2.1) and assume that the pair (A, B) is controllable. Let $(u_{[0,T-1]}, x_{[0,T-1]}, y_{[0,T-1]})$ be an input/state/output trajectory of (2.1). Assume that the input $u_{[0,T-1]}$ is persistently exciting of order $n + L$. Then the following statements hold:

(i) The matrix

$$\begin{bmatrix} \mathcal{H}_1(x_{[0,T-L]}) \\ \mathcal{H}_L(u_{[0,T-1]}) \end{bmatrix} = \begin{bmatrix} x(0) & x(1) & \cdots & x(T-L) \\ u(0) & u(1) & \cdots & u(T-L) \\ \vdots & \vdots & & \vdots \\ u(L-1) & u(L) & \cdots & u(T-1) \end{bmatrix} \tag{2.4}$$

has full row rank.

(ii) Every length L input/output trajectory of (2.1) can be expressed in terms of $u_{[0,T-1]}$ and $y_{[0,T-1]}$ as follows: $(\bar{u}_{[0,L-1]}, \bar{y}_{[0,L-1]})$ is an input/output trajectory of (2.1) if and only if

$$\begin{bmatrix} \bar{u}_{[0,L-1]} \\ \bar{y}_{[0,L-1]} \end{bmatrix} = \begin{bmatrix} \mathcal{H}_L(u_{[0,T-1]}) \\ \mathcal{H}_L(y_{[0,T-1]}) \end{bmatrix} g, \tag{2.5}$$

for some real vector g.

Statement (i) has appeared first in the original paper by Willems and coworkers, c.f. [241, Cor. 2(iii)]. The result is intriguing since a rank condition on *both* input and state matrices can be inposed by injecting a sufficiently exciting input sequence. This rank condition is important from a design perspective and plays a fundamental role in MOESP type subspace algorithms, c.f. [227, Sec. 3.3]. Also, in the case that $L = 1$, full row rank of (2.4) has been shown to be instrumental for the construction of state feedback controllers from data [47]. In our work, statement (i) is used to prove the second statement of Theorem 2.1. Statement (ii) is a reformulation of [241, Thm. 1]. In what follows, we provide a self-contained and elementary proof of the fundamental lemma in a state-space context.

Proof. Statement (ii) has been proven assuming statement (i) in [47, Lem. 2]. It therefore remains to be shown that (2.4) has full row rank. Let $\begin{bmatrix} \xi & \eta \end{bmatrix}$ be a vector in the left kernel of (2.4), where $\xi^\top \in \mathbb{R}^n$ and $\eta^\top \in \mathbb{R}^{mL}$. We will first show that ξ and η can be used to construct $n+1$ vectors in the left kernel of the "deeper" Hankel matrix

$$\begin{bmatrix} \mathcal{H}_1(x_{[0,T-n-L]}) \\ \mathcal{H}_{n+L}(u_{[0,T-1]}) \end{bmatrix}. \tag{2.6}$$

First, by definition of ξ and η, it is clear that

$$\begin{bmatrix} \xi & \eta & 0_{nm} \end{bmatrix} \begin{bmatrix} \mathcal{H}_1(x_{[0,T-n-L]}) \\ \mathcal{H}_{n+L}(u_{[0,T-1]}) \end{bmatrix} = 0.$$

Next, by the laws of system (2.1a) we have

$$\mathcal{H}_1(x_{[1,T-n-L+1]}) = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} \mathcal{H}_1(x_{[0,T-n-L]}) \\ \mathcal{H}_1(u_{[0,T-n-L]}) \end{bmatrix}.$$

Using this fact, we see that

$$\begin{bmatrix} \xi A & \xi B & \eta & 0_{(n-1)m} \end{bmatrix} \begin{bmatrix} \mathcal{H}_1(x_{[0,T-n-L]}) \\ \mathcal{H}_{n+L}(u_{[0,T-1]}) \end{bmatrix} = \begin{bmatrix} \xi & \eta \end{bmatrix} \begin{bmatrix} \mathcal{H}_1(x_{[1,T-n-L+1]}) \\ \mathcal{H}_L(u_{[1,T-n]}) \end{bmatrix} = 0,$$

where the latter equality holds by definition of ξ and η. Now, by repeatedly exploiting the laws of (2.1a) and using the same arguments we find that the $n+1$ vectors

$$\begin{aligned} w_0 &:= \begin{bmatrix} \xi & \eta & 0_{nm} \end{bmatrix} \\ w_1 &:= \begin{bmatrix} \xi A & \xi B & \eta & 0_{(n-1)m} \end{bmatrix} \\ w_2 &:= \begin{bmatrix} \xi A^2 & \xi AB & \xi B & \eta & 0_{(n-2)m} \end{bmatrix} \\ &\vdots \\ w_n &:= \begin{bmatrix} \xi A^n & \xi A^{n-1}B & \cdots & \xi B & \eta \end{bmatrix} \end{aligned} \tag{2.7}$$

are all contained in the left kernel of the matrix (2.6). By persistency of excitation, $\mathcal{H}_{n+L}(u_{[0,T-1]})$ has full row rank, and hence the left kernel of (2.6) has dimension at most n. Therefore, the $n+1$ vectors in (2.7) are linearly dependent. We claim that this implies $\eta = 0$. To prove this claim, partition $\eta = \begin{bmatrix} \eta_1 & \eta_2 & \cdots & \eta_L \end{bmatrix}$, where $\eta_1^\top, \eta_2^\top, \dots, \eta_L^\top \in \mathbb{R}^m$. Since the last m entries of the vectors $w_0, w_1, \dots, w_{n-1}$ are zero, the linear dependence of the vectors (2.7) implies $\eta_L = 0$ by inspection of w_n. We substitute this equation in η and conclude that the last $2m$ entries of $w_0, w_1, \dots, w_{n-1}$ are zero. As such, also $\eta_{L-1} = 0$. We can proceed with these substitutions to show that $\eta_1 = \eta_2 = \cdots \eta_L = 0$, i.e., $\eta = 0$. Next, by Cayley-Hamilton theorem, $\sum_{i=0}^n \alpha_i A^i = 0$ where $\alpha_i \in \mathbb{R}$ for all $i = 0, 1, \dots, n$, and $\alpha_n = 1$. Define the linear combination $v := \sum_{i=0}^n \alpha_i w_i$. By (2.7) and by substitution of $\eta = 0$, the vector v is equal to

$$\begin{bmatrix} 0_n & \sum_{i=1}^n \alpha_i \xi A^{i-1} B & \sum_{i=2}^n \alpha_i \xi A^{i-2} B & \cdots & \alpha_n \xi B & 0_{mL} \end{bmatrix}.$$

This implies that the vector

$$\begin{bmatrix} \sum_{i=1}^n \alpha_i \xi A^{i-1} B & \sum_{i=2}^n \alpha_i \xi A^{i-2} B & \cdots & \alpha_n \xi B \end{bmatrix}$$

is contained in the left kernel of $\mathcal{H}_n(u_{[0,T-L-1]})$, which is zero by persistency of excitation. In other words,

$$\begin{aligned}
0 &= \alpha_1 \xi B + \cdots + \alpha_n \xi A^{n-1} B \\
0 &= \alpha_2 \xi B + \cdots + \alpha_n \xi A^{n-2} B \\
&\vdots \\
0 &= \alpha_{n-1} \xi B + \alpha_n \xi A B \\
0 &= \alpha_n \xi B.
\end{aligned}$$

Since $\alpha_n = 1$ it follows from the last equation that $\xi B = 0$. Substitution in the second to last equation then results in $\xi AB = 0$. We continue by backward substitution to obtain $\xi B = \xi AB = \cdots = \xi A^{n-1} B = 0$. Controllability of (A, B) hence results in $\xi = 0$. We therefore conclude that (2.4) has full row rank, which proves the theorem. □

2.3 EXTENSION TO MULTIPLE TRAJECTORIES

In this section we propose an extension of the fundamental lemma that is applicable to the case in which *multiple* system trajectories are given. Our approach will require the notion of *collective* persistency of excitation.

Definition 2.2. Consider the input sequences $u^i_{[0,T_i-1]}$ for $i = 1, 2, \ldots, q$, where q is the number of data sets. Let k be a positive integer such that $k \leqslant T_i$ for all i. The input sequences $u^i_{[0,T_i-1]}$ for $i = 1, 2, \ldots, q$ are called *collectively persistently exciting* of order k if the mosaic-Hankel matrix

$$\begin{bmatrix} \mathcal{H}_k(u^1_{[0,T_1-1]}) & \mathcal{H}_k(u^2_{[0,T_2-1]}) & \cdots & \mathcal{H}_k(u^q_{[0,T_q-1]}) \end{bmatrix} \tag{2.8}$$

has full row rank.

Collective persistency of excitation is more flexible than the persistency of excitation of a single input sequence. Indeed, for the input sequences $u^i_{[0,T_i]}$ to be collectively persistently exciting, it is sufficient that at least one of them is persistently exciting. However, this is clearly not necessary: the sequences $u^i_{[0,T_i]}$ may be collectively persistently exciting even when none of the individual input sequences is persistently exciting. The added flexibility of collective persistency of excitation is also apparent from the *length* of the input sequences. Indeed, a single $u_{[0,T-1]}$ can only be persistently exciting of order k if $T \geqslant k(m+1) - 1$. In comparison, for collective persistency of excitation of order k it is necessary that $\sum_{i=1}^q T_i \geqslant k(m+q) - q$. This means that collective persistency of excitation can be achieved by input sequences having length T_i as short as k, assuming the number of data sets q is sufficiently large. In the next theorem we extend the fundamental lemma to the case of multiple data sets.

Theorem 2.2. Consider system (2.1) and assume that the pair (A, B) is controllable. Let $(u^i_{[0,T_i-1]}, x^i_{[0,T_i-1]}, y^i_{[0,T_i-1]})$ be an input/state/output trajectory of (2.1) for $i = 1, 2, \dots, q$. Assume that the inputs $u^i_{[0,T_i-1]}$ are collectively persistently exciting of order $n + L$. Then the following statements hold:

(i) The matrix

$$\begin{bmatrix} \mathcal{H}_1(x^1_{[0,T_1-L]}) & \mathcal{H}_1(x^2_{[0,T_2-L]}) & \cdots & \mathcal{H}_1(x^q_{[0,T_q-L]}) \\ \mathcal{H}_L(u^1_{[0,T_1-1]}) & \mathcal{H}_L(u^2_{[0,T_2-1]}) & \cdots & \mathcal{H}_L(u^q_{[0,T_q-1]}) \end{bmatrix} \tag{2.9}$$

has full row rank.

(ii) Every length L input/output trajectory of (2.1) can be expressed in terms of $u^i_{[0,T_i-1]}$ and $y^i_{[0,T_i-1]}$ $(i = 1, 2, \dots, q)$ as follows: $(\bar{u}_{[0,L-1]}, \bar{y}_{[0,L-1]})$ is an input/output trajectory of (2.1) if and only if

$$\begin{bmatrix} \bar{u}_{[0,L-1]} \\ \bar{y}_{[0,L-1]} \end{bmatrix} = \begin{bmatrix} \mathcal{H}_L(u^1_{[0,T_1-1]}) & \cdots & \mathcal{H}_L(u^q_{[0,T_q-1]}) \\ \mathcal{H}_L(y^1_{[0,T_1-1]}) & \cdots & \mathcal{H}_L(y^q_{[0,T_q-1]}) \end{bmatrix} g, \tag{2.10}$$

for some real vector g.

Note that if $q = 1$ and $T_1 = T$ we deal with a single experiment, and in this case Theorem 2.2 recovers Theorem 2.1.

Proof. We first prove that (2.9) has full row rank. Let $\begin{bmatrix} \xi & \eta \end{bmatrix}$ be a vector in the left kernel of (2.9), where $\xi^\top \in \mathbb{R}^n$ and $\eta^\top \in \mathbb{R}^{mL}$. By exploiting the laws of the system (2.1a) we see that the vectors

$$\begin{aligned} w_0 &:= \begin{bmatrix} \xi & \eta & 0_{nm} \end{bmatrix} \\ w_1 &:= \begin{bmatrix} \xi A & \xi B & \eta & 0_{(n-1)m} \end{bmatrix} \\ w_2 &:= \begin{bmatrix} \xi A^2 & \xi AB & \xi B & \eta & 0_{(n-2)m} \end{bmatrix} \\ &\vdots \\ w_n &:= \begin{bmatrix} \xi A^n & \xi A^{n-1}B & \cdots & \xi B & \eta \end{bmatrix} \end{aligned} \tag{2.11}$$

are contained in the left kernel of the matrix

$$\begin{bmatrix} \mathcal{H}_1(x^1_{[0,T_1-n-L]}) & \cdots & \mathcal{H}_1(x^q_{[0,T_q-n-L]}) \\ \mathcal{H}_{n+L}(u^1_{[0,T_1-1]}) & \cdots & \mathcal{H}_{n+L}(u^q_{[0,T_q-1]}) \end{bmatrix}. \tag{2.12}$$

By the persistency of excitation assumption, the matrix

$$\begin{bmatrix} \mathcal{H}_{n+L}(u^1_{[0,T_1-1]}) & \cdots & \mathcal{H}_{n+L}(u^q_{[0,T_q-1]}) \end{bmatrix}$$

has full row rank, and hence the left kernel of (2.12) has dimension at most n. Therefore, the $n + 1$ vectors in (2.11) are linearly dependent. This yields $\eta = 0$

following the same argument as in the proof of Theorem 2.1. Next, by Cayley-Hamilton theorem, $\sum_{i=0}^{n} \alpha_i A^i = 0$ where $\alpha_i \in \mathbb{R}$ for $i = 0, 1, \ldots, n$ and $\alpha_n = 1$. We define the linear combination $v := \sum_{i=0}^{n} \alpha_i w_i$. Clearly, the vector v is equal to

$$\begin{bmatrix} 0_n & \sum_{i=1}^{n} \alpha_i \xi A^{i-1} B & \sum_{i=2}^{n} \alpha_i \xi A^{i-2} B & \cdots & \alpha_n \xi B & 0_{mL} \end{bmatrix}.$$

Hence, the vector

$$\begin{bmatrix} \sum_{i=1}^{n} \alpha_i \xi A^{i-1} B & \sum_{i=2}^{n} \alpha_i \xi A^{i-2} B & \cdots & \alpha_n \xi B \end{bmatrix}$$

is contained in the left kernel of

$$\begin{bmatrix} \mathcal{H}_n(u^1_{[0,T_1-L-1]}) & \cdots & \mathcal{H}_n(u^q_{[0,T_q-L-1]}) \end{bmatrix},$$

which is zero by collective persistency of excitation. Following the same steps as in the proof of Theorem 2.1 we conclude by backward substitution that $\xi B = \xi AB = \cdots = \xi A^{n-1} B = 0$. By controllability of (A, B) we have $\xi = 0$, proving statement (i).

Next, we prove statement (ii). Let $\bar{u}_{[0,L-1]}$ and $\bar{y}_{[0,L-1]}$ be vectors such that (2.10) is satisfied for some g. Then

$$\begin{bmatrix} \bar{u}_{[0,L-1]} \\ \bar{y}_{[0,L-1]} \end{bmatrix}$$

is a linear combination of length L trajectories of (2.1) and hence, by linearity, itself an input/output trajectory of (2.1). Conversely, let $(\bar{u}_{[0,L-1]}, \bar{y}_{[0,L-1]})$ be an input/output trajectory of (2.1) and denote by $\bar{x}_0$ a corresponding initial state at time 0. We have the relation

$$\begin{bmatrix} \bar{u}_{[0,L-1]} \\ \bar{y}_{[0,L-1]} \end{bmatrix} = \begin{bmatrix} 0 & I \\ \mathcal{O}_L & \mathcal{T}_L \end{bmatrix} \begin{bmatrix} \bar{x}_0 \\ \bar{u}_{[0,L-1]} \end{bmatrix}, \tag{2.13}$$

where $\mathcal{T}_L$ and $\mathcal{O}_L$ are defined as

$$\mathcal{T}_L := \begin{bmatrix} D & 0 & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & D & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{L-2}B & CA^{L-3}B & CA^{L-4}B & \cdots & D \end{bmatrix}, \tag{2.14}$$

$$\mathcal{O}_L := \begin{bmatrix} C^\top & (CA)^\top & (CA^2)^\top & \cdots & (CA^{L-1})^\top \end{bmatrix}^\top. \tag{2.15}$$

Since (2.9) has full row rank, there exists a vector g such that

$$\begin{bmatrix} \bar{x}_0 \\ \bar{u}_{[0,L-1]} \end{bmatrix} = \begin{bmatrix} \mathcal{H}_1(x^1_{[0,T_1-L]}) & \cdots & \mathcal{H}_1(x^q_{[0,T_q-L]}) \\ \mathcal{H}_L(u^1_{[0,T_1-1]}) & \cdots & \mathcal{H}_L(u^q_{[0,T_q-1]}) \end{bmatrix} g.$$

Substitution of the latter expression into (2.13) and using the fact that

$$\begin{bmatrix} 0 & I \\ \mathcal{O}_L & \mathcal{T}_L \end{bmatrix} \begin{bmatrix} \mathcal{H}_1(x^i_{[0,T_i-L]}) \\ \mathcal{H}_L(u^i_{[0,T_i-1]}) \end{bmatrix} = \begin{bmatrix} \mathcal{H}_L(u^i_{[0,T_i-1]}) \\ \mathcal{H}_L(y^i_{[0,T_i-1]}) \end{bmatrix}$$

for all $i = 1, 2, \ldots, q$ yields (2.10), as desired. □

2.4 EXAMPLES

2.4.1 Identification with missing data samples

In this section we treat an example in which we want to identify a system model from a measured trajectory with missing data samples. As we will see, it is possible to apply Theorem 2.2(ii) in this context.

Suppose that we have access to the following, partially corrupted, input/output trajectory of length $T = 20$:

t	0	1	2	3	4	5	6	7	8	9
$u(t)$	1	0	2	−1	0	×	1	1	−1	−5
$y(t)$	3	3	7	6	11	×	18	21	23	24

t	10	11	12	13	14	15	16	17	18	19
$u(t)$	0	−1	×	1	−6	2	−2	0	1	×
$y(t)$	33	31	×	30	20	26	14	10	3	×

The data are generated by a minimal LTI system of (unknown) state-space dimension $n = 2$. Note that some of the samples are *missing*, which we indicate by ×. Our goal is to identify an LTI system that is compatible with the observed data.

In this problem, we have access to three input/output system trajectories, namely $(u_{[0,4]}, y_{[0,4]})$, $(u_{[6,11]}, y_{[6,11]})$ and $(u_{[13,18]}, y_{[13,18]})$. It is not difficult to verify that the input sequences $u_{[0,4]}$, $u_{[6,11]}$ and $u_{[13,18]}$ are collectively persistently exciting of order 5. It can be easily verified that no LTI system of dimension 0 or 1 can explain the data. Thus we consider LTI systems of dimension 2. Since the inputs are collectively persistently exciting of order 5, and since the data-generating system has dimension $n = 2$, by Theorem 2.2(ii) every length $L = 3$ input/output trajectory of the system can be written as linear combination of the columns of

$$\mathcal{D} := \begin{bmatrix} \mathcal{H}_3(u_{[0,4]}) & \mathcal{H}_3(u_{[6,11]}) & \mathcal{H}_3(u_{[13,18]}) \\ \mathcal{H}_3(y_{[0,4]}) & \mathcal{H}_3(y_{[6,11]}) & \mathcal{H}_3(y_{[13,18]}) \end{bmatrix}. \tag{2.16}$$

We exploit this result by computing, as a function of $\mathcal{D}$, the length 7 system trajectory

$$\bar{u}_{[-2,4]} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}^\top \tag{2.17}$$

$$\bar{y}_{[-2,4]} = \begin{bmatrix} 0 & 0 & ? & ? & ? & ? & ? \end{bmatrix}^\top, \tag{2.18}$$

where question marks denote to-be-computed values. The idea is as follows: if the “past" inputs $\bar{u}(-2), \bar{u}(-1)$ and “past" outputs $\bar{y}(-2), \bar{y}(-1)$ are zero, the state $\bar{x}(0) \in \mathbb{R}^2$ corresponding to $(\bar{u}_{[-2,4]}, \bar{y}_{[-2,4]})$ is unique, and equal to zero. This means that $\bar{u}_{[0,4]}$ is an impulse, applied to a system of the form (2.1) with zero initial state. Consequently, the output $\bar{y}_{[0,4]}$ simply consists of the first Markov parameters of (2.1), that is, $\bar{y}_{[0,4]} = \begin{bmatrix} D & CB & CAB & CA^2B & CA^3B \end{bmatrix}$. From these Markov parameters it is straightforward to compute a state-space realization, e.g., using the Ho-Kalman algorithm [228, Sec. 3.4.4].

Therefore, our remaining task is to compute $\bar{y}_{[0,4]}$. Inspired by [127], we will compute this trajectory iteratively by computing multiple length 3 trajectories as linear combinations of the columns of (2.16). To begin with, we compute the first unknown in (2.18), which is $\bar{y}(0)$. To do so, we have to solve the system of linear equations[2]

$$\mathcal{D}g = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & \bar{y}(0) \end{bmatrix}^\top \tag{2.19}$$

in the unknowns g and $\bar{y}(0)$. One possible approach [125, Alg. 1] is to obtain a solution $\bar{g}$ to the first five linear equations in (2.19). Subsequently, $\bar{y}(0)$ is obtained by multiplication of the last row of $\mathcal{D}$ with $\bar{g}$. We do this to find $\bar{y}(0) = 1$. Next, to find $\bar{y}(1)$ we complete the length 3 trajectory $(\bar{u}_{[-1,1]}, \bar{y}_{[-1,1]})$ by solving the system of equations

$$\mathcal{D}g = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & \bar{y}(1) \end{bmatrix}^\top,$$

which results in $\bar{y}(1) = 0$. Repeating this process, we obtain $\bar{y}(2) = 1$, $\bar{y}(3) = 2$ and $\bar{y}(4) = 3$, meaning that

$$D = 1,\ CB = 0,\ CAB = 1,\ CA^2B = 2,\ CA^3B = 3.$$

Finally, it is not difficult to obtain a state-space realization of these Markov parameters as

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix},\ B = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ C = \begin{bmatrix} 0 & 1 \end{bmatrix},\ D = 1.$$

The approach outlined in this section is generally also applicable in the case that multiple consecutive data samples are missing. Even in the case that the number of consecutive missing samples is *unknown*, we can apply Theorem 2.2 to the partial trajectories. Note that we require a sufficient number of partial trajectories of length at least 5 to guarantee collective persistency of excitation of order 5. In the case of missing data with larger frequency, it may still be possible to identify the system by computation of the left kernels of submatrices of the Hankel matrix [121].

[2] Note that the the solution g is not unique in general, but $\bar{y}(0)$ *is* unique. The reason is that the initial state $\bar{x}(0) = 0$ is uniquely specified by the "past" inputs $\bar{u}(-2), \bar{u}(-1)$ and outputs $\bar{y}(-2), \bar{y}(-1)$. In turn, the initial state $\bar{x}(0)$ and input $\bar{u}(0)$ uniquely specify the output $\bar{y}(0)$. Also see [125, Prop. 1].

2.4.2 Data-driven LQR of an unstable system

Consider the unstable batch reactor system [230], which we have discretized using a sampling time of 0.5s to obtain a system of the form (2.1a) with

$$A = \begin{bmatrix} 2.622 & 0.320 & 1.834 & -1.066 \\ -0.238 & 0.187 & -0.136 & 0.202 \\ 0.161 & 0.789 & 0.286 & 0.606 \\ -0.104 & 0.764 & 0.089 & 0.736 \end{bmatrix}, \quad B = \begin{bmatrix} 0.465 & -1.550 \\ 1.314 & 0.085 \\ 2.055 & -0.673 \\ 2.023 & -0.160 \end{bmatrix}.$$

The goal of this example is the data-based design of an optimal control input u^* that minimizes the cost functional

$$J := \sum_{t=0}^{\infty} x^\top(t)Qx(t) + u^\top(t)Ru(t)$$

under the zero endpoint constraint $\lim_{t\to\infty} x(t) = 0$. Here Q and R are state and input weight matrices, respectively. Under standard assumptions on A, B, Q and R [221, Thm. 23], the optimal input exists, is unique, and is generated by the feedback law $u^* = Kx$, where

$$K = -(R + B^\top P^+ B)^{-1} B^\top P^+ A$$

and where P^+ is the largest real symmetric solution to the algebraic Riccati equation

$$P = A^\top PA - A^\top PB(R + B^\top PB)^{-1} B^\top PA + Q.$$

In [47, Thm. 4] an attractive design procedure is introduced to obtain K directly from input/state data. The idea is to inject an input sequence $u_{[0,T-1]}$ that is persistently exciting of order $n+1$ such that the matrix[3]

$$\begin{bmatrix} X_- \\ \hline U_- \end{bmatrix} := \begin{bmatrix} x(0) & x(1) & \cdots & x(T-1) \\ \hline u(0) & u(1) & \cdots & u(T-1) \end{bmatrix} \tag{2.20}$$

has full row rank by Theorem 2.1(i). Subsequently, K is found by solving a semidefinite program involving the data $x_{[0,T]}$ and $u_{[0,T-1]}$ alone; see [47, Eq. 27]. It was shown in [221, Thm. 26] that full row rank of (2.20) is actually also *necessary* for obtaining K from input/state data. In addition, another semidefinite program was introduced [221, Thm. 29] to obtain P^+ and K from input/state data. Both semidefinite programs of [47] and [221] are applicable to this example, but we will follow the method of [221] since it involves less decision variables, c.f. [221, Rem. 31]. We will compare the approach based on a *single* measured trajectory of the system with the one based on *multiple* trajectories. In both the approaches, we take Q and R as the identity matrices of appropriate dimensions.

First, we compute K on the basis of a *single* measured trajectory of (2.1a). We choose a random initial state and random input sequence of length $T = 20$,

[3] Note that $X_- := \mathcal{H}_1(x_{[0,T-1]})$ and $U_- := \mathcal{H}_1(u_{[0,T-1]})$.

generated using the Matlab command `rand`. This input is persistently exciting of order 5. Finally, we let X_- and U_- as in (2.20), and define $X_+ := \mathcal{H}_1(x_{[1,T]})$. By [221, Thm. 29] (see also Chapter 3), the largest solution P^+ to the algebraic Riccati equation is the unique solution to the optimization problem

$$\begin{aligned} &\text{maximize} \ \operatorname{tr} P \\ &\text{subject to } P = P^\top \geqslant 0 \text{ and } \mathcal{L}(P) \leqslant 0, \end{aligned} \tag{2.21}$$

where $\mathcal{L}(P) := X_-^\top P X_- - X_+^\top P X_+ - X_-^\top Q X_- - U_-^\top R U_-$. We use Yalmip with Sedumi 1.3 as LMI solver. Because of the large magnitude of the data samples (reaching $||x(19)|| = 1.049 \cdot 10^8$), the solver runs into numerical problems and returns a matrix P_{sing} that does not resemble P^+. In fact, comparing P_{sing} with the “true" matrix P^+ obtained via the (model-based) Matlab command `dare`, we see

$$P_{\text{sing}} = \begin{bmatrix} 0.002 & 0.013 & -0.005 & 0.015 \\ 0.013 & 0.075 & 0.017 & 0.067 \\ -0.005 & 0.017 & 0.823 & 0.066 \\ 0.015 & 0.067 & 0.066 & 0.010 \end{bmatrix}$$
$$P^+ = \begin{bmatrix} 3.604 & 0.049 & 1.762 & -1.306 \\ 0.049 & 1.170 & 0.072 & 0.142 \\ 1.762 & 0.072 & 2.202 & -0.845 \\ -1.306 & 0.142 & -0.845 & 1.823 \end{bmatrix}.$$

To overcome this problem, we consider *multiple* short experiments, demonstrating the effectiveness of this second approach. We collect $q = 5$ data sets of length $T_i = 6$ for $i = 1, 2, 3, 4, 5$. The input sequences $u^i_{[0,T_i-1]}$ of these sets are again chosen randomly, and are verified to be collectively persistently exciting of order 5. Similar as before, we use the notation $X^i_- := \mathcal{H}_1(x^i_{[0,T_i-1]})$, $X^i_+ := \mathcal{H}_1(x^i_{[1,T_i]})$ and $U^i_- := \mathcal{H}_1(u^i_{[0,T_i-1]})$ for all i. In addition, we concatenate these data matrices and define

$$\begin{aligned} X_- &:= \begin{bmatrix} X^1_- & X^2_- & \cdots & X^5_- \end{bmatrix} \\ X_+ &:= \begin{bmatrix} X^1_+ & X^2_+ & \cdots & X^5_+ \end{bmatrix} \\ U_- &:= \begin{bmatrix} U^1_- & U^2_- & \cdots & U^5_- \end{bmatrix}. \end{aligned}$$

With these data matrices, we solve again (2.21). This result in the solution P_{mult} with $||P_{\text{mult}} - P^+|| = 7.849 \cdot 10^{-10}$. Next, we continue the design procedure of [221, Thm. 29] by computing a right inverse $X_-^\dagger$ of X_- such that $\mathcal{L}(P_{\text{mult}})X_-^\dagger = 0$. The optimal control gain is then computed as

$$K_{\text{mult}} := U_- X_-^\dagger = \begin{bmatrix} 0.163 & -0.292 & 0.046 & -0.328 \\ 1.418 & 0.116 & 0.984 & -0.625 \end{bmatrix}.$$

The error between between K_{mult} and the true optimal gain K obtained via the command `dare` is small. In fact, we have $||K_{\text{mult}} - K|| = 7.083 \cdot 10^{-11}$. The closed-loop matrix $A + BK_{\text{mult}}$ is stable and its spectral radius is 0.188.

The approach that uses multiple trajectories overall requires more samples than the one using a single trajectory. Indeed, as explained in Section 2.3, a necessary condition for collective persistency of excitation of order k is that

$$\sum_{i=1}^{q} T_i \geqslant k(m+q) - q.$$

This means that $\sum_{i=1}^{5} T_i \geqslant 30$ in our example. In comparison, a necessary condition for persistency of excitation of order 5 of a single trajectory is $T \geqslant 14$. Nonetheless, as shown in this example, the use of multiple short trajectories enables the accurate computation of feedback gains even for unstable systems while this may be problematic when using a single long trajectory.

2.5 CONCLUSIONS

Willems *et al.*'s fundamental lemma is a beautiful result that asserts that all trajectories of a linear system can be parameterized by a single, persistently exciting one. In this chapter we have extended the fundamental lemma to the scenario where multiple trajectories are given instead of a single one. To this end, we have introduced a notion of collective persistency of excitation. Subsequently, we have shown that all trajectories of a linear system can be parameterized by a finite number of them, assuming these are collectively persistently exciting. We have shown that this result enables the identification of linear systems from data sets with missing data samples. We have also shown that the result can be used to construct controllers of unstable systems from multiple measured trajectories, even when this is not possible from a single trajectory.

3 DATA INFORMATIVITY FOR ANALYSIS AND CONTROL

In the previous chapter we saw that persistently exciting trajectories can be used to parameterize all trajectories that a linear system can produce. This also means that a system model can be obtained from persistently exciting (and sufficiently long) system trajectories. The current chapter focuses on identifying system properties and controllers -rather than models- from measured trajectories. A natural question is whether a system model (and the ability to obtain one from data) is still necessary if one is interested only in an *aspect* of the underlying system, such as a property or a controller. To get a grip on this question, we introduce a general notion of "data informativity" for data-driven analysis and control.

3.1 INTRODUCTION

One of the main paradigms in the field of systems and control is that of model-based control. Indeed, many control design techniques rely on a system model, represented by e.g. a state-space system or transfer function. In practice, system models are rarely known a priori and have to be identified from measured data using system identification methods such as prediction error [114] or subspace identification [217]. As a consequence, the use of model-based control techniques inherently leads to a two-step control procedure consisting of system identification followed by control design.

Direct data-driven control aims to bypass this two-step procedure by constructing controllers directly from data, without (explicitly) identifying a system model. This approach is not only attractive from a conceptual point of view but can also be useful in situations where system identification is difficult or even impossible because the data do not give sufficient information.

The first contribution to data-driven control is often attributed to Ziegler and Nichols for their work on tuning PID controllers [251]. Adaptive control [7], iterative feedback tuning [85, 86] and unfalsified control [182] can also be regarded as classical data-driven control techniques. More recently, the problem of finding optimal controllers from data has received considerable attention [1, 4, 10, 23, 56, 62, 71, 125, 150, 162, 193, 197]. The proposed solutions to this problem are quite varied, ranging from the use of batch-form Riccati equations [197] to approaches that apply reinforcement learning [23]. Additional noteworthy data-driven control problems include predictive control [40, 55, 183], model reference control [25, 60] and (intelligent) PID control [59, 99]. For more references and classifications of data-driven control techniques, we refer to the survey [89].

In addition to control problems, also *analysis* problems have been studied within a data-based framework. The authors of [164] analyze the stability of an input/output system using time series data. The papers [111, 155, 232, 248] deal with data-based controllability and observability analysis. Moreover, the problem of verifying dissipativity on the basis of measured system trajectories has been studied in [16, 131, 178, 179].

A result that is becoming increasingly popular in the study of data-driven problems is the fundamental lemma by Willems and coworkers [241], see also Chapter 2. This result roughly states that all possible trajectories of a linear time-invariant system can be obtained from any given trajectory whose input component is persistently exciting. The fundamental lemma has clear implications for system identification. Indeed, it provides criteria under which the data are sufficiently informative to uniquely identify the system model within a given model class. In addition, the result has also been applied to data-driven control problems. The idea is that control laws can be obtained directly from data, with the underlying mechanism that the system is represented implicitly by the Hankel matrix of a measured trajectory. This framework has led to several interesting control strategies, first in a behavioral setting [124, 125, 132], and more recently in the context of state-space systems [16, 18, 40, 47, 90, 178].

The above approaches all use persistently exciting data in the control design, meaning that one could (hypothetically) identify the system model from the same data. An intriguing question is therefore the following: is it possible to obtain a controller from data that are *not* informative enough to uniquely identify the system? An affirmative answer would be remarkable, since it would highlight situations in which direct data-driven control is more powerful than the combination of system identification and model-based control. On the other hand, a negative answer would also be significant, as it would give a theoretic justification for the use of persistently exciting data for data-driven analysis and control.

To address the above question, this chapter introduces a general framework to study data informativity problems for data-driven analysis and control. Specifically, our contributions are the following:

1. Inspired by the concept of data informativity in system identification [65, 66, 114], we introduce a general notion of informativity for data-driven analysis and control.

2. We study the data-driven analysis of several system theoretic properties like stability, stabilizability and controllability. For each of these problems, we provide necessary and sufficient conditions under which the data are informative for this property, i.e., conditions required to ascertain the system's property from data.

3. We study data-driven control problems such as stabilization by state feedback, stabilization by dynamic measurement feedback, deadbeat control

and linear quadratic regulation. In each of the cases, we give conditions under which the data are informative for controller design.

4. For each of the studied control problems, we develop methods to compute a controller from data, assuming that the informativity conditions are satisfied.

Our work has multiple noteworthy implications. First of all, we show that for problems like stabilization by state feedback, the corresponding informativity conditions on the data are *weaker* than those for system identification. This implies that a stabilizing feedback can be obtained from data that are not sufficiently informative to uniquely identify the system.

Moreover, for problems such as linear quadratic regulation (LQR), we show that the informativity conditions are essentially the same as for system identification. Therefore, our results provide a theoretic justification for imposing the strong persistency of excitation conditions in prior work on the LQR problem, such as [124] and [47].

The chapter is organized as follows. In Section 3.2 we introduce the problem at a conceptual level. Subsequently, in Section 3.3 we provide data informativity conditions for controllability and stabilizability. Section 3.4 deals with data-driven control problems with input/state data. Next, Section 3.5 discusses control problems where ouput data plays a role. Finally, Section 3.6 contains our conclusions and suggestions for future work.

3.2 PROBLEM FORMULATION

In this section we will first introduce the *informativity framework* for data-driven analysis and control in a fairly abstract manner.

Let $\mathcal{M}$ be a model class, i.e. a given set of systems containing the "true" system denoted by $\mathcal{S}$. We assume that the true system $\mathcal{S}$ is not known but that we have access to a set of data, $\mathcal{D}$, which is generated by this system. In this chapter we are interested in assessing system-theoretic properties of $\mathcal{S}$ and designing control laws for it from the data $\mathcal{D}$.

Given the data $\mathcal{D}$, we define $\Sigma_{\mathcal{D}} \subseteq \mathcal{M}$ to be the set of all systems that are consistent with the data $\mathcal{D}$, i.e., that could also have generated these data.

We first focus on data-driven analysis. Let $\mathcal{P}$ be a system-theoretic property. We will denote the set of all systems within $\mathcal{M}$ having this property by $\Sigma_{\mathcal{P}}$.

Now suppose we are interested in the question whether our true system $\mathcal{S}$ has the property $\mathcal{P}$. As the only information we have to base our answer on are the data $\mathcal{D}$ obtained from the system, we can only conclude that the true system has property $\mathcal{P}$ if *all* systems consistent with the data $\mathcal{D}$ have the property $\mathcal{P}$. This leads to the following definition:

Definition 3.1 (Informativity)**.** We say that the data $\mathcal{D}$ are *informative* for property $\mathcal{P}$ if $\Sigma_{\mathcal{D}} \subseteq \Sigma_{\mathcal{P}}$.

Example 3.1. For given n and m, let $\mathcal{M}$ be the set of all discrete-time linear input/state systems of the form

$$\boldsymbol{x}(t+1) = A\boldsymbol{x}(t) + B\boldsymbol{u}(t)$$

where $\boldsymbol{x}$ is the n-dimensional state and $\boldsymbol{u}$ is the m-dimensional input. Let the true system $\mathcal{S}$ be represented by the matrices (A_s, B_s).

An example of a data set $\mathcal{D}$ arises when considering data-driven problems on the basis of input and state measurements. Suppose that we collect input/state data on q time intervals $\{0, 1, \dots, T_i\}$ for $i = 1, 2, \dots, q$. Let

$$U_-^i := \begin{bmatrix} u^i(0) & u^i(1) & \cdots & u^i(T_i - 1) \end{bmatrix}, \tag{3.1a}$$

$$X^i := \begin{bmatrix} x^i(0) & x^i(1) & \cdots & x^i(T_i) \end{bmatrix} \tag{3.1b}$$

denote the input and state data on the i-th interval. By defining

$$X_-^i := \begin{bmatrix} x^i(0) & x^i(1) & \cdots & x^i(T_i - 1) \end{bmatrix}, \tag{3.2a}$$

$$X_+^i := \begin{bmatrix} x^i(1) & x^i(2) & \cdots & x^i(T_i) \end{bmatrix}, \tag{3.2b}$$

we clearly have $X_+^i = A_s X_-^i + B_s U_-^i$ for each i because the true system is assumed to generate the data. Now, introduce the notation

$$U_- := \begin{bmatrix} U_-^1 & \cdots & U_-^q \end{bmatrix}, \quad X := \begin{bmatrix} X^1 & \cdots & X^q \end{bmatrix}, \tag{3.3a}$$

$$X_- := \begin{bmatrix} X_-^1 & \cdots & X_-^q \end{bmatrix}, \quad X_+ := \begin{bmatrix} X_+^1 & \cdots & X_+^q \end{bmatrix}. \tag{3.3b}$$

We then define the data as $\mathcal{D} := (U_-, X)$. In this case, the set $\Sigma_\mathcal{D}$ is equal to $\Sigma_{(U_-,X)}$ defined by

$$\Sigma_{(U_-,X)} := \left\{ (A, B) \mid X_+ = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix} \right\}. \tag{3.4}$$

Clearly, we have $(A_s, B_s) \in \Sigma_\mathcal{D}$.

Suppose that we are interested in the system-theoretic property $\mathcal{P}$ of *stabilizability*. The corresponding set $\Sigma_\mathcal{P}$ is then equal to Σ_{stab} defined by

$$\Sigma_{\text{stab}} := \{(A, B) \mid (A, B) \text{ is stabilizable}\}.$$

Then, the data (U_-, X) are informative for stabilizability if $\Sigma_{(U_-,X)} \subseteq \Sigma_{\text{stab}}$. That is, if all systems consistent with the input/state measurements are stabilizable.

In general, if the true system $\mathcal{S}$ can be uniquely determined from the data $\mathcal{D}$, that is $\Sigma_\mathcal{D} = \{\mathcal{S}\}$ and $\mathcal{S}$ has the property $\mathcal{P}$, then it is evident that the data $\mathcal{D}$ are informative for $\mathcal{P}$. However, the converse may not be true: $\Sigma_\mathcal{D}$ might contain many systems, all of which have property $\mathcal{P}$. In this chapter, we are interested in necessary *and* sufficient conditions for informativity of the data. Such conditions reveal the minimal amount of information required to assess the property $\mathcal{P}$. A natural problem statement is therefore the following:

Problem 3.1 (Informativity problem)**.** Provide necessary and sufficient conditions on $\mathcal{D}$ under which the data are informative for property $\mathcal{P}$.

The above gives us a general framework to deal with data-driven analysis problems. Such analysis problems will be the main focus of Section 3.3.

This chapter also deals with data-driven control problems. The objective in such problems is the data-based design of controllers such that the closed loop system, obtained from the interconnection of the true system $\mathcal{S}$ and the controller, has a specified property.

As for the analysis problem, we have only the information from the data to base our design on. Therefore, we can only guarantee our control objective if the designed controller imposes the specified property when interconnected with *any* system from the set $\Sigma_{\mathcal{D}}$.

For the framework to allow for data-driven control problems, we will consider a system-theoretic property $\mathcal{P}(\mathcal{K})$ that depends on a given controller $\mathcal{K}$. For properties such as these, we have the following variant of informativity:

Definition 3.2 (Informativity for control)**.** We say that the data $\mathcal{D}$ are *informative* for the property $\mathcal{P}(\cdot)$ if there exists a controller $\mathcal{K}$ such that $\Sigma_{\mathcal{D}} \subseteq \Sigma_{\mathcal{P}(\mathcal{K})}$.

Example 3.2. For systems and data like in Example 3.1, we can take the controller $\mathcal{K} = K \in \mathbb{R}^{m \times n}$ and the property $\mathcal{P}(\mathcal{K})$: "interconnection with the state feedback K yields a stable closed loop system". The corresponding set of systems $\Sigma_{\mathcal{P}(\mathcal{K})}$ is equal to Σ_K defined by

$$\Sigma_K = \{(A, B) \mid A + BK \text{ is stable}^{1}\}.$$

The first step in any data-driven control problem is to determine whether it is possible to obtain a suitable controller from given data. This leads to the following informativity problem:

Problem 3.2 (Informativity problem for control)**.** Provide necessary and sufficient conditions on $\mathcal{D}$ under which there exists a controller $\mathcal{K}$ such that the data are informative for property $\mathcal{P}(\mathcal{K})$.

The second step of data-driven control involves the design of a suitable controller. In terms of our framework, this can be stated as:

Problem 3.3 (Control design problem)**.** Under the assumption that the data $\mathcal{D}$ are informative for property $\mathcal{P}(\cdot)$, find a controller $\mathcal{K}$ such that $\Sigma_{\mathcal{D}} \subseteq \Sigma_{\mathcal{P}(\mathcal{K})}$.

As stated in the introduction, we will highlight the strength of this framework by solving multiple problems. We stress that throughout the chapter it is assumed that the data are *given* and are *not corrupted by noise*.

[1] We say that a matrix is *stable* if all its eigenvalues are contained in the open unit disk.

3.3 DATA-DRIVEN ANALYSIS

In this section, we will study data-driven analysis of controllability and stabilizability given input and state measurements. As in Example 3.1, consider the discrete-time linear system

$$\boldsymbol{x}(t+1) = A_s\boldsymbol{x}(t) + B_s\boldsymbol{u}(t). \tag{3.5}$$

We will consider data consisting of input and state measurements. We define the matrices U_- and X as in (3.3a) and define X_- and X_+ as in (3.3b). The set of all systems compatible with these data was introduced in (3.4). In order to stress that we deal with input/state data, we define

$$\Sigma_{\mathrm{i/s}} := \left\{(A,B) \mid X_+ = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix}\right\}. \tag{3.6}$$

Note that the defining equation of (3.6) is a system of linear equations in the unknowns A and B. The solution space of the corresponding homogeneous equations is denoted by $\Sigma^0_{\mathrm{i/s}}$ and is equal to

$$\Sigma^0_{\mathrm{i/s}} := \left\{(A_0,B_0) \mid 0 = \begin{bmatrix} A_0 & B_0 \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix}\right\}. \tag{3.7}$$

We consider the problem of data-driven analysis for systems of the form (3.5). If (A_s, B_s) is the only system that explains the data, data-driven analysis could be performed by first identifying this system and then analyzing its properties. It is therefore of interest to know under which conditions there is only one system that explains the data.

Definition 3.3. We say that the data (U_-, X) are *informative for system identification* if $\Sigma_{\mathrm{i/s}} = \{(A_s, B_s)\}$.

It is straightforward to derive the following result:

Proposition 3.1. The data (U_-, X) are informative for system identification if and only if

$$\operatorname{rank} \begin{bmatrix} X_- \\ U_- \end{bmatrix} = n + m. \tag{3.8}$$

Furthermore, if (3.8) holds, there exists a right inverse[2] $\begin{bmatrix} V_1 & V_2 \end{bmatrix}$ such that

$$\begin{bmatrix} X_- \\ U_- \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}, \tag{3.9}$$

and for any such right inverse $A_s = X_+V_1$ and $B_s = X_+V_2$.

[2] Note that $\begin{bmatrix} V_1 & V_2 \end{bmatrix}$ is not unique whenever $T > n + m$.

As we will show in this section, the condition (3.8) is not necessary for data-driven analysis in general. We now proceed by studying data-driven analysis of controllability and stabilizability. Recall the Hautus test [208, Thm. 3.13] for controllability: a system (A, B) is controllable if and only if

$$\operatorname{rank}\begin{bmatrix} A-\lambda I & B \end{bmatrix} = n \tag{3.10}$$

for all $\lambda \in \mathbb{C}$. For stabilizability, we require that (3.10) holds for all λ outside the open unit disc.

Now, we introduce the following sets of systems:

$$\begin{aligned} \Sigma_{\text{cont}} &:= \{(A,B) \mid (A,B) \text{ is controllable}\} \\ \Sigma_{\text{stab}} &:= \{(A,B) \mid (A,B) \text{ is stabilizable}\}. \end{aligned}$$

Using Definition 3.1, we obtain the notions of *informativity for controllability* and *stabilizability*. To be precise:

Definition 3.4. We say that the data (U_-, X) are *informative for controllability* if $\Sigma_{\text{i/s}} \subseteq \Sigma_{\text{cont}}$ and *informative for stabilizability* if $\Sigma_{\text{i/s}} \subseteq \Sigma_{\text{stab}}$.

In the following theorem, we give necessary and sufficient conditions for the above notions of informativity. The result is remarkable as only data matrices are used to assess controllability and stabilizability.

Theorem 3.1 (Data-driven Hautus tests)**.** The data (U_-, X) are informative for controllability if and only if

$$\operatorname{rank}(X_+ - \lambda X_-) = n \quad \forall \lambda \in \mathbb{C}. \tag{3.11}$$

Similarly, the data (U_-, X) are informative for stabilizability if and only if

$$\operatorname{rank}(X_+ - \lambda X_-) = n \quad \forall \lambda \in \mathbb{C} \text{ with } |\lambda| \geqslant 1. \tag{3.12}$$

Before proving the theorem, we will discuss some of its implications. We begin with computational issues.

Remark 3.1. Similar to the classical Hautus test, (3.11) and (3.12) can be verified by checking the rank for finitely many complex numbers λ. Indeed, (3.11) is equivalent to $\operatorname{rank}(X_+) = n$ and

$$\operatorname{rank}(X_+ - \lambda X_-) = n$$

for all $\lambda \neq 0$ with $\lambda^{-1} \in \sigma(X_- X_+^\dagger)$, where $X_+^\dagger$ is any right inverse of X_+. Here, $\sigma(M)$ denotes the spectrum, i.e. set of eigenvalues of the matrix M. Similarly, (3.12) is equivalent to $\operatorname{rank}(X_+ - X_-) = n$ and

$$\operatorname{rank}(X_+ - \lambda X_-) = n$$

for all $\lambda \neq 1$ with $(\lambda - 1)^{-1} \in \sigma(X_-(X_+ - X_-)^\dagger)$, where $(X_+ - X_-)^\dagger$ is any right inverse of $X_+ - X_-$.

A noteworthy point to mention is that there are situations in which we can conclude controllability/stabilizability from the data without being able to identify the true system uniquely, as illustrated next.

Example 3.3. Suppose that $n = 2$, $m = 1$, $q = 1$, $T_1 = 2$ and we obtain the data

$$X = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } U_- = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

This implies that

$$X_+ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } X_- = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Clearly, by Theorem 3.1 we see that these data are informative for controllability, as

$$\operatorname{rank} \begin{bmatrix} 1 & -\lambda \\ 0 & 1 \end{bmatrix} = 2 \quad \forall \lambda \in \mathbb{C}.$$

As therefore all systems explaining the data are controllable, we conclude that the true system is controllable. Note that the data are not informative for system identification, as

$$\Sigma_{\mathrm{i/s}} = \left\{ \left(\begin{bmatrix} 0 & a_1 \\ 1 & a_2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \mid a_1, a_2 \in \mathbb{R} \right\}. \tag{3.13}$$

Proof of Theorem 3.1. We will only prove the characterization of informativity for controllability. The proof for stabilizability uses very similar arguments, and is hence omitted.

Note that the condition (3.11) is equivalent to the implication:

$$z \in \mathbb{C}^n, \lambda \in \mathbb{C} \text{ and } z^* X_+ = \lambda z^* X_- \implies z = 0. \tag{3.14}$$

Suppose that the implication (3.14) holds. Let $(A, B) \in \Sigma_{\mathrm{i/s}}$ and suppose that $z^* \begin{bmatrix} A - \lambda I & B \end{bmatrix} = 0$. We want to prove that $z = 0$. Note that $z^* \begin{bmatrix} A - \lambda I & B \end{bmatrix} = 0$ implies that

$$z^* \begin{bmatrix} A - \lambda I & B \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix} = 0,$$

or equivalently $z^* X_+ = \lambda z^* X_-$. This means that $z = 0$ by (3.14). We conclude that (A, B) is controllable, i.e., (U_-, X) are informative for controllability.

Conversely, suppose that (U_-, X) are informative for controllability. Let $z \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$ be such that $z^* X_+ = \lambda z^* X_-$. This implies that for all $(A, B) \in \Sigma_{\mathrm{i/s}}$, we have $z^* \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix} = \lambda z^* X_-$. In other words,

$$z^* \begin{bmatrix} A - \lambda I & B \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix} = 0. \tag{3.15}$$

We now distinguish two cases, namely the case that λ is real, and the case that λ is complex. First suppose that λ is real. Without loss of generality, z is real. We want to prove that $z = 0$. Suppose on the contrary that $z \neq 0$ and $z^\top z = 1$. We define the (real) matrices

$$\bar{A} := A - zz^\top (A - \lambda I) \text{ and } \bar{B} := B - zz^\top B.$$

In view of (3.15), we find that $(\bar{A}, \bar{B}) \in \Sigma_{\mathrm{i/s}}$. Moreover,

$$z^\top \bar{A} = z^\top A - z^\top (A - \lambda I) = \lambda z^\top$$

and

$$z^\top \bar{B} = z^\top B - z^\top B = 0.$$

This means that

$$z^\top \begin{bmatrix} \bar{A} - \lambda I & \bar{B} \end{bmatrix} = 0.$$

However, this is a contradiction as $(\bar{A}, \bar{B})$ is controllable by the hypothesis that (U_-, X) are informative for controllability. We conclude that $z = 0$ which shows that (3.14) holds for the case that λ is real.

Secondly, consider the case that λ is complex. We write z as $z = p + iq$, where $p, q \in \mathbb{R}^n$ and i denotes the imaginary unit. If p and q are linearly dependent, then $p = \alpha q$ or $q = \beta p$ for $\alpha, \beta \in \mathbb{R}$. If $p = \alpha q$ then substitution of $z = (\alpha + i)q$ into $z^* X_+ = \lambda z^* X_-$ yields

$$(\alpha - i) q^\top X_+ = \lambda (\alpha - i) q^\top X_-,$$

that is, $q^\top X_+ = \lambda q^\top X_-$. As $q^\top X_+$ is real and λ is complex, we must have $q^\top X_+ = 0$ and $q^\top X_- = 0$. This means that $z^* X_+ = z^* X_- = 0$, hence $z^* X_+ = \mu z^* X_-$ for any real μ, which means that $z = 0$ by case 1. Using the same arguments, we can show that $z = 0$ if $q = \beta p$.

Now, it suffices to prove that p and q are linearly dependent. Suppose on the contrary that p and q are linearly independent. Since λ is complex, $n \geqslant 2$. Therefore, by linear independence of p and q there exist $\eta, \zeta \in \mathbb{R}^n$ such that

$$\begin{bmatrix} p^\top \\ q^\top \end{bmatrix} \begin{bmatrix} \eta & \zeta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

We now define the real matrices $\bar{A}$ and $\bar{B}$ as

$$\begin{bmatrix} \bar{A} & \bar{B} \end{bmatrix} := \begin{bmatrix} A & B \end{bmatrix} - \begin{bmatrix} \eta & \zeta \end{bmatrix} \begin{bmatrix} \operatorname{Re}\left(z^* \begin{bmatrix} A - \lambda I & B \end{bmatrix} \right) \\ \operatorname{Im}\left(z^* \begin{bmatrix} A - \lambda I & B \end{bmatrix} \right) \end{bmatrix}.$$

By (3.15) we have $(\bar{A}, \bar{B}) \in \Sigma_{\mathrm{i/s}}$. Next, we compute

$$\begin{aligned} z^* \begin{bmatrix} \bar{A} & \bar{B} \end{bmatrix} &= z^* \begin{bmatrix} A & B \end{bmatrix} - \begin{bmatrix} 1 & i \end{bmatrix} \begin{bmatrix} \operatorname{Re}\left(z^* \begin{bmatrix} A - \lambda I & B \end{bmatrix} \right) \\ \operatorname{Im}\left(z^* \begin{bmatrix} A - \lambda I & B \end{bmatrix} \right) \end{bmatrix} \\ &= z^* \begin{bmatrix} A & B \end{bmatrix} - z^* \begin{bmatrix} A - \lambda I & B \end{bmatrix} \\ &= z^* \begin{bmatrix} \lambda I & 0 \end{bmatrix}. \end{aligned}$$

This implies that $z^* \begin{bmatrix} \bar{A} - \lambda I & \bar{B} \end{bmatrix} = 0$. Using the fact that $(\bar{A}, \bar{B})$ is controllable, we conclude that $z = 0$. This is a contradiction with the fact that p and q are linearly independent. Thus p and q are linearly dependent and therefore implication (3.14) holds. This proves the theorem. □

In addition to controllability and stabilizability, we can also study the *stability* of an autonomous system of the form

$$\boldsymbol{x}(t+1) = A_s \boldsymbol{x}(t). \tag{3.16}$$

To this end, let X denote the matrix of state measurements obtained from (3.16), as defined in (3.3a). The set of all autonomous systems compatible with these data is

$$\Sigma_s := \{A \mid X_+ = AX_-\}.$$

Then, we say the data X are *informative for stability* if any matrix $A \in \Sigma_s$ is stable, i.e. Schur. Using Theorem 3.1 we can show that stability can only be concluded if the true system can be uniquely identified.

Corollary 3.1. The data X are informative for stability if and only if X_- has full row rank and $X_+X_-^\dagger$ is stable for any right inverse $X_-^\dagger$, equivalently $\Sigma_s = \{A_s\}$ and $A_s = X_+X_-^\dagger$ is stable.

Proof. Since the "if" part is evident, we only prove the "only if" part. By taking $B = 0$, it follows from Theorem 3.1 that the data X are informative for stability if and only if

$$\operatorname{rank}(X_+ - \lambda X_-) = n \quad \forall \lambda \in \mathbb{C} \text{ with } |\lambda| \geqslant 1. \tag{3.17}$$

Let z be such that $z^\top X_- = 0$. Take $A \in \Sigma_s$ and $\lambda > 1$ such that λ is not an eigenvalue of A. Note that

$$z^\top (A - \lambda I)^{-1}(X_+ - \lambda X_-) = z^\top X_- = 0.$$

Since $\operatorname{rank}(X_+ - \lambda X_-) = n$, we may conclude that $z = 0$. Hence, X_- has full row rank. Therefore, $\Sigma_s = \{A_s\}$ where $A_s = X_+X_-^\dagger$ for any right inverse $X_-^\dagger$ and A_s is stable. □

Note that there is a subtle but important difference between the characterizations (3.12) and (3.17). For the first the data X are assumed to be generated by a system with inputs, whereas the data for the second characterization are generated by an autonomous system.

3.4 CONTROL USING INPUT AND STATE DATA

In this section we will consider various state feedback control problems on the basis of input/state measurements. First, we will consider the problem of data-driven stabilization by static state feedback, where the data consist of input and state measurements. As described in the problem statement we will look at the informativity and design problems separately as special cases of Problem 3.2 and Problem 3.3. We will then use similar techniques to obtain a result for *deadbeat control*.

After this, we will shift towards the linear quadratic regulator problem, where we wish to find a stabilizing feedback that additionally minimizes a specified quadratic cost.

3.4.1 Stabilization by state feedback

In what follows, we will consider the problem of finding a stabilizing controller for the system (3.5), using only the data (U_-, X). To this end, we define the set of systems (A, B) that are stabilized by a given K:

$$\Sigma_K := \{(A, B) \mid A + BK \text{ is stable}\}.$$

In addition, recall the set $\Sigma_{\mathrm{i/s}}$ as defined in (3.6) and $\Sigma^0_{\mathrm{i/s}}$ from (3.7). In line with Definition 3.2 we obtain the following notion of informativity for stabilization by state feedback.

Definition 3.5. We say that the data (U_-, X) are *informative for stabilization by state feedback* if there exists a feedback gain K such that $\Sigma_{\mathrm{i/s}} \subseteq \Sigma_K$.

Remark 3.2. At this point, one may wonder about the relation between informativity for stabilizability (as in Section 3.3) and informativity for stabilization. It is clear that (U_-, X) are informative for stabilizability if (U_-, X) are informative for stabilization by state feedback. However, the reverse statement does not hold in general. This is due to the fact that all systems (A, B) in $\Sigma_{\mathrm{i/s}}$ may be stabilizable, but there may not be a *common* feedback gain K such that $A + BK$ is stable for all of these systems. Note that the existence of a common stabilizing K for all systems in $\Sigma_{\mathrm{i/s}}$ is essential, since there is no way to distinguish between the systems in $\Sigma_{\mathrm{i/s}}$ based on the given data (U_-, X).

The following example further illustrates the difference between informativity for stabilizability and informativity for stabilization.

Example 3.4. Consider the scalar system

$$\boldsymbol{x}(t+1) = \boldsymbol{u}(t),$$

where $\boldsymbol{x}, \boldsymbol{u} \in \mathbb{R}$. Suppose that $q = 1$, $T_1 = 1$ and $x(0) = 0$, $u(0) = 1$ and $x(1) = 1$. This means that $U_- = \begin{bmatrix} 1 \end{bmatrix}$ and $X = \begin{bmatrix} 0 & 1 \end{bmatrix}$. It can be shown that $\Sigma_{\mathrm{i/s}} = \{(a, 1) \mid a \in \mathbb{R}\}$. Clearly, all systems in $\Sigma_{\mathrm{i/s}}$ are stabilizable, i.e., $\Sigma_{\mathrm{i/s}} \subseteq \Sigma_{\mathrm{stab}}$. Nonetheless, the data are not informative for *stabilization*. This is because the systems $(-1, 1)$ and $(1, 1)$ in $\Sigma_{\mathrm{i/s}}$ cannot be stabilized by the *same* controller of the form $u(t) = Kx(t)$. We conclude that informativity of the data for stabilizability does not imply informativity for stabilization by state feedback.

The notion of informativity for stabilization by state feedback is a specific example of informativity for control. As described in Problem 3.2, we will first find necessary and sufficient conditions for informativity for stabilization by state feedback. After this, we will design a corresponding controller, as described in Problem 3.3.

In order to be able to characterize informativity for stabilization, we first state the following lemma.

Lemma 3.1. Suppose that the data (U_-, X) are informative for stabilization by state feedback, and let K be a feedback gain such that $\Sigma_{\mathrm{i/s}} \subseteq \Sigma_K$. Then $A_0 + B_0 K = 0$ for all $(A_0, B_0) \in \Sigma^0_{\mathrm{i/s}}$. Equivalently,

$$\operatorname{im} \begin{bmatrix} I \\ K \end{bmatrix} \subseteq \operatorname{im} \begin{bmatrix} X_- \\ U_- \end{bmatrix}.$$

Proof. We first prove that $A_0 + B_0 K$ is *nilpotent* for all $(A_0, B_0) \in \Sigma^0_{\mathrm{i/s}}$. By hypothesis, $A + BK$ is stable for all $(A, B) \in \Sigma_{\mathrm{i/s}}$. Let $(A, B) \in \Sigma_{\mathrm{i/s}}$ and $(A_0, B_0) \in \Sigma^0_{\mathrm{i/s}}$ and define the matrices $F := A + BK$ and $F_0 := A_0 + B_0 K$. Then, the matrix $F + \alpha F_0$ is stable for all $\alpha \geqslant 0$. By dividing by α, it follows that, for all $\alpha \geqslant 1$, the spectral radius of the matrix

$$M_\alpha := \frac{1}{\alpha} F + F_0$$

is smaller than $1/\alpha$. From the continuity of the spectral radius by taking the limit as α tends to infinity, we see that $F_0 = A_0 + B_0 K$ is nilpotent for all $(A_0, B_0) \in \Sigma^0_{\mathrm{i/s}}$. Note that we have

$$((A_0 + B_0 K)^T A_0, (A_0 + B_0 K)^T B_0) \in \Sigma^0_{\mathrm{i/s}}$$

whenever $(A_0, B_0) \in \Sigma^0_{\mathrm{i/s}}$. This means that $(A_0 + B_0 K)^T (A_0 + B_0 K)$ is nilpotent. Since the only symmetric nilpotent matrix is the zero matrix, we see that $A_0 + B_0 K = 0$ for all $(A_0, B_0) \in \Sigma^0_{\mathrm{i/s}}$. This is equivalent to

$$\ker \begin{bmatrix} X_-^\top & U_-^\top \end{bmatrix} \subseteq \ker \begin{bmatrix} I & K^\top \end{bmatrix}$$

which is equivalent to $\operatorname{im} \begin{bmatrix} I \\ K \end{bmatrix} \subseteq \operatorname{im} \begin{bmatrix} X_- \\ U_- \end{bmatrix}$. □

The previous lemma is instrumental in proving the following theorem that gives necessary and sufficient conditions for informativity for stabilization by state feedback.

Theorem 3.2. The data (U_-, X) are informative for stabilization by state feedback if and only if the matrix X_- has full row rank and there exists a right inverse $X_-^\dagger$ of X_- such that $X_+ X_-^\dagger$ is stable.

Moreover, K is such that $\Sigma_{\mathrm{i/s}} \subseteq \Sigma_K$ if and only if $K = U_- X_-^\dagger$, where $X_-^\dagger$ satisfies the above properties.

Proof. To prove the "if" part of the first statement, suppose that X_- has full row rank and there exists a right inverse $X_-^\dagger$ of X_- such that $X_+X_-^\dagger$ is stable. We define $K := U_-X_-^\dagger$. Next, we see that

$$X_+X_-^\dagger = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix} X_-^\dagger = A + BK, \tag{3.18}$$

for all $(A, B) \in \Sigma_{\mathrm{i/s}}$. Therefore, $A + BK$ is stable for all $(A, B) \in \Sigma_{\mathrm{i/s}}$, i.e., $\Sigma_{\mathrm{i/s}} \subseteq \Sigma_K$. We conclude that the data (U_-, X) are informative for stabilization by state feedback, proving the "if" part of the first statement. Since $K = U_-X_-^\dagger$ is such that $\Sigma_{\mathrm{i/s}} \subseteq \Sigma_K$, we have also proven the "if" part of the second statement as a byproduct.

Next, to prove the "only if" part of the first statement, suppose that the data (U_-, X) are informative for stabilization by state feedback. Let K be such that $A + BK$ is stable for all $(A, B) \in \Sigma_{\mathrm{i/s}}$. By Lemma 3.1 we know that

$$\operatorname{im} \begin{bmatrix} I \\ K \end{bmatrix} \subseteq \operatorname{im} \begin{bmatrix} X_- \\ U_- \end{bmatrix}.$$

This implies that X_- has full row rank and there exists a right inverse $X_-^\dagger$ such that

$$\begin{bmatrix} I \\ K \end{bmatrix} = \begin{bmatrix} X_- \\ U_- \end{bmatrix} X_-^\dagger. \tag{3.19}$$

By (3.18), we obtain $A + BK = X_+X_-^\dagger$, which shows that $X_+X_-^\dagger$ is stable. This proves the "only if" part of the first statement. Finally, by (3.19), the stabilizing feedback gain K is indeed of the form $K = U_-X_-^\dagger$, which also proves the "only if" part of the second statement. □

Theorem 3.2 gives a characterization of all data that are informative for stabilization by state feedback and provides a stabilizing controller. Nonetheless, the procedure to compute this controller might not be entirely satisfactory since it is not clear how to find a right inverse of X_- that makes $X_+X_-^\dagger$ stable. In general, X_- has many right inverses, and $X_+X_-^\dagger$ can be stable or unstable depending on the particular right inverse $X_-^\dagger$. To deal with this problem and to solve the design problem, we give a characterization of informativity for stabilization in terms of linear matrix inequalities (LMI's). The feasibility of such LMI's can be verified using standard tools.

Theorem 3.3. The data (U_-, X) are informative for stabilization by state feedback if and only if there exists a matrix $\Theta \in \mathbb{R}^{T \times n}$ satisfying

$$X_-\Theta = (X_-\Theta)^\top \quad \text{and} \quad \begin{bmatrix} X_-\Theta & X_+\Theta \\ \Theta^\top X_+^\top & X_-\Theta \end{bmatrix} > 0. \tag{3.20}$$

Moreover, K satisfies $\Sigma_{\mathrm{i/s}} \subseteq \Sigma_K$ if and only if $K = U_-\Theta(X_-\Theta)^{-1}$ for some matrix Θ satisfying (3.20).

Remark 3.3. To the best of our knowledge, LMI conditions for data-driven stabilization were first studied in [47]. In fact, the linear matrix inequality (3.20) is the same as that of [47, Thm. 3]. However, an important difference is that the results in [47] assume that the input u is persistently exciting of sufficiently high order. In contrast, Theorem 3.3, as well as Theorem 3.2, do not require such conditions. The characterization (3.20) provides the minimal conditions on the data under which it is possible to obtain a stabilizing controller.

Example 3.5. Consider an unstable system of the form (3.5), where A_s and B_s are given by

$$A_s = \begin{bmatrix} 1.5 & 0 \\ 1 & 0.5 \end{bmatrix}, \quad B_s = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

We collect data from this system on a single time interval from $t = 0$ until $t = 2$, which results in the data matrices

$$X = \begin{bmatrix} 1 & 0.5 & -0.25 \\ 0 & 1 & 1 \end{bmatrix}, \quad U_- = \begin{bmatrix} -1 & -1 \end{bmatrix}.$$

Clearly, the matrix X_- is square and invertible, and it can be verified that

$$X_+ X_-^{-1} = \begin{bmatrix} 0.5 & -0.5 \\ 1 & 0.5 \end{bmatrix}$$

is stable, since its eigenvalues are $\frac{1}{2}(1 \pm \sqrt{2}i)$. We conclude by Theorem 3.2 that the data (U_-, X) are informative for stabilization by state feedback. The same conclusion can be drawn from Theorem 3.3 since

$$\Theta = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}$$

solves (3.20). Next, we can conclude from either Theorem 3.2 or Theorem 3.3 that the stabilizing feedback gain in this example is unique, and given by $K = U_- X_-^{-1} = \begin{bmatrix} -1 & -0.5 \end{bmatrix}$. Finally, it is worth noting that the data are not informative for system identification. In fact, $(A, B) \in \Sigma_{\mathrm{i/s}}$ if and only if

$$A = \begin{bmatrix} 1.5 + a_1 & 0.5a_1 \\ 1 + a_2 & 0.5 + 0.5a_2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 + a_1 \\ a_2 \end{bmatrix}$$

for some $a_1, a_2 \in \mathbb{R}$.

Proof of Theorem 3.3. To prove the "if" part of the first statement, suppose that there exists a Θ satisfying (3.20). In particular, this implies that $X_-\Theta$ is symmetric positive definite. Therefore, X_- has full row rank. By taking a Schur complement and multiplying by -1, we obtain

$$X_+\Theta(X_-\Theta)^{-1}(X_-\Theta)(X_-\Theta)^{-1}\Theta^\top X_+^\top - X_-\Theta < 0.$$

Since $X_-\Theta$ is positive definite, this implies that $X_+\Theta(X_-\Theta)^{-1}$ is stable. In other words, there exists a right inverse $X_-^\dagger := \Theta(X_-\Theta)^{-1}$ of X_- such that $X_+X_-^\dagger$ is stable. By Theorem 3.2, we conclude that (U_-, X) are informative for stabilization by state feedback, proving the "if" part of the first statement. Using Theorem 3.2 once more, we see that $K := U_-\Theta(X_-\Theta)^{-1}$ stabilizes all systems in $\Sigma_{\mathrm{i/s}}$, which in turn proves the "if" part of the second statement.

Subsequently, to prove the "only if" part of the first statement, suppose that the data (U_-, X) are informative for stabilization by state feedback. Let K be any feedback gain such that $\Sigma_{\mathrm{i/s}} \subseteq \Sigma_K$. By Theorem 3.2, X_- has full row rank and K is of the form $K = U_- X_-^\dagger$, where $X_-^\dagger$ is a right inverse of X_- such that $X_+ X_-^\dagger$ is stable. The stability of $X_+ X_-^\dagger$ implies the existence of a symmetric positive definite matrix P such that

$$(X_+ X_-^\dagger) P (X_+ X_-^\dagger)^\top - P < 0.$$

Next, we define $\Theta := X_-^\dagger P$ and note that

$$X_+ \Theta P^{-1} (X_+ \Theta)^\top - P < 0.$$

Via the Schur complement we conclude that

$$\begin{bmatrix} P & X_+\Theta \\ \Theta^\top X_+^\top & P \end{bmatrix} > 0.$$

Since $X_- X_-^\dagger = I$, we see that $P = X_-\Theta$, which proves the "only if" part of the first statement. Finally, by definition of Θ, we have $X_-^\dagger = \Theta P^{-1} = \Theta(X_-\Theta)^{-1}$. Recall that $K = U_- X_-^\dagger$, which shows that K is of the form $K = U_-\Theta(X_-\Theta)^{-1}$ for Θ satisfying (3.20). This proves the "only if" part of the second statement and hence the proof is complete. □

In addition to the stabilizing controllers discussed in Theorems 3.2 and 3.3, we may also look for a controller of the form $\boldsymbol{u}(t) = K\boldsymbol{x}(t)$ that stabilizes the system in *finite time*. Such a controller is called a *deadbeat controller* and is characterized by the property that $(A_s + B_s K)^t x_0 = 0$ for all $t \geqslant n$ and all $x_0 \in \mathbb{R}^n$. Thus, K is a deadbeat controller if and only if $A_s + B_s K$ is nilpotent. Now, for a given matrix K define

$$\Sigma_K^{\mathrm{nil}} := \{(A, B) \mid A + BK \text{ is nilpotent}\}.$$

Then, analogous to the definition of informativity for stabilization by state feedback, we have the following definition of informativity for deadbeat control.

Definition 3.6. We say that the data (U_-, X) are *informative for deadbeat control* if there exists a feedback gain K such that $\Sigma_{\mathrm{i/s}} \subseteq \Sigma_K^{\mathrm{nil}}$.

Similarly to Theorem 3.2, we obtain the following necessary and sufficient conditions for informativity for deadbeat control.

Theorem 3.4. The data (U_-, X) are informative for deadbeat control if and only if the matrix X_- has full row rank and there exists a right inverse $X_-^\dagger$ of X_- such that $X_+ X_-^\dagger$ is nilpotent.

Moreover, if this condition is satisfied then the feedback gain $K := U_- X_-^\dagger$ yields a deadbeat controller, that is, $\Sigma_{\mathrm{i/s}} \subseteq \Sigma_K^{\mathrm{nil}}$.

Remark 3.4. In order to compute a suitable right inverse $X_-^\dagger$ such that $X_+X_-^\dagger$ is nilpotent, we can proceed as follows. Since X_- has full row rank, we have $T \geqslant n$. We now distinguish two cases: $T = n$ and $T > n$. In the former case, X_- is nonsingular and hence $X_+X_-^{-1}$ is nilpotent. In the latter case, there exist matrices $F \in \mathbb{R}^{T\times n}$ and $G \in \mathbb{R}^{T\times(T-n)}$ such that $\begin{bmatrix} F & G \end{bmatrix}$ is nonsingular and $X_- \begin{bmatrix} F & G \end{bmatrix} = \begin{bmatrix} I_n & 0_{n\times(T-n)} \end{bmatrix}$. Note that $X_-^\dagger$ is a right inverse of X_- if and only if $X_-^\dagger = F + GH$ for some $H \in \mathbb{R}^{(T-n)\times n}$. Finding a right inverse $X_-^\dagger$ such that $X_+X_-^\dagger$ is nilpotent, therefore, amounts to finding H such that $X_+F + X_+GH$ is nilpotent, i.e. has only zero eigenvalues. Such a matrix H can be computed by invoking [208, Thm. 3.29 and Thm. 3.32] for the pair (X_+F, X_+G) and the stability domain $\mathbb{C}_g = \{0\}$.

3.4.2 Informativity for linear quadratic regulation

Consider the discrete-time linear system (3.5). Let $x_{x_0,u}(\cdot)$ be the state sequence of (3.5) resulting from the input $u(\cdot)$ and initial condition $x(0) = x_0$. We omit the subscript and simply write $x(\cdot)$ whenever the dependence on x_0 and u is clear from the context.

Associated to system (3.5), we define the quadratic cost functional

$$J(x_0, u) = \sum_{t=0}^{\infty} x^\top(t)Qx(t) + u^\top(t)Ru(t), \tag{3.21}$$

where $Q = Q^\top$ is positive semidefinite and $R = R^\top$ is positive definite. Then, the linear quadratic regulator (LQR) problem is the following:

Problem 3.4 (LQR)**.** Determine for every initial condition x_0 an input u^*, such that $\lim_{t\to\infty} x_{x_0,u^*}(t) = 0$, and the cost functional $J(x_0, u)$ is minimized under this constraint.

Such an input u^* is called optimal for the given x_0. Of course, an optimal input does not necessarily exist for all x_0. We say that the linear quadratic regulator problem is *solvable* for (A, B, Q, R) if for every x_0 there exists an input u^* such that

1. The cost $J(x_0, u^*)$ is finite.
2. The limit $\lim_{t\to\infty} x_{x_0,u^*}(t) = 0$.
3. The input u^* minimizes the cost functional, i.e.,

$$J(x_0, u^*) \leqslant J(x_0, \bar{u})$$

for all $\bar{u}$ such that $\lim_{t\to\infty} x_{x_0,\bar{u}}(t) = 0$.

In the sequel, we will require the notion of observable eigenvalues. Recall from e.g. [208, Sec. 3.5] that an eigenvalue λ of A is (Q, A)-observable if

$$\operatorname{rank} \begin{pmatrix} A - \lambda I \\ Q \end{pmatrix} = n.$$

The following theorem provides necessary and sufficient conditions for the solvability of the linear quadratic regulator problem for (A, B, Q, R). This theorem is the discrete-time analogue to the continuous-time case stated in [208, Thm. 10.18].

Theorem 3.5. Let $Q = Q^\top$ be positive semidefinite and $R = R^\top$ be positive definite. Then the following statements hold:

(i) If (A, B) is stabilizable, there exists a unique largest real symmetric solution P^+ to the discrete-time algebraic Riccati equation (DARE)

$$P = A^\top PA - A^\top PB(R + B^\top PB)^{-1}B^\top PA + Q, \tag{3.22}$$

in the sense that $P^+ \geqslant P$ for every real symmetric P satisfying (3.22). The matrix P^+ is positive semidefinite.

(ii) If, in addition to stabilizability of (A, B), every eigenvalue of A on the unit circle is (Q, A)-observable then for every x_0 a unique optimal input u^* exists. Furthermore, this input sequence is generated by the feedback law $\boldsymbol{u} = K\boldsymbol{x}$, where

$$K := -(R + B^\top P^+ B)^{-1}B^\top P^+ A. \tag{3.23}$$

Moreover, the matrix $A + BK$ is stable.

(iii) In fact, the linear quadratic regulator problem is solvable for (A, B, Q, R) if and only if (A, B) is stabilizable and every eigenvalue of A on the unit circle is (Q, A)-observable.

If the LQR problem is solvable for (A, B, Q, R), we say that K given by (3.23) is the optimal feedback gain for (A, B, Q, R).

Now, for any given K we define $\Sigma_K^{Q,R}$ as the set of all systems of the form (3.5) for which K is the optimal feedback gain corresponding to Q and R, that is,

$$\Sigma_K^{Q,R} := \{(A, B) \mid K \text{ is the optimal gain for } (A, B, Q, R)\}.$$

This gives rise to another notion of informativity in line with Definition 3.2. Again, let $\Sigma_{\mathrm{i/s}}$ be given by (3.6).

Definition 3.7. Given matrices Q and R, we say that the data (U_-, X) are *informative for linear quadratic regulation* if there exists K such that $\Sigma_{\mathrm{i/s}} \subseteq \Sigma_K^{Q,R}$.

In order to provide necessary and sufficient conditions for the corresponding informativity problem, we need the following auxiliary lemma.

Lemma 3.2. Let $Q = Q^\top$ be positive semidefinite and $R = R^\top$ be positive definite. Suppose the data (U_-, X) are informative for linear quadratic regulation. Let K

be such that $\Sigma_{\mathrm{i/s}} \subseteq \Sigma_K^{Q,R}$. Then, there exist a square matrix M and a symmetric positive semidefinite matrix P^+ such that for all $(A,B) \in \Sigma_{\mathrm{i/s}}$

$$M = A + BK, \tag{3.24}$$

$$P^+ = A^\top P^+ A - A^\top P^+ B(R + B^\top P^+ B)^{-1} B^\top P^+ A + Q, \tag{3.25}$$

$$P^+ - M^\top P^+ M = K^\top R K + Q, \tag{3.26}$$

$$K = -(R + B^\top P^+ B)^{-1} B^\top P^+ A. \tag{3.27}$$

Proof. Since the data (U_-, X) are informative for linear quadratic regulation, $A + BK$ is stable for every $(A,B) \in \Sigma_{\mathrm{i/s}}$. By Lemma 3.1, this implies that $A_0 + B_0 K = 0$ for all $(A_0, B_0) \in \Sigma^0_{\mathrm{i/s}}$. Thus, there exists M such that $M = A + BK$ for all $(A,B) \in \Sigma_{\mathrm{i/s}}$. For the rest, note that Theorem 3.5 implies that for every $(A,B) \in \Sigma_{\mathrm{i/s}}$ there exists $P^+_{(A,B)}$ satisfying the DARE

$$P^+_{(A,B)} = A^\top P^+_{(A,B)} A - A^\top P^+_{(A,B)} B(R + B^\top P^+_{(A,B)} B)^{-1} B^\top P^+_{(A,B)} A + Q \tag{3.28}$$

such that

$$K = -(R + B^\top P^+_{(A,B)} B)^{-1} B^\top P^+_{(A,B)} A. \tag{3.29}$$

It is important to note that, although K is independent of the choice of (A,B), the matrix $P^+_{(A,B)}$ might depend on (A,B). We will, however, show that also $P^+_{(A,B)}$ is independent of the choice of (A,B).

By rewriting (3.28), we see that

$$P^+_{(A,B)} - M^\top P^+_{(A,B)} M = K^\top R K + Q. \tag{3.30}$$

Since M is stable, $P^+_{(A,B)}$ is the unique solution to the discrete-time Lyapunov equation (3.30), see e.g. [196, Sec. 6]. Moreover, since M and K do not depend on the choice of $(A,B) \in \Sigma_{\mathrm{i/s}}$, it indeed follows that $P^+_{(A,B)}$ does not depend on (A,B). It follows from (3.28)–(3.30) that $P^+ := P^+_{(A,B)}$ satisfies (3.25)–(3.27). □

The following theorem solves the informativity problem for linear quadratic regulation.

Theorem 3.6. Let $Q = Q^\top$ be positive semidefinite and $R = R^\top$ be positive definite. Then, the data (U_-, X) are informative for linear quadratic regulation if and only if at least one of the following two conditions hold:

(i) The data (U_-, X) are informative for system identification, that is, $\Sigma_{\mathrm{i/s}} = \{(A_s, B_s)\}$, and the linear quadratic regulator problem is solvable for the tuple (A_s, B_s, Q, R). In this case, the optimal feedback gain K is of the form (3.23) where P^+ is the largest real symmetric solution to (3.22).

(ii) For all $(A,B) \in \Sigma_{\mathrm{i/s}}$ we have $A = A_s$. Moreover, A_s is stable, $QA_s = 0$, and the optimal feedback gain is given by $K = 0$.

Remark 3.5. Condition (ii) of Theorem 3.6 is a pathological case in which A is stable and $QA = 0$ for all matrices A that are compatible with the data. Since $x(t) \in \operatorname{im} A$ for all $t > 0$, we have $Qx(t) = 0$ for all $t > 0$ if the input function is chosen as $u = 0$. Additionally, since A is stable, this shows that the optimal input is equal to $u^* = 0$. If we set aside condition (ii), the implication of Theorem 3.6 is the following: if the data are informative for linear quadratic regulation they are also informative for system identification.

At first sight, this might seem like a negative result in the sense that data-driven LQR is only possible with data that are also informative enough to uniquely identify the system. However, at the same time, Theorem 3.6 can be viewed as a positive result in the sense that it provides fundamental justification for the data conditions imposed in e.g. [47]. Indeed, in [47] the data-driven infinite horizon LQR problem[3] is solved using input/state data under the assumption that the input is persistently exciting of sufficiently high order. Under the latter assumption, the input/state data are informative for system identification, i.e., the matrices A_s and B_s can be uniquely determined from data. Theorem 3.6 justifies such a strong assumption on the richness of data in data-driven linear quadratic regulation.

The data-driven *finite* horizon LQR problem was solved under a persistency of excitation assumption in [124]. Our results suggest that also in this case informativity for system identification is necessary for data-driven LQR, although further analysis is required to prove this claim.

Proof of Theorem 3.6. We first prove the "if" part. Sufficiency of the condition (i) readily follows from Theorem 3.5. To prove the sufficiency of the condition (ii), assume that the matrix A is stable and $QA = 0$ for all $(A, B) \in \Sigma_{\mathrm{i/s}}$. By the discussion following Theorem 3.6, this implies that $u^* = 0$ for all $(A, B) \in \Sigma_{\mathrm{i/s}}$. Hence, for $K = 0$ we have $\Sigma_{\mathrm{i/s}} \subseteq \Sigma_K^{Q,R}$, i.e., the data are informative for linear quadratic regulation.

To prove the "only if" part, suppose that the data (U_-, X) are informative for linear quadratic regulation. From Lemma 3.2, we know that there exist M and P^+ satisfying (3.24)–(3.27) for all $(A, B) \in \Sigma_{\mathrm{i/s}}$. By substituting (3.27) into (3.25) and using (3.24), we obtain

$$A^\top P^+ M = P^+ - Q. \tag{3.31}$$

In addition, it follows from (3.27) that $-(R + B^\top P^+ B)K = B^\top P^+ A$. By using (3.24), we have

$$B^\top P^+ M = -RK. \tag{3.32}$$

Since (3.31) and (3.32) hold for all $(A, B) \in \Sigma_{\mathrm{i/s}}$, we have that

$$\begin{bmatrix} A_0^\top \\ B_0^\top \end{bmatrix} P^+ M = 0$$

[3] Note that the authors of [47] formulate this problem as the minimization of the H_2-norm of a certain transfer matrix.

for all $(A_0, B_0) \in \Sigma^0_{\text{i/s}}$. Note that $(FA_0, FB_0) \in \Sigma^0_{\text{i/s}}$ for all $F \in \mathbb{R}^{n\times n}$ whenever $(A_0, B_0) \in \Sigma^0_{\text{i/s}}$. This means that

$$\begin{bmatrix} A_0^\top \\ B_0^\top \end{bmatrix} F^\top P^+ M = 0$$

for all $F \in \mathbb{R}^{n\times n}$. Therefore, either $\begin{bmatrix} A_0 & B_0 \end{bmatrix} = 0$ for all $(A_0, B_0) \in \Sigma^0_{\text{i/s}}$ or $P^+M = 0$. The former is equivalent to $\Sigma^0_{\text{i/s}} = \{0\}$. In this case, we see that the data (U_-, X) are informative for system identification, equivalently $\Sigma_{\text{i/s}} = \{(A_s, B_s)\}$, and the LQR problem is solvable for (A_s, B_s, Q, R). Therefore, condition (i) holds. On the other hand, if $P^+M = 0$ then we have

$$\begin{aligned} 0 = P^+M &= P^+(A + BK) \\ &= P^+\big(A - B(R + B^\top P^+ B)^{-1} B^\top P^+ A\big) \\ &= \big(I - P^+ B(R + B^\top P^+ B)^{-1} B^\top\big) P^+ A. \end{aligned}$$

for all $(A, B) \in \Sigma_{\text{i/s}}$. From the identity

$$(I + P^+ B R^{-1} B^\top)^{-1} = I - P^+ B (R + B^\top P^+ B)^{-1} B^\top,$$

we see that $P^+A = 0$ for all $(A, B) \in \Sigma_{\text{i/s}}$. Then, it follows from (3.27) that $K = 0$. Since $A_0 + B_0 K = 0$ for all $(A_0, B_0) \in \Sigma^0_{\text{i/s}}$ due to Lemma 3.1, we see that A_0 must be zero. Hence, we have $A = A_s$ for all $(A, B) \in \Sigma_{\text{i/s}}$ and A_s is stable. Moreover, it follows from (3.31) that $P^+ = Q$. Therefore, $QA_s = 0$. In other words, condition (ii) is satisfied, which proves the theorem. □

Theorem 3.6 gives necessary and sufficient conditions under which the data are informative for linear quadratic regulation. However, it might not be directly clear how these conditions can be verified given input/state data. Therefore, in what follows we rephrase the conditions of Theorem 3.6 in terms of the data matrices X and U_-.

Theorem 3.7. Let $Q = Q^\top$ be positive semidefinite and $R = R^\top$ be positive definite. Then, the data (U_-, X) are informative for linear quadratic regulation if and only if at least one of the following two conditions hold:

(i) The data (U_-, X) are informative for system identification. Equivalently, there exists $\begin{bmatrix} V_1 & V_2 \end{bmatrix}$ such that (3.9) holds. Moreover, the linear quadratic regulator problem is solvable for (A_s, B_s, Q, R), where $A_s = X_+ V_1$ and $B_s = X_+ V_2$.

(ii) There exists $\Theta \in \mathbb{R}^{T\times n}$ such that $X_-\Theta = (X_-\Theta)^\top$, $U_-\Theta = 0$,

$$\begin{bmatrix} X_-\Theta & X_+\Theta \\ \Theta^\top X_+^\top & X_-\Theta \end{bmatrix} > 0. \tag{3.33}$$

and $QX_+\Theta = 0$.

Proof. The equivalence of condition (i) of Theorem 3.6 and condition (i) of Theorem 3.7 is obvious. It remains to be shown that condition (ii) of Theorem 3.6 and condition (ii) of Theorem 3.7 are equivalent as well. To this end, suppose that there exists a matrix $\Theta \in \mathbb{R}^{T\times n}$ such that the conditions of (ii) holds. By Theorem 3.3, we have $\Sigma_{\mathrm{i/s}} \subseteq \Sigma_K$ for $K = 0$, that is, A is stable for all $(A, B) \in \Sigma_{\mathrm{i/s}}$. In addition, note that

$$QX_+\Theta(X_-\Theta)^{-1} = Q\begin{bmatrix} A & B \end{bmatrix}\begin{bmatrix} X_- \\ U_- \end{bmatrix}\Theta(X_-\Theta)^{-1} = QA \tag{3.34}$$

for all $(A, B) \in \Sigma_{\mathrm{i/s}}$. This shows that $QA = 0$ and therefore that condition (ii) of Theorem 3.6 holds. Conversely, suppose that A is stable and $QA = 0$ for all $(A, B) \in \Sigma_{\mathrm{i/s}}$. This implies that $K = 0$ is a stabilizing controller for all $(A, B) \in \Sigma_{\mathrm{i/s}}$. By Theorem 3.3, there exists a matrix $\Theta \in \mathbb{R}^{T\times n}$ satisfying the first three conditions of (ii). Finally, it follows from $QA = 0$ and (3.34) that Θ also satisfies the fourth equation of (ii). This proves the theorem. □

3.4.3 From data to LQ gain

In this section our goal is to devise a method in order to compute the optimal feedback gain K directly from the data. For this, we will employ ideas from the study of Riccati inequalities (see e.g [173]).

The following theorem asserts that P^+ as in Lemma 3.2 can be found as the unique solution to an optimization problem involving only the data. Furthermore, the optimal feedback gain K can subsequently be found by solving a set of linear equations.

Theorem 3.8. Let $Q = Q^\top \geqslant 0$ and $R = R^\top > 0$. Suppose that the data (U_-, X) are informative for linear quadratic regulation. Consider the linear operator $P \mapsto \mathcal{L}(P)$ defined by

$$\mathcal{L}(P) := X_-^\top P X_- - X_+^\top P X_+ - X_-^\top Q X_- - U_-^\top R U_-.$$

Let P^+ be as in Lemma 3.2. The following statements hold:

(i) The matrix P^+ is equal to the unique solution to the optimization problem

$$\begin{aligned} &\text{maximize} \quad \operatorname{tr} P \\ &\text{subject to} \quad P = P^\top \geqslant 0 \;\text{ and }\; \mathcal{L}(P) \leqslant 0. \end{aligned}$$

(ii) There exists a right inverse $X_-^\dagger$ of X_- such that

$$\mathcal{L}(P^+)X_-^\dagger = 0. \tag{3.35}$$

Moreover, if $X_-^\dagger$ satisfies (3.35), then the optimal feedback gain is given by $K = U_-X_-^\dagger$.

Remark 3.6. From a design viewpoint, the optimal feedback gain K can be found in the following way. First solve the semidefinite program in Theorem 3.8(i). Subsequently, compute a solution $X_-^\dagger$ to the linear equations $X_- X_-^\dagger = I$ and (3.35). Then, the optimal feedback gain is given by $K = U_- X_-^\dagger$.

Remark 3.7. The data-driven LQR problem was also solved using semidefinite programming in [47, Thm. 4]. There, the optimal feedback gain was found by minimizing the trace of a weighted sum of two matrix variables, subject to two LMI constraints. The semidefinite program in Theorem 3.8 is attractive since the dimension of the unknown P is (only) $n \times n$. In comparison, the dimensions of the two unknowns in [47, Thm. 4] are $T \times n$ and $m \times m$, respectively. In general, the number of samples T is much larger[4] than n. An additional attractive feature of Theorem 3.8 is that P^+ is obtained from the data. This is useful since the minimal cost associated to any initial condition x_0 can be computed as $x_0^\top P^+ x_0$.

The data-driven LQR approach in [71] is quite different from Theorem 3.8 since the solution to the Riccati equation is approximated using a batch-form solution to the *Riccati difference equation*. A similar approach was used in [1,62,193,197] for the *finite horizon* data-driven LQR/LQG problem. In the setup of [71], the approximate solution to the Riccati equation is exact only if the number of data points tends to infinity. The main difference between our approach and the one in [71] is hence that the solution P^+ to the Riccati equation can be obtained exactly from *finite* data via Theorem 3.8.

Proof of Theorem 3.8. We begin with proving the first statement. Note that

$$\mathcal{L}(P) = \begin{bmatrix} X_- \\ U_- \end{bmatrix}^\top \begin{bmatrix} P - A^\top PA - Q & -A^\top PB \\ -B^\top PA & -(R + B^\top PB) \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix}$$

for all $(A, B) \in \Sigma_{\mathrm{i/s}}$. We claim that the following implication holds:

$$P = P^\top \geqslant 0 \ \text{ and } \ \mathcal{L}(P) \leqslant 0 \implies P^+ \geqslant P. \tag{3.36}$$

To prove this claim, let P be such that $P = P^\top \geqslant 0$ and $\mathcal{L}(P) \leqslant 0$. Since the data are informative for linear quadratic regulation, they are also informative for stabilization by state feedback. Therefore, the optimal feedback gain K satisfies

$$\operatorname{im} \begin{bmatrix} I \\ K \end{bmatrix} \subseteq \operatorname{im} \begin{bmatrix} X_- \\ U_- \end{bmatrix}$$

due to Lemma 3.1. Therefore, the above expression for $\mathcal{L}(P)$ implies that

$$\begin{bmatrix} I \\ K \end{bmatrix}^\top \begin{bmatrix} P - A^\top PA - Q & -A^\top PB \\ -B^\top PA & -(R + B^\top PB) \end{bmatrix} \begin{bmatrix} I \\ K \end{bmatrix} \leqslant 0$$

for all $(A, B) \in \Sigma_{\mathrm{i/s}}$. This yields

$$P - M^\top PM \leqslant K^\top RK + Q$$

[4] In fact, this is always the case under the persistency of excitation conditions imposed in [47] as such conditions can only be satisfied provided that $T \geqslant nm + n + m$.

where M is as in Lemma 3.2. By subtracting this from (3.26), we obtain

$$(P^+ - P) - M^\top (P^+ - P) M \geqslant 0.$$

Since M is stable, this discrete-time Lyapunov inequality implies that $P^+ - P \geqslant 0$ and hence $P^+ \geqslant P$. This proves the claim (3.36).

Note that $R + B^\top P^+ B$ is positive definite. Then, it follows from (3.25) that

$$\begin{bmatrix} P^+ - A^\top P^+ A - Q & -A^\top P^+ B \\ -B^\top P^+ A & -(R + B^\top P^+ B) \end{bmatrix} \leqslant 0$$

via a Schur complement argument. Therefore, $\mathcal{L}(P^+) \leqslant 0$. Since $P^+ \geqslant P$, we have $\operatorname{tr} P^+ \geqslant \operatorname{tr} P$. Together with (3.36), this shows that P^+ is a solution to the optimization problem stated in the theorem.

Next, we prove uniqueness. Let $\bar{P}$ be another solution of the optimization problem. Then, we have that $\bar{P} = \bar{P}^\top \geqslant 0$, $\mathcal{L}(\bar{P}) \leqslant 0$, and $\operatorname{tr} \bar{P} = \operatorname{tr} P^+$. From (3.36), we see that $P^+ \geqslant \bar{P}$. In particular, this implies that $(P^+)_{ii} \geqslant \bar{P}_{ii}$ for all i. Together with $\operatorname{tr} \bar{P} = \operatorname{tr} P^+$, this implies that $(P^+)_{ii} = \bar{P}_{ii}$ for all i. Now, for any i and j, we have

$$(e_i - e_j)^\top P^+ (e_i - e_j) \geqslant (e_i - e_j)^\top \bar{P} (e_i - e_j) \text{ and}$$
$$(e_i + e_j)^\top P^+ (e_i + e_j) \geqslant (e_i + e_j)^\top \bar{P} (e_i + e_j),$$

where e_i denotes the i-th standard basis vector. This leads to $(P^+)_{ij} \leqslant \bar{P}_{ij}$ and $(P^+)_{ij} \geqslant \bar{P}_{ij}$, respectively. We conclude that $(P^+)_{ij} = \bar{P}_{ij}$ for all i, j. This proves uniqueness.

Finally, we prove the second statement. It follows from (3.25) and (3.27) that

$$\mathcal{L}(P^+) = -\left(U_- - KX_-\right)^\top \left(R + B^\top P^+ B\right) \left(U_- - KX_-\right). \tag{3.37}$$

The optimal feedback K is stabilizing, therefore it follows from Theorem 3.2 that K can be written as $K = U_- \Gamma$, where Γ is some right inverse of X_-. Note that this implies the existence of a right inverse $X_-^\dagger$ of X_- satisfying (3.35). Indeed, $X_-^\dagger := \Gamma$ is such a matrix by (3.37). Moreover, if $X_-^\dagger$ is a right inverse of X_- satisfying (3.35) then $(U_- - KX_-)X_-^\dagger = 0$ by (3.37) and positive definiteness of R. We conclude that the optimal feedback gain is equal to $K = U_- X_-^\dagger$, which proves the second statement. □

3.5 CONTROL USING INPUT AND OUTPUT DATA

In this section, we will consider problems where system outputs play a role. In particular, we will consider the problem of stabilization by dynamic measurement feedback. We will first consider this problem based on input, state and output measurements. Subsequently, we will turn our attention to the case of input/output data.

Consider the true system

$$\begin{aligned} \boldsymbol{x}(t+1) &= A_s\boldsymbol{x}(t) + B_s\boldsymbol{u}(t) && (3.38a)\\ \boldsymbol{y}(t) &= C_s\boldsymbol{x}(t) + D_s\boldsymbol{u}(t). && (3.38b) \end{aligned}$$

We want to design a stabilizing dynamic controller of the form

$$\begin{aligned} \boldsymbol{w}(t+1) &= K\boldsymbol{w}(t) + L\boldsymbol{y}(t) && (3.39a)\\ \boldsymbol{u}(t) &= M\boldsymbol{w}(t) && (3.39b) \end{aligned}$$

such that the closed-loop system, given by

$$\begin{bmatrix} \boldsymbol{x}(t+1) \\ \boldsymbol{w}(t+1) \end{bmatrix} = \begin{bmatrix} A_s & B_sM \\ LC_s & K+LD_sM \end{bmatrix} \begin{bmatrix} \boldsymbol{x}(t) \\ \boldsymbol{w}(t) \end{bmatrix},$$

is stable. This is equivalent to the condition that

$$\begin{bmatrix} A_s & B_sM \\ LC_s & K+LD_sM \end{bmatrix} \tag{3.40}$$

is a stable matrix.

3.5.1 Stabilization using input, state and output data

Suppose that we collect input/state/output data on ℓ time intervals $\{0, 1, \dots, T_i\}$ for $i = 1, 2, \dots, q$. Let U_-, X, X_-, and X_+ be defined as in (3.3) and let Y_- be defined in a similar way as U_-. Then, we have

$$\begin{bmatrix} X_+ \\ Y_- \end{bmatrix} = \begin{bmatrix} A_s & B_s \\ C_s & D_s \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix} \tag{3.41}$$

relating the data and the true system (3.38). The set of all systems that are consistent with these data is then given by:

$$\Sigma_{\mathrm{i/s/o}} := \left\{ (A, B, C, D) \mid \begin{bmatrix} X_+ \\ Y_- \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix} \right\}. \tag{3.42}$$

In addition, for given K, L and M, we define the set of systems that are stabilized by the dynamic controller (3.39) by

$$\Sigma_{K,L,M} := \left\{ (A, B, C, D) \mid \begin{bmatrix} A & BM \\ LC & K+LDM \end{bmatrix} \text{ is stable} \right\}.$$

Subsequently, in line with Definition 3.2, we consider the following notion of informativity:

Definition 3.8. We say the data (U_-, X, Y_-) are *informative for stabilization by dynamic measurement feedback* if there exist matrices K, L and M such that $\Sigma_{\mathrm{i/s/o}} \subseteq \Sigma_{K,L,M}$.

As in the general case of informativity for control, we consider two consequent problems: First, to characterize informativity for stabilization in terms of necessary and sufficient conditions on the data and next to design a controller based on these data. To aid in solving these problems, we will first investigate the case where U_- does not have full row rank. In this case, we will show that the problem can be “reduced" to the full row rank case.

For this, we start with the observation that any $U_- \in \mathbb{R}^{m\times T}$ of row rank $k < m$ can be decomposed as $U_- = S\hat{U}_-$, where S has full column rank and $\hat{U}_- \in \mathbb{R}^{k\times T}$ has full row rank. We now have the following lemma:

Lemma 3.3. Consider the data (U_-, X, Y_-) and the corresponding set $\Sigma_{\mathrm{i/s/o}}$. Let S be a matrix of full column rank such that $U_- = S\hat{U}_-$ with $\hat{U}_-$ a matrix of full row rank. Let $S^\dagger$ be a left inverse of S.

Then the data (U_-, X, Y_-) are informative for stabilization by dynamic measurement feedback if and only if the data $(\hat{U}_-, X, Y_-)$ are informative for stabilization by dynamic measurement feedback.

In particular, if we let $\hat{\Sigma}_{\mathrm{i/s/o}}$ be the set of systems consistent with the “reduced" data set $(\hat{U}_-, X, Y_-)$, and if $\hat{K}$ $\hat{L}$ and $\hat{M}$ are real matrices of appropriate dimensions, then:

$$\Sigma_{\mathrm{i/s/o}} \subseteq \Sigma_{K,L,M} \implies \hat{\Sigma}_{\mathrm{i/s/o}} \subseteq \Sigma_{K,L,S^\dagger M}, \tag{3.43}$$

$$\hat{\Sigma}_{\mathrm{i/s/o}} \subseteq \Sigma_{\hat{K},\hat{L},\hat{M}} \implies \Sigma_{\mathrm{i/s/o}} \subseteq \Sigma_{\hat{K},\hat{L},S\hat{M}}. \tag{3.44}$$

Proof. First note that

$$\hat{\Sigma}_{\mathrm{i/s/o}} = \left\{ (\hat{A}, \hat{B}, \hat{C}, \hat{D}) \mid \begin{bmatrix} X_+ \\ Y_- \end{bmatrix} = \begin{bmatrix} \hat{A} & \hat{B} \\ \hat{C} & \hat{D} \end{bmatrix} \begin{bmatrix} X_- \\ \hat{U}_- \end{bmatrix} \right\}.$$

We will start by proving the following two implications:

$$(A, B, C, D) \in \Sigma_{\mathrm{i/s/o}} \implies (A, BS, C, DS) \in \hat{\Sigma}_{\mathrm{i/s/o}}, \tag{3.45}$$

$$(\hat{A}, \hat{B}, \hat{C}, \hat{D}) \in \hat{\Sigma}_{\mathrm{i/s/o}} \implies (\hat{A}, \hat{B}S^\dagger, \hat{C}, \hat{D}S^\dagger) \in \Sigma_{\mathrm{i/s/o}}. \tag{3.46}$$

To prove implication (3.45), assume that $(A, B, C, D) \in \Sigma_{\mathrm{i/s/o}}$. Then, by definition

$$\begin{bmatrix} X_+ \\ Y_- \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix}.$$

From the definition of S, we have $U_- = S\hat{U}_-$. Substitution of this results in

$$\begin{bmatrix} X_+ \\ Y_- \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X_- \\ S\hat{U}_- \end{bmatrix} = \begin{bmatrix} A & BS \\ C & DS \end{bmatrix} \begin{bmatrix} X_- \\ \hat{U}_- \end{bmatrix}.$$

This implies that $(A, BS, C, DS) \in \hat{\Sigma}_{\mathrm{i/s/o}}$. The implication (3.46) can be proven similarly by substitution of $\hat{U}_- = S^\dagger U_-$.

To prove the lemma, suppose that the data (U_-, X, Y_-) are informative for stabilization by dynamic measurement feedback. This means that there exist K, L, and M such that

$$\begin{bmatrix} A & BM \\ LC & K + LDM \end{bmatrix}$$

is stable for all $(A,B,C,D) \in \Sigma_{\mathrm{i/s/o}}$. In particular, if $(\hat{A},\hat{B},\hat{C},\hat{D}) \in \hat{\Sigma}_{\mathrm{i/s/o}}$ then $(\hat{A},\hat{B}S^\dagger,\hat{C},\hat{D}S^\dagger) \in \Sigma_{\mathrm{i/s/o}}$ by (3.46). This means that the matrix

$$\begin{bmatrix} \hat{A} & \hat{B}S^\dagger M \\ L\hat{C} & K + L\hat{D}S^\dagger M \end{bmatrix}$$

is stable for all $(\hat{A},\hat{B},\hat{C},\hat{D}) \in \hat{\Sigma}_{\mathrm{i/s/o}}$. In other words, $\hat{\Sigma}_{\mathrm{i/s/o}} \subseteq \Sigma_{K,L,S^\dagger M}$ and hence implication (3.43) holds and the data $(\hat{U}_-,X,Y_-)$ are informative for stabilization by dynamic measurement feedback. The proofs of (3.44) and the "if" part of the theorem are analogous and hence omitted. □

We will now solve the informativity and design problems under the condition that U_- has full row rank.

Theorem 3.9. Consider the data (U_-,X,Y_-) and assume that U_- has full row rank. Then (U_-,X,Y_-) are informative for stabilization by dynamic measurement feedback if and only if the following conditions are satisfied:

(i) We have

$$\operatorname{rank} \begin{bmatrix} X_- \\ U_- \end{bmatrix} = n + m.$$

Equivalently, there exists $\begin{bmatrix} V_1 & V_2 \end{bmatrix}$ such that (3.9) holds. This means that

$$\Sigma_{\mathrm{i/s/o}} = \{(X_+V_1, X_+V_2, Y_-V_1, Y_-V_2)\}.$$

(ii) The pair (X_+V_1, X_+V_2) is stabilizable and (Y_-V_1, X_+V_1) is detectable.

Moreover, if the above conditions are satisfied, a stabilizing controller (K,L,M) can be constructed as follows:

(a) Select a matrix M such that $X_+(V_1 + V_2M)$ is stable.

(b) Choose a matrix L such that $(X_+ - LY_-)V_1$ is stable.

(c) Define $K := (X_+ - LY_-)(V_1 + V_2M)$.

Remark 3.8. Under the condition that U_- has full row rank, Theorem 3.9 asserts that in order to construct a stabilizing dynamic controller, it is necessary that the data are rich enough to identify the system matrices A_s, B_s, C_s and D_s uniquely. The controller proposed in (a), (b), (c) is a so-called *observer-based* controller, see e.g. [208, Sec. 3.12]. The feedback gains M and L can be computed using standard methods, for example via pole placement or LMI's.

Proof of Theorem 3.9. To prove the "if" part, suppose that conditions (i) and (ii) are satisfied. This implies the existence of the matrices (K,L,M) as defined in items (a), (b) and (c). We will now show that these matrices indeed constitute a stabilizing controller. Note that by condition (i), $\Sigma_{\mathrm{i/s/o}} = \{(A_s, B_s, C_s, D_s)\}$ with

$$\begin{bmatrix} A_s & B_s \\ C_s & D_s \end{bmatrix} = \begin{bmatrix} X_+V_1 & X_+V_2 \\ Y_-V_1 & Y_-V_2 \end{bmatrix}. \tag{3.47}$$

By definition of K, L and M, the matrices $A_s + B_sM$ and $A_s - LC_s$ are stable and $K = A_s + B_sM - LC_s - LD_sM$. This implies that (3.40) is stable since the matrices

$$\begin{bmatrix} A_s & B_sM \\ LC_s & A_s + B_sM - LC_s \end{bmatrix} \text{ and } \begin{bmatrix} A_s + B_sM & B_sM \\ 0 & A_s - LC_s \end{bmatrix}$$

are similar [208, Sec. 3.12]. We conclude that (U_-, X, Y_-) are informative for stabilization by dynamic measurement feedback and that the recipe given by (a), (b) and (c) leads to a stabilizing controller (K, L, M).

It remains to prove the “only if" part. To this end, suppose that the data (U_-, X, Y_-) are informative for stabilization by dynamic measurement feedback. Let (K, L, M) be such that $\Sigma_{\mathrm{i/s/o}} \subseteq \Sigma_{K,L,M}$. This means that

$$\begin{bmatrix} A & BM \\ LC & K + LDM \end{bmatrix}$$

is stable for all $(A, B, C, D) \in \Sigma_{\mathrm{i/s/o}}$. Let $\zeta \in \mathbb{R}^n$ and $\eta \in \mathbb{R}^m$ be such that

$$\begin{bmatrix} \zeta^\top & \eta^\top \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix} = 0.$$

Note that $(A + \zeta\zeta^\top, B + \zeta\eta^\top, C, D) \in \Sigma_{\mathrm{i/s/o}}$ if $(A, B, C, D) \in \Sigma_{\mathrm{i/s/o}}$. Therefore, the matrix

$$\begin{bmatrix} A & BM \\ LC & K + LDM \end{bmatrix} + \alpha \begin{bmatrix} \zeta\zeta^\top & \zeta\eta^\top M \\ 0 & 0 \end{bmatrix}$$

is stable for all $\alpha \in \mathbb{R}$. We conclude that the spectral radius of the matrix

$$W_\alpha := \frac{1}{\alpha} \begin{bmatrix} A & BM \\ LC & K + LDM \end{bmatrix} + \begin{bmatrix} \zeta\zeta^\top & \zeta\eta^\top M \\ 0 & 0 \end{bmatrix}$$

is smaller than $1/\alpha$. By taking the limit as $\alpha \to \infty$, we see that the spectral radius of $\zeta\zeta^\top$ must be zero due to the continuity of spectral radius. Therefore, ζ must be zero. Since U_- has full column rank, we can conclude that η must be zero too. This proves that condition (i) and therefore $\Sigma_{\mathrm{i/s/o}} = \{(A_s, B_s, C_s, D_s)\}$. Since the controller (K, L, M) stabilizes (A_s, B_s, C_s, D_s), the pair (A_s, B_s) is stabilizable and (C_s, A_s) is detectable. By (3.47) we conclude that condition (ii) is also satisfied. This proves the theorem. □

The following corollary follows from Lemma 3.3 and Theorem 3.9 and gives necessary and sufficient conditions for informativity for stabilization by dynamic measurement feedback. Note that we do not make any a priori assumptions on the rank of U_-.

Corollary 3.2. Let S be any full column rank matrix such that $U_- = S\hat{U}_-$ with $\hat{U}_-$ full row rank k. The data (U_-, X, Y_-) are informative for stabilization by dynamic measurement feedback if and only if the following two conditions are satisfied:

(i) We have

$$\operatorname{rank} \begin{bmatrix} X_- \\ \hat{U}_- \end{bmatrix} = n + k.$$

Equivalently, there exists a matrix $\begin{bmatrix} V_1 & V_2 \end{bmatrix}$ such that

$$\begin{bmatrix} X_- \\ \hat{U}_- \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}.$$

(ii) The pair (X_+V_1, X_+V_2) is stabilizable and (Y_-V_1, X_+V_1) is detectable.

Moreover, if the above conditions are satisfied, a stabilizing controller (K, L, M) is constructed as follows:

(a) Select a matrix $\hat{M}$ such that $X_+(V_1 + V_2\hat{M})$ is stable. Define $M := S\hat{M}$.

(b) Choose a matrix L such that $(X_+ - LY_-)V_1$ is stable.

(c) Define $K := (X_+ - LY_-)(V_1 + V_2\hat{M})$.

Remark 3.9. In the previous corollary it is clear that the system matrices of the data-generating system are related to the data via

$$\begin{bmatrix} A_s & B_sS \\ C_s & D_sS \end{bmatrix} = \begin{bmatrix} X_+ \\ Y_- \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix}.$$

Therefore the corollary shows that informativity for stabilization by dynamic measurement feedback requires that A_s and C_s can be identified uniquely from the data. However, this does not hold for B_s and D_s in general.

3.5.2 Stabilization using input and output data

Recall that we consider a system of the form (3.38). When given input, state and output data, any system (A, B, C, D) consistent with these data satisfies

$$\begin{bmatrix} X_+ \\ Y_- \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix}. \tag{3.48}$$

In this section, we will consider the situation where we have access to input and output measurements only. Moreover, we assume that the data are collected on a single time interval, i.e. $q = 1$. This means that our data are of the form (U_-, Y_-), where

$$U_- := \begin{bmatrix} u(0) & u(1) & \cdots & u(T-1) \end{bmatrix} \tag{3.49a}$$

$$Y_- := \begin{bmatrix} y(0) & y(1) & \cdots & y(T-1) \end{bmatrix}. \tag{3.49b}$$

Again, we are interested in informativity of the data, this time given by (U_-, Y_-). Therefore we wish to consider the set of all systems of the form (3.38) with the state space dimension[5] n that admit the same input/output data. This leads to the following set of consistent systems:

$$\Sigma_{\mathrm{i/o}} := \left\{ (A, B, C, D) \mid \exists X \in \mathbb{R}^{n \times (T+1)} \text{ s.t. } (3.48) \text{ holds} \right\}.$$

As in the previous section, we wish to find a controller of the form (3.39) that stabilizes the system. This means that, in line with Definition 3.2, we have the following notion of informativity:

Definition 3.9. We say the data (U_-, Y_-) are *informative for stabilization by dynamic measurement feedback* if there exist matrices K, L and M such that $\Sigma_{\mathrm{i/o}} \subseteq \Sigma_{K,L,M}$.

In order to obtain conditions under which (U_-, Y_-) are informative for stabilization, it may be tempting to follow the same steps as in Section 3.5.1. In that section we first proved that we can assume without loss of generality that U_- has full row rank. Subsequently, Theorem 3.9 and Corollary 3.2 characterize informativity for stabilization by dynamic measurement feedback based on input, state and output data. It turns out that we can perform the first of these two steps for input/output data as well. Indeed, in line with Lemma 3.3, we can state the following:

Lemma 3.4. Consider the data (U_-, Y_-) and the corresponding set $\Sigma_{\mathrm{i/o}}$. Let S be a matrix of full column rank such that $U_- = S\hat{U}_-$ with $\hat{U}_-$ a matrix of full row rank.

Then the data (U_-, Y_-) are informative for stabilization by dynamic measurement feedback if and only if the data $(\hat{U}_-, Y_-)$ are informative for stabilization by dynamic measurement feedback.

The proof of this lemma is analogous to that of Lemma 3.3 and therefore omitted. Lemma 3.4 implies that without loss of generality we can consider the case where U_- has full row rank.

In contrast to the first step, the second step in Section 3.5.1 relies heavily on the affine structure of the considered set $\Sigma_{\mathrm{i/s/o}}$. Indeed, the proof of Theorem 3.9 makes use of the fact that $\Sigma^0_{\mathrm{i/s/o}}$ is a subspace. However, the set $\Sigma_{\mathrm{i/o}}$ is not an affine set. This means that it is not straightforward to extend the results of Corollary 3.2 to the case of input/output measurements.

Nonetheless, under certain conditions on the input/output data it is possible to construct the corresponding state sequence X of (3.38) up to similarity transformation. In fact, state reconstruction is one of the main themes of subspace identification, see e.g. [143, 217]. The construction of a state sequence would allow us to reduce the problem of stabilization using input/output data to that with input, state and output data. The following result gives sufficient conditions on the data (U_-, Y_-) for state construction.

[5] The state space dimension of the system may be known a priori. In the case that it is not, it can be computed using subspace identification methods, see e.g. [217, Thm. 2].

To state the result, we will first require a bit of notation. First, recall that for a signal $f(0), \dots, f(T-1)$ and $\ell < T$, the *Hankel matrix of depth* ℓ is defined as

$$\mathcal{H}_\ell(f) = \begin{bmatrix} f(0) & f(1) & \cdots & f(T-\ell) \\ f(1) & f(2) & \cdots & f(T-\ell+1) \\ \vdots & \vdots & & \vdots \\ f(\ell-1) & f(\ell) & \cdots & f(T-1) \end{bmatrix}.$$

Given input and output data of the form (3.49), and k such that $2k < T$ we consider $\mathcal{H}_{2k}(u)$ and $\mathcal{H}_{2k}(y)$. Next, we partition our data into so-called *"past"* and *"future"* data as

$$\mathcal{H}_{2k}(u) = \begin{bmatrix} U_p \\ U_f \end{bmatrix}, \quad \mathcal{H}_{2k}(y) = \begin{bmatrix} Y_p \\ Y_f \end{bmatrix},$$

where U_p, U_f, Y_p and Y_f all have k block rows. Let $x(0), \dots, x(T)$ denote the state trajectory of (3.38) compatible with a given (U_-, Y_-). We now denote

$$\begin{aligned} X_p &= \begin{bmatrix} x(0) & \cdots & x(T-2k) \end{bmatrix}, \\ X_f &= \begin{bmatrix} x(k) & \cdots & x(T-k) \end{bmatrix}. \end{aligned}$$

Lastly, let $\operatorname{rs}(M)$ denote the row space of the matrix M. Now we have the following result, which is a rephrasing of [143, Thm. 3].

Theorem 3.10. Consider the system (3.38) and assume it is minimal. Let the input/output data (U_-, Y_-) be as in (3.49). Assume that k is such that $n < k < \frac{1}{2}T$. If

$$\operatorname{rank} \begin{bmatrix} \mathcal{H}_{2k}(u) \\ \mathcal{H}_{2k}(y) \end{bmatrix} = 2km + n, \tag{3.50}$$

then

$$\operatorname{rs}(X_f) = \operatorname{rs}\left(\begin{bmatrix} U_p \\ Y_p \end{bmatrix} \right) \cap \operatorname{rs}\left(\begin{bmatrix} U_f \\ Y_f \end{bmatrix} \right),$$

and this row space is of dimension n.

Under the conditions of this theorem, we can now find the true state sequence X_f up to similarity transformation. That is, we can find $\bar{X} = SX_f$ for some unknown invertible matrix S. This means that, under these conditions, we obtain an input/state/output trajectory given by the matrices

$$\bar{U}_- = \begin{bmatrix} u(k) & u(k+1) & \cdots & u(T-k-1) \end{bmatrix}, \tag{3.51a}$$

$$\bar{Y}_- = \begin{bmatrix} y(k) & y(k+1) & \cdots & y(T-k-1) \end{bmatrix}, \tag{3.51b}$$

$$\bar{X} = S \begin{bmatrix} x(k) & x(k+1) & \cdots & x(T-k) \end{bmatrix}. \tag{3.51c}$$

We can now state the following sufficient condition for informativity for stabilization with input/output data.

Corollary 3.3. Consider the system (3.38) and assume it is minimal. Let the input/output data (U_-, Y_-) be as in (3.49). Assume that k is such that $n < k < \frac{1}{2}T$. Then the data (U_-, Y_-) are informative for stabilization by dynamic measurement feedback if the following two conditions are satisfied:

(i) The rank condition (3.50) holds.

(ii) The data $(\bar{U}_-, \bar{X}, \bar{Y}_-)$, as defined in (3.51), are informative for stabilization by dynamic measurement feedback.

Moreover, if these conditions are satisfied, a stabilizing controller (K, L, M) such that $\Sigma_{\mathrm{i/o}} \subseteq \Sigma_{K,L,M}$ can be found by applying Corollary 3.2 (a),(b),(c) to the data $(\bar{U}_-, \bar{X}, \bar{Y}_-)$.

The conditions provided in Corollary 3.3 are sufficient, but not necessary for informativity for stabilization by dynamic measurement feedback. In addition, it can be shown that data satisfying these conditions are also informative for system identification, in the sense that $\Sigma_{\mathrm{i/o}}$ contains only the true system (3.38) and all systems similar to it.

An interesting question is whether the conditions of Corollary 3.3 can be sharpened to necessary and sufficient conditions. In this case it would be of interest to investigate whether such conditions are weaker than those for informativity for system identification.

At this moment, we do not have a conclusive answer to the above question. However, we note that even for subspace identification there are no known necessary and sufficient conditions for data to be informative, although several sufficient conditions exist, e.g. [143, Thm. 3 and 5], [217, Thm. 2] and [227, Thm. 3 and 4].

3.6 CONCLUSIONS AND FUTURE WORK

Results in data-driven control should clearly highlight the differences and possible advantages as compared to system identification paired with model-based control. One clear advantage of data-driven control is its capability of solving problems in the presence of data that are not informative for system identification. Therefore, informativity is a very important concept for data-driven analysis and control.

In this chapter we have introduced a comprehensive framework for studying informativity problems. We have applied this framework to analyze several system-theoretic properties on the basis of data. The same framework was used to solve multiple data-driven control problems.

After solving these problems, we have made the comparison between our data-driven methods, and the “classical" combination of identification and model-based control. We have shown that for many analysis and control problems, such as controllability analysis and stabilization, the data-driven approach can indeed be performed on data that are not informative for system identification. On the other hand, for data-driven linear quadratic regulation it has been shown that informativity for system identification is a necessary condition. This effectively means that for this data-driven control problem, we have given a theoretic justification for the use of persistently exciting data.

Future work

Due to the generality of the introduced framework, many different problems can be studied in a similar fashion: one could consider different types of data, where more results based on only input and output data would be particularly interesting. Many other system-theoretic properties could be considered as well, for example, analyzing passivity or tackling robust control problems based on data.

It would also be of interest to generalize the model class under consideration. One could, for instance, consider larger classes of systems like differential algebraic or polynomial systems. On the other hand, the class under consideration can also be made smaller by prior knowledge of the system. For example, the system might have an observed network structure, or could in general be parametrized.

A framework similar to ours could be employed in the presence of disturbances, which is a problem of practical interest. A study of data-driven control problems in this situation is particularly interesting, because system identification is less straightforward. We note that data-driven stabilization under measurement noise has been studied in [47] and under unknown disturbances in [17]. Additionally, the data-driven LQR problem is popular in the machine learning community, where it is typically assumed that the system is influenced by (Gaussian) process noise, see e.g. [50].

In this chapter, we have assumed that the data are given. Yet another problem of practical interest is that of *experiment design*, where inputs need to be chosen such that the resulting data are informative. In system identification, this problem led to the notion of persistence of excitation. For example, it is shown in [241] that the rank condition (3.8) can be imposed by injecting an input sequence that is persistently exciting of order $n+1$. However, as we have shown, this rank condition is not necessary for some data-driven control problems, like stabilization by state feedback. The question therefore arises whether we can find tailor-made conditions on the input only, that guarantee informativity for data-driven control.

4 DATA-BASED PARAMETERIZATIONS OF SUBOPTIMAL CONTROLLERS

In the previous chapter we established that the conditions for data informativity are dependent on the particular controller that we want to design. In particular, the conditions for obtaining a linear quadratic regulator (LQR) from data are more stringent than those for stabilization. In this chapter we investigate the middle ground between these two controllers. In particular, we are interested in deriving suboptimal controllers from data. We will focus on the suboptimal LQR and $\mathcal{H}_2$ problems.

4.1 INTRODUCTION

In the field of systems and control, the majority of control techniques is model-based, meaning that these methods require knowledge of a plant model, for example in the form of a transfer function or state-space system. Such system models are rarely known a priori and typically have to be identified using measured data. The aim of (direct) data-driven control is to bypass this system identification step, and to design control laws for dynamical systems directly on the basis of data. Contributions to data-driven control can roughly be divided in on- and offline techniques.

Methods in the former class are iterative and make use of multiple online experiments. Examples include direct adaptive control [7], iterative feedback tuning [85] and methods based on reinforcement learning [4, 23]. Offline techniques construct controllers on the basis of data (typically a single system trajectory) that is collected offline. The paper [197] considers optimal control using a batch-form solution to the Riccati equation. Virtual reference feedback tuning was introduced in [26]. Moreover, the authors of [25] cast the problem of designing model reference controllers in the prediction error framework. The paper [10] designs minimum energy controls using data. The fundamental lemma [241] has also been leveraged for data-driven control in a behavioral setting [125], and in the context of state-space systems to design model predictive controllers [40], stabilizing and optimal controllers [47] and robust controllers [17].

An important persisting problem is to understand the relative merits of data-driven control and combined system identification and model-based control, see e.g. [209]. A recent paper sheds some light on this issue by studying data-driven control from the perspective of *data informativity*. In particular, [221] provides conditions under which given data contain enough information for control design. For control problems such as stabilization, these conditions do not require that the underlying system can be uniquely identified. As such, one can generally stabilize an unknown system without learning its dynamics exactly. For the linear quadratic regulator problem, however, it was shown that the data essentially need to be rich enough for system identification.

Inspired by the above results, it is our goal to study data-driven *suboptimal* control problems. Intuitively, we expect that the data requirements for such suboptimal problems are *weaker* than those for their optimal counterparts. We will focus on data-driven versions of the suboptimal linear quadratic regulator (LQR) problem and the $\mathcal{H}_2$ suboptimal control problem. Both of these problems involve the data-guided design of controllers that stabilize the unknown system and render the (LQR or $\mathcal{H}_2$) cost smaller than a given tolerance.

Our main results are the following. First, for both suboptimal problems, we establish necessary and sufficient conditions under which the data are informative for control design. These conditions do not require that the underlying system can be identified uniquely. Secondly, for both problems we give a parameterization of all suboptimal controllers in terms of data-driven linear matrix inequalities.

This chapter is structured as follows. In Section 4.2 we provide some preliminaries. In Section 4.3 we state the problem. Next, Section 4.4 and Section 4.5 contain our main results. An illustrative example is given in Section 4.6. Finally, Section 4.7 contains our conclusions.

4.2 SUBOPTIMAL CONTROL PROBLEMS

The purpose of this section is to review two (model-based) suboptimal control problems whose data-driven versions will be the main topic of this chapter.

4.2.1 The suboptimal LQR problem

Consider the linear system

$$\boldsymbol{x}(t+1) = A\boldsymbol{x}(t) + B\boldsymbol{u}(t), \tag{4.1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the state, $\boldsymbol{u} \in \mathbb{R}^m$ is the input and A and B are real matrices of appropriate dimensions. We will occasionally use the shorthand notation (A, B) to refer to system (4.1). Associated with (4.1), we consider the infinite-horizon cost functional

$$J(x_0, u) = \sum_{t=0}^{\infty} x^\top(t)Qx(t) + u^\top(t)Ru(t), \tag{4.2}$$

where x_0 is the initial state and $Q = Q^\top \geqslant 0$ and $R = R^\top > 0$ are real matrices. Whenever the input function u results from a state feedback law $\boldsymbol{u} = K\boldsymbol{x}$, we will write $J(x_0, K)$ instead of $J(x_0, u)$. The suboptimal linear quadratic regulator problem can be formulated as follows. Given an initial condition $x_0 \in \mathbb{R}^n$ and tolerance $\gamma > 0$, find (if it exists) a feedback law $\boldsymbol{u} = K\boldsymbol{x}$ such that $A + BK$ is stable, and the cost satisfies $J(x_0, K) < \gamma$. Such a K is called a *suboptimal feedback gain* for the system (A, B). The following proposition gives necessary and sufficient conditions under which a given matrix K is a suboptimal feedback gain.

Proposition 4.1. Let $x_0 \in \mathbb{R}^n$ and $\gamma > 0$. The matrix K is a suboptimal feedback gain if and only if there exists a matrix $P = P^\top > 0$ such that

$$(A+BK)^\top P(A+BK) - P + Q + K^\top RK < 0 \tag{4.3}$$

$$x_0^\top P x_0 < \gamma. \tag{4.4}$$

4.2.2 The $\mathcal{H}_2$ suboptimal control problem

Consider the system

$$\boldsymbol{x}(t+1) = A\boldsymbol{x}(t) + B\boldsymbol{u}(t) + E\boldsymbol{w}(t) \tag{4.5a}$$

$$\boldsymbol{z}(t) = C\boldsymbol{x}(t) + D\boldsymbol{u}(t), \tag{4.5b}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ denotes the state, $\boldsymbol{u} \in \mathbb{R}^m$ is the control input, $\boldsymbol{w} \in \mathbb{R}^d$ is a disturbance input and $\boldsymbol{z} \in \mathbb{R}^p$ is the performance output. The real matrices A, B, C, D and E are of appropriate dimensions. The feedback law $\boldsymbol{u} = K\boldsymbol{x}$ yields the closed-loop system

$$\boldsymbol{x}(t+1) = (A+BK)\boldsymbol{x}(t) + E\boldsymbol{w}(t) \tag{4.6a}$$

$$\boldsymbol{z}(t) = (C+DK)\boldsymbol{x}(t). \tag{4.6b}$$

Associated with (4.6), we consider the $\mathcal{H}_2$ cost functional

$$J_{\mathcal{H}_2}(K) := \sum_{t=0}^{\infty} \operatorname{tr}\left(T_K^\top(t) T_K(t)\right),$$

where $T_K(t) := (C+DK)(A+BK)^t E$ is the closed-loop impulse response from $\boldsymbol{w}$ to $\boldsymbol{z}$ and tr denotes trace. The cost $J_{\mathcal{H}_2}(K)$ equals the squared $\mathcal{H}_2$ norm of the transfer function from $\boldsymbol{w}$ to $\boldsymbol{z}$ of (4.6). It is well-known that the $\mathcal{H}_2$ cost of a given stabilizing K can be computed using the observability Gramian. Indeed for a stabilizing K, the unique solution P to the Lyapunov equation

$$(A+BK)^\top P(A+BK) - P + (C+DK)^\top(C+DK) = 0 \tag{4.7}$$

is related to the $\mathcal{H}_2$ cost by $\operatorname{tr}(E^\top PE) = J_{\mathcal{H}_2}(K)$. For a given $\gamma > 0$, the $\mathcal{H}_2$ suboptimal control problem amounts to finding a gain K (if it exists) such that $A + BK$ is stable and $J_{\mathcal{H}_2}(K) < \gamma$. Such a K is called an $\mathcal{H}_2$ *suboptimal feedback gain*. Similar to Proposition 4.1 the following proposition gives conditions under which a given K is an $\mathcal{H}_2$ suboptimal feedback gain.

Proposition 4.2. Let $\gamma > 0$. The matrix K is an $\mathcal{H}_2$ suboptimal feedback gain if and only if there exists a matrix $P = P^\top > 0$ such that

$$(A+BK)^\top P(A+BK) - P + (C+DK)^\top(C+DK) < 0$$

$$\operatorname{tr}(E^\top PE) < \gamma.$$

Clearly, the LQR suboptimal control problem can be viewed as a special case of the $\mathcal{H}_2$ suboptimal control problem. Indeed, the $\mathcal{H}_2$ problem boils down to the LQR problem if $E = x_0$, $C^\top C = Q$, $D^\top D = R$ and $C^\top D = 0$. However, as we will see in the next section, the data-driven versions of these problems are different in the way that data is collected.

4.3 PROBLEM FORMULATION

In this section we formulate our problems. We will start by introducing the data-driven suboptimal LQR problem. Consider the linear system

$$\boldsymbol{x}(t+1) = A_s \boldsymbol{x}(t) + B_s \boldsymbol{u}(t), \tag{4.8}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ denotes the state, $\boldsymbol{u} \in \mathbb{R}^m$ is the input and A_s and B_s are real matrices of appropriate dimensions. We refer to (4.8) as the "true" system. Suppose that the system matrices A_s and B_s of the true system are unknown, but we have access to a finite set of data[1]

$$\begin{aligned} U_- &:= \begin{bmatrix} u(0) & u(1) & \cdots & u(T-1) \end{bmatrix} \\ X &:= \begin{bmatrix} x(0) & x(1) & \cdots & x(T) \end{bmatrix}, \end{aligned}$$

generated by system (4.8). By partitioning the state data as

$$\begin{aligned} X_- &:= \begin{bmatrix} x(0) & x(1) & \cdots & x(T-1) \end{bmatrix} \\ X_+ &:= \begin{bmatrix} x(1) & x(2) & \cdots & x(T) \end{bmatrix}, \end{aligned}$$

we can relate the data and (A_s, B_s) via

$$X_+ = \begin{bmatrix} A_s & B_s \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix}.$$

The set of all systems that explain the input/state data (U_-, X) is given by

$$\Sigma_{\mathrm{i/s}} := \left\{ (A, B) \mid X_+ = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix} \right\}.$$

Associated with system (4.8) we consider the cost functional (4.2), where the matrices $Q = Q^\top \geqslant 0$ and $R = R^\top > 0$ and the initial condition[2] x_0 are assumed to be given. We want to design a suboptimal feedback gain for the unknown (A_s, B_s) on the basis of the data. Given (U_-, X), it is impossible to distinguish between the systems in $\Sigma_{\mathrm{i/s}}$, and therefore we can only guarantee that K is a suboptimal gain for (A_s, B_s) if it is a suboptimal gain *for all* systems in $\Sigma_{\mathrm{i/s}}$. With this in mind, we introduce the following notion of data informativity.

[1] We assume a single trajectory is measured. Our results are also applicable in case multiple (short) trajectories are measured, which can be beneficial if A_s is unstable [220].

[2] We emphasize that the initial condition x_0 is not necessarily the same as the first measured state sample $x(0)$.

Definition 4.1. Let $x_0 \in \mathbb{R}^n$ and $\gamma > 0$. The data (U_-, X) are *informative for suboptimal linear quadratic regulation* if there exists a matrix K that is a suboptimal feedback gain for all $(A, B) \in \Sigma_{\mathrm{i/s}}$.

We want to find conditions under which the data are informative for suboptimal LQR, and we want to obtain suboptimal controllers from data. These problems are stated more formally as follows.

Problem 4.1. Let $x_0 \in \mathbb{R}^n$ and $\gamma > 0$. Provide necessary and sufficient conditions under which the data (U_-, X) are informative for suboptimal linear quadratic regulation. Moreover, for data (U_-, X) that are informative, find a feedback gain K that is suboptimal for all $(A, B) \in \Sigma_{\mathrm{i/s}}$.

Subsequently, we turn our attention to the $\mathcal{H}_2$ suboptimal control problem. For this, consider the system

$$\boldsymbol{x}(t+1) = A_s \boldsymbol{x}(t) + B_s \boldsymbol{u}(t) + E_s \boldsymbol{w}(t) \tag{4.9}$$

$$\boldsymbol{z}(t) = C\boldsymbol{x}(t) + D\boldsymbol{u}(t), \tag{4.10}$$

where the system matrices A_s, B_s and E_s are unknown, but the matrices C and D defining the performance output are known. We collect the data X and U_- as before, as well as the corresponding measurements of the disturbance

$$W_- := \begin{bmatrix} w(0) & w(1) & \cdots & w(T-1) \end{bmatrix}.$$

The assumption that W_- is available is reasonable in applications such as aircraft control, where gust disturbances can be measured via on-board LIDAR measurement systems, see e.g., [199]. In this setup, all triples of system matrices (A, B, E) that explain the data (U_-, W_-, X) are given by

$$\Sigma_{\mathrm{i/d/s}} := \left\{ (A, B, E) \mid X_+ = \begin{bmatrix} A & B & E \end{bmatrix} \begin{bmatrix} X_- \\ U_- \\ W_- \end{bmatrix} \right\}.$$

We can now state the following notion of data informativity for $\mathcal{H}_2$ suboptimal control.

Definition 4.2. Let $\gamma > 0$. The data (U_-, W_-, X) are *informative for $\mathcal{H}_2$ suboptimal control* if there exists a K that is an $\mathcal{H}_2$ suboptimal feedback gain for all $(A, B, E) \in \Sigma_{\mathrm{i/d/s}}$.

As before, we are interested in both data informativity conditions and a control design procedure. We formalize this in the following problem.

Problem 4.2. Let $\gamma > 0$. Provide necessary and sufficient conditions under which the data (U_-, W_-, X) are informative for $\mathcal{H}_2$ suboptimal control. Moreover, for data (U_-, W_-, X) that are informative, find a feedback gain K that is $\mathcal{H}_2$ suboptimal for all $(A, B) \in \Sigma_{\mathrm{i/s}}$.

Remark 4.1. We note that the data-driven $\mathcal{H}_2$ *optimal* control problem was studied by [47] in the case that $E_s = I$ and (U_-, X) data are collected in the absence of disturbances. Sufficient data conditions were given for this problem via the concept of persistency of excitation. Moreover, [17] aims to design data-driven controllers that minimize a quadratic performance specification (with the $\mathcal{H}_\infty$ problem as a special case). The authors provide sufficient data conditions in the scenario that E is known and w is unmeasured.

4.4 DATA-DRIVEN SUBOPTIMAL LQR

In this section we report our solution to Problem 4.1. Before we start, we need some results from [221]. We say that (U_-, X) are *informative for stabilization by state feedback* if there exists a K such that $A + BK$ is stable for all $(A, B) \in \Sigma_{\mathrm{i/s}}$. The following result was proven in [221, Thm. 16] (see also Lemma 3.1 of this thesis).

Lemma 4.1. The data (U_-, X) are informative for stabilization by state feedback if and only if there exists a right inverse $X_-^\dagger$ of X_- such that $X_+ X_-^\dagger$ is stable.

Moreover, K is a stabilizing feedback for all systems in $\Sigma_{\mathrm{i/s}}$ if and only if $K = U_- X_-^\dagger$ for some $X_-^\dagger$ satisfying the above properties.

Next, we characterize the informativity of data for suboptimal LQR in terms of data-driven matrix inequalities.

Theorem 4.1. Let $x_0 \in \mathbb{R}^n$ and $\gamma > 0$. The data (U_-, X) are informative for suboptimal linear quadratic regulation if and only if there exists a matrix $P = P^\top > 0$ and a right inverse $X_-^\dagger$ of X_- such that

$$(X_+ X_-^\dagger)^\top P X_+ X_-^\dagger - P + Q + (U_- X_-^\dagger)^\top R U_- X_-^\dagger < 0 \tag{4.11}$$

$$x_0^\top P x_0 < \gamma. \tag{4.12}$$

Moreover, K is a suboptimal feedback gain for all systems $(A, B) \in \Sigma_{\mathrm{i/s}}$ if and only if it is of the form $K = U_- X_-^\dagger$ for some right inverse $X_-^\dagger$ satisfying (4.11) and (4.12).

Proof. To prove the "if" parts of both statements, suppose that there exists a matrix $P = P^\top > 0$ and a right inverse $X_-^\dagger$ such that (4.11) and (4.12) are satisfied. Define the controller $K := U_- X_-^\dagger$. For any $(A, B) \in \Sigma_{\mathrm{i/s}}$ we have $X_+ = AX_- + BU_-$, which implies that $X_+ X_-^\dagger = A + BK$. Substitution of the latter expression into (4.11) yields

$$(A + BK)^\top P(A + BK) - P + Q + K^\top RK < 0,$$

which shows that there exists a K and $P = P^\top > 0$ satisfying (4.3) and (4.4) for all $(A, B) \in \Sigma_{\mathrm{i/s}}$. By Proposition 4.1, the data are informative for suboptimal LQR.

To prove the "only if" parts of both statements, suppose that the data (U_-, X) are informative for suboptimal linear quadratic regulation. This means that there exists a feedback gain K and a matrix $P_{(A,B)} = P_{(A,B)}^\top > 0$ such that

$$\begin{aligned}(A+BK)^\top P_{(A,B)}(A+BK) - P_{(A,B)} + Q + K^\top RK < 0\\ x_0^\top P_{(A,B)} x_0 < \gamma\end{aligned}$$

for all $(A, B) \in \Sigma_{\text{i/s}}$. We emphasize that the matrix $P_{(A,B)}$ may depend on the particular system (A, B), but the feedback gain K is fixed by definition. Since K is such that $A + BK$ is stable for all $(A, B) \in \Sigma_{\text{i/s}}$, we obtain by Lemma 4.1 that K is of the form $K = U_- X_-^\dagger$ for some right inverse $X_-^\dagger$ of X_-. This yields $A + BK = X_+ X_-^\dagger$. The matrix $A + BK$ is therefore the same for all $(A, B) \in \Sigma_{\text{i/s}}$. This implies the existence of a (common) $P = P^\top > 0$ such that (4.11) and (4.12) are satisfied. □

Note that the conditions of Theorem 4.1 are not ideal from computational point of view since (4.11) depends nonlinearly on P and $X_-^\dagger$. Nonetheless, it is straightforward to reformulate these conditions in terms of linear matrix inequalities. This is described in the following corollary.

Corollary 4.1. Let $Q = C^\top C$, $R = D^\top D$ and $C^\top D = 0$, and let $x_0 \in \mathbb{R}^n$ and $\gamma > 0$. The data (U_-, X) are informative for suboptimal linear quadratic regulation if and only if there exist $Y = Y^\top \in \mathbb{R}^{n\times n}$ and $\Theta \in \mathbb{R}^{T\times n}$ such that

$$\begin{bmatrix} Y & \Theta^\top X_+^\top & \Theta^\top Z_-^\top \\ X_+\Theta & Y & 0 \\ Z_-\Theta & 0 & I \end{bmatrix} > 0 \tag{4.13}$$

$$\begin{bmatrix} \gamma & x_0^\top \\ x_0 & Y \end{bmatrix} > 0 \tag{4.14}$$

$$X_-\Theta = Y. \tag{4.15}$$

Here $Z_- := CX_- + DU_-$. Moreover, K is a suboptimal feedback gain for all $(A, B) \in \Sigma_{\text{i/s}}$ if and only if $K = U_-\Theta Y^{-1}$ for some Y and Θ satisfying (4.13), (4.14) and (4.15).

Corollary 4.1 follows from Theorem 4.1 via a few well-known tricks, see e.g. [188]. First a congruence transformation P^{-1} is applied to (4.11), after which a Schur complement argument and change of variables $Y := P^{-1}$ and $\Theta := X_-^\dagger Y$ yields (4.13), (4.14) and (4.15).

Remark 4.2. It is noteworthy that the conditions of Theorem 4.1 and Corollary 4.1 do not require that the data (U_-, X) contain enough information to uniquely identify the system matrices (A_s, B_s). Quite naturally, the conditions do become more difficult to satisfy for decreasing values of γ. Clearly, Theorem 4.1 and Corollary 4.1 require the matrix X_- to have full row rank. This means that at least $T \geqslant n$ samples are needed to obtain a suboptimal controller from data. In comparison, note that to uniquely identify A_s and B_s, it is necessary that the rank condition

$$\operatorname{rank}\begin{bmatrix} X_- \\ U_- \end{bmatrix} = n + m$$

is satisfied, which is only possible if $T \geqslant n + m$. In Section 4.6 we will illustrate Corollary 4.1 in detail by numerical examples.

4.5 DATA-DRIVEN $\mathcal{H}_2$ SUBOPTIMAL CONTROL

In this section we study the data-driven $\mathcal{H}_2$ suboptimal control problem as formulated in Problem 4.2. As a first step, we extend Lemma 4.1 to systems with disturbances. We say the data (U_-, W_-, X) are *informative for stabilization by state feedback* if there exists K such that $A + BK$ is stable for all $(A, B, E) \in \Sigma_{\mathrm{i/d/s}}$.

Lemma 4.2. The data (U_-, W_-, X) are informative for stabilization by state feedback if and only if there exists a right inverse $X_-^\dagger$ of X_- with the properties that $X_+ X_-^\dagger$ is stable and $W_- X_-^\dagger = 0$.

Moreover, K is a stabilizing controller for all systems in $\Sigma_{\mathrm{i/d/s}}$ if and only if $K = U_- X_-^\dagger$, where $X_-^\dagger$ satisfies the above properties.

Proof. The proof follows a similar line as that of [221, Thm. 16]. To prove the "if" part of both statements, suppose that there exists a right inverse $X_-^\dagger$ such that $X_+ X_-^\dagger$ is stable and $W_- X_-^\dagger = 0$. Define $K := U_- X_-^\dagger$. Then $X_+ X_-^\dagger = A + BK$ for all $(A, B, E) \in \Sigma_{\mathrm{i/d/s}}$. Hence $A + BK$ is stable for all $(A, B, E) \in \Sigma_{\mathrm{i/d/s}}$ and $K = U_- X_-^\dagger$ is stabilizing.

To prove the "only if" parts, suppose that the data are informative for stabilization by state feedback. Let K be stabilizing for all systems in $\Sigma_{\mathrm{i/d/s}}$. Define the subspace

$$\Sigma^0_{\mathrm{i/d/s}} := \left\{ (A_0, B_0, E_0) \mid 0 = \begin{bmatrix} A_0 & B_0 & E_0 \end{bmatrix} \begin{bmatrix} X_- \\ U_- \\ W_- \end{bmatrix} \right\}.$$

The matrix $A + BK + \alpha(A_0 + B_0 K)$ is stable for all $\alpha \in \mathbb{R}$ and all $(A_0, B_0, E_0) \in \Sigma^0_{\mathrm{i/d/s}}$. Thus we have

$$\rho\left(\frac{1}{\alpha}(A + BK) + A_0 + B_0 K\right) \leqslant \frac{1}{\alpha} \quad \forall\, \alpha \geqslant 1,$$

where $\rho(\cdot)$ denotes spectral radius. We take the limit as $\alpha \to \infty$, and conclude by continuity of the spectral radius that $A_0 + B_0 K$ is nilpotent for all $(A_0, B_0, E_0) \in \Sigma^0_{\mathrm{i/d/s}}$. Note that $(A_0, B_0, E_0) \in \Sigma^0_{\mathrm{i/d/s}}$ implies that

$$\left((A_0 + B_0 K)^\top A_0, (A_0 + B_0 K)^\top B_0, (A_0 + B_0 K)^\top E_0\right)$$

is also a member of $\Sigma^0_{\mathrm{i/d/s}}$. This implies that the matrix $(A_0 + B_0K)^\top(A_0 + B_0K)$ is nilpotent for all (A_0, B_0, E_0). The only symmetric nilpotent matrix is zero, thus $A_0 + B_0K = 0$ for all $(A_0, B_0, E_0) \in \Sigma^0_{\mathrm{i/d/s}}$. We conclude that

$$\ker \begin{bmatrix} X_-^\top & U_-^\top & W_-^\top \end{bmatrix} \subseteq \ker \begin{bmatrix} I & K^\top & 0 \end{bmatrix},$$

equivalently,

$$\operatorname{im} \begin{bmatrix} I \\ K \\ 0 \end{bmatrix} \subseteq \operatorname{im} \begin{bmatrix} X_- \\ U_- \\ W_- \end{bmatrix}.$$

This means that there exists a right inverse $X_-^\dagger$ of X_- such that $K = U_-X_-^\dagger$ and $W_-X_-^\dagger = 0$. Clearly, $X_+X_-^\dagger = A + BK$ for all $(A, B, E) \in \Sigma_{\mathrm{i/d/s}}$, hence $X_+X_-^\dagger$ is stable. □

The following theorem provides necessary and sufficient conditions for data informativity for the $\mathcal{H}_2$ problem. It also characterizes all suboptimal controllers in terms of the data. Recall that Z_- was defined as $Z_- = CX_- + DU_-$.

Theorem 4.2. Let $\gamma > 0$. The data (U_-, W_-, X) are informative for $\mathcal{H}_2$ suboptimal control if and only if at least one of the following two conditions is satisfied:

(i) There exists a right inverse $X_-^\dagger$ such that $X_+X_-^\dagger$ is stable and

$$\begin{bmatrix} W_- \\ Z_- \end{bmatrix} X_-^\dagger = 0.$$

(ii) There exist right inverses $X_-^\dagger$ and $W_-^\dagger$ such that $X_+X_-^\dagger$ is stable, $W_-X_-^\dagger = 0$,

$$\begin{bmatrix} X_- \\ U_- \end{bmatrix} W_-^\dagger = 0,$$

and the unique solution P to

$$(X_-^\dagger)^\top \left(X_+^\top P X_+ - X_-^\top P X_- + Z_-^\top Z_- \right) X_-^\dagger = 0 \tag{4.16}$$

has the property that

$$\operatorname{tr}\left((X_+W_-^\dagger)^\top P X_+ W_-^\dagger \right) < \gamma. \tag{4.17}$$

Moreover, K is an $\mathcal{H}_2$ suboptimal controller for all $(A, B, E) \in \Sigma_{\mathrm{i/d/s}}$ if and only if $K = U_-X_-^\dagger$, where $X_-^\dagger$ satisfies the conditions of (i) or (ii).

Remark 4.3. The interpretation of Theorem 4.2 is as follows. Note that both condition (i) and (ii) require the existence of $X_-^\dagger$ such that $X_+X_-^\dagger$ is stable and $W_-X_-^\dagger = 0$. These conditions are necessary for the existence of a stabilizing controller by Lemma 4.2. In condition (i) it is further required that $X_-^\dagger$ satisfies $Z_-X_-^\dagger = 0$, which means that the output of all systems in $\Sigma_{\mathrm{i/d/s}}$ can be made identically equal to zero (hence the $\mathcal{H}_2$ norm is zero). In condition (ii), the properties of $W_-^\dagger$ imply that $E_s = X_+W_-^\dagger$ can be uniquely identified from the data. Similar to the suboptimal LQR problem, it is generally not required that A_s and B_s can be uniquely identified from the data.

Proof. We first prove the "if" parts of both statements. Suppose that condition (i) is satisfied and let $K := U_-X_-^\dagger$. By Lemma 4.2, $A + BK$ is stable for all $(A, B, E) \in \Sigma_{\mathrm{i/d/s}}$. As $Z_-X_-^\dagger = 0$ we have $C + DU_-X_-^\dagger = C + DK = 0$. This means that the $\mathcal{H}_2$ norm of (4.6) is zero for all $(A, B, E) \in \Sigma_{\mathrm{i/d/s}}$. We conclude that the data are informative for $\mathcal{H}_2$ suboptimal control and K is an $\mathcal{H}_2$ suboptimal controller.

Next suppose that condition (ii) is satisfied, and let $K := U_-X_-^\dagger$ where $X_-^\dagger$ satisfies the conditions of (ii). Clearly, $A + BK = X_+X_-^\dagger$ is stable for all $(A, B, E) \in \Sigma_{\mathrm{i/d/s}}$. By the properties of $W_-^\dagger$, $(A, B, E) \in \Sigma_{\mathrm{i/d/s}}$ implies $E = E_s$. In view of (4.16) and (4.17) we see that for any $(A, B, E_s) \in \Sigma_{\mathrm{i/d/s}}$ the unique solution P to (4.7) satisfies $\operatorname{tr}(E_s^\top PE_s) < \gamma$. Therefore, the data are informative for $\mathcal{H}_2$ suboptimal control and K is $\mathcal{H}_2$ suboptimal.

Subsequently, we prove the "only if" parts of both statements. Suppose that the data are informative for $\mathcal{H}_2$ suboptimal control and let K be an $\mathcal{H}_2$ suboptimal controller for all $(A, B, E) \in \Sigma_{\mathrm{i/d/s}}$. By Lemma 4.2, there exists a right inverse $X_-^\dagger$ such that $X_+X_-^\dagger$ is stable and $W_-X_-^\dagger = 0$. Also, the feedback K is of the form $K = U_-X_-^\dagger$. The solution P to (4.16) exists and is unique by stability of $X_+X_-^\dagger$. The matrix P satisfies $\operatorname{tr}(E^\top PE) < \gamma$ for all $(A, B, E) \in \Sigma_{\mathrm{i/d/s}}$. Therefore, we have

$$\operatorname{tr}\Big((E + \alpha E_0)^\top P(E + \alpha E_0)\Big) < \gamma \tag{4.18}$$

for all $(A, B, E) \in \Sigma_{\mathrm{i/d/s}}$, $(A_0, B_0, E_0) \in \Sigma^0_{\mathrm{i/d/s}}$ and $\alpha \in \mathbb{R}$. We divide both sides of (4.18) by α^2 and take the limit as $\alpha \to \infty$. Then, by continuity of the trace we obtain $\operatorname{tr}(E_0^\top PE_0) = 0$, which yields $PE_0 = 0$ for all $(A_0, B_0, E_0) \in \Sigma^0_{\mathrm{i/d/s}}$. We claim that this implies that either $P = 0$ or $E_0 = 0$ for all $(A_0, B_0, E_0) \in \Sigma^0_{\mathrm{i/d/s}}$. Suppose that this claim is not true. Then $P \neq 0$ and there exists a triple $(A_0, B_0, E_0) \in \Sigma^0_{\mathrm{i/d/s}}$ such that $E_0 \neq 0$. Note that $(FA_0, FB_0, FE_0) \in \Sigma^0_{\mathrm{i/d/s}}$ for any $F \in \mathbb{R}^{n\times n}$. Clearly, there exists an F such that $PFE_0 \neq 0$. This is a contradiction, which proves our claim. Now, in the case that $P = 0$ we obtain $Z_-X_-^\dagger$ and condition (i) is satisfied. In the case that $E_0 = 0$ for all $(A_0, B_0, E_0) \in \Sigma^0_{\mathrm{i/d/s}}$, there exists a right inverse $W_-^\dagger$ such that $X_-W_-^\dagger = 0$ and $U_-W_-^\dagger = 0$. This means that $(A, B, E) \in \Sigma_{\mathrm{i/d/s}}$ implies $E = E_s = X_+W_-^\dagger$. Hence (4.17), and therefore (ii), holds. In both cases, the controller K is of the form $K = U_-X_-^\dagger$, where $X_-^\dagger$ satisfies either (i) or (ii). □

Similar to Corollary 4.1 we can reformulate Theorem 4.2 in terms of linear matrix inequalities using Proposition 4.2.

Corollary 4.2. Let $\gamma > 0$. The data (U_-, W_-, X) are informative for $\mathcal{H}_2$ suboptimal control if and only if at least one of the following two conditions is satisfied:

(i) There exists a $\Theta \in \mathbb{R}^{T\times n}$ such that $X_-\Theta = (X_-\Theta)^\top$,

$$\begin{bmatrix} W_- \\ Z_- \end{bmatrix} \Theta = 0 \text{ and } \begin{bmatrix} X_-\Theta & \Theta^\top X_+^\top \\ X_+\Theta & X_-\Theta \end{bmatrix} > 0.$$

(ii) There exists a right inverse $W_-^\dagger$, a $Y = Y^\top \in \mathbb{R}^{n\times n}$ and $\Theta \in \mathbb{R}^{T\times n}$ such that $X_-\Theta$ is symmetric, the matrices $W_-\Theta$, $X_-W_-^\dagger$ and $U_-W_-^\dagger$ are zero, and

$$\begin{bmatrix} X_-\Theta & \Theta^\top X_+^\top & \Theta^\top Z_-^\top \\ X_+\Theta & X_-\Theta & 0 \\ Z_-\Theta & 0 & I \end{bmatrix} > 0$$

$$\begin{bmatrix} Y & (W_-^\dagger)^\top X_+^\top \\ X_+W_-^\dagger & X_-\Theta \end{bmatrix} > 0$$

$$\operatorname{tr}(Y) < \gamma.$$

Moreover, K is an $\mathcal{H}_2$ suboptimal controller for all $(A, B, E) \in \Sigma_{\mathrm{i/d/s}}$ if and only if $K = U_-\Theta(X_-\Theta)^{-1}$, where Θ satisfies the conditions of (i) or (ii).

4.6 ILLUSTRATIVE EXAMPLE

We study steered consensus dynamics of the form

$$\boldsymbol{x}(t+1) = (I - 0.15L)\,\boldsymbol{x}(t) + B\boldsymbol{u}(t), \tag{4.19}$$

where $\boldsymbol{x} \in \mathbb{R}^{20}$, $\boldsymbol{u} \in \mathbb{R}^{10}$, L is the Laplacian matrix of the graph G in Figure 4.1, and $B = \begin{bmatrix} I & 0 \end{bmatrix}^\top$, meaning that inputs are applied to the first 10 nodes. The goal of this example is to apply the theory from Section 4.4 to construct suboptimal controllers for (4.19) using data. We choose the weight matrices as $Q = I$ and $R = I$, and define $x_0 \in \mathbb{R}^{20}$ entry-wise as $(x_0)_i = i$.

We start with a time horizon of $T = 20$ and collect data (U_-, X) where the entries of U_- and the initial state of the experiment $x(0)$ are drawn uniformly at random from $(0, 1)$. Given these data, we attempt to solve a semidefinite program (SDP) where the objective is to minimize γ subject to the constraints (4.13), (4.14) and (4.15). We use Yalmip, with Mosek as a solver. Next, we collect one additional sample of the input and state, and we solve the SDP again for the augmented data set. We continue this process up to a time horizon of $T = 30$.

We repeat this entire experiment for 100 trials and display the results in Figures 4.2 and 4.3. Figure 4.2 depicts the fraction of successful trials in which the

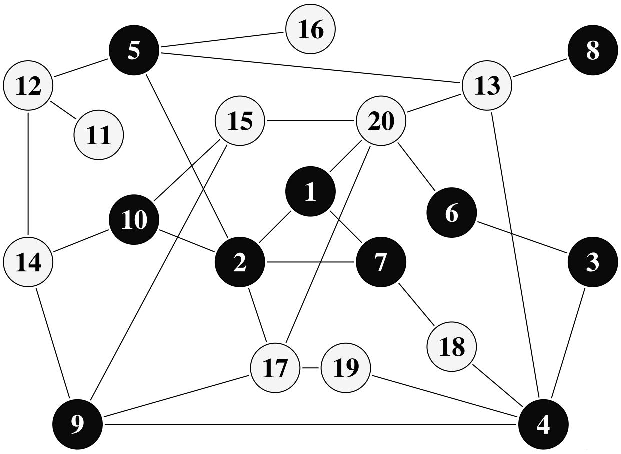

Figure 4.1: Graph G with leader vertices colored black.

constraints (4.13), (4.14) and (4.15) were feasible and a stabilizing controller was found. Note that a stabilizing controller was only found in 2 out of the 100 trials for $T = 20$. This fraction rapidly increases to 0.88 for $T = 22$, while 100% of the trials were successful for $T \geqslant 24$. Figure 4.3 displays the minimum cost γ of the controller, averaged over all successful trials. The cost is very large for small sample size $(T = 20)$ but decreases rapidly as the number of samples increases. Figure 4.3 therefore highlights an interesting trade-off between the sample size and the cost. Note that for $T = 30$, γ coincides with the optimal cost obtained from the (model-based) solution to the Riccati equation. This is as expected since $30 = n + m$ is the minimum number of samples from which the state and input matrices can be uniquely identified.

4.7 CONCLUSIONS

In this chapter we have studied the data-driven suboptimal LQR and $\mathcal{H}_2$ problems. For both problems, we have presented conditions under which a given data set contains sufficient information for control design. We have also given a parameterization of all suboptimal controllers in terms of data-driven linear matrix inequalities. Finally, we have illustrated these results by numerical simulations, which reveal a trade-off between the number of collected data samples and the achieved controller performance.

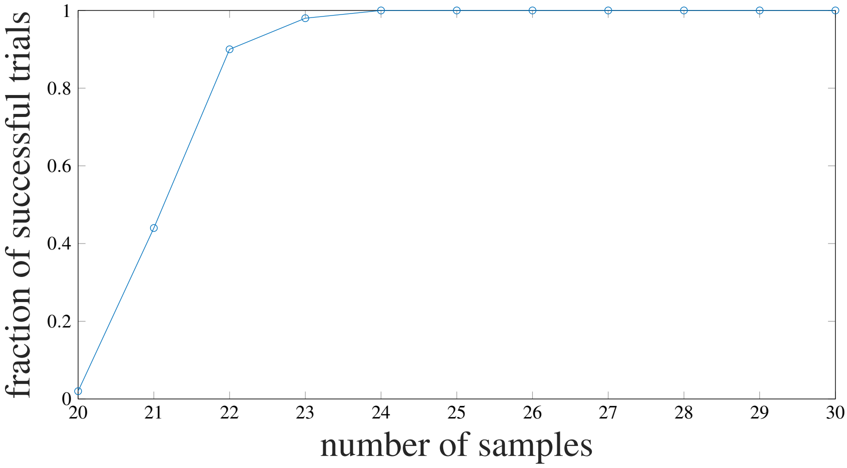

Figure 4.2: Fraction of successful trials as a function of T.

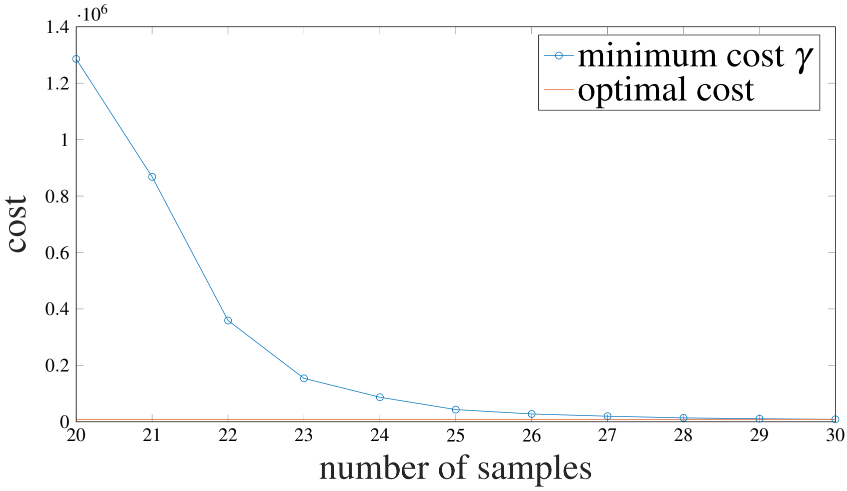

Figure 4.3: Average minimum cost as a function of T.

5 CONTROL FROM NOISY DATA VIA THE MATRIX S-LEMMA

So far, we have solved multiple data-driven analysis and control problems using *noiseless* data. The purpose of this chapter is to further extend these results and to consider control using *noisy* data. It will become apparent that to do so, we need a generalization of the S-lemma to matrix variables. Hence, the contributions of this chapter consist of a new matrix S-lemma, and the application of this result to data-driven control.

5.1 INTRODUCTION

In this chapter we study the problem of designing control laws for an unknown dynamical system using noisy data. This general problem exists for a long time, but has seen a renewed surge of interest over the last few years. The problem can be approached via different angles, for example using combined system identification and model-based control, or by computing control laws from data without the intermediate modeling step. We will contribute to the second category of methods, aiming at control design directly from noisy data.

One of the main challenges in this area is to come up with robust control laws that guarantee stability and performance of the unknown system despite the inherent uncertainty caused by noisy data. Even though there are several recent contributions addressing this issue, there are multiple open questions. In fact, one of the unsolved problems is to come up with *non-conservative* control design strategies using only a finite number of data samples.

We will tackle this problem by providing necessary and sufficient conditions on noisy data under which controllers can be obtained. As a consequence, our ensuing control technique is non-conservative, and also shown to be tractable from a computational point of view. The technical ingredient that enables our design is a new generalization of the classical S-lemma [169, 243], which will be proven in this chapter. We will formulate our control problems using the general data informativity framework as introduced in [221]. As such, the results developed in this chapter can be seen as a natural extension of those in [221] to noisy data.

Literature on data-driven control

The literature on data-driven control is expanding rapidly. Our account of previous work is therefore not exhaustive, but we note that additional references can be found in the survey [89]. We mention contributions to data-driven optimal control [1, 4, 10, 50, 56, 62, 71, 150, 162, 193, 197, 222], PID control [59, 99], predictive control [6, 55, 83, 90, 183], and nonlinear control [21, 44, 75, 203, 204]. Some of these techniques are iterative in nature: the controller is updated online when new data are presented. Examples of this include policy iteration methods [23] and iterative feedback tuning [85]. Other methods are one-shot in the sense that the controller is constructed offline from a batch of data. We mention, for instance, virtual reference feedback tuning [26] and methods based on Willems' fundamental lemma [241] (see also [220]). The latter line of work has been quite fruitful, with contributions ranging from output matching [125] and control by interconnection [132], to data-enabled predictive control [40] and a data-based closed-loop system parameterization [47]. This parameterization has been used for stabilizing and optimal control design using data-based linear matrix inequalities [47]. We also mention the extension [48] studying LQR using noisy data, and the paper [17] for a closed-loop parameterization using noisy data. Additional recent research directions include data-driven control of networks [5, 9] and the interplay between data-guided control and model reduction [140].

Review of the S-lemma

First proven by Yakubovich [243], the *S-lemma* is a classical result in control theory and optimization [169]. The result revolves around the question when the non-negativity of one quadratic function implies that of another one. The crux of the S-lemma is that this -seemingly difficult- implication is equivalent to the feasibility of a linear matrix inequality. The act of replacing the implication by a linear matrix inequality is often referred to as the *S-procedure*. A notable fact about the S-lemma is that the involved quadratic functions are not required to be convex; as such, the result can be interpreted as a non-convex theorem of alternatives.

For reasons that will become clear in Section 5.2, we need a type of S-lemma that is applicable to quadratic functions of *matrix* variables. Such a result has been reported for specific quadratic functions [116, Thm. 3.3], but a general theorem is to the best of our knowledge still missing. Another related result is the *full block S-procedure* [186, 187]. Our matrix S-lemma is different in nature from the full block S-procedure in the sense that it directly generalizes the S-lemma and (like the classical result) also involves a single scalar multiplier. Other differences are the lack of a boundedness assumption in our matrix S-lemma, and the fact that we consider both strict and non-strict inequalities.

Our contributions

The core of our approach is to formulate data-driven control as the problem of deciding whether one quadratic matrix inequality is implied by another one. Our first contribution is to extend the classical S-lemma to quadratic matrix inequalities. Actually, we prove multiple variants of this *matrix S-lemma*, for both strict and non-strict inequalities. Our second contribution is to apply these results to data-driven control. In particular, we come up with non-conservative design procedures for quadratic stabilization, $\mathcal{H}_2$ control and $\mathcal{H}_\infty$ control.

Throughout the chapter we will assume no statistics on the noise, but we will work with general bounded disturbances. We are thereby inspired by recent papers [17,47] that formalize the assumption of bounded disturbances in terms of quadratic matrix inequalities. In fact, we will work with an assumption on the noise that is closely related to that of [17], and is more general than the assumption in [47]. In terms of control design, our approach completely differs from the above papers. In fact, instead of working with data-based parameterizations of closed-loop systems [17,47,48], we will work with a representation of all *open-loop* systems explaining the data, akin to the framework of [221]. We believe that our approach is attractive for the following reasons:

1. We provide robust guarantees on the stability and performance of the unknown data-generating system. The design involves data-guided LMI's that are tractable from a computational point of view and are easy to implement.

2. Our approach is applicable to general bounded disturbances. This is an advantage when the noise does not behave according to a known probability distribution. On the other side of the spectrum, we do note that assumptions like Gaussian noise lead to sample complexity results [50], that are instead more difficult (or even impossible) to derive in the bounded setting.

3. By virtue of our matrix S-lemma, the design method is *non-conservative*. This is in contrast with previous LMI formulations in [17,47] that provide sufficient conditions under which controllers can be obtained from data. In fact, we believe that our result is the first non-conservative control design procedure using a finite number of noisy data samples.

4. Last but not least, the variables involved in our method are independent of the time horizon of the experiment. Our approach is thus applicable to large data sets. This is an advantage over closed-loop system parameterizations [17,47], that become computationally intractable when applied to big data.

Outline of the chapter

In Section 5.2 we will formulate the problem. Section 5.3 contains our results on the matrix S-lemma. These results are then applied to data-driven stabilization in Section 5.4, and to data-driven $\mathcal{H}_2$ control and $\mathcal{H}_\infty$ control in Section 5.5. In Section 5.6 we provide simulation examples. Finally, our conclusions are provided in Section 5.7.

5.2 DATA-DRIVEN STABILIZATION

Consider the linear time-invariant system

$$\boldsymbol{x}(t+1) = A_s\boldsymbol{x}(t) + B_s\boldsymbol{u}(t) + \boldsymbol{w}(t), \tag{5.1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ denotes the state, $\boldsymbol{u} \in \mathbb{R}^m$ is the input and $\boldsymbol{w} \in \mathbb{R}^n$ is an unknown noise term. The matrices $A_s \in \mathbb{R}^{n\times n}$ and $B_s \in \mathbb{R}^{n\times m}$ denote the unknown state and input matrices. Our goal is to design controllers for (5.1) on the basis of a finite number of measurements of the state and input of the system. To this end, suppose that we measure state and input data on a time interval[1], and collect these samples in the matrices

$$\begin{aligned} X &:= \begin{bmatrix} x(0) & x(1) & \cdots & x(T) \end{bmatrix}, \\ U_- &:= \begin{bmatrix} u(0) & u(1) & \cdots & u(T-1) \end{bmatrix}. \end{aligned}$$

By defining the matrices

$$\begin{aligned} X_+ &:= \begin{bmatrix} x(1) & x(2) & \cdots & x(T) \end{bmatrix}, \\ X_- &:= \begin{bmatrix} x(0) & x(1) & \cdots & x(T-1) \end{bmatrix}, \\ W_- &:= \begin{bmatrix} w(0) & w(1) & \cdots & w(T-1) \end{bmatrix}, \end{aligned}$$

we clearly have

$$X_+ = A_sX_- + B_sU_- + W_-. \tag{5.4}$$

We emphasize that the system matrices A_s and B_s as well as the noise term W_- are *unknown*, while X and U_- are measured. Before we introduce the problem we will explain our assumption on the noise W_-.

5.2.1 Assumption on the noise

We will formalize our assumption on the noise in terms of a quadratic matrix inequality.

Assumption 1. The noise samples $w(0), w(1), \ldots, w(T-1)$, collected in the matrix W_-, satisfy the bound

$$\begin{bmatrix} I \\ W_-^\top \end{bmatrix}^\top \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{bmatrix} \begin{bmatrix} I \\ W_-^\top \end{bmatrix} \geqslant 0, \tag{5.5}$$

for known matrices $\Phi_{11} = \Phi_{11}^\top$, Φ_{12} and $\Phi_{22} = \Phi_{22}^\top < 0$.

Note that the negative definiteness of Φ_{22} ensures that the set of noise matrices W_- satisfying (5.5) is bounded. In the special case $\Phi_{12} = 0$ and $\Phi_{22} = -I$, (5.5) reduces to

$$W_-W_-^\top = \sum_{t=0}^{T-1} w(t)w(t)^\top \leqslant \Phi_{11}. \tag{5.6}$$

[1] All our results are still true for data collected on multiple intervals, see [221, Ex. 2] for more details on how to arrange the data matrices in this case.

The inequality (5.6) has the interpretation that the energy of $\boldsymbol{w}$ is bounded on the finite time interval $[0, T-1]$. Note that [47, Ass. 5] is a special case of Assumption 1 for the choices $\Phi_{11} = \gamma X_+ X_+^\top$ with $\gamma > 0$, $\Phi_{12} = 0$ and $\Phi_{22} = -I$. If $\boldsymbol{w}$ is a random variable, its *sample covariance matrix* is given by

$$\frac{1}{T-1} W_- (I - \frac{1}{T} J) W_-^\top,$$

where J is the matrix of ones. Thus, (5.5) can also capture known bounds on the sample covariance by the choices $\Phi_{12} = 0$ and $\Phi_{22} = -\frac{1}{T-1}(I - \frac{1}{T}J)$. We emphasize that in this chapter we do not make any assumptions on the statistics of $\boldsymbol{w}$ and work with the general bound (5.5) instead. We remark that norm bounds on the individual noise samples $w(t)$ also give rise to bounds of the form (5.5), although this may lead to some conservatism. Indeed, note that $\|w(t)\|_2^2 \leqslant \epsilon$ implies that $w(t)w(t)^\top \leqslant \epsilon I$ for all t. As such, the bound (5.6) is satisfied for $\Phi_{11} = T\epsilon I$.

Remark 5.1. Note that the noise model in Assumption 1 is the “transposed" of the model in [17], in the sense that we penalize, e.g., the term $W_- \Phi_{22} W_-^\top$ instead of a term $W_-^\top Q_w W_-$. In some cases, these two different noise models are actually equivalent. For example, if $\Phi_{11} > 0$ and $\Phi_{12} = 0$ then (5.5) can be written via a Schur complement argument as

$$\begin{bmatrix} \Phi_{11} & W_- \\ W_-^\top & -\Phi_{22}^{-1} \end{bmatrix} \geqslant 0.$$

In turn, this is equivalent to $-\Phi_{22}^{-1} - W_-^\top \Phi_{11}^{-1} W_- \geqslant 0$, which is of the same form as [17].

Remark 5.2. In some cases, we may know a priori that the noise $\boldsymbol{w}$ does not directly affect the entire state-space, but is contained in a subspace. This prior knowledge can be captured by the noise model in Assumption 1. Indeed, W_- is of the form $W_- = E\hat{W}_-$ for some $\hat{W}_- \in \mathbb{R}^{r \times T}$ satisfying

$$\begin{bmatrix} I \\ \hat{W}_-^\top \end{bmatrix}^\top \begin{bmatrix} \hat{\Phi}_{11} & \hat{\Phi}_{12} \\ \hat{\Phi}_{12}^\top & \hat{\Phi}_{22} \end{bmatrix} \begin{bmatrix} I \\ \hat{W}_-^\top \end{bmatrix} \geqslant 0$$

if and only if

$$\begin{bmatrix} I \\ W_-^\top \end{bmatrix}^\top \begin{bmatrix} E\hat{\Phi}_{11}E^\top & E\hat{\Phi}_{12} \\ \hat{\Phi}_{12}^\top E^\top & \hat{\Phi}_{22} \end{bmatrix} \begin{bmatrix} I \\ W_-^\top \end{bmatrix} \geqslant 0.$$

Thus, the conclusion is that we can incorporate the knowledge that $W_- \in \operatorname{im} E$ by appropriate choices of the Φ-matrices in (5.5). Showing the above claim is straightforward: note that the “only if" statement follows by pre- and post-multiplication with E and $E^\top$, respectively. The “if" part follows by noting that $x \in \ker E^\top$ implies $x^\top W_- \hat{\Phi}_{22} W_-^\top x \geqslant 0$, thus $W_-^\top x = 0$. Hence, $\ker E^\top \subseteq \ker W_-^\top$, equivalently, $\operatorname{im} W_- \subseteq \operatorname{im} E$.

5.2.2 Problem formulation

We will follow the general framework for data-driven analysis and control in [221]. To this end, we define the set of all systems (A, B) explaining the data (U_-, X), i.e., all (A, B) satisfying

$$X_+ = AX_- + BU_- + W_- \tag{5.7}$$

for some W_- satisfying (5.5). We denote this set by Σ:

$$\Sigma := \{(A, B) \mid (5.7) \text{ holds for some } W_- \text{ satisfying } (5.5)\}.$$

We can only guarantee that a state feedback $\boldsymbol{u} = K\boldsymbol{x}$ stabilizes the true system (A_s, B_s) if it stabilizes *all* systems in Σ. This motivates the following definition of *informative* data. Loosely speaking, data are called informative if they enable the design of a controller that stabilizes all systems in Σ (and thus, the unknown (A_s, B_s)).

Definition 5.1. The data (U_-, X) are called *informative for quadratic stabilization* if there exists a feedback gain K and a matrix $P = P^\top > 0$ such that

$$P - (A + BK)P(A + BK)^\top > 0 \tag{5.8}$$

for all $(A, B) \in \Sigma$.

Note that in particular, we are interested in *quadratic stabilization* and we ask for a *common* Lyapunov matrix P for all $(A, B) \in \Sigma$. We will not treat (A, B)-dependent Lyapunov matrices in this chapter, but consider this case for future work instead.

Definition 5.1 leads to two natural problems. First, we are interested in the question under which conditions the data are informative. We formalize this in the following problem.

Problem 5.1 (Informativity)**.** Find necessary and sufficient conditions under which the data (U_-, X) are informative for quadratic stabilization.

The second problem is a design issue: we are interested in procedures to come up with a feedback that stabilizes all systems in Σ.

Problem 5.2 (Control design)**.** Given informative data (U_-, X), find a stabilizing feedback gain K such that (5.8) is satisfied for all $(A, B) \in \Sigma$.

In addition to data-driven stabilization, we are also interested in including performance specifications. Natural extensions to Problems 5.1 and 5.2 will be discussed in Section 5.5.

5.2.3 Our approach

In what follows, we will outline our strategy for solving Problems 5.1 and 5.2. Let $(A, B) \in \Sigma$ and rewrite (5.7) as

$$W_- = X_+ - AX_- - BU_-. \tag{5.9}$$

Recall that by Assumption 1, we have

$$\begin{bmatrix} I \\ W_-^\top \end{bmatrix}^\top \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{bmatrix} \begin{bmatrix} I \\ W_-^\top \end{bmatrix} \geqslant 0.$$

By substitution of (5.9), this yields

$$\begin{bmatrix} I \\ A^\top \\ B^\top \end{bmatrix}^\top \begin{bmatrix} I & X_+ \\ 0 & -X_- \\ 0 & -U_- \end{bmatrix} \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{bmatrix} \begin{bmatrix} I & X_+ \\ 0 & -X_- \\ 0 & -U_- \end{bmatrix}^\top \begin{bmatrix} I \\ A^\top \\ B^\top \end{bmatrix} \geqslant 0. \tag{5.10}$$

This shows that A and B satisfy a *quadratic matrix inequality* (QMI) of the form (5.10)[2]. In fact, the set Σ of all systems explaining the data can be equivalently characterized in terms of (5.10), as asserted in the following lemma.

Lemma 5.1. We have that $\Sigma = \{(A, B) \mid (5.10) \text{ is satisfied}\}$.

Proof. Suppose that $(A, B) \in \Sigma$. Then (5.9) is satisfied for some W_- satisfying (5.5). This means that (5.10) holds. As such

$$\Sigma \subseteq \{(A, B) \mid (5.10) \text{ is satisfied}\}.$$

To prove the reverse inclusion, let (A, B) be such that (5.10) is satisfied. Define $W_- := X_+ - AX_- - BU_-$. By (5.10), W_- satisfies the assumption (5.5). Since (5.7) holds for (A, B) by construction, we conclude that $(A, B) \in \Sigma$. □

By Lemma 5.1 the set Σ of systems explaining the data is characterized by a quadratic matrix inequality in (A, B). Next, we turn our attention to the design condition (5.8). Suppose that we fix[3] a Lyapunov matrix $P = P^\top > 0$ and a feedback gain K. Note that the inequality (5.8) is equivalent to

$$\begin{bmatrix} I \\ A^\top \\ B^\top \end{bmatrix}^\top \begin{bmatrix} P & 0 & 0 \\ 0 & -P & -PK^\top \\ 0 & -KP & -KPK^\top \end{bmatrix} \begin{bmatrix} I \\ A^\top \\ B^\top \end{bmatrix} > 0, \tag{5.11}$$

which is yet another quadratic matrix inequality in A and B. Therefore, Problem 5.1 essentially boils down to understanding under which conditions the quadratic matrix inequality (5.11) holds for all (A, B) satisfying the quadratic matrix inequality (5.10). Data-driven stabilization thus naturally leads to the following fundamental question:

When does one QMI imply another QMI?

The familiar reader will immediately recognize the similarity between the above question and the statement of the so-called *S-lemma* [169]. In fact, the S-lemma provides conditions under which the non-negativity of one quadratic function implies that of another one. This motivates the following section, in which we generalize the S-lemma to matrix variables.

[2] We note that similar quadratic uncertainty descriptions also arise in the papers [91, 212] on data-driven control, where it is assumed that w is a normally distributed process noise.

[3] We make this hypothesis purely to explain the ideas behind our approach. In fact, in Section 5.4 we show how P and K can be computed from data.

5.3 THE MATRIX-VALUED S-LEMMA

In this section we present a new S-lemma with matrix variables. Before we do so, we provide a brief recap on the classical S-lemma.

5.3.1 Recap of the classical S-lemma

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *quadratic* if it can be written in the form

$$f(x) = \begin{bmatrix} 1 \\ x \end{bmatrix}^\top \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^\top & M_{22} \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}, \tag{5.12}$$

for some $M_{11} \in \mathbb{R}$, $M_{12} \in \mathbb{R}^{1\times n}$ and $M_{22} = M_{22}^\top \in \mathbb{R}^{n\times n}$. A homogeneous quadratic function of the form $f(x) = x^\top M_{22} x$ is called a *quadratic form*. The following theorem describes the celebrated S-lemma, proven by Yakubovich in [243], see also [169, Thm. 2.2].

Theorem 5.1 (S-lemma)**.** Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be quadratic functions. Suppose that there exists $\bar{x} \in \mathbb{R}^n$ such that $g(\bar{x}) > 0$. Then $f(x) \geqslant 0$ for all $x \in \mathbb{R}^n$ such that $g(x) \geqslant 0$ if and only if there exists a scalar $\alpha \geqslant 0$ such that

$$f(x) - \alpha g(x) \geqslant 0 \quad \forall x \in \mathbb{R}^n. \tag{5.13}$$

We note that the functions f and g are not assumed to be convex. As such, it appears to be difficult to check the condition $f(x) \geqslant 0$ for all $x \in \mathbb{R}^n$ satisfying $g(x) \geqslant 0$. The importance of the S-lemma lies in the fact that the characterization (5.13) of this condition is equivalent to a *linear matrix inequality*

$$\begin{bmatrix} M_{11} & M_{12} \\ M_{12}^\top & M_{22} \end{bmatrix} - \alpha \begin{bmatrix} N_{11} & N_{12} \\ N_{12}^\top & N_{22} \end{bmatrix} \geqslant 0$$

in the scalar variable $\alpha \geqslant 0$. Here the matrices $N_{11} \in \mathbb{R}$, $N_{12} \in \mathbb{R}^{1\times n}$ and $N_{22} \in \mathbb{R}^{n\times n}$ define the quadratic function g analogous to (5.12).

The scalar α is called a *multiplier* and the assumption $g(\bar{x}) > 0$ for some $\bar{x} \in \mathbb{R}^n$ is often referred to as the *Slater condition*. This assumption is necessary in the sense that Theorem 5.1 is false without it. To show this by means of an example, one can take, e.g., $f(x) = x^\top A x$ and $g(x) = -x^\top B x$ with A and B as in the example of [250, Page 4476]. A version of the S-lemma where g satisfies a strict inequality has been presented in [169, Thm. 7.8]. We will reformulate the result in the following theorem.

Theorem 5.2 (Strict S-lemma)**.** Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be quadratic forms. Suppose that there exists an $\bar{x} \in \mathbb{R}^n$ such that $g(\bar{x}) > 0$. Then $f(x) \geqslant 0$ for all $x \in \mathbb{R}^n$ such that $g(x) > 0$ if and only if there exists a scalar $\alpha \geqslant 0$ such that

$$f(x) - \alpha g(x) \geqslant 0 \quad \forall x \in \mathbb{R}^n.$$

Note that Theorem 5.2 is stated with two multipliers in [169, Thm. 7.8]. However, the inclusion of the Slater condition allows us to state Theorem 5.2 with a single multiplier α.

5.3.2 S-lemma with matrix variables

Next, we aim at generalizing Theorems 5.1 and 5.2 to quadratic functions of the form

$$\begin{bmatrix} I \\ X \end{bmatrix}^\top \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^\top & M_{22} \end{bmatrix} \begin{bmatrix} I \\ X \end{bmatrix},$$

where $X \in \mathbb{R}^{n\times k}$ is a *matrix variable*, $M_{11} = M_{11}^\top \in \mathbb{R}^{k\times k}$, $M_{12} \in \mathbb{R}^{k\times n}$ and $M_{22} = M_{22}^\top \in \mathbb{R}^{n\times n}$. As our first step, the following theorem provides an S-lemma for homogeneous quadratic functions of the form $X^\top MX$. Naturally, instead of the non-negativity of functions in the classical S-lemma, we now consider the positive (semi)definiteness of quadratic functions of matrix variables.

Theorem 5.3 (Homogeneous matrix S-lemma)**.** Let $M, N \in \mathbb{R}^{n\times n}$ be symmetric matrices and assume that $\bar{X}^\top N\bar{X} > 0$ for some $\bar{X} \in \mathbb{R}^{n\times k}$. The following statements are equivalent:

(i) $X^\top MX \geqslant 0$ for all $X \in \mathbb{R}^{n\times k}$ such that $X^\top NX \geqslant 0$.

(ii) $X^\top MX \geqslant 0$ for all $X \in \mathbb{R}^{n\times k}$ such that $X^\top NX > 0$.

(iii) There exists a scalar $\alpha \geqslant 0$ such that $M - \alpha N \geqslant 0$.

Remark 5.3. The assumption on the existence of $\bar{X}$ such that $\bar{X}^\top N\bar{X} > 0$ is a natural generalization of the Slater condition in Theorems 5.1 and 5.2. The assumption is again necessary in the sense that Theorem 5.3 is false without it. Nonetheless, it can be shown that the assumption can be weakened if one is interested only in the equivalence of (i) and (iii). In fact, one can show using similar arguments as in the proof of Theorem 5.3 that (i) $\iff$ (iii) under the assumption that $\exists \bar{x} \in \mathbb{R}^n$ such that $\bar{x}^\top N\bar{x} > 0$, i.e., under the “standard" Slater condition.

Proof of Theorem 5.3. It is clear that (i) $\implies$ (ii) and (iii) $\implies$ (i). As such, it suffices to prove the implication (ii) $\implies$ (iii). To this end, suppose that (ii) holds. Let $x \in \mathbb{R}^n$ be such that $x^\top Nx > 0$. We want to prove that $x^\top Mx \geqslant 0$ so that we can apply Theorem 5.2. Choose a vector $v \in \mathbb{R}^k$ such that $\|v\| = 1$. Next, we define the matrix $X \in \mathbb{R}^{n\times k}$ as $X := \epsilon\bar{X} + xv^\top$ for $\epsilon \neq 0$. Clearly, $X^\top NX$ is equal to

$$\epsilon^2\bar{X}^\top N\bar{X} + \epsilon\left(\bar{X}^\top Nxv^\top + vx^\top N\bar{X}\right) + (x^\top Nx)vv^\top.$$

We claim that $X^\top NX$ is positive definite for ϵ sufficiently small. To prove this claim, first suppose that $y \in \mathbb{R}^k$ is nonzero and $v^\top y = 0$. Then we obtain

$$y^\top X^\top NXy = \epsilon^2 y^\top\bar{X}^\top N\bar{X}y > 0.$$

Secondly, suppose that $y \in \mathbb{R}^k$ is nonzero and $v^\top y =: \beta \neq 0$. Then $y^\top X^\top NXy$ is equal to

$$y^\top\left(\epsilon^2\bar{X}^\top N\bar{X} + \epsilon\left(\bar{X}^\top Nxv^\top + vx^\top N\bar{X}\right)\right)y + (x^\top Nx)\beta^2,$$

which is positive for ϵ sufficiently small since $\beta \neq 0$ and $x^\top Nx > 0$. We conclude that $X^\top NX > 0$ for ϵ sufficiently small. Now, by (ii) we conclude that $X^\top MX \geqslant 0$. Multiplication of the latter inequality from left by $v^\top$ and right by v yields the inequality

$$\epsilon^2 v^\top \bar{X}^\top M\bar{X}v + \epsilon \left(v^\top \bar{X}^\top Mx + x^\top M\bar{X}v \right) + x^\top Mx \geqslant 0. \tag{5.14}$$

This implies that $x^\top Mx \geqslant 0$. Indeed, if $x^\top Mx < 0$ then there exists a sufficiently small $\epsilon \neq 0$ such that

$$\epsilon^2 v^\top \bar{X}^\top M\bar{X}v + \epsilon \left(v^\top \bar{X}^\top Mx + x^\top M\bar{X}v \right) + x^\top Mx < 0,$$

which contradicts (5.14). To conclude, we have shown that $x^\top Mx \geqslant 0$ for all $x \in \mathbb{R}^n$ such that $x^\top Nx > 0$. By Theorem 5.2, the condition (iii) is satisfied. This proves the theorem. □

Next, we build on Theorem 5.3 by introducing a general (inhomogeneous) S-lemma with matrix variables. The following theorem is one of the main results of this section.

Theorem 5.4 (Matrix S-lemma)**.** Let $M, N \in \mathbb{R}^{(k+n)\times(k+n)}$ be symmetric matrices and assume that there exists some matrix $\bar{Z} \in \mathbb{R}^{n\times k}$ such that

$$\begin{bmatrix} I \\ \bar{Z} \end{bmatrix}^\top N \begin{bmatrix} I \\ \bar{Z} \end{bmatrix} > 0. \tag{5.15}$$

Then the following statements are equivalent:

(I) $\begin{bmatrix} I \\ Z \end{bmatrix}^\top M \begin{bmatrix} I \\ Z \end{bmatrix} \geqslant 0$ for all $Z \in \mathbb{R}^{n\times k}$ such that $\begin{bmatrix} I \\ Z \end{bmatrix}^\top N \begin{bmatrix} I \\ Z \end{bmatrix} \geqslant 0$.

(II) $\begin{bmatrix} I \\ Z \end{bmatrix}^\top M \begin{bmatrix} I \\ Z \end{bmatrix} \geqslant 0$ for all $Z \in \mathbb{R}^{n\times k}$ such that $\begin{bmatrix} I \\ Z \end{bmatrix}^\top N \begin{bmatrix} I \\ Z \end{bmatrix} > 0$.

(III) There exists a scalar $\alpha \geqslant 0$ such that $M - \alpha N \geqslant 0$.

Note that for $k = 1$, the assumption (5.15) reduces to the standard Slater condition. In this case, Theorem 5.4 recovers Theorems 5.1 and 5.2 in the following sense: the equivalence of (I) and (III) is the statement of Theorem 5.1. The equivalence of (II) and (III) generalizes Theorem 5.2 for quadratic forms to general quadratic functions.

Proof of Theorem 5.4. Clearly, (I) $\implies$ (II) and (III) $\implies$ (I). Thus, it suffices to prove that (II) $\implies$ (III). Our strategy will be to show that (II) implies statement (ii) of Theorem 5.3. To this end, suppose that (II) holds and let $X \in \mathbb{R}^{(k+n)\times k}$ be such that $X^\top NX > 0$. Partition X as

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix},$$

where $X_1 \in \mathbb{R}^{k \times k}$ and $X_2 \in \mathbb{R}^{n \times k}$. Clearly, for all sufficiently small $\epsilon > 0$ we have

$$\begin{bmatrix} X_1 + \epsilon I \\ X_2 \end{bmatrix}^\top N \begin{bmatrix} X_1 + \epsilon I \\ X_2 \end{bmatrix} > 0.$$

Also note that $X_1 + \epsilon I$ is nonsingular for all sufficiently small $\epsilon > 0$. This implies that

$$\begin{bmatrix} I \\ X_2(X_1 + \epsilon I)^{-1} \end{bmatrix}^\top N \begin{bmatrix} I \\ X_2(X_1 + \epsilon I)^{-1} \end{bmatrix} > 0.$$

By (II), we have

$$\begin{bmatrix} I \\ X_2(X_1 + \epsilon I)^{-1} \end{bmatrix}^\top M \begin{bmatrix} I \\ X_2(X_1 + \epsilon I)^{-1} \end{bmatrix} \geqslant 0,$$

equivalently,

$$\begin{bmatrix} X_1 + \epsilon I \\ X_2 \end{bmatrix}^\top M \begin{bmatrix} X_1 + \epsilon I \\ X_2 \end{bmatrix} \geqslant 0$$

for all $\epsilon > 0$ sufficiently small. By taking the limit $\epsilon \downarrow 0$ we conclude that $X^\top M X \geqslant 0$. Therefore, statement (ii) (equivalently, statement (iii)) of Theorem 5.3 is satisfied. This means that (III) holds, which proves the theorem. □

As a special case of Theorem 5.4 we recover the following result by Luo, Sturm and Zhang.

Corollary 5.1 (Theorem 3.3 of [116]). The quadratic matrix inequality

$$M_{11} + M_{12}Z + Z^\top M_{12}^\top + Z^\top M_{22} Z \geqslant 0$$

holds for all $Z \in \mathbb{R}^{n \times k}$ satisfying $I - Z^\top D Z \geqslant 0$ if and only if there exists a scalar $\alpha \geqslant 0$ such that

$$\begin{bmatrix} M_{11} & M_{12} \\ M_{12}^\top & M_{22} \end{bmatrix} - \alpha \begin{bmatrix} I & 0 \\ 0 & -D \end{bmatrix} \geqslant 0.$$

Proof. Note that the generalized Slater condition (5.15) is satisfied (one can choose e.g., $\bar{Z} = 0$). Thus, the statement follows from Theorem 5.4. □

Theorem 5.4 provides a natural generalization of the S-lemma to matrix variables. However, note that for the application that we have in mind, we need a slightly different version of the theorem. Indeed, note that in the data-driven context of Section 5.2, a *strict* inequality (5.11) must hold for all (A, B) satisfying a non-strict inequality (5.10). As such, we need to extend Theorem 5.4 to the case when the inequality involving M is *strict*. Before we do so we introduce the shorthand notation

$$\mathcal{S}_N := \left\{ Z \in \mathbb{R}^{n \times k} \mid \begin{bmatrix} I \\ Z \end{bmatrix}^\top N \begin{bmatrix} I \\ Z \end{bmatrix} \geqslant 0 \right\}.$$

Theorem 5.5 (Strict matrix S-lemma)**.** Let M and N by symmetric matrices in $\mathbb{R}^{(k+n)\times(k+n)}$. Assume that $\mathcal{S}_N$ is bounded and that there exists some matrix $\bar{Z} \in \mathbb{R}^{n\times k}$ satisfying (5.15). Then we have that

$$\begin{bmatrix} I \\ Z \end{bmatrix}^\top M \begin{bmatrix} I \\ Z \end{bmatrix} > 0 \text{ for all } Z \in \mathcal{S}_N \tag{5.16}$$

if and only if there exists $\alpha \geqslant 0$ such that $M - \alpha N > 0$.

Proof. The "if" part is clear, so we focus on proving the "only if" part. Suppose that (5.16) holds. We claim that there exists an $\epsilon > 0$ such that

$$\begin{bmatrix} I \\ Z \end{bmatrix}^\top (M - \epsilon I) \begin{bmatrix} I \\ Z \end{bmatrix} > 0 \text{ for all } Z \in \mathcal{S}_N. \tag{5.17}$$

Suppose that this is not the case. Then there exists a sequence $\{\epsilon_i\}$ with $\epsilon_i \to 0$ as $i \to \infty$ with the property that for each i there exists $Z_i \in \mathcal{S}_N$ such that

$$\begin{bmatrix} I \\ Z_i \end{bmatrix}^\top (M - \epsilon_i I) \begin{bmatrix} I \\ Z_i \end{bmatrix} \not> 0.$$

Since $\mathcal{S}_N$ is bounded, the sequence $\{Z_i\}$ is bounded. As such, by the Bolzano-Weierstrass theorem, it contains a converging subsequence with limit, say, Z^*. We conclude that

$$\begin{bmatrix} I \\ Z^* \end{bmatrix}^\top M \begin{bmatrix} I \\ Z^* \end{bmatrix} \not> 0.$$

Note that $\mathcal{S}_N$ is closed and thus $Z^* \in \mathcal{S}_N$. Since (5.16) holds we arrive at a contradiction. Therefore, we conclude that there exists an $\epsilon > 0$ such that (5.17) holds. In particular, this implies the existence of $\epsilon > 0$ such that

$$\begin{bmatrix} I \\ Z \end{bmatrix}^\top (M - \epsilon I) \begin{bmatrix} I \\ Z \end{bmatrix} \geqslant 0 \text{ for all } Z \in \mathcal{S}_N.$$

Now, by Theorem 5.4 there exists an $\alpha \geqslant 0$ such that

$$(M - \epsilon I) - \alpha N \geqslant 0.$$

We conclude that $M - \alpha N > 0$ which proves the theorem. □

It turns out that we can even state Theorem 5.5 without the boundedness assumption if some more structure on the matrices M and N is given. In what follows we partition M and N in the natural way as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^\top & M_{22} \end{bmatrix}, \quad N = \begin{bmatrix} N_{11} & N_{12} \\ N_{12}^\top & N_{22} \end{bmatrix}. \tag{5.18}$$

We then have the following result.

Theorem 5.6. Let $M, N \in \mathbb{R}^{(k+n)\times(k+n)}$ by symmetric matrices, partitioned as in (5.18). Assume that $M_{22} \leqslant 0$, $N_{22} \leqslant 0$ and $\ker N_{22} \subseteq \ker N_{12}$. Suppose that there exists some matrix $\bar{Z} \in \mathbb{R}^{n\times k}$ satisfying (5.15). Then we have that

$$\begin{bmatrix} I \\ Z \end{bmatrix}^\top M \begin{bmatrix} I \\ Z \end{bmatrix} > 0 \text{ for all } Z \in \mathcal{S}_N \tag{5.19}$$

if and only if there exist $\alpha \geqslant 0$ and $\beta > 0$ such that

$$M - \alpha N \geqslant \begin{bmatrix} \beta I & 0 \\ 0 & 0 \end{bmatrix}.$$

Proof. The "if" part is clear so we focus on proving the "only if" statement. Suppose that (5.19) holds. We will first prove that $\ker N_{22} \subseteq \ker M_{22}$ and $\ker N_{22} \subseteq \ker M_{12}$. Let $Z \in \mathcal{S}_N$ and $\hat{Z} \in \mathbb{R}^{n\times k}$ be such that $N_{22}\hat{Z} = 0$. By the hypothesis $\ker N_{22} \subseteq \ker N_{12}$ we have $Z + \gamma\hat{Z} \in \mathcal{S}_N$ for any $\gamma \in \mathbb{R}$. Thus, we obtain

$$\begin{bmatrix} I \\ Z \end{bmatrix}^\top M \begin{bmatrix} I \\ Z \end{bmatrix} + \gamma(M_{12}\hat{Z} + (M_{12}\hat{Z})^\top) + \gamma^2\hat{Z}^\top M_{22}\hat{Z} > 0. \tag{5.20}$$

This implies that $M_{22}\hat{Z} = 0$. Indeed, recall that $M_{22} \leqslant 0$. Thus, if $M_{22}\hat{Z} \neq 0$ then there exists a sufficiently large γ such that (5.20) is violated. Similarly, we conclude that $M_{12}\hat{Z} = 0$. Therefore, we have shown that

$$\ker N_{22} \subseteq \ker M_{22} \text{ and } \ker N_{22} \subseteq \ker M_{12}. \tag{5.21}$$

Subsequently, we claim that there exists a $\beta > 0$ such that

$$\begin{bmatrix} I \\ Z \end{bmatrix}^\top \left(M - \begin{bmatrix} \beta I & 0 \\ 0 & 0 \end{bmatrix} \right) \begin{bmatrix} I \\ Z \end{bmatrix} > 0 \text{ for all } Z \in \mathcal{S}_N. \tag{5.22}$$

If this claim is not true, then there exists a sequence $\{\beta_i\}$ such that $\beta_i \to 0$ and for all i there exists $Z_i \in \mathcal{S}_N$ such that

$$\begin{bmatrix} I \\ Z_i \end{bmatrix}^\top \left(M - \begin{bmatrix} \beta_i I & 0 \\ 0 & 0 \end{bmatrix} \right) \begin{bmatrix} I \\ Z_i \end{bmatrix} \not> 0. \tag{5.23}$$

Define $\mathcal{V} := \{Z \in \mathbb{R}^{n\times k} \mid N_{22}Z = 0\}$. Write Z_i as $Z_i = Z_i^1 + Z_i^2$, where $Z_i^1 \in \mathcal{V}^\perp$ and $Z_i^2 \in \mathcal{V}$. By the hypothesis $\ker N_{22} \subseteq \ker N_{12}$ we see that $Z_i^1 \in \mathcal{S}_N$. Next, we claim that the sequence $\{Z_i^1\}$ is bounded. We will prove this claim by contradiction. Thus, suppose that $\{Z_i^1\}$ is unbounded. Clearly, the sequence

$$\left\{ \frac{Z_i^1}{\|Z_i^1\|} \right\}$$

is bounded. By the Bolzano-Weierstrass theorem it thus has a convergent subsequence with limit, say Z_*. Note that

$$\frac{1}{\|Z_i^1\|^2} \left(N_{11} + N_{12}Z_i^1 + (N_{12}Z_i^1)^\top + (Z_i^1)^\top N_{22}Z_i^1 \right) \geqslant 0.$$

By taking the limit along the subsequence as $i \to \infty$, we get $Z_*^\top N_{22} Z_* \geqslant 0$. Using the fact that $N_{22} \leqslant 0$ we conclude that $Z_* \in \mathcal{V}$. Since $Z_i^1 \in \mathcal{V}^\perp$ for all i, also $\frac{Z_i^1}{\|Z_i^1\|} \in \mathcal{V}^\perp$ and thus $Z_* \in \mathcal{V}^\perp$. Therefore, we conclude that both $Z_* \in \mathcal{V}$ and $Z_* \in \mathcal{V}^\perp$, i.e., $Z_* = 0$. This is a contradiction since $\frac{Z_i^1}{\|Z_i^1\|}$ has norm 1 for all i. We conclude that the sequence $\{Z_i^1\}$ is bounded. It thus contains a convergent subsequence with limit, say Z_∞. Note that $\mathcal{S}_N$ is closed and thus $Z_\infty \in \mathcal{S}_N$. By (5.21) and (5.23) we conclude that

$$\begin{bmatrix} I \\ Z_i^1 \end{bmatrix}^\top \left(M - \begin{bmatrix} \beta_i I & 0 \\ 0 & 0 \end{bmatrix} \right) \begin{bmatrix} I \\ Z_i^1 \end{bmatrix} \not> 0$$

for all i. We take the limit as $i \to \infty$, which yields

$$\begin{bmatrix} I \\ Z_\infty \end{bmatrix}^\top M \begin{bmatrix} I \\ Z_\infty \end{bmatrix} \not> 0.$$

As $Z_\infty \in \mathcal{S}_N$ this contradicts (5.19). As such, we conclude that there exists $\beta > 0$ such that (5.22) holds. In particular, there exists $\beta > 0$ such that

$$\begin{bmatrix} I \\ Z \end{bmatrix}^\top \left(M - \begin{bmatrix} \beta I & 0 \\ 0 & 0 \end{bmatrix} \right) \begin{bmatrix} I \\ Z \end{bmatrix} \geqslant 0 \text{ for all } Z \in \mathcal{S}_N.$$

The theorem now follows by application of Theorem 5.4. □

5.4 DATA-DRIVEN STABILIZATION REVISITED

In this section, we apply the theory from Section 5.3 to data-driven stabilization, i.e., to Problems 5.1 and 5.2 defined in Section 5.2. To this end, for given $P = P^\top > 0$ and K we define the partitioned matrices

$$M = \left[\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{12}^\top & M_{22} \end{array}\right] := \left[\begin{array}{c|cc} P & 0 & 0 \\ \hline 0 & -P & -PK^\top \\ 0 & -KP & -KPK^\top \end{array}\right], \tag{5.24}$$

$$\begin{aligned} N &= \left[\begin{array}{c|c} N_{11} & N_{12} \\ \hline N_{12}^\top & N_{22} \end{array}\right] \\ &:= \left[\begin{array}{cc} I & X_+ \\ \hline 0 & -X_- \\ 0 & -U_- \end{array}\right] \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{bmatrix} \left[\begin{array}{cc} I & X_+ \\ \hline 0 & -X_- \\ 0 & -U_- \end{array}\right]^\top . \end{aligned} \tag{5.25}$$

Recall from Section 5.2 that data-driven stabilization entails deciding whether (5.11) holds for all (A, B) satisfying (5.10). In terms of the matrices M and N as defined above, we thus have to decide whether

$$\begin{bmatrix} I \\ Z \end{bmatrix}^\top M \begin{bmatrix} I \\ Z \end{bmatrix} > 0 \quad \text{for all } Z \in \mathbb{R}^{(n+m)\times n} \text{ such that } \begin{bmatrix} I \\ Z \end{bmatrix}^\top N \begin{bmatrix} I \\ Z \end{bmatrix} \geqslant 0. \tag{5.26}$$

Here Z is given by

$$Z := \begin{bmatrix} A^\top \\ B^\top \end{bmatrix}.$$

The idea is now to apply Theorem 5.6. To this end, we have to verify its assumptions. In particular, we will check that $M_{22} \leqslant 0$, $N_{22} \leqslant 0$ and $\ker N_{22} \subseteq \ker N_{12}$. Note that

$$M_{22} = -\begin{bmatrix} I \\ K \end{bmatrix} P \begin{bmatrix} I \\ K \end{bmatrix}^\top \leqslant 0, \; N_{22} = \begin{bmatrix} X_- \\ U_- \end{bmatrix} \Phi_{22} \begin{bmatrix} X_- \\ U_- \end{bmatrix}^\top \leqslant 0,$$

because $P > 0$ and $\Phi_{22} < 0$. Since Φ_{22} is nonsingular, we also see that

$$\begin{aligned} \ker N_{22} &= \ker \begin{bmatrix} X_- \\ U_- \end{bmatrix}^\top, \\ \ker N_{12} &= \ker \left((\Phi_{12} + X_+ \Phi_{22}) \begin{bmatrix} X_- \\ U_- \end{bmatrix}^\top \right), \end{aligned}$$

and thus $\ker N_{22} \subseteq \ker N_{12}$. We conclude that the assumptions of Theorem 5.6 are satisfied. We assume that the generalized Slater condition (5.15) holds for N in (5.25). Then, Theorem 5.6 asserts that (5.26) holds if and only if there exist scalars $\alpha \geqslant 0$ and $\beta > 0$ such that

$$M - \alpha N \geqslant \begin{bmatrix} \beta I & 0 \\ 0 & 0 \end{bmatrix}. \tag{5.27}$$

From a design point of view, the matrices P and K that appear in M are not given. However, the idea is now to *compute* matrices P, K and scalars α and β such that (5.27) holds. In fact, by the above discussion, the data (U_-, X) are informative for quadratic stabilization *if and only if* there exists an $n \times n$ matrix $P = P^\top > 0$, a $K \in \mathbb{R}^{m \times n}$ and two scalars $\alpha \geqslant 0$ and $\beta > 0$ such that (5.27) holds. We note that (5.27) (in particular, M) is not linear in P and K. Nonetheless, by a rather standard change of variables and a Schur complement argument, we can transform (5.27) into a linear matrix inequality. We summarize our progress in the following theorem, which is the main result of this section.

Theorem 5.7. Assume that the generalized Slater condition (5.15) holds for N in (5.25) and some $\bar{Z} \in \mathbb{R}^{(n+m) \times n}$. Then the data (U_-, X) are informative for quadratic stabilization if and only if there exists an $n \times n$ matrix $P = P^\top > 0$, an $L \in \mathbb{R}^{m \times n}$ and scalars $\alpha \geqslant 0$ and $\beta > 0$ satisfying

$$\begin{bmatrix} P - \beta I & 0 & 0 & 0 \\ 0 & -P & -L^\top & 0 \\ 0 & -L & 0 & L \\ 0 & 0 & L^\top & P \end{bmatrix} - \alpha \begin{bmatrix} I & X_+ \\ 0 & -X_- \\ 0 & -U_- \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{bmatrix} \begin{bmatrix} I & X_+ \\ 0 & -X_- \\ 0 & -U_- \\ 0 & 0 \end{bmatrix}^\top \geqslant 0. \tag{FS}$$

Moreover, if P and L satisfy (FS) then $K := LP^{-1}$ is a stabilizing feedback gain for all $(A, B) \in \Sigma$.

Proof. To prove the "if" statement, suppose that there exist P, L, α and β satisfying (FS). Define $K := LP^{-1}$. By computing the Schur complement of (FS) with respect to its fourth diagonal block, we obtain (5.27). As such, (5.26) holds. We conclude that the data (U_-, X) are informative for quadratic stabilization and $K = LP^{-1}$ is indeed a stabilizing controller for all $(A, B) \in \Sigma$.

Conversely, to prove the "only if" statement, suppose that the data (U_-, X) are informative for quadratic stabilization. This means that there exist $P = P^\top > 0$ and K such that (5.26) holds. By Theorem 5.6 there exist $\alpha \geqslant 0$ and $\beta > 0$ satisfying (5.27). Finally, by defining $L := KP$ and using a Schur complement argument, we conclude that (FS) is feasible. □

Theorem 5.7 provides a powerful necessary *and* sufficient condition under which quadratically stabilizing controllers can be obtained from noisy data. The assumption (5.15) puts a mild condition on the data matrices appearing in (5.25). It is satisfied whenever N has at least n positive eigenvalues, a condition that is simple to verify from given data. So far, this condition was satisfied in all of our numerical experiments, see Section 5.6 for more details[4]. Theorem 5.7 leads to an effective design procedure for obtaining stabilizing controllers directly from data. Indeed, the approach entails solving the linear matrix inequality (FS) for P, L, α and β and computing a controller as $K = LP^{-1}$. Before we prove Theorem 5.7 we discuss some of the features of our control design procedure.

1. First of all, we stress that the procedure is *non-conservative* since Theorem 5.7 provides a necessary and sufficient condition for obtaining quadratically stabilizing controllers from data. To the best of our knowledge, this is the first non-conservative design procedure for quadratic stabilization from a finite number of noisy data samples.

2. We believe that our approach based on the set Σ of *open-loop* systems provides a valuable alternative to the data-based closed-loop system parameterizations of [47, Thm. 2] and [17, Thm. 4]. Indeed, in the case of noisy data, it was recognized that certain linear constraints [17, Eq. (3)] defining these closed-loop systems were difficult to incorporate in the control design[5]. Our design procedure does not suffer from the above problem. In fact, the constraint [17, Eq. (3)] is *automatically incorporated* in our control design approach.

3. The variables P, L, α and β are *independent* of the time horizon T of the experiment. In fact, note that $P \in \mathbb{R}^{n \times n}$, $L \in \mathbb{R}^{m \times n}$ and $\alpha, \beta \in \mathbb{R}$. Also, the LMI (FS) is of dimension $(3n + m) \times (3n + m)$ and thus independent of T. As such, our approach fundamentally differs from the design methods in [17, 47] where certain decision variables have dimension $T \times n$, c.f. [47, Thm. 6] and [17, Cor. 6]. We believe that our T-independent design method

[4] In addition, we remark that even if the generalized Slater condition does not hold, the "if" statement of Theorem 5.7 remains true.

[5] In fact, it was mentioned in [101] that involving the condition [17, Eq. (3)] in design procedures is still an open problem.

will play a crucial role in control design from larger data sets. We note that the collection of big data sets is often unavoidable, for example because the signal-to-noise ratio is small, or because the data-generating system is large-scale.

Remark 5.4. We note that under the extra assumption

$$\operatorname{rank} \begin{bmatrix} X_- \\ U_- \end{bmatrix} = n + m \tag{5.28}$$

it is possible to prove a variant Theorem 5.7 in which the non-strict inequality is replaced by a strict inequality, and the term $-\beta I$ is removed. This can be done by invoking Theorem 5.5, which is possible since (5.28) implies that the set Σ is bounded. The reason is that the coefficient matrix N_{22} defining the quadratic term in (5.10) is negative definite if (5.28) holds.

Here we chose to state and prove Theorem 5.7 in the slightly more general setting without assuming (5.28). In the discussion preceding Theorem 5.7 we have verified the assumptions of Theorem 5.6 for M and N in (5.24), (5.25). In particular, this implies that the subspace inclusions (5.21) hold and thus $\ker \begin{bmatrix} X_-^\top & U_-^\top \end{bmatrix} \subseteq \ker \begin{bmatrix} I & K^\top \end{bmatrix}$, equivalently

$$\operatorname{im} \begin{bmatrix} I \\ K \end{bmatrix} \subseteq \operatorname{im} \begin{bmatrix} X_- \\ U_- \end{bmatrix}. \tag{5.29}$$

Therefore, any controller K that stabilizes the systems in Σ is necessarily of the form (5.29). This generalizes [221, Lem. 15] (see also Lemma 3.1 of this thesis) to the case of noisy data.

5.5 INCLUSION OF PERFORMANCE SPECIFICATIONS

In this section we extend our data-driven stabilization result by including different performance specifications. In particular, we will treat the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control problems, thereby illustrating the general applicability of the theory in Section 5.3.

5.5.1 $\mathcal{H}_2$ control

As before, consider the the unknown system (5.1). We associate to (5.1) a performance output

$$\boldsymbol{z}(t) = C\boldsymbol{x}(t) + D\boldsymbol{u}(t), \tag{5.30}$$

where $\boldsymbol{z} \in \mathbb{R}^p$, and C and D are known matrices that specify the performance. For any $(A, B) \in \Sigma$ explaining the data, the feedback law $\boldsymbol{u} = K\boldsymbol{x}$ yields the closed-loop system

$$\begin{aligned} \boldsymbol{x}(t+1) &= (A + BK)\boldsymbol{x}(t) + \boldsymbol{w}(t) \\ \boldsymbol{z}(t) &= (C + DK)\boldsymbol{x}(t). \end{aligned} \tag{5.31}$$

The transfer matrix from w to z of (5.31) is given by

$$G(z) := (C + DK)(zI - (A + BK))^{-1},$$

and its $\mathcal{H}_2$ norm is denoted by $\|G(z)\|_{\mathcal{H}_2}$. Let $\gamma > 0$. It is well-known that $A + BK$ is stable and $\|G(z)\|_{\mathcal{H}_2} < \gamma$ if and only if there exists a matrix $P = P^\top > 0$ such that

$$\begin{aligned} P &> (A + BK)^\top P(A + BK) + (C + DK)^\top (C + DK) \\ \operatorname{tr} P &< \gamma^2. \end{aligned} \tag{5.32}$$

The data-driven $\mathcal{H}_2$ problem entails the computation of a feedback gain K from data such that $\|G(z)\|_{\mathcal{H}_2} < \gamma$ *for all* $(A, B) \in \Sigma$. Similar to our results for quadratic stabilization, we restrict the attention to a matrix P that is common for all (A, B). This leads to the following natural definition.

Definition 5.2. The data (U_-, X) are *informative for $\mathcal{H}_2$ control* with performance γ if there exist matrices $P = P^\top > 0$ and K such that (5.32) holds for all $(A, B) \in \Sigma$.

With the theory of Section 5.3 in place, characterizing informativity for $\mathcal{H}_2$ control essentially boils down to massaging the inequalities (5.32) such that they are amenable to design. To this end, note that the first inequality of (5.32) is equivalent to

$$Y - A_{Y,L}^\top P A_{Y,L} - C_{Y,L}^\top C_{Y,L} > 0,$$

where we defined $A_{Y,L} := AY + BL$ and $C_{Y,L} := CY + DL$ with $Y := P^{-1}$ and $L := KY$. Using a Schur complement argument, this is equivalent to

$$\begin{bmatrix} Y - C_{Y,L}^\top C_{Y,L} & A_{Y,L}^\top \\ A_{Y,L} & Y \end{bmatrix} > 0, \tag{5.33}$$

Now, (5.33) holds if and only if

$$Y - C_{Y,L}^\top C_{Y,L} > 0, \tag{5.34}$$

$$Y - A_{Y,L}(Y - C_{Y,L}^\top C_{Y,L})^{-1} A_{Y,L}^\top > 0. \tag{5.35}$$

Note that (5.34) is independent of A and B. In turn, we can write (5.35) as

$$\begin{bmatrix} I \\ \hdashline A^\top \\ B^\top \end{bmatrix}^\top \underbrace{\left[\begin{array}{c:c} Y & 0 \\ \hdashline 0 & -\begin{bmatrix} Y \\ L \end{bmatrix} (Y - C_{Y,L}^\top C_{Y,L})^{-1} \begin{bmatrix} Y \\ L \end{bmatrix}^\top \end{array}\right]}_{=:M} \begin{bmatrix} I \\ \hdashline A^\top \\ B^\top \end{bmatrix} > 0. \tag{5.36}$$

Note that the inequality (5.36) is of a form where A and B appear on the left and their transposes appear on the right, analogous to (5.11). As such, we are in a position to apply Theorem 5.6. In fact, we derive the following theorem.

Theorem 5.8. Assume that the generalized Slater condition (5.15) holds for N in (5.25) and some $\bar{Z} \in \mathbb{R}^{(n+m)\times n}$. Then the data (U_-, X) are informative for $\mathcal{H}_2$ control with performance γ if and only if there exist matrices $Y = Y^\top > 0$, $Z = Z^\top$ and L, and scalars $\alpha \geqslant 0$ and $\beta > 0$ satisfying

$$\begin{bmatrix} Y-\beta I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & Y & 0 \\ 0 & 0 & 0 & L & 0 \\ 0 & Y & L^\top & Y & C_{Y,L}^\top \\ 0 & 0 & 0 & C_{Y,L} & I \end{bmatrix} - \alpha \begin{bmatrix} I & X_+ \\ 0 & -X_- \\ 0 & -U_- \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{bmatrix} \begin{bmatrix} I & X_+ \\ 0 & -X_- \\ 0 & -U_- \\ 0 & 0 \\ 0 & 0 \end{bmatrix}^\top \geqslant 0$$

$$\begin{bmatrix} Y & C_{Y,L}^\top \\ C_{Y,L} & I \end{bmatrix} > 0, \quad \begin{bmatrix} Z & I \\ I & Y \end{bmatrix} \geqslant 0, \quad \operatorname{tr} Z < \gamma^2. \tag{$\mathcal{H}_2$}$$

Moreover, if Y and L satisfy ($\mathcal{H}_2$) then $K := LY^{-1}$ is such that $A + BK$ is stable and $\|G(z)\|_{\mathcal{H}_2} < \gamma$ for all $(A, B) \in \Sigma$.

Proof. Suppose that ($\mathcal{H}_2$) is feasible and define $P := Y^{-1}$ and $K := LP$. The last two inequalities of ($\mathcal{H}_2$) imply that $\operatorname{tr} P < \gamma^2$. We now compute the Schur complement of the first LMI in ($\mathcal{H}_2$) with respect to the diagonal block

$$\begin{bmatrix} Y & C_{Y,L}^\top \\ C_{Y,L} & I \end{bmatrix}.$$

We thereby make use of the fact that this block is nonsingular by the second LMI of ($\mathcal{H}_2$). The computation of the Schur complement results in

$$M - \alpha N \geqslant \begin{bmatrix} \beta I & 0 \\ 0 & 0 \end{bmatrix}, \tag{5.37}$$

where M is defined in (5.36) and N is defined in (5.25). We thus conclude that the inequality (5.36) is satisfied for all $(A, B) \in \Sigma$. As such, (5.35) holds for all $(A, B) \in \Sigma$. Note that (5.34) holds by the second LMI of ($\mathcal{H}_2$). Therefore, we conclude that (5.32) holds for all $(A, B) \in \Sigma$. In other words, the data (U_-, X) are informative for $\mathcal{H}_2$ control with performance γ, and $K = LY^{-1}$ is a suitable controller.

Conversely, suppose that the data (U_-, X) are informative for $\mathcal{H}_2$ control with performance γ. Then there exist matrices $P = P^\top > 0$ and K such that (5.32) holds for all $(A, B) \in \Sigma$. Define $Y := P^{-1}$, $L := KY$ and $Z := P$. Clearly, the last two inequalities of ($\mathcal{H}_2$) are satisfied by definition of Z. In addition, we know that (5.34) and (5.35) hold for all $(A, B) \in \Sigma$. By (5.34), the second LMI of ($\mathcal{H}_2$) is satisfied. To prove that the first LMI of ($\mathcal{H}_2$) also holds, we want to apply Theorem 5.6. Note that we have already verified the assumptions of this theorem for the matrix N in (5.25), see the discussion preceding Theorem 5.7. In addition, we note that

$$M_{22} = -\begin{bmatrix} Y \\ L \end{bmatrix} (Y - C_{Y,L}^\top C_{Y,L})^{-1} \begin{bmatrix} Y \\ L \end{bmatrix}^\top \leqslant 0$$

since $Y - C_{Y,L}^\top C_{Y,L} > 0$. Hence, Theorem 5.6 is applicable. We conclude that there exist $\alpha \geqslant 0$ and $\beta > 0$ such that (5.37) holds. Using a Schur complement argument, we see that Y, L, α and β satisfy the first LMI of ($\mathcal{H}_2$). Thus, ($\mathcal{H}_2$) is feasible which proves the theorem. □

Remark 5.5. If we know a priori that the noise $\boldsymbol{w}$ is contained in a subspace, say $\operatorname{im} E$, then this information can easily be exploited in the $\mathcal{H}_2$ controller design. In fact, we only need to replace the LMI involving Z by

$$\begin{bmatrix} Z & E^\top \\ E & Y \end{bmatrix} \geqslant 0.$$

We recall that prior knowledge of $\boldsymbol{w} \in \operatorname{im} E$, if available, can also be captured by our noise model, see Remark 5.2. A natural choice is thus to use E both in the noise model (5.5) as well as in the LMI ($\mathcal{H}_2$). However, we remark that this is not necessary: the noise in the experiment may come from a different subspace than the disturbances that are attenuated by the $\mathcal{H}_2$ controller.

5.5.2 $\mathcal{H}_\infty$ control

In this section we will turn our attention to the $\mathcal{H}_\infty$ control problem. As before, consider system (5.1) with performance output (5.30). For any $(A, B) \in \Sigma$, the feedback $\boldsymbol{u} = K\boldsymbol{x}$ yields the system (5.31) with transfer matrix from $\boldsymbol{w}$ to $\boldsymbol{z}$ given by $G(z)$. We will denote the $\mathcal{H}_\infty$ norm of $G(z)$ by $\|G(z)\|_{\mathcal{H}_\infty}$. Let $\gamma > 0$. By [198, Thm. 4.6.6(iii)], the matrix $A + BK$ is stable and $\|G(z)\|_{\mathcal{H}_\infty} < \gamma$ if and only if there exists a matrix $P = P^\top > 0$ such that

$$P - A_K^\top (P^{-1} - \frac{1}{\gamma^2} I)^{-1} A_K - C_K^\top C_K > 0, \tag{5.38}$$

$$P^{-1} - \frac{1}{\gamma^2} I > 0, \tag{5.39}$$

where we have defined $A_K := A + BK$ and $C_K := C + DK$. We now have the following definition of informativity for $\mathcal{H}_\infty$ control.

Definition 5.3. The data (U_-, X) are *informative for $\mathcal{H}_\infty$ control* with performance γ if there exist matrices $P = P^\top > 0$ and K such that (5.38) and (5.39) hold for all $(A, B) \in \Sigma$.

By pre- and postmultiplication of (5.38) by P^{-1} we obtain

$$Y - A_{Y,L}^\top (Y - \frac{1}{\gamma^2} I)^{-1} A_{Y,L} - C_{Y,L}^\top C_{Y,L} > 0,$$

$$Y - \frac{1}{\gamma^2} I > 0,$$

where the matrices $Y := P^{-1}$, $L := KY$, $A_{Y,L} := AY + BL$ and $C_{Y,L} := CY + DL$ are defined as in the $\mathcal{H}_2$ problem. Note that the first of these inequalities can again be written in the -by now familiar- form

$$\begin{bmatrix} I \\ \hdashline A^\top \\ B^\top \end{bmatrix}^\top \left[\begin{array}{c:c} Y - C_{Y,L}^\top C_{Y,L} & 0 \\ \hdashline 0 & -\begin{bmatrix} Y \\ L \end{bmatrix} Z \begin{bmatrix} Y \\ L \end{bmatrix}^\top \end{array}\right] \begin{bmatrix} I \\ \hdashline A^\top \\ B^\top \end{bmatrix} > 0,$$

where $Z := (Y - \frac{1}{\gamma^2} I)^{-1}$. We thus have the following theorem.

Theorem 5.9. Assume that the generalized Slater condition (5.15) holds for N in (5.25) and some $\bar{Z} \in \mathbb{R}^{(n+m)\times n}$. Then the data (U_-, X) are informative for $\mathcal{H}_\infty$ control with performance γ if and only if there exist matrices $Y = Y^\top > 0$ and L, and scalars $\alpha \geqslant 0$ and $\beta > 0$ satisfying

$$\begin{bmatrix} Y - \beta I & 0 & 0 & 0 & C_{Y,L}^\top \\ 0 & 0 & 0 & Y & 0 \\ 0 & 0 & 0 & L & 0 \\ 0 & Y & L^\top & Y - \frac{1}{\gamma^2} I & 0 \\ C_{Y,L} & 0 & 0 & 0 & I \end{bmatrix} - \alpha \begin{bmatrix} I & X_+ \\ 0 & -X_- \\ 0 & -U_- \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{bmatrix} \begin{bmatrix} I & X_+ \\ 0 & -X_- \\ 0 & -U_- \\ 0 & 0 \\ 0 & 0 \end{bmatrix}^\top \geqslant 0$$

$$Y - \frac{1}{\gamma^2} I > 0. \tag{$\mathcal{H}_\infty$}$$

Moreover, if Y and L satisfy ($\mathcal{H}_\infty$) then $K := LY^{-1}$ is such that $A + BK$ is stable and $\|G(z)\|_{\mathcal{H}_\infty} < \gamma$ for all $(A, B) \in \Sigma$.

The proof of Theorem 5.9 is based on Theorem 5.6. It follows similar steps as the proof of Theorem 5.8, and is therefore not reported here.

5.6 SIMULATION EXAMPLES

In this section we illustrate our theoretical results by numerical simulations.

5.6.1 Stabilization using bounds on the noise samples

Consider an unstable system of the form (5.1) with A_s and B_s given by

$$A_s = \begin{bmatrix} 0.850 & -0.038 & -0.380 \\ 0.735 & 0.815 & 1.594 \\ -0.664 & 0.697 & -0.064 \end{bmatrix}, B_s = \begin{bmatrix} 1.431 & 0.705 \\ 1.620 & -1.129 \\ 0.913 & 0.369 \end{bmatrix}.$$

In this example, we assume that the noise samples $w(t)$ are bounded in norm as $\|w(t)\|_2^2 \leqslant \epsilon$ for all t. As explained in Section 5.2, we can capture this prior knowledge using the noise model (5.5) with $\Phi_{11} = T\epsilon I$, $\Phi_{12} = 0$ and $\Phi_{22} - I$.

We pick a time horizon of $T = 20$ and draw the entries of the inputs and initial state randomly from a Gaussian distribution with zero mean and unit variance. The noise samples are drawn uniformly at random from the ball $\{w \in \mathbb{R}^3 \mid \|w\|_2^2 \leqslant \epsilon\}$. We aim at constructing stabilizing controllers from the input/state data for various values of ϵ. In particular, we investigate six different noise levels: $\epsilon \in \{0.5, 1, 1.5, 2, 2.2, 2.4\}$. For each noise level, we generate 100 data sets using the method described above. We check the generalized Slater condition (5.15) by verifying that N in (5.25) has 3 positive eigenvalues; this turns out to be true for all 600 data sets. For each noise level, we record the percentage of data sets from which a stabilizing controller was found for (A_s, B_s) using the formulation (FS). We display the results in the following table.

$\epsilon = 0.5$	$\epsilon = 1$	$\epsilon = 1.5$	$\epsilon = 2$	$\epsilon = 2.2$	$\epsilon = 2.4$
100%	96%	90%	82%	75%	73%

For $\epsilon = 0.5$ we find a stabilizing controller in all 100 cases. When the noise level increases, the percentage of data sets for which the LMI (FS) is feasible decreases. The interpretation is that by increasing the noise we enlarge the set of explaining systems Σ. It thus becomes harder to simultaneously stabilize the systems in Σ. Nonetheless, even for the larger noise level of $\epsilon = 2.4$ we find a stabilizing controller in 73 out of the 100 data sets.

5.6.2 $\mathcal{H}_2$ control of a fighter aircraft

We consider a state-space model of a fighter aircraft [198, Ex. 10.1.2]. In particular, we discretize the model of [198] using a sampling time of 0.01, which results in the (unstable) system of the form (5.1) with A_s and B_s given by

$$\begin{bmatrix} 1.000 & -0.374 & -0.190 & -0.321 & 0.056 & -0.026 \\ 0.000 & 0.982 & 0.010 & -0.000 & -0.003 & 0.001 \\ 0.000 & 0.115 & 0.975 & -0.000 & -0.269 & 0.191 \\ 0.000 & 0.001 & 0.010 & 1.000 & -0.001 & 0.001 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.741 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.741 \end{bmatrix},$$

$$\begin{bmatrix} 0.007 & 0.000 & -0.043 & 0.000 & 0.259 & 0.000 \\ -0.003 & 0.000 & 0.030 & 0.000 & 0.000 & 0.259 \end{bmatrix}^\top,$$

respectively. We consider the performance output as in (5.30) with

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and $D = 0$. First, we look for the smallest γ such that (5.32) is feasible for (A_s, B_s). This minimum value of γ is 1.000 and can be regarded as a benchmark: no data-driven method can perform better than the model-based solution using full knowledge of (A_s, B_s).

Of course, our goal is not to use the knowledge of (A_s, B_s) but to seek a data-driven solution instead. Therefore, we collect $T = 750$ input and state samples of (5.1). The entries of the inputs and initial state were drawn randomly from a Gaussian distribution with zero mean and unit variance. Also the noise samples were drawn randomly from a Gaussian distribution, with zero mean and variance σ^2 with $\sigma = 0.005$. In this example, we assume knowledge of a bound on the energy of the noise as

$$W_- W_-^\top \leqslant 1.35 T \sigma^2 I. \tag{5.40}$$

We verified that this bound is satisfied for the generated noise sequence. In addition, we verified that the matrix N in (5.25) has 6 positive eigenvalues, thus the generalized Slater condition (5.15) holds.

Next, we want to compute an $\mathcal{H}_2$ controller for the unknown system using the generated data. We do so by minimizing γ subject to ($\mathcal{H}_2$). This is a semidefinite program that we solve in Matlab, using Yalmip [115] with Mosek as an LMI solver. The obtained controller K is given by

$$\begin{bmatrix} -0.023 & 1.413 & 0.695 & 0.227 & -1.591 & 0.090 \\ 0.001 & -0.041 & -0.028 & -0.034 & 0.010 & -2.723 \end{bmatrix}.$$

This controller stabilizes the original system (A_s, B_s). In addition, the system, in feedback with K, has an $\mathcal{H}_2$ norm of γ_s where $\gamma_s^2 = 1.007$. We note that this is almost identical to the smallest possible $\mathcal{H}_2$ norm of 1.000.

Subsequently, we repeat the above experiment using only a *part* of our data set. In particular, we compute an $\mathcal{H}_2$ controller via the semidefinite program as before, using only the first i samples of X_+, X_- and U_- for $i = 50, 100, \dots, 750$. We display the results in Figure 5.1.

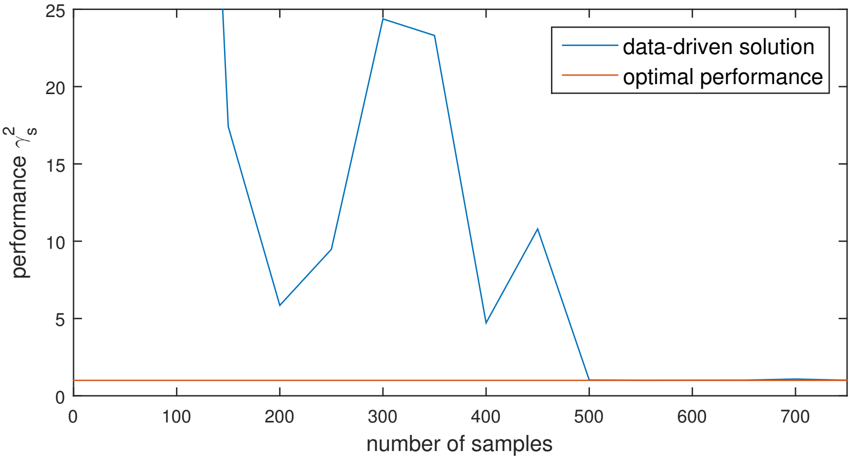

Figure 5.1: Achieved $\mathcal{H}_2$ performance of the true system in feedback with a data-based controller (blue) and the optimal (model-based) performance of the true system (red).

In each of the cases a stabilizing controller was found from data. However, the performance of these controllers when applied to the true system varies, and is quite poor for $i < 500$. Starting from $i = 500$ and onward, the performance is close to the optimal performance of the true system.

Next, we investigate what happens when we increase the variance σ^2 of the noise. First, we take $\sigma = 0.05$. We again generate 750 data samples, and assume the same bound on the noise. The $\mathcal{H}_2$ controller we obtain is given by

$$\begin{bmatrix} -0.007 & 0.179 & 0.464 & -0.284 & -1.411 & 0.100 \\ 0.005 & -0.014 & -0.363 & 0.184 & 0.123 & -1.514 \end{bmatrix},$$

and achieves a performance of $\gamma_s^2 = 1.146$ when interconnected to the true system. Increasing the variance of the noise has the effect that the set Σ of explaining systems becomes larger. As such, it is more difficult to control all systems in Σ resulting in a slightly larger γ_s. This behavior becomes even more apparent when increasing the variance of the noise to $\sigma = 0.5$. In this case we obtain the controller

$$\begin{bmatrix} -0.002 & -0.001 & 0.234 & 0.016 & -0.553 & 0.020 \\ 0.001 & -0.071 & -0.122 & -0.002 & 0.141 & -0.550 \end{bmatrix}$$

which yields a performance of $\gamma_s^2 = 3.579$. Increasing σ even more to $\sigma = 1$ results in infeasibility of the LMI's ($\mathcal{H}_2$) for any γ; the set of explaining systems has become too large for a quadratically stabilizing controller to exist.

We remark that the size of the set Σ does not only depend on the variance of the noise, but also on the available bound on the noise. Throughout this example, we have used the bound (5.40). However, if we reconsider the case of $\sigma = 0.5$ with the tighter bound $W_- W_-^\top \leqslant 1.22T\sigma^2 I$ we obtain a controller with better performance $\gamma_s^2 = 2.706$. This illustrates the simple fact that data-driven controllers not only depend on the particular design strategy, but also on the *prior knowledge* on the noise.

We conclude the example with a remark on the dimension of the variables involved in the formulation ($\mathcal{H}_2$). The symmetric matrices Y and Z both have 21 free variables. The matrix L contains 12 variables, and α and β are both scalar variables. Thus, the total number of variables is 56. The size of the largest LMI in ($\mathcal{H}_2$) is 21×21. We emphasize that our approach is based directly on the set Σ of open-loop systems and avoids the parameterization of closed-loop systems, as employed in [17, 47]. Such parameterizations involve decision variables of dimension $T \times n$, which would result in at least 4500 variables in this example.

5.7 DISCUSSION AND CONCLUSIONS

We have studied the problem of obtaining feedback controllers from noisy data. The essence of our approach has been to formulate data-driven control as the problem of determining when one quadratic matrix inequality implies another one. To get a grip on this fundamental question, we have generalized the classical S-lemma [169] to matrix variables. The implication involving quadratic matrix inequalities is thereby *equivalent* to a linear matrix inequality in a scalar variable. We have established several versions of the matrix S-lemma, for both strict and non-strict inequalities. These matrix S-lemmas are interesting in their own right, and generalize existing S-lemmas [169] as well as a theorem involving quadratic matrix inequalities [116].

We have followed up by applying our matrix S-lemma to data-driven control. In particular, we have given necessary and sufficient conditions under which stabilizing, $\mathcal{H}_2$, and $\mathcal{H}_\infty$ controllers can be obtained from noisy data. Our control design revolves around data-guided linear matrix inequalities, which can be solved efficiently using modern LMI solvers. In addition to being non-conservative, an attractive feature of our design procedure is that decision variables are *independent* of the time horizon of the experiment.

So far, we have only applied the matrix S-lemma involving a strict inequality (Thms. 5.5, 5.6) to data-driven control. However, we are convinced that also the matrix S-lemma with *non-strict* inequalities (Thm. 5.4) will find applications, for example, in the verification of dissipativity properties from data [101].

The noise model that we have employed is flexible, and can describe, e.g., constant disturbances, energy bounded noise and norm bounds on noise samples. If one is only interested in the latter, however, we expect that more specific control techniques are possible. In fact, analogous to (5.10), we can write the inequality $w(t)^\top w(t) \leqslant \epsilon$ as

$$\begin{bmatrix} I \\ A^\top \\ B^\top \end{bmatrix}^\top \underbrace{\begin{bmatrix} I & x(t+1) \\ 0 & -x(t) \\ 0 & -u(t) \end{bmatrix} \begin{bmatrix} \epsilon I & 0 \\ 0 & -I \end{bmatrix} \begin{bmatrix} I & x(t+1) \\ 0 & -x(t) \\ 0 & -u(t) \end{bmatrix}^\top}_{:=N_t} \begin{bmatrix} I \\ A^\top \\ B^\top \end{bmatrix} \geqslant 0.$$

In the spirit of the S-procedure, one could thus design a stabilizing controller by computing[6] matrices $P = P^\top > 0$ and K, and *multiple* non-negative scalars $\alpha_0, \alpha_1, \dots, \alpha_{T-1}$ such that

$$M - \sum_{t=0}^{T-1} \alpha_t N_t > 0,$$

with M given by (5.24). We will consider norm bounded noise samples in more detail in future work.

[6] This procedure is likely to be conservative, however, since the classical S-lemma is in general conservative for more than two quadratic functions [169].

6 DATA INFORMATIVITY FOR DISSIPATIVITY

In this chapter we focus on assessing dissipativity properties of a linear system from measured data. We will study this problem both in the case that the data are exact and noisy. In the case of exact data, we are able to show that one can only verify dissipativity of a linear system from given data if the system is uniquely identifiable from these data. In the case of noisy data we will see that the matrix S-lemma, as established in the previous chapter, will play an important role in characterizing informative data for dissipativity.

6.1 INTRODUCTION

The theory of dissipativity was first introduced by Willems in [240]. Dissipativity plays an important role in control design and the study of interconnections of dynamical systems, see e.g. [188, 215]. In the case that the dynamics of the system are known, dissipativity can be verified using well-known tests involving extremal storage functions. In addition, if the dynamics of the system are linear, one can formulate a test for dissipativity in terms of linear matrix inequalities involving the system matrices. However, in many situations the system dynamics are not known a priori. In such situations, the question arises whether we can verify dissipativity using a measured system trajectory, instead of a system model.

The problem of inferring dissipativity properties from data has been considered in several recent publications. In [179], the set of supply rates with a given structure with respect to which a (not necessarily linear) system is dissipative is computed on the basis of a finite number of its input-output trajectories. In [180] an iterative procedure is illustrated to compute the input feedforward passivity index and the shortage of passivity for discrete-time linear systems.

The most relevant publications for the problem studied in this chapter are [100, 101, 132, 178]. The notion of (finite-horizon) L-dissipativity was introduced in [132] and also studied in [178]. In both these contributions, a crucial assumption is that the input trajectory is *persistently exciting* of a sufficiently high order (see [241]). This property of the input sequence can be shown to imply that the data-generating system is uniquely identifiable on the basis of the data.

In this chapter we study the more classical notion of dissipativity for linear systems, rather than L-dissipativity. We consider a similar setup as the one in [100, 101]. In these papers, sufficient conditions were given under which dissipativity of a system can be ascertained using data. The main difference between this chapter and [100,101] is that the conditions we provide are necessary *and* sufficient for data-driven dissipativity, both in the cases of noiseless and noisy data. An additional difference is that in our setting, the matrices defining the system's output are a priori unknown.

Specifically, our contributions are the following. First, we prove that dissipativity of an unknown linear system can only be ascertained if a data matrix involving measured states and inputs has full rank. In the noiseless data case, this implies that we can only verify dissipativity from data if the data-generating system is the only one that explains the data. In this case, dissipativity of the unknown system can be ascertained if and only if a data-based linear matrix inequality is feasible. In the noisy data case, we need a new type of dualization lemma that we prove in this chapter. We combine this dualization lemma with the matrix S-lemma (see [218] and Chapter 5), to provide a neat data-driven test for dissipativity in the noisy data setting.

The outline of this chapter is as follows. In Section 6.2 we revisit dissipativity of discrete-time linear time-invariant systems. In Section 6.3 we set up the problems. Next, Section 6.4 contains our main results. Finally, we provide conclusions in Section 6.5 and proofs of auxiliary results in Section 6.6.

Notation

The *inertia* of a symmetric matrix S is denoted by $\operatorname{In}(S) = (\rho_-, \rho_0, \rho_+)$ where ρ_-, ρ_0, and ρ_+ respectively denote the number of negative, zero, and positive eigenvalues of S. The *interior* of a set V is denoted by $\operatorname{int}(V)$.

6.2 DISSIPATIVITY OF LINEAR SYSTEMS

Consider a linear discrete-time input/state/output system

$$\boldsymbol{x}(t+1) = A\boldsymbol{x}(t) + B\boldsymbol{u}(t) \tag{6.1a}$$

$$\boldsymbol{y}(t) = C\boldsymbol{x}(t) + D\boldsymbol{u}(t) \tag{6.1b}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, and $D \in \mathbb{R}^{p\times m}$.

Let $S = S^\top \in \mathbb{R}^{(m+p)\times(m+p)}$. The system (6.1) is said to be *dissipative* with respect to the *supply rate*

$$s(u,y) = \begin{bmatrix} u \\ y \end{bmatrix}^\top S \begin{bmatrix} u \\ y \end{bmatrix} \tag{6.2}$$

if there exists $P \in \mathbb{R}^{n\times n}$ with $P = P^\top \geqslant 0$ such that the *dissipation inequality*

$$\boldsymbol{x}(t)^\top P\boldsymbol{x}(t) + s\big(\boldsymbol{u}(t), \boldsymbol{y}(t)\big) \geqslant \boldsymbol{x}(t+1)^\top P\boldsymbol{x}(t+1) \tag{6.3}$$

holds for all $t \geqslant 0$ and for all trajectories $(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{y}) : \mathbb{N} \to \mathbb{R}^{m+n+p}$ of (6.1).

It follows from (6.3) that dissipativity with respect to the supply rate (6.2) is equivalent with the feasibility of the linear matrix inequalities $P = P^\top \geqslant 0$ and

$$\begin{bmatrix} I & 0 \\ A & B \end{bmatrix}^\top \begin{bmatrix} P & 0 \\ 0 & -P \end{bmatrix} \begin{bmatrix} I & 0 \\ A & B \end{bmatrix} + \begin{bmatrix} 0 & I \\ C & D \end{bmatrix}^\top S \begin{bmatrix} 0 & I \\ C & D \end{bmatrix} \geqslant 0. \tag{6.4}$$

6.3 INFORMATIVITY: A VOCABULARY

Consider the linear discrete-time input/state/output system

$$\boldsymbol{x}(t+1) = A_s\boldsymbol{x}(t) + B_s\boldsymbol{u}(t) + \boldsymbol{w}(t) \tag{6.5a}$$
$$\boldsymbol{y}(t) = C_s\boldsymbol{x}(t) + D_s\boldsymbol{u}(t) + \boldsymbol{z}(t) \tag{6.5b}$$

where $(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{m+n+p}$ are the input, state and output, and $(\boldsymbol{w}, \boldsymbol{z}) \in \mathbb{R}^{n+p}$ are noise terms. Throughout the chapter, we assume that the "true" system matrices (A_s, B_s, C_s, D_s) and the noise $(\boldsymbol{w}, \boldsymbol{z})$ are *unknown*. What is known instead are a finite number of input/state/output measurements harvested from the true system (6.5):

$$\begin{aligned} &u(0), u(1), \ldots, u(T-1) \\ &x(0), x(1), \ldots, x(T) \\ &y(0), y(1), \ldots, y(T-1). \end{aligned}$$

We collect these data in the matrices

$$\begin{aligned} X &:= \begin{bmatrix} x(0) & x(1) & \cdots & x(T) \end{bmatrix} \\ X_- &:= \begin{bmatrix} x(0) & x(1) & \cdots & x(T-1) \end{bmatrix} \\ X_+ &:= \begin{bmatrix} x(1) & x(2) & \cdots & x(T) \end{bmatrix} \\ U_- &:= \begin{bmatrix} u(0) & u(1) & \cdots & u(T-1) \end{bmatrix} \\ Y_- &:= \begin{bmatrix} y(0) & y(1) & \cdots & y(T-1) \end{bmatrix}. \end{aligned}$$

Remark 6.1. To make progress on the general problem of inferring dissipativity properties from input/output data, it makes sense to consider the simpler one in which the state is directly measured, even though experiments usually only measure inputs and outputs. As a matter of fact, in this chapter we show that in this way definitive conclusions can be drawn about the possibility of ascertaining dissipativity from data.

Our goal is to infer dissipativity properties of the true system from the data (U_-, X, Y_-). We define

$$\Sigma^{\mathcal{N}} = \left\{ (A, B, C, D) \mid \begin{bmatrix} X_+ \\ Y_- \end{bmatrix} - \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix} \in \mathcal{N} \right\},$$

where $\mathcal{N} \subseteq \mathbb{R}^{(n+p)\times T}$ is a set associated with the noise model to be specified below. We assume that

$$(A_s, B_s, C_s, D_s) \in \Sigma^{\mathcal{N}}. \tag{6.7}$$

In the sequel, we will consider three types of noise models. The first one will capture noise-free situations in which the measurements (U_-, X, Y_-) are exact:

$$\mathcal{N}_0 := \{0\}. \tag{6.8}$$

The second noise model is defined by

$$\mathcal{N}_1 := \left\{ V \in \mathbb{R}^{(n+p)\times T} \mid \begin{bmatrix} I \\ V^\top \end{bmatrix}^\top \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{bmatrix} \begin{bmatrix} I \\ V^\top \end{bmatrix} \geqslant 0 \right\}, \tag{6.9}$$

where $\Phi_{11} = \Phi_{11}^\top \in \mathbb{R}^{(n+p)\times(n+p)}$, $\Phi_{12} \in \mathbb{R}^{(n+p)\times T}$, $\Phi_{22} = \Phi_{22}^\top \in \mathbb{R}^{T\times T}$. This noise model was studied before [218] in the context of state feedback control, see also Chapter 5.

The third noise model is defined by

$$\mathcal{N}_2 := \{ V \in \mathbb{R}^{(n+p)\times T} \mid \begin{bmatrix} I \\ V \end{bmatrix}^\top \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{12}^\top & \Theta_{22} \end{bmatrix} \begin{bmatrix} I \\ V \end{bmatrix} \geqslant 0 \} \tag{6.10}$$

where $\Theta_{11} = \Theta_{11}^\top \in \mathbb{R}^{T\times T}$, $\Theta_{12} \in \mathbb{R}^{T\times(n+p)}$, $\Theta_{22} = \Theta_{22}^\top \in \mathbb{R}^{(n+p)\times(n+p)}$. This noise model was also studied in [17, 101] in the contexts of state feedback control and dissipativity.

We now define the property of informativity for dissipativity.

Definition 6.1. Let a noise model $\mathcal{N}$ be given. The data (U_-, X, Y_-) are *informative for dissipativity* with respect to the supply rate (6.2) if there exists $P = P^\top \geqslant 0$ such that the LMI (6.4) holds for every system $(A, B, C, D) \in \Sigma^{\mathcal{N}}$.

The following assumptions will be valid throughout the chapter:

(A1) The matrix S has inertia $\text{In}(S) = (p, 0, m)$.

(A2) The sets $\mathcal{N}_1$ and $\mathcal{N}_2$ are bounded and have nonempty interior.

Assumption (A1) is satisfied, for example, for the positive-real and bounded-real case [188]. Indeed, in the positive-real case we have that $m = p$ and

$$S = \begin{bmatrix} 0 & I_m \\ I_m & 0 \end{bmatrix},$$

so that $\text{In}(S) = (m, 0, m)$. In the bounded-real case we have

$$S = \begin{bmatrix} \gamma^2 I_m & 0 \\ 0 & -I_p \end{bmatrix}$$

for $\gamma > 0$, which implies that $\operatorname{In}(S) = (p, 0, m)$. Assumption (A2) also turns out to be instrumental in computing storage functions of the "dual" system from those of the primal one; see Proposition 6.2.

Assumption (A2) can be verified straightforwardly using the following lemma which we prove in Section 6.6.1.

Lemma 6.1. Let $\Psi_{11} = \Psi_{11}^\top \in \mathbb{R}^{q \times q}$, $\Psi_{12} \in \mathbb{R}^{q \times r}$, and $\Psi_{22} = \Psi_{22}^\top \in \mathbb{R}^{r \times r}$. Then, the set

$$\mathcal{N} := \left\{ R \in \mathbb{R}^{r \times q} \mid \begin{bmatrix} I \\ R \end{bmatrix}^\top \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{12}^\top & \Psi_{22} \end{bmatrix} \begin{bmatrix} I \\ R \end{bmatrix} \geqslant 0 \right\}$$

is bounded and has nonempty interior if and only if $\Psi_{22} < 0$ and $\Psi_{11} - \Psi_{12}\Psi_{22}^{-1}\Psi_{12}^\top > 0$.

Moreover, it is always possible to convert noise model $\mathcal{N}_1$ to $\mathcal{N}_2$ and vice versa. To do this, one can use the following type of dualization lemma involving nonstrict inequalities and matrix variables (see also [188, Lem. 4.9] for a "standard" dualization lemma). We postpone the proof of Lemma 6.2 to Section 6.6.2.

Lemma 6.2. Let

$$\Psi = \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{12}^\top & \Psi_{22} \end{bmatrix}$$

where $\Psi_{11} = \Psi_{11}^\top \in \mathbb{R}^{q \times q}$, $\Psi_{12} \in \mathbb{R}^{q \times r}$, and $\Psi_{22} = \Psi_{22}^\top \in \mathbb{R}^{r \times r}$ be such that $\Psi_{22} < 0$ and $\Psi_{11} - \Psi_{12}\Psi_{22}^{-1}\Psi_{12}^\top > 0$. Define

$$\Xi := \begin{bmatrix} 0 & -I_r \\ I_q & 0 \end{bmatrix} \Psi^{-1} \begin{bmatrix} 0 & -I_q \\ I_r & 0 \end{bmatrix}.$$

Let $R \in \mathbb{R}^{r \times q}$. Then,

$$\begin{bmatrix} I \\ R \end{bmatrix}^\top \Psi \begin{bmatrix} I \\ R \end{bmatrix} \geqslant 0 \tag{6.11}$$

if and only if

$$\begin{bmatrix} I \\ R^\top \end{bmatrix}^\top \Xi \begin{bmatrix} I \\ R^\top \end{bmatrix} \geqslant 0. \tag{6.12}$$

In the next section we will provide necessary and sufficient conditions for data informativity as defined in Definition 6.1 for the noise models $\mathcal{N}_0$, $\mathcal{N}_1$, and $\mathcal{N}_2$.

6.4 MAIN RESULTS

6.4.1 A necessary condition for informativity

We begin with a necessary condition for informativity, that applies to all three noise models.

Theorem 6.1. Let a noise model $\mathcal{N}$ be given. If the data (U_-, X, Y_-) are informative for dissipativity with respect to the supply rate (6.2), then

$$\begin{bmatrix} X_- \\ U_- \end{bmatrix} \text{ has full row rank.} \tag{6.13}$$

Proof. Suppose that (6.13) does not hold. Then, there exist $\xi \in \mathbb{R}^n$ and $\eta \in \mathbb{R}^m$ such that $\xi^\top \xi + \eta^\top \eta = 1$ and

$$\begin{bmatrix} \xi^\top & \eta^\top \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix} = 0. \tag{6.14}$$

The set $\Gamma = \{u \mid \exists y \text{ such that } s(u,y) < 0\}$ has nonempty interior since there exists $(\hat{u}, \hat{y})$ with $s(\hat{u}, \hat{y}) < 0$ due to Assumption (A1). We claim that there exist $x \in \mathbb{R}^n$ and $u \in \Gamma$ such that

$$\xi^\top x + \eta^\top u = 1. \tag{6.15}$$

Indeed, if $\xi \neq 0$, then one can construct x and u by selecting $u \in \Gamma$ arbitrarily, and by defining $x := \frac{1-\eta^\top u}{\xi^\top \xi} \xi$. If $\xi = 0$ then $x \in \mathbb{R}^n$ can be selected arbitrarily. In this case, we can choose u as follows. Since Γ has nonempty interior, there exists $\bar{u} \in \Gamma$ such that $\eta^\top \bar{u} \neq 0$. Note that $\alpha \bar{u} \in \Gamma$ for all nonzero $\alpha \in \mathbb{R}$. As such, there exists an $\alpha \in \mathbb{R}$ such that $u := \alpha \bar{u} \in \Gamma$ and $\eta^\top u = 1$. For this u, we obtain (6.15) which proves our claim.

Since $u \in \Gamma$, there exists y such that $s(u,y) < 0$. Let $(A_0, B_0, C_0, D_0) \in \Sigma^{\mathcal{N}}$. Define

$$\zeta := x - A_0 x - B_0 u \quad \text{and} \quad \theta := y - C_0 x - D_0 u, \tag{6.16}$$

and

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} := \begin{bmatrix} A_0 & B_0 \\ C_0 & D_0 \end{bmatrix} + \begin{bmatrix} \zeta \\ \theta \end{bmatrix} \begin{bmatrix} \xi^\top & \eta^\top \end{bmatrix}.$$

It follows from (6.14) that $(A, B, C, D) \in \Sigma^{\mathcal{N}}$. Since the data are informative for dissipativity with respect to the supply rate (6.2), there must exist $P = P^\top \geqslant 0$ such that

$$\begin{bmatrix} I & 0 \\ A & B \end{bmatrix}^\top \begin{bmatrix} P & 0 \\ 0 & -P \end{bmatrix} \begin{bmatrix} I & 0 \\ A & B \end{bmatrix} + \begin{bmatrix} 0 & I \\ C & D \end{bmatrix}^\top S \begin{bmatrix} 0 & I \\ C & D \end{bmatrix} \geqslant 0. \tag{6.17}$$

Note that

$$\begin{bmatrix} I & 0 \\ A & B \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} x \\ x \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & I \\ C & D \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} u \\ y \end{bmatrix}$$

due to (6.15) and (6.16). Therefore, the following inequality holds:

$$\begin{aligned} &\begin{bmatrix} x \\ u \end{bmatrix}^\top \left(\begin{bmatrix} I & 0 \\ A & B \end{bmatrix}^\top \begin{bmatrix} P & 0 \\ 0 & -P \end{bmatrix} \begin{bmatrix} I & 0 \\ A & B \end{bmatrix} + \begin{bmatrix} 0 & I \\ C & D \end{bmatrix}^\top S \begin{bmatrix} 0 & I \\ C & D \end{bmatrix} \right) \begin{bmatrix} x \\ u \end{bmatrix} \\ &= \begin{bmatrix} x \\ x \end{bmatrix}^\top \begin{bmatrix} P & 0 \\ 0 & -P \end{bmatrix} \begin{bmatrix} x \\ x \end{bmatrix} + \begin{bmatrix} u \\ y \end{bmatrix}^\top S \begin{bmatrix} u \\ y \end{bmatrix} = s(u,y) < 0. \end{aligned}$$

However, this contradicts (6.17). Consequently, (6.13) holds. □

6.4.2 Informativity and noiseless data

We now give a characterization of informativity for dissipativity for the noiseless case.

Theorem 6.2. Consider the noise model $\mathcal{N}_0$. The data (U_-, X, Y_-) are informative for dissipativity with respect to the supply rate (6.2) if and only if

$$\operatorname{rank}\begin{bmatrix} X_- \\ U_- \end{bmatrix} = n+m \tag{6.18}$$

and there exists $P = P^\top \geqslant 0$ such that

$$\begin{bmatrix} X_- \\ X_+ \end{bmatrix}^\top \begin{bmatrix} P & 0 \\ 0 & -P \end{bmatrix} \begin{bmatrix} X_- \\ X_+ \end{bmatrix} + \begin{bmatrix} U_- \\ Y_- \end{bmatrix}^\top S \begin{bmatrix} U_- \\ Y_- \end{bmatrix} \geqslant 0. \tag{6.19}$$

Proof. To prove the "if" part, note that (6.18) implies that $\Sigma^{\mathcal{N}_0}$ is a singleton. It follows from (6.7) that

$$\Sigma^{\mathcal{N}_0} = \{(A_s, B_s, C_s, D_s)\}$$

and hence

$$\begin{bmatrix} X_+ \\ Y_- \end{bmatrix} = \begin{bmatrix} A_s & B_s \\ C_s & D_s \end{bmatrix} \begin{bmatrix} X_- \\ U_- \end{bmatrix}.$$

Define

$$L := \begin{bmatrix} I & 0 \\ A_s & B_s \end{bmatrix}^\top \begin{bmatrix} P & 0 \\ 0 & -P \end{bmatrix} \begin{bmatrix} I & 0 \\ A_s & B_s \end{bmatrix} + \begin{bmatrix} 0 & I \\ C_s & D_s \end{bmatrix}^\top S \begin{bmatrix} 0 & I \\ C_s & D_s \end{bmatrix}.$$

Then (6.19) implies

$$\begin{bmatrix} X_- \\ U_- \end{bmatrix}^\top L \begin{bmatrix} X_- \\ U_- \end{bmatrix} \geqslant 0. \tag{6.20}$$

It follows again from (6.18) that $L \geqslant 0$. By (6.4), this means that the system (A_s, B_s, C_s, D_s) is dissipative with respect to the supply rate (6.2).

To prove the "only if" part, note that it follows from Theorem 6.1 that (6.18) holds. Hence we have

$$\Sigma^{\mathcal{N}_0} = \{(A_s, B_s, C_s, D_s)\}.$$

Since the data are informative for dissipativity for the given $\mathcal{N}_0$, there exists $P = P^\top \geqslant 0$ such that

$$\begin{bmatrix} I & 0 \\ A_s & B_s \end{bmatrix}^\top \begin{bmatrix} P & 0 \\ 0 & -P \end{bmatrix} \begin{bmatrix} I & 0 \\ A_s & B_s \end{bmatrix} + \begin{bmatrix} 0 & I \\ C_s & D_s \end{bmatrix}^\top S \begin{bmatrix} 0 & I \\ C_s & D_s \end{bmatrix} \geqslant 0.$$

By post- and pre-multiplying this expression by $\begin{bmatrix} X_- \\ U_- \end{bmatrix}$ and its transpose, we conclude that (6.19) holds. □

Remark 6.2. Condition (6.18) implies that even if the state is measured (a more advantageous situation than knowing only the input-output data, as is typically assumed in data-driven applications), it is *only* possible to ascertain dissipativity from data if the plant is *uniquely* identifiable, i.e., if $|\Sigma^{\mathcal{N}_0}| = 1$. Consequently, in the noise-free setting, methods for determining dissipativity directly from data are conceptually equivalent with indirect ones consisting of a system identification stage, followed by a second one involving a check on the solvability of an LMI (condition (6.4)).

6.4.3 Informativity and noisy data

We first consider the noise model $\mathcal{N}_1$ defined in (6.9). Define

$$N_1 := \left[\begin{array}{c|c} I & \begin{matrix} X_+ \\ Y_- \end{matrix} \\ \hline 0 & \begin{matrix} -X_- \\ -U_- \end{matrix} \end{array}\right] \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{bmatrix} \left[\begin{array}{c|c} I & \begin{matrix} X_+ \\ Y_- \end{matrix} \\ \hline 0 & \begin{matrix} -X_- \\ -U_- \end{matrix} \end{array}\right]^\top. \tag{6.21}$$

Note that $(A, B, C, D) \in \Sigma^{\mathcal{N}_1}$ if and only if

$$\left[\begin{array}{cc} \multicolumn{2}{c}{I} \\ \hline A^\top & C^\top \\ B^\top & D^\top \end{array}\right]^\top N_1 \left[\begin{array}{cc} \multicolumn{2}{c}{I} \\ \hline A^\top & C^\top \\ B^\top & D^\top \end{array}\right] \geqslant 0. \tag{6.22}$$

Partition

$$S = \begin{bmatrix} F & G \\ G^\top & H \end{bmatrix},$$

where $F \in \mathbb{R}^{m\times m}$, $G \in \mathbb{R}^{m\times p}$, $H \in \mathbb{R}^{p\times p}$, and define

$$M_1 := \begin{bmatrix} P & 0 & 0 & 0 \\ 0 & F & 0 & G \\ 0 & 0 & -P & 0 \\ 0 & G^\top & 0 & H \end{bmatrix}.$$

With this notation in place, the problem of informativity for dissipativity thus boils down to the question under which conditions the inequality

$$\left[\begin{array}{cc} \multicolumn{2}{c}{I} \\ \hline A & B \\ C & D \end{array}\right]^\top M_1 \left[\begin{array}{cc} \multicolumn{2}{c}{I} \\ \hline A & B \\ C & D \end{array}\right] \geqslant 0 \tag{6.23}$$

holds for all (A, B, C, D) satisfying

$$\left[\begin{array}{cc} \multicolumn{2}{c}{I} \\ \hline A^\top & C^\top \\ B^\top & D^\top \end{array}\right]^\top N_1 \left[\begin{array}{cc} \multicolumn{2}{c}{I} \\ \hline A^\top & C^\top \\ B^\top & D^\top \end{array}\right] \geqslant 0. \tag{6.24}$$

Our strategy to get a grip on this question is to invoke the matrix S-lemma [218], see also Chapter 5. We recall the result in what follows.

Proposition 6.1 (Matrix S-lemma)**.** Let $M, N \in \mathbb{R}^{(q+r)\times(q+r)}$ be symmetric matrices. Assume that there exists a matrix $\bar{Z} \in \mathbb{R}^{r\times q}$ such that

$$\begin{bmatrix} I \\ \bar{Z} \end{bmatrix}^\top N \begin{bmatrix} I \\ \bar{Z} \end{bmatrix} > 0. \tag{6.25}$$

Then we have that

$$\begin{bmatrix} I \\ Z \end{bmatrix}^\top M \begin{bmatrix} I \\ Z \end{bmatrix} \geqslant 0 \quad \text{for all } \; Z \in \mathbb{R}^{r\times q} \text{ with } \begin{bmatrix} I \\ Z \end{bmatrix}^\top N \begin{bmatrix} I \\ Z \end{bmatrix} \geqslant 0$$

if and only if there exists a scalar $\alpha \geqslant 0$ such that $M - \alpha N \geqslant 0$.

Before we can apply Proposition 6.1 we note that the inequality (6.23) is in terms of (A, B, C, D) while the inequality (6.24) is in terms of the *transposed* matrices $(A^\top, C^\top, B^\top, D^\top)$. Therefore, we will need an additional *dualization* result that we formulate in the following proposition.

Proposition 6.2. Let

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathbb{R}^{(n+p)\times(n+m)} \text{ and } S = S^\top \in \mathbb{R}^{(m+p)\times(m+p)}.$$

Suppose that S satisfies Assumption (A1). Define

$$\hat{S} := \begin{bmatrix} 0 & -I_p \\ I_m & 0 \end{bmatrix} S^{-1} \begin{bmatrix} 0 & -I_m \\ I_p & 0 \end{bmatrix}. \tag{6.26}$$

A real $n \times n$ matrix $P = P^\top > 0$ satisfies

$$L_1 := \begin{bmatrix} I & 0 \\ A & B \end{bmatrix}^\top \begin{bmatrix} P & 0 \\ 0 & -P \end{bmatrix} \begin{bmatrix} I & 0 \\ A & B \end{bmatrix} + \begin{bmatrix} I & 0 \\ C & D \end{bmatrix}^\top S \begin{bmatrix} I & 0 \\ C & D \end{bmatrix} \geqslant 0 \tag{6.27}$$

if and only if

$$L_2 := \begin{bmatrix} I & 0 \\ A^\top & C^\top \end{bmatrix}^\top \begin{bmatrix} P^{-1} & 0 \\ 0 & -P^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ A^\top & C^\top \end{bmatrix} + \begin{bmatrix} 0 & I \\ B^\top & D^\top \end{bmatrix}^\top \hat{S} \begin{bmatrix} 0 & I \\ B^\top & D^\top \end{bmatrix} \geqslant 0. \tag{6.28}$$

The proof of Proposition 6.2 is postponed to Section 6.6.3. Next, we will use Propositions 6.1 and 6.2 to prove the following characterization of informativity for dissipativity, given the noise model $\mathcal{N}_1$.

Theorem 6.3. Suppose that there exists $V \in \mathbb{R}^{(n+m)\times(n+p)}$ such that

$$\begin{bmatrix} I \\ V \end{bmatrix}^\top N_1 \begin{bmatrix} I \\ V \end{bmatrix} > 0. \tag{6.29}$$

Partition

$$\begin{bmatrix} \hat{F} & \hat{G} \\ \hat{G}^\top & \hat{H} \end{bmatrix} := -S^{-1},$$

where $\hat{F} = \hat{F}^\top \in \mathbb{R}^{m\times m}$, $\hat{G} \in \mathbb{R}^{m\times p}$, and $\hat{H} = \hat{H}^\top \in \mathbb{R}^{p\times p}$.

Given the noise model $\mathcal{N}_1$, the data (U_-, X, Y_-) are informative for dissipativity with respect to the supply rate (6.2) if and only if there exist a real $n \times n$ matrix $Q = Q^\top > 0$ and a scalar $\alpha \geqslant 0$ such that (LMI) holds.

Proof. To prove the "if" statement, let $(A, B, C, D) \in \Sigma^{\mathcal{N}_1}$. We multiply (LMI) from right and left by

$$\begin{bmatrix} I & 0 \\ 0 & I \\ A^\top & C^\top \\ B^\top & D^\top \end{bmatrix}$$

and its transpose. By the assumption on the noise (see Equation (6.9)), this leads to

$$\begin{bmatrix} I & 0 \\ A^\top & C^\top \end{bmatrix}^\top \begin{bmatrix} Q & 0 \\ 0 & -Q \end{bmatrix} \begin{bmatrix} I & 0 \\ A^\top & C^\top \end{bmatrix} + \begin{bmatrix} 0 & I \\ B^\top & D^\top \end{bmatrix}^\top \hat{S} \begin{bmatrix} 0 & I \\ B^\top & D^\top \end{bmatrix} \geqslant 0,$$

where $\hat{S}$ is related to S via (6.26). Finally, by Proposition 6.2 we conclude that (6.27) holds for $P = Q^{-1}$. That is, the data (U_-, X, Y_-) are informative for dissipativity with respect to the supply rate (6.2).

To prove the "only if" part, let $P = P^\top \geqslant 0$ satisfy

$$\begin{bmatrix} I & 0 \\ A & B \end{bmatrix}^\top \begin{bmatrix} P & 0 \\ 0 & -P \end{bmatrix} \begin{bmatrix} I & 0 \\ A & B \end{bmatrix} + \begin{bmatrix} 0 & I \\ C & D \end{bmatrix}^\top S \begin{bmatrix} 0 & I \\ C & D \end{bmatrix} \geqslant 0 \tag{6.30}$$

for all $(A, B, C, D) \in \Sigma^{\mathcal{N}_1}$. Let $\xi \in \ker P$. It follows from (6.30) that

$$\begin{bmatrix} \alpha \\ \eta \end{bmatrix}^\top \left(- \begin{bmatrix} \xi^\top A^\top \\ B^\top \end{bmatrix} P \begin{bmatrix} A\xi & B \end{bmatrix} + \begin{bmatrix} 0 & I \\ C\xi & D \end{bmatrix}^\top S \begin{bmatrix} 0 & I \\ C\xi & D \end{bmatrix} \right) \begin{bmatrix} \alpha \\ \eta \end{bmatrix} \geqslant 0$$

for all $\alpha \in \mathbb{R}$, $\eta \in \mathbb{R}^m$, and $(A, B, C, D) \in \Sigma^{\mathcal{N}_1}$. This implies that

$$R := \begin{bmatrix} 0 & I \\ C\xi & D \end{bmatrix}^\top S \begin{bmatrix} 0 & I \\ C\xi & D \end{bmatrix} \geqslant 0$$

$$\begin{bmatrix} Q & 0 & 0 & 0 \\ 0 & \hat{H} & 0 & -\hat{G}^\top \\ 0 & 0 & -Q & 0 \\ 0 & -\hat{G} & 0 & \hat{F} \end{bmatrix} - \alpha \left[\begin{array}{c|c} I & \begin{matrix} X_+ \\ Y_- \end{matrix} \\ \hline 0 & \begin{matrix} -X_- \\ -U_- \end{matrix} \end{array}\right] \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{bmatrix} \left[\begin{array}{c|c} I & \begin{matrix} X_+ \\ Y_- \end{matrix} \\ \hline 0 & \begin{matrix} -X_- \\ -U_- \end{matrix} \end{array}\right]^\top \geqslant 0. \tag{LMI}$$

for every $(A,B,C,D) \in \Sigma^{\mathcal{N}_1}$. It follows from [45, Theorem 3.1] that

$$\dim(\ker R) \geqslant 1.$$

Therefore, $C\xi = 0$ for every $(A,B,C,D) \in \Sigma^{\mathcal{N}_1}$. The hypothesis (6.29) implies that the set $\Sigma^{\mathcal{N}_1}$ has nonempty interior. Consequently, we can conclude that $\xi = 0$ and hence $P > 0$.

Now, by Proposition 6.2 it follows that (6.28) holds for all $(A,B,C,D) \in \Sigma^{\mathcal{N}_1}$. We define $Q := P^{-1}$. By rearranging terms in (6.28) we see that

$$\begin{bmatrix} I & 0 \\ 0 & I \\ A^\top & C^\top \\ B^\top & D^\top \end{bmatrix}^\top \begin{bmatrix} Q & 0 & 0 & 0 \\ 0 & \hat{H} & 0 & -\hat{G}^\top \\ 0 & 0 & -Q & 0 \\ 0 & -\hat{G} & 0 & \hat{F} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & I \\ A^\top & C^\top \\ B^\top & D^\top \end{bmatrix} \geqslant 0$$

holds for all $(A,B,C,D) \in \Sigma^{\mathcal{N}_1}$, i.e., for all (A,B,C,D) satisfying (6.24). Finally, by the matrix S-lemma, Proposition 6.1, there exists a scalar $\alpha \geqslant 0$ such that (LMI) holds. This completes the proof. □

Remark 6.3. Under Assumption (A2) we can always transform the noise model $\mathcal{N}_1$ to $\mathcal{N}_2$ and vice versa by Lemma 6.2. Therefore, by combining Theorem 6.3 and Lemma 6.2 we can also come up with necessary and sufficient conditions for informativity for dissipativity given the noise model $\mathcal{N}_2$. This again results in a data-based LMI condition for dissipativity, analogous to (LMI).

6.5 CONCLUSIONS

In this chapter we have provided methods to verify dissipativity properties of linear systems directly from measured data. We have focused both on exact and noisy data. In the case of exact data, we have proven that one can only ascertain dissipativity of a system from given data if the system can be uniquely identified from the data. If this is the case, dissipativity can be verified by means of a data-based linear matrix inequality. In the case of noisy data, we have leveraged the matrix S-lemma and a type of dualization lemma to characterize data informativity for dissipativity. Also in this setting, dissipativity properties of the data-generating system can be ascertained if a data-based LMI is solvable.

6.6 PROOFS OF AUXILIARY RESULTS

6.6.1 Proof of Lemma 6.1

Proof. Let

$$\Psi = \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{12}^\top & \Psi_{22} \end{bmatrix}.$$

To prove sufficiency, note that $\Psi_{22} < 0$ implies that $\mathcal{N}$ is bounded. Define $R := -\Psi_{22}^{-1}\Psi_{12}^\top$ and note that $\Psi_{11} - \Psi_{12}\Psi_{22}^{-1}\Psi_{12}^\top > 0$ implies

$$f(R) := \begin{bmatrix} I \\ R \end{bmatrix}^\top \Psi \begin{bmatrix} I \\ R \end{bmatrix} > 0.$$

This means that $\mathcal{N}$ has nonempty interior.

To prove necessity, we first show that $\Psi_{22} < 0$. Let $R \in \operatorname{int}(\mathcal{N})$ and let $\xi \in \mathbb{R}^r$ be such that $\xi^\top \Psi_{22}\xi \geqslant 0$. Then $f(R + \alpha\xi\xi^\top(\Psi_{12}^\top + \Psi_{22}R)) \geqslant 0$ for all $\alpha \geqslant 0$. Since $\mathcal{N}$ is bounded, we conclude that

$$\xi^\top(\Psi_{12}^\top + \Psi_{22}R) = 0. \tag{6.31}$$

Since $R \in \operatorname{int}(\mathcal{N})$ is arbitrary, it follows that for all $R_1, R_2 \in \operatorname{int}(\mathcal{N})$ the equality $\xi^\top\Psi_{22}(R_1 - R_2) = 0$ holds. This implies that $\xi^\top\Psi_{22} = 0$ and, by (6.31), also $\xi^\top\Psi_{12}^\top = 0$. Now, observe that $f(R + \alpha\xi\xi^\top) = f(R) \geqslant 0$ for all $\alpha \in \mathbb{R}$ and $R \in \mathcal{N}$. By boundedness of $\mathcal{N}$, this implies that $\xi = 0$. Therefore, $\Psi_{22} < 0$.

To prove the rest of the claim, let $\zeta \in \mathbb{R}^q$ and $\eta \in \mathbb{R}^r$ be such that

$$\Psi \begin{bmatrix} \zeta \\ \eta \end{bmatrix} = 0. \tag{6.32}$$

If $R \in \mathcal{N}$ then

$$\begin{aligned} 0 \leqslant \zeta^\top f(R)\zeta &= \begin{bmatrix} \zeta \\ R\zeta \end{bmatrix}^\top \Psi \begin{bmatrix} \zeta \\ R\zeta \end{bmatrix} \\ &= \left(\begin{bmatrix} 0 \\ R\zeta - \eta \end{bmatrix} + \begin{bmatrix} \zeta \\ \eta \end{bmatrix} \right)^\top \Psi \left(\begin{bmatrix} 0 \\ R\zeta - \eta \end{bmatrix} + \begin{bmatrix} \zeta \\ \eta \end{bmatrix} \right), \end{aligned}$$

and thus $0 \leqslant (R\zeta - \eta)^\top\Psi_{22}(R\zeta - \eta)$. Since $\Psi_{22} < 0$, we conclude that $R\zeta - \eta = 0$ for all $R \in \mathcal{N}$. This implies that $(R_1 - R_2)\zeta = 0$ for all $R_1, R_2 \in \mathcal{N}$. Since the interior of $\mathcal{N}$ is nonempty, we conclude that $\zeta = 0$. Thus, (6.32) leads to $\Psi_{22}\eta = 0$ and since $\Psi_{22} < 0$, we conclude $\eta = 0$. Therefore, Ψ is nonsingular. By Haynsworth's inertia formula (see [19, Fact 6.5.5]),

$$\operatorname{In}(\Psi) = \operatorname{In}(\Psi_{22}) + \operatorname{In}(\Psi_{11} - \Psi_{12}\Psi_{22}^{-1}\Psi_{12}^\top). \tag{6.33}$$

Let ν be the number of negative eigenvalues of Ψ. From $\Psi_{22} < 0$ and (6.33) we see that $\nu \geqslant r$. Since $\mathcal{N}$ is nonempty, it follows from [45, Thm. 3.1] that $\nu \leqslant r$. Therefore, we conclude that $\nu = r$. Since Ψ is nonsingular, (6.33) implies that $\Psi_{11} - \Psi_{12}\Psi_{22}^{-1}\Psi_{12}^\top > 0$, proving the claim. □

6.6.2 Proof of Lemma 6.2

Proof. Let

$$\begin{bmatrix} \hat{\Psi}_{11} & \hat{\Psi}_{12} \\ \hat{\Psi}_{12}^\top & \hat{\Psi}_{22} \end{bmatrix} := -\Psi^{-1}$$

where $\hat{\Psi}_{11} \in \mathbb{R}^{q\times q}$, $\hat{\Psi}_{12} \in \mathbb{R}^{q\times r}$, and $\hat{\Psi}_{22} \in \mathbb{R}^{r\times r}$. Also let $R \in \mathbb{R}^{r\times q}$ and define

$$M_R := \begin{bmatrix} 0 & I & R^\top \\ I & \hat{\Psi}_{11} & \hat{\Psi}_{12} \\ R & \hat{\Psi}_{12}^\top & \hat{\Psi}_{22} \end{bmatrix}.$$

By Haynsworth's inertia theorem (see [19, Fact 6.5.5]) we have

$$\operatorname{In}(M_R) = \operatorname{In}(-\Psi^{-1}) + \operatorname{In}\left(\begin{bmatrix} I \\ R \end{bmatrix}^\top \Psi \begin{bmatrix} I \\ R \end{bmatrix}\right). \tag{6.34}$$

Next, we define

$$N := \begin{bmatrix} 0 & I \\ I & \hat{\Psi}_{11} \end{bmatrix}.$$

Note that

$$N^{-1} = \begin{bmatrix} -\hat{\Psi}_{11} & I \\ I & 0 \end{bmatrix}.$$

By [117, Lemma 5.1] the matrix N has inertia $\operatorname{In}(N) = (q, 0, q)$. We also have that the Schur complement of M_R with respect to N is given by

$$\begin{aligned} \hat{\Psi}_{22} - \begin{bmatrix} R & \hat{\Psi}_{12}^\top \end{bmatrix} \begin{bmatrix} -\hat{\Psi}_{11} & I \\ I & 0 \end{bmatrix} \begin{bmatrix} R^\top \\ \hat{\Psi}_{12} \end{bmatrix} &= \\ \hat{\Psi}_{22} + R\hat{\Psi}_{11}R^\top - R\hat{\Psi}_{12} - \hat{\Psi}_{12}^\top R^\top &= \\ \begin{bmatrix} I \\ R^\top \end{bmatrix}^\top \begin{bmatrix} \hat{\Psi}_{22} & -\hat{\Psi}_{12}^\top \\ -\hat{\Psi}_{12} & \hat{\Psi}_{11} \end{bmatrix} \begin{bmatrix} I \\ R^\top \end{bmatrix} &= \\ \begin{bmatrix} I \\ R^\top \end{bmatrix}^\top \Xi \begin{bmatrix} I \\ R^\top \end{bmatrix}&. \end{aligned}$$

This implies that

$$\operatorname{In}(M_R) = \operatorname{In}(N) + \operatorname{In}\left(\begin{bmatrix} I \\ R^\top \end{bmatrix}^\top \Xi \begin{bmatrix} I \\ R^\top \end{bmatrix}\right). \tag{6.35}$$

Since $\Psi_{22} < 0$ and $\Psi_{11} - \Psi_{12}\Psi_{22}^{-1}\Psi_{12}^\top > 0$ by our hypotheses, we know that $\operatorname{In}(\Psi) = (r, 0, q)$. Therefore, (6.34) and (6.35) imply

$$(q, 0, r) + \operatorname{In}\left(\begin{bmatrix} I \\ R \end{bmatrix}^\top \Psi \begin{bmatrix} I \\ R \end{bmatrix}\right) = (q, 0, q) + \operatorname{In}\left(\begin{bmatrix} I \\ R^\top \end{bmatrix}^\top \Xi \begin{bmatrix} I \\ R^\top \end{bmatrix}\right).$$

This means that

$$\begin{bmatrix} I \\ R \end{bmatrix}^\top \Psi \begin{bmatrix} I \\ R \end{bmatrix} \geqslant 0$$

if and only if

$$\begin{bmatrix} I \\ R^\top \end{bmatrix}^\top \Xi \begin{bmatrix} I \\ R^\top \end{bmatrix} \geqslant 0,$$

which proves the lemma. □

6.6.3 Proof of Proposition 6.2

Proof. Denote

$$\begin{bmatrix} \hat{F} & \hat{G} \\ \hat{G}^\top & \hat{H} \end{bmatrix} := -S^{-1}$$

where $\hat{F} = \hat{F}^\top \in \mathbb{R}^{m\times m}$, $\hat{G} \in \mathbb{R}^{m\times p}$, and $\hat{H} = \hat{H}^\top \in \mathbb{R}^{p\times p}$. Define the matrix

$$M := \begin{bmatrix} P & 0 & A^\top & 0 & C^\top \\ 0 & 0 & B^\top & I_m & D^\top \\ A & B & P^{-1} & 0 & 0 \\ 0 & I_m & 0 & \hat{F} & \hat{G} \\ C & D & 0 & \hat{G}^\top & \hat{H} \end{bmatrix}.$$

In addition, let L_1 be the Schur complement of M with respect to its submatrix

$$\begin{bmatrix} P^{-1} & 0 & 0 \\ 0 & \hat{F} & \hat{G} \\ 0 & \hat{G}^\top & \hat{H} \end{bmatrix} =: N_1.$$

Apply Haynsworth's inertia theorem (see [19, Fact 6.5.5]) to M, and conclude that

$$\begin{aligned} \operatorname{In}(M) &= \operatorname{In}(N_1) + \operatorname{In}(L_1) \\ &= \operatorname{In}(P) + \operatorname{In}(-S) + \operatorname{In}(L_1). \end{aligned} \tag{6.36}$$

Now consider the following submatrix of M:

$$N_2 := \begin{bmatrix} P & 0 & 0 \\ 0 & 0 & I_m \\ 0 & I_m & \hat{F} \end{bmatrix}.$$

It follows from [117, Lemma 5.1] that

$$\operatorname{In}(N_2) = \operatorname{In}(P) + (m, 0, m).$$

The Schur complement of M with respect to N_2 is given by

$$\begin{bmatrix} P^{-1} & 0 \\ 0 & \hat{H} \end{bmatrix} - \begin{bmatrix} A & B & 0 \\ C & D & \hat{G}^\top \end{bmatrix} \begin{bmatrix} P^{-1} & 0 & 0 \\ 0 & -\hat{F} & I_m \\ 0 & I_m & 0 \end{bmatrix} \begin{bmatrix} A & B & 0 \\ C & D & \hat{G}^\top \end{bmatrix}^\top.$$

By rearranging terms, this expression can be rewritten as

$$L_2 := \begin{bmatrix} I_n & 0 \\ A^\top & C^\top \end{bmatrix}^\top \begin{bmatrix} P^{-1} & 0 \\ 0 & -P^{-1} \end{bmatrix} \begin{bmatrix} I_n & 0 \\ A^\top & C^\top \end{bmatrix} + \begin{bmatrix} 0 & I_p \\ B^\top & D^\top \end{bmatrix}^\top \hat{S} \begin{bmatrix} 0 & I_p \\ B^\top & D^\top \end{bmatrix}.$$

Applying Haynsworth's inertia theorem again, we conclude that

$$\begin{aligned} \operatorname{In}(M) &= \operatorname{In}(N_2) + \operatorname{In}(L_2) \\ &= \operatorname{In}(P) + (m, 0, m) + \operatorname{In}(L_2). \end{aligned} \tag{6.37}$$

Since $\operatorname{In}(-S) = (m, 0, p)$, (6.36) and (6.37) imply that $\operatorname{In}(L_2) = (0, 0, p - m) + \operatorname{In}(L_1)$. Then, it follows that $L_1 \geqslant 0$ if and only if $L_2 \geqslant 0$. □

7 TOPOLOGY IDENTIFICATION OF HETEROGENEOUS NETWORKS

In Chapter 2 we saw how to identify state-space models from one or multiple input/output trajectories. The focus of this chapter is on the identification of state-space models with a *certain structure*. In particular, we will be interested in identifying a state-space model that consists of several known subsystems that are interconnected through an unknown topology. The problem then boils down to identifying the unknown interactions between the subsystems from input/output data. We will focus both on an identifiability aspect (i.e., the question when identification is conceptually possible), and the topology identification problem itself.

7.1 INTRODUCTION

Graph structure plays an important role in the overall behavior of dynamical networks. Indeed, it is well-known that the convergence rate of consensus algorithms depends on the connectivity of the network topology. In addition, many properties of dynamical networks, like controllability, can be assessed on the basis of the network graph [29,96,113]. Unfortunately, the graph structure of dynamical networks is often unknown. This problem is particularly apparent in biology, for example in neural networks and genetic networks [97], but also emerges in other areas such as power grids [28].

To deal with this problem, several topology identification methods have been developed. Such methods aim at reconstructing the topology (and weights) of a dynamical network on the basis of measured data obtained from the network.

The paper [70] studies necessary and sufficient conditions for dynamical structure reconstruction, see also [246]. A node-knockout scheme for topology identification was introduced in [153] and further investigated in [202]. Moreover, the paper [184] studies topology identification using compressed sensing, while [130] considers network reconstruction using Wiener filtering. A distributed algorithm for network reconstruction has also been studied [145]. The paper [190] studies topology identification using power spectral analysis. In [226], the network topology was reconstructed by solving certain Lyapunov equations. A Bayesian approach to the network identification problem was investigated in [32]. The network topology was inferred from multiple independent observations of consensus dynamics in [189]. The paper [41] studies topology identification via subspace methods. There are also several results for topology reconstruction of nonlinear systems, see e.g., [192,206,231] albeit in this case few guarantees on the accuracy of identification can be given. In addition, we remark that the complementary problem of identifying the nodes dynamics assuming a *known* topology has also been studied, see e.g. [31,76,81,172,214,223,224], along with the joint topology and dynamics recovery problem [92,229].

The goal of this chapter is to provide a comprehensive treatment of topology identification for linear MIMO heterogeneous networks, with no assumptions on the network structure such as sparsity or regularity. Most existing work on topology identification emphasizes the role of the network topology by considering relatively simple node dynamics. For example, networks of single integrators have been studied in [79,145,153,226]. In addition, the papers [202] and [190] consider homogeneous networks comprised of identical single-input single-output systems. Nonetheless, there are many examples of networks in which the subsystems are not necessarily the same, for example, mass-spring-damper networks [102], where the masses at the nodes can be distinct. Heterogeneity in the node dynamics has also been studied in the detail in synchronization problems, see e.g. [238,245].

We study topology identification for the general class of *heterogeneous* networks, where the node dynamics are modelled by general, possibly distinct, MIMO linear systems. We divide our analysis in two parts, namely the study of *identifiability* and the development of *identification algorithms*. The study of identifiability of the network topology deals with the question whether there exists a data set from which the topology can be uniquely identified. Identifiability of the topology is hence a property of the node systems and the network graph, and is *independent* of any data. Topological identifiability is an important property. Indeed, if it is not satisfied, then it is impossible to uniquely identify the network topology, regardless of the amount and richness of the data. After studying topological identifiability, we will turn our attention towards identification algorithms. Our two main contributions are hence the following:

1. We provide conditions for topological identifiability of general heterogeneous networks. Our results recover an identifiability result for the special case of networks of single integrators [163, 226]. We will also see that homogeneous networks of single-input single-output systems have quite special identifiability properties that do not extend to the general case of heterogeneous networks.

2. We establish a topology identification scheme for heterogeneous networks. The idea of the method is to reconstruct the interconnection matrix of the network by solving a *generalized Sylvester equation* involving the Markov parameters of the network. We prove that the network topology can be uniquely reconstructed in this way, under the assumptions of topological identifiability and persistency of excitation of the input data.

The chapter is organized as follows. In Section 7.2 we formulate the problem. Section 7.3 contains our results on topological identifiability. Subsequently, we describe our topology identification method in Section 7.4. Finally, we state our conclusions in Section 7.5.

Notation

The *direct sum* of matrices $A_1, A_2, \ldots, A_k$ is the block diagonal matrix defined by

$$\bigoplus_{i=1}^{k} A_i := \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_k \end{bmatrix}.$$

Let $A(z)$ be an $n \times m$ rational matrix. Then the *constant kernel* of $A(z)$ is $\operatorname{cker} A(z) := \{w \in \mathbb{R}^m \mid A(z)w = 0\}$.

7.2 PROBLEM FORMULATION

We consider a network model similar to the one studied by Fuhrmann and Helmke [61, Ch. 9]. Specifically, we consider networks composed of N discrete-time systems of the form

$$\begin{aligned} x_i(t+1) &= A_i x_i(t) + B_i v_i(t) \\ w_i(t) &= C_i x_i(t), \end{aligned} \tag{7.1}$$

where $x_i(t) \in \mathbb{R}^{n_i}$ is the state of the i-th node system, $v_i(t) \in \mathbb{R}^{m_i}$ is its input and $w_i(t) \in \mathbb{R}^{p_i}$ is its output for $i = 1, 2, \ldots, N$. The real matrices A_i, B_i and C_i are of appropriate dimensions. We occasionally use the shorthand notation (A_i, B_i, C_i) to denote (7.1). The coupling between nodes is realized by the inputs $v_i(t)$, which are specified as

$$v_i(t) = \sum_{j=1}^{N} Q_{ij} w_j(t) + R_i u(t),$$

where $u(t) \in \mathbb{R}^m$ is the external network input and Q_{ij} and R_i are real matrices of appropriate dimensions. In addition, let S_i be a real $p \times p_i$ matrix and consider the external network output $y(t) \in \mathbb{R}^p$, defined by

$$y(t) = \sum_{i=1}^{N} S_i w_i(t).$$

Then, by introducing the block diagonal matrices

$$A = \bigoplus_{i=1}^{N} A_i,\ B = \bigoplus_{i=1}^{N} B_i,\ \text{and}\ C = \bigoplus_{i=1}^{N} C_i, \tag{7.2}$$

and the matrices

$$Q = \begin{bmatrix} Q_{11} & \cdots & Q_{1N} \\ \vdots & \ddots & \vdots \\ Q_{N1} & \cdots & Q_{NN} \end{bmatrix},\ R = \begin{bmatrix} R_1 \\ \vdots \\ R_N \end{bmatrix},\ S^\top = \begin{bmatrix} S_1^\top \\ \vdots \\ S_N^\top \end{bmatrix},$$

we can represent the network dynamics compactly as

$$\begin{aligned} x(t+1) &= (A+BQC)x(t) + BRu(t) \\ y(t) &= SCx(t). \end{aligned} \tag{7.3}$$

Here $x(t) = \mathrm{col}(x_1(t), x_2(t), \dots, x_N(t)) \in \mathbb{R}^n$ where n is defined as $n := \sum_{i=1}^N n_i$. We emphasize that the coupling of the node dynamics is induced by the matrix Q, which we will hence call the *interconnection matrix*.

There are a few important special cases of node dynamics (7.1) and resulting network dynamics (7.3). If $A_i = A_0$, $B_i = B_0$ and $C_i = C_0$ for all $i = 1, 2, \dots, N$, the dynamics of all nodes in the network are the same and the resulting dynamical network is called *homogeneous*. The more general setting in which the node dynamics are not necessarily the same is referred to as a *heterogeneous* network. Another special case of node dynamics occurs when $m_i = p_i = 1$ for all $i = 1, 2, \dots, N$. In this case, the node systems are single-input single-output (SISO) systems, and the resulting dynamical network is referred to as a *SISO network*[1]. Topology identification of homogeneous SISO networks has been studied in [202] and [190]. In addition, topology identification has been well-studied (see e.g [70, 79, 153, 226]) for networks of so-called *single-integrators*, in which the node dynamics are described by $\dot{x}_i(t) = v_i(t)$. This type of node dynamics can be seen continuous-time counterpart of (7.1) where $A_i = 0$, $B_i = 1$ and $C_i = 1$ for $i = 1, 2, \dots, N$.

The purpose of this chapter is to study topology identification for general, heterogeneous dynamical networks of the form (7.3). Although we focus on discrete-time systems, our results can be stated for continuous-time systems as well. In order to make the problem more precise, we first explain what we mean by the topology of (7.3). Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a weighted directed graph with $\mathcal{V} = \{1, 2, \dots, N\}$ and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ such that $(j, i) \in \mathcal{E}$ if and only if $Q_{ij} \neq 0$. Each edge $(j, i) \in \mathcal{E}$ is weighted by the nonzero matrix Q_{ij}. We refer to $\mathcal{G}$ as the *topology* of the dynamical network (7.3). With this in mind, the problem of *topology identification* concerns finding $\mathcal{G}$ (equivalently, finding Q) using measurements of the input $u(t)$ and output $y(t)$ of (7.3). We assume knowledge of the local node dynamics (i.e., the matrices A, B and C) as well as the external input/output matrices R and S[2].

At this point, we may ask the following natural question: is it possible to *uniquely* reconstruct the topology of (7.3) from input/output data? To formalize and answer this question, we define the notion of *topological identifiability*. Let $y_{u,x_0,Q}(t)$ denote the output of (7.3) at time t, where the subscript emphasizes the dependence on the input $u(\cdot)$, the initial condition $x_0 = x(0)$ and interconnection matrix Q. The following definition is inspired by [73] and defines the notion of *distinguishability* of interconnection matrices.

[1] Here we emphasize that "SISO" refers to the node systems of the network. The overall network dynamics (7.3) can still have multiple external inputs and outputs.

[2] This assumption is standard in the literature on topology identification, see, e.g., [190] and [202]. Without knowledge of the node dynamics, topology identification becomes a full system identification problem.

Definition 7.1. Let $y_{u,x_0,Q}(\cdot)$ and $y_{u,\bar{x}_0,\bar{Q}}(\cdot)$ denote the output trajectories of two systems of the form (7.3) with interconnection matrices Q and $\bar{Q}$ and initial conditions x_0 and $\bar{x}_0$, respectively. We say that Q and $\bar{Q}$ are *indistinguishable* if there exist initial conditions $x_0, \bar{x}_0 \in \mathbb{R}^n$ such that

$$y_{u,x_0,Q}(\cdot) = y_{u,\bar{x}_0,\bar{Q}}(\cdot)$$

for all input functions u. Moreover, Q and $\bar{Q}$ are said to be *distinguishable* if they are not indistinguishable.

With this in mind, the topology of (7.3) is said to be identifiable if Q is distinguishable from all other interconnection matrices. More formally, we have the following definition.

Definition 7.2. Consider system (7.3) with interconnection matrix Q. The topology of system (7.3) is said to be *identifiable* if Q and $\bar{Q}$ are distinguishable for all real $\bar{Q} \neq Q$.

The importance of topological identifiability lies in the fact that unique reconstruction of Q from input/output data is *only* possible if the topology of (7.3) is identifiable. Indeed, if this is not the case, there exists some $\bar{Q} \neq Q$ that is indistinguishable from Q, meaning that both Q and $\bar{Q}$ explain *any* input/output trajectory of (7.3). Topological identifiability is hence a structural property of the system (7.3) that is independent of a particular data sequence and that is *necessary* for the unique reconstruction of Q from data.

Following [73], it is straightforward to characterize topological identifiability in terms of the transfer matrix from u to y. This transfer function will be denoted by

$$F_Q(z) := SC(zI - A - BQC)^{-1}BR. \tag{7.4}$$

Proposition 7.1. The topology of the networked system (7.3) is identifiable if and only if the following implication holds:

$$F_Q(z) = F_{\bar{Q}}(z) \text{ for real } \bar{Q} \implies Q = \bar{Q}.$$

Although Proposition 7.1 provides a necessary and sufficient condition for topological identifiability, the condition involves the arbitrary matrix $\bar{Q}$. Hence, it is not clear how to verify the condition of Proposition 7.1. Instead, in this chapter we want to establish conditions for topological identifiability in terms of the local system matrices A, B and C and the matrices Q, R and S. This is formalized in the following problem.

Problem 7.1. Find necessary and sufficient conditions on the node dynamics A, B, C, the external input/output matrices R, S and the interconnection matrix Q under which the topology of (7.3) is identifiable.

Our second goal is to identify Q from input/output data.

Problem 7.2. Develop a methodology to identify the interconnection matrix Q from measurements of the input $u(\cdot)$ and output $y(\cdot)$ of system (7.3).

7.3 CONDITIONS FOR TOPOLOGICAL IDENTIFIABILITY

In this section we state our solution to Problem 7.1 by providing necessary and sufficient conditions for topological identifiability. We start by providing an overview of the results that are proven in this section. In the following table, "**N**" denotes necessary and "**S**" denotes sufficient.

Thm. 7.1	General **N**-**S** conditions
Thm. 7.2	**N** condition; also **S** if R has full rank
Thm. 7.3	**N** condition for homogeneous SISO networks
Thm. 7.4	**N**-**S** conditions for homog. SISO networks

For analysis purposes, we first rewrite the network transfer matrix $F_Q(z)$. Note that

$$zI - A = (zI - A - BQC) + BQC.$$

Premultiplication by $(zI - A)^{-1}$ and postmultiplication by $(zI - A - BQC)^{-1}$ yields

$$(zI - A - BQC)^{-1} = (zI - A)^{-1} + (zI - A)^{-1} BQC(zI - A - BQC)^{-1}.$$

This means that

$$C(zI - A - BQC)^{-1}B = G(z) + G(z)QC(zI - A - BQC)^{-1}B,$$

where $G(z) = C(zI - A)^{-1}B$ is a block diagonal matrix containing the transfer matrices of all node systems. Finally, by rearranging terms we obtain

$$C(zI - A - BQC)^{-1}B = (I - G(z)Q)^{-1} G(z). \tag{7.5}$$

Note that the inverse of $I - G(z)Q$ exists as a rational matrix. Indeed, since $(zI - A)^{-1}$ is strictly proper we see that $\lim_{z\to\infty}(I - G(z)Q) = I$. Therefore, we conclude by (7.5) that the transfer matrix $F_Q(z)$ equals

$$F_Q(z) = S\,(I - G(z)Q)^{-1}\, G(z)R. \tag{7.6}$$

We remark that (7.6) is an attractive representation of the network transfer matrix, since the matrices A, B and C describing the local system dynamics are grouped and contained in the transfer matrix $G(z)$.

Remark 7.1. By (7.6), we see that the networked system (7.3) can be represented by the block diagram in Figure 7.1. Hence, the problem of topology identification can be viewed as the identification of the *static output feedback gain* Q, assuming knowledge of the system $G(z)$ and the external input/output matrices R and S.

The following theorem gives necessary and sufficient conditions for topological identifiability. We will use the notation $G_i(z) := C_i(zI - A_i)^{-1}B_i$ to denote the transfer matrix from v_i to w_i of node system $i \in \mathcal{V}$.

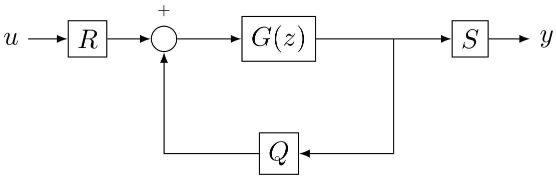

Figure 7.1: Block diagram of the networked system (7.3).

Theorem 7.1. Consider the networked system (7.3) and assume that the matrix S has full column rank. The topology of (7.3) is identifiable if and only if

$$\operatorname{cker}\left(G_i(z) \otimes H_Q^\top(z)\right) = \{0\} \text{ for all } i \in \mathcal{V}, \tag{7.7}$$

where $H_Q(z) := (I - G(z)Q)^{-1} G(z)R$.

Proof. Suppose that $F_Q(z) = F_{\bar{Q}}(z)$, where $\bar{Q}$ is real. Then, from (7.6) we have

$$S\,(I - G(z)Q)^{-1} G(z)R = S\,(I - G(z)\bar{Q})^{-1} G(z)R.$$

By hypothesis, S has full column rank and hence

$$(I - G(z)Q)^{-1} G(z)R = (I - G(z)\bar{Q})^{-1} G(z)R. \tag{7.8}$$

We define $\Delta := Q - \bar{Q}$. Then, (7.8) is equivalent to each of the following statements:

$$\begin{aligned} (I - G(z)\bar{Q})\,(I - G(z)Q)^{-1} G(z)R &= G(z)R \\ (I - G(z)(Q - \Delta))\,(I - G(z)Q)^{-1} G(z)R &= G(z)R \\ G(z)\Delta\,(I - G(z)Q)^{-1} G(z)R &= 0 \\ G(z)\Delta H_Q(z) &= 0. \end{aligned}$$

Equivalently,

$$H_Q^\top(z)\Delta^\top G^\top(z) = 0. \tag{7.9}$$

Next, let $\operatorname{vec}(M)$ denote the vectorization of a matrix M. Then (7.9) is equivalent to

$$(G(z) \otimes H_Q^\top(z)) \operatorname{vec}(\Delta^\top) = 0. \tag{7.10}$$

By (7.10) it is clear that the topology of (7.3) is identifiable if and only if the constant kernel of $G(z) \otimes H_Q^\top(z)$ is zero. Finally, by the block diagonal structure of $G(z)$, this is equivalent to (7.7) which proves the theorem. □

By Theorem 7.1, topological identifiability is equivalent to $G_i(z) \otimes H_Q^\top(z)$ having zero constant kernel for all i. Note that this condition generally depends on the -a priori unknown- matrix Q. Notably, identifiability is *independent* of the particular matrix Q whenever all node inputs are excited and all node outputs are measured, as stated in the following theorem.

Theorem 7.2. Consider the networked system (7.3). If the topology of (7.3) is identifiable then

$$\operatorname{cker}\left(G_i^\top(z) \otimes G_j(z)\right) = \{0\} \tag{7.11}$$

for all $i, j \in \mathcal{V}$. In addition, suppose that S has full column rank and R has full row rank. Then the topology of (7.3) is identifiable *if and only if* (7.11) holds.

The importance of Theorem 7.2 lies in the fact that the identifiability condition (7.11) can be verified *without knowledge* of Q. This means that, whenever the rank conditions on S and R hold, one can check for topological identifiability before collecting data from the system.

Remark 7.2. A proper transfer matrix $T(z)$ has constant kernel $\{0\}$ if and only if the matrix $\operatorname{col}(M_0, M_1, \dots, M_r)$ has full column rank. Here $M_0, M_1, \dots, M_r$ are the Markov parameters of $T(z)$ and r is greater or equal to the order of $T(z)$. As such, the conditions of Theorems 7.1 and 7.2 can be verified by computing the rank of the Markov parameter matrices associated to the transfer matrices in (7.7) and (7.11).

Proof. We first prove the second statement. Suppose that S has full column rank and R has full row rank. Then $F_Q(z) = F_{\bar{Q}}(z)$ is equivalent to

$$(I - G(z)Q)^{-1} G(z) = (I - G(z)\bar{Q})^{-1} G(z).$$

We define $\Delta := Q - \bar{Q}$. Then, $F_Q(z) = F_{\bar{Q}}(z)$ is equivalent to

$$G(z)\Delta(I - G(z)Q)^{-1} G(z) = 0,$$

In other words, $G(z)\Delta G(z)(I - QG(z))^{-1} = 0$. This in turn is equivalent to $G(z)\Delta G(z) = 0$. In other words, $\left(G^\top(z) \otimes G(z)\right) \operatorname{vec}(\Delta) = 0$. Exploiting the block diagonal structure of $G(z)$, we conclude that the topology of (7.3) is identifiable if and only if (7.11) holds. □

A consequence of Theorem 7.2 is that identifiability of the topology of (7.3) implies that the constant kernel of both $G_i^\top(z)$ and $G_i(z)$ is zero for all $i \in \mathcal{V}$. Based on this fact, we relate topological identifiability and output controllability of the node systems.

Definition 7.3. Consider the system

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t), \end{aligned} \tag{7.12}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$ and $y \in \mathbb{R}^p$, and let $y_{u,x_0}(\cdot)$ denote the output trajectory of (7.12) for a given initial condition x_0 and input $u(\cdot)$. System (7.12) is called *output controllable* if for every $x_0 \in \mathbb{R}^n$ and $y_1 \in \mathbb{R}^p$ there exists an input $u(\cdot)$ and time instant $T \in \mathbb{N}$ such that $y_{x_0,u}(T) = y_1$.

Corollary 7.1. If the topology of (7.3) is identifiable then the systems (A_i, B_i, C_i) and $(A_i^\top, C_i^\top, B_i^\top)$ are output controllable for all $i \in \mathcal{V}$.

Proof. By Theorem 7.2, identifiability of the topology of (7.3) implies that the constant kernel of $G_i^\top(z)$ is zero for all $i \in \mathcal{V}$. Now, for $w \in \mathbb{R}^{p_i}$ we have $w^\top G_i(z) = 0$ if and only if $w^\top C_i A_i^k B_i = 0$ for all $k = 0, 1, \ldots$, equivalently, $w^\top C_i A_i^k B_i = 0$ for all $k = 0, 1, \ldots, n_i - 1$. Hence,

$$w^\top \begin{bmatrix} C_i B_i & C_i A_i B_i & \cdots & C_i A_i^{n-1} B_i \end{bmatrix} = 0 \implies w = 0.$$

The latter implication holds if and only if the output controllability matrix of (A_i, B_i, C_i) has full row rank, equivalently (A_i, B_i, C_i) is output controllable [208, Ex. 3.22]. The proof for the necessity of output controllability of $(A_i^\top, C_i^\top, B_i^\top)$ is analogous and hence omitted. □

Remark 7.3. Output controllability of (A_i, B_i, C_i) can be interpreted as an "excitability" condition. Indeed, it guarantees that we have enough freedom in steering the output $w_i(t)$ of each node $i \in \mathcal{V}$.

Example 7.1. We will now illustrate Theorems 7.1 and 7.2. Consider a network of $N = 10$ oscillators of the form

$$\begin{aligned} x_i(t+1) &= \begin{bmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{bmatrix} x_i(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} v_i(t) \\ w_i(t) &= \begin{bmatrix} 1 & 0 \end{bmatrix} x_i(t), \end{aligned}$$

where $\theta_i \in \mathbb{R}$ is a constant, given by $\theta_i = (0.2 + 0.01i)\pi$ for $i = 1, 2, \ldots, N$. The network topology is a *cycle graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ (with self-loops), defined by $\mathcal{V} := \{1, 2, \ldots, N\}$ and $\mathcal{E} := \{(i, j) \mid i - j \equiv -1, 0, 1 (\operatorname{mod} N)\}$. Here mod denotes the modulo operation and $\equiv$ denotes congruence. The network nodes are diffusively coupled, and an external input is applied to node 1, that is,

$$v_i(t) = \begin{cases} \frac{1}{2} \sum_{j \in \mathcal{N}_i} (w_j(t) - w_i(t)) + u(t) & \text{if } i = 1 \\ \frac{1}{2} \sum_{j \in \mathcal{N}_i} (w_j(t) - w_i(t)) & \text{otherwise,} \end{cases}$$

where $\mathcal{N}_i := \{j \mid (j, i) \in \mathcal{E}\}$. This means that the interconnection matrix Q is defined element-wise as

$$Q_{ij} = \begin{cases} 1 & \text{if } i = j \\ -\frac{1}{2} & \text{if } i \neq j \text{ and } (j, i) \in \mathcal{E} \\ 0 & \text{otherwise.} \end{cases}$$

Since we only externally influence the first node system, the corresponding matrix R is given by the first column of I. We assume that we externally measure all node outputs, meaning that $S = I$.

Using Theorem 7.1, we want to show that the topology of (7.3) is identifiable. First, note that the transfer function $G_i(z)$ of node system i is given by

$$G_i(z) = \frac{z - \cos\theta_i}{z^2 - 2z\cos\theta_i + 1},$$

which is nonzero for all $i \in \mathcal{V}$. Since $G_i(z)$ is scalar, Theorem 7.1 implies that the topology of (7.3) is identifiable if and only if $\operatorname{cker} H_Q^\top(z) = \{0\}$. This is equivalent to the output controllability of the system $(A + BQC, BR, C)$. It can be easily verified that the output controllability matrix

$$\begin{bmatrix} CBR & C(A+BQC)BR & \cdots & C(A+BQC)^{N-1}BR \end{bmatrix}$$

has full row rank. We therefore conclude by Theorem 7.1 that the topology of (7.3) is identifiable. Note that the rank of the output controllability matrix (and hence, identifiability) depends on the interconnection matrix Q.

Next, we discuss the scenario in which $R = I$. In this case, we can externally influence all nodes. Now, identifiability can be checked without knowledge of Q. In fact, by Theorem 7.2, the topology of (7.3) is identifiable if and only if $\operatorname{cker}\left(G_i^\top(z) \otimes G_j(z)\right) = \{0\}$. This condition is satisfied, since all local transfer functions are nonzero scalars.

So far, we have provided a general condition for identifiability in Theorem 7.1, and we have discussed some of the implications of this result in Theorem 7.2 and Corollary 7.1. However, possible criticism of the results may arise from the full rank condition on S in Theorem 7.1, which, until now, has been left rather unjustified.

It turns out that full column rank of S (or the dual, full row rank of R) is *necessary* for topological identifiability in case the networked system is homogeneous and SISO. For this important class of networked systems, the rank condition on S in Theorem 7.1 is hence not restrictive.

Theorem 7.3. Consider a homogeneous SISO network, that is, a system of the form (7.3) with $m_i = p_i = 1$ and $A_i = A_0$, $B_i = B_0$ and $C_i = C_0$ for all $i \in \mathcal{V}$. If the topology of (7.3) is identifiable then $\operatorname{rank} S = N$ or $\operatorname{rank} R = N$.

Remark 7.4. Theorem 7.3 generalizes several known results (see [163, 225, 226]) for networks of single-integrators. Indeed, in the special case that $A_0 = 0$, $B_0 = C_0 = 1$, the node output $w_i(t)$ equals the node state $x_i(t)$ for all $i \in \mathcal{V}$, and Theorem 7.3 asserts that either full state measurement or full state excitation is necessary for identifiability. This fact has been observed in different setups in [163, Thm. 1], [226, Rem. 2], and [225, Thm. 5].

Before proving Theorem 7.3, we state the following lemma.

Lemma 7.1. Suppose that $m_i = p_i = 1$ and $A_i = A_0$, $B_i = B_0$ and $C_i = C_0$ for all $i \in \mathcal{V}$. If the topology of (7.3) is identifiable then (Q, R) is controllable and (S, Q) is observable.

Proof. Suppose on the contrary that (S, Q) is unobservable. Let $v \in \mathbb{R}^N$ be a nonzero vector in the unobservable subspace of (S, Q), i.e.,

$$SQ^k v = 0 \text{ for all } k \in \mathbb{N}.$$

This implies that $SQ^k = S(Q+vv^\top)^k$ for all $k \in \mathbb{N}$. By (7.6), the network transfer matrix is given by

$$F_Q(z) = S(I - G_0(z)Q)^{-1}G_0(z)R,$$

where $G_0(z) := C_0(zI - A_0)^{-1}B_0$ is a scalar transfer function. Next, by expanding $F_Q(z)$ as a formal series

$$F_Q(z) = S\left(\sum_{k=0}^{\infty}(QG_0(z))^k\right)G_0(z)R,$$

it is clear that $F_Q(z) = F_{\bar{Q}}(z)$, where the matrix $\bar{Q}$ is defined as $\bar{Q} := Q + vv^\top$. Since $v \neq 0$, the matrices Q and $\bar{Q}$ are distinct. Hence, the topology of (7.3) is not identifiable. The proof for necessity of controllability of (Q, R) is analogous and therefore omitted. □

Proof of Theorem 7.3: Suppose on the contrary that $\operatorname{rank} R < N$ and $\operatorname{rank} S < N$. Then there exist nonzero vectors $v_1, v_2 \in \mathbb{R}^N$ such that $Sv_1 = 0$ and $v_2^\top R = 0$. We assume without loss of generality that v_2 is such that $v_2^\top v_1 \neq -1$. Next, we define $T := I + v_1 v_2^\top$. By the Sherman-Morrison formula, T is invertible if and only if $1 + v_2^\top v_1 \neq 0$, equivalently, $v_2^\top v_1 \neq -1$. By our assumption on v_2, the matrix T is hence invertible, and

$$T^{-1} = I - \frac{v_1 v_2^\top}{1 + v_2^\top v_1}.$$

We define the matrix

$$\bar{Q} := T^{-1}QT = \left(I - \frac{v_1 v_2^\top}{1 + v_2^\top v_1}\right)Q(I + v_1 v_2^\top). \tag{7.13}$$

Now, we distinguish two cases: $Q \neq \bar{Q}$ and $Q = \bar{Q}$. First suppose that $Q \neq \bar{Q}$. Since we have $\bar{Q} = T^{-1}QT$, $TR = R$ and $ST^{-1} = S$, we obtain

$$\begin{aligned} \mathcal{T}(I \otimes A_0 + Q \otimes B_0 C_0)\mathcal{T}^{-1} &= I \otimes A_0 + \bar{Q} \otimes B_0 C_0 \\ \mathcal{T}(I \otimes B_0)R &= (I \otimes B_0)R \\ S(I \otimes C_0)\mathcal{T}^{-1} &= S(I \otimes C_0), \end{aligned}$$

where $\mathcal{T} := T \otimes I$. Here we have used the fact that $p_i = m_i = 1$ for all $i \in \mathcal{V}$, as well as the property $(X_1 \otimes Y_1)(X_2 \otimes Y_2) = (X_1 X_2) \otimes (Y_1 Y_2)$ for matrices X_1, X_2, Y_1, Y_2 of compatible dimensions. We conclude that $F_Q(z) = F_{\bar{Q}}(z)$, i.e., the topology of (7.3) is not identifiable.

Secondly, suppose that $Q = \bar{Q}$. It follows from (7.13) that

$$Qv_1 v_2^\top - \frac{v_1 v_2^\top}{1 + v_2^\top v_1}Q - \frac{v_1 v_2^\top}{1 + v_2^\top v_1}Qv_1 v_2^\top = 0,$$

equivalently,

$$(1 + v_2^\top v_1)Qv_1 v_2^\top - v_1 v_2^\top Q - v_1 v_2^\top Q v_1 v_2^\top = 0.$$

Multiply from right by v_2 and rearrange terms to obtain

$$(1 + v_2^\top v_1)v_2^\top v_2 Q v_1 = (v_2^\top Q v_2 + v_2^\top Q v_1 v_2^\top v_2)v_1.$$

This means that v_1 is an eigenvector of Q contained in the kernel of S. Therefore, (S, Q) is unobservable (cf. [208, Ch. 3]). By the previous lemma, this implies that the topology of (7.3) is not identifiable. □

Theorem 7.3 is interesting because it shows that the ability to measure all node outputs or to excite all node inputs is *necessary* for identifiability in the case of homogeneous SISO networks. This result allows us to sharpen Theorem 7.1 for this particular class of networks.

Theorem 7.4. Consider a homogeneous SISO network, that is, a system of the form (7.3) with $m_i = p_i = 1$ and $A_i = A_0$, $B_i = B_0$ and $C_i = C_0$ for all $i \in \mathcal{V}$. The topology of (7.3) is identifiable if and only if $G_0(z) := C_0(zI - A_0)^{-1}B_0 \neq 0$ and at least one of the following two conditions holds:

(i) $\operatorname{rank} S = N$ and (Q, R) is controllable

(ii) $\operatorname{rank} R = N$ and (S, Q) is observable.

Proof. To prove the “if"-statement, we first assume that $G_0(z)$ is nonzero, $\operatorname{rank} S = N$ and (Q, R) is controllable. By Theorem 7.1, the topology of (7.3) is identifiable if and only if $\operatorname{cker} H_Q^\top(z) = \{0\}$, where $H_Q(z)$ is given by $H_Q(z) = (I - G_0(z)Q)^{-1}G_0(z)R$. We expand the latter matrix as a formal series as

$$(I - G_0(z)Q)^{-1}G_0(z)R = \left(\sum_{k=0}^{\infty}(G_0(z)Q)^k\right)G_0(z)R. \tag{7.14}$$

We claim that by strict properness of $G_0(z)$, the powers $G_0^k(z)$ $(k = 0, 1, 2, \dots)$ are linearly independent over the reals. Indeed, suppose $\alpha_1 G_0^{k_1}(z) + \cdots + \alpha_r G_0^{k_r}(z) = 0$ for $\alpha_1, \dots, \alpha_r \in \mathbb{R}$ and $k_1 < \cdots < k_r$. Let $G_0(z) = \frac{p_0(z)}{q_0(z)}$ where p_0 and q_0 are polynomials. If $\alpha_1 \neq 0$ then

$$\frac{p_0^{k_1}(z)q_0^{k_r-k_1}(z)}{q_0^{k_r}(z)} = -\frac{1}{\alpha_1}\sum_{i=2}^{r}\alpha_i \frac{p_0^{k_i}(z)q_0^{k_r-k_i}(z)}{q_0^{k_r}(z)}. \tag{7.15}$$

By strict properness of $G_0(z)$, this is a contradiction since every numerator on the right hand side of (7.15) has degree less than $p_0^{k_1}(z)q_0^{k_r-k_1}(z)$. Thus $\alpha_1 = 0$. In fact, we can repeat the same argument to show $\alpha_1 = \cdots = \alpha_r = 0$, proving the claim of independence. It follows from (7.14) that $v \in \mathbb{R}^N$ satisfies $v^\top H_Q(z) = 0$ if and only if

$$\sum_{k=0}^{\infty} G_0^k(z) v^\top Q^k R = 0,$$

where we leveraged the hypothesis that $G_0(z)$ is nonzero. Now, using the fact that $G_0^k(z)$ $(k = 0, 1, 2, \dots)$ are linearly independent, we obtain $v^\top Q^k R = 0$

for all $k \in \mathbb{N}$. We conclude by controllability of the pair (Q, R) that $v = 0$, hence cker $H_Q^\top(z) = \{0\}$. In other words, the topology of (7.3) is identifiable. The sufficiency of the three conditions $G_0(z) \neq 0$, rank $R = N$ and (S, Q) is observable is proven in a similar fashion and thus omitted.

To prove the "only if"-statement, suppose that the topology of (7.3) is identifiable. Clearly, this implies that $G_0(z) \neq 0$. Indeed, if $G_0(z) = 0$ then $F_Q(z) = 0$ and any $\bar{Q}$ satisfies $F_Q(z) = F_{\bar{Q}}(z)$. By Lemma 7.1, (Q, R) is controllable and (S, Q) is observable. Furthermore, by Theorem 7.3, either S or R has full rank. □

It is noteworthy that full rank of either R or S is not necessary for topological identifiability of heterogeneous networks, as demonstrated next.

Example 7.2. Consider a networked system (7.3) consisting of two nodes $A_1 = 0, B_1 = 1$, and $C_1 = 1$, and

$$A_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \; B_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \; C_2 = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

In addition, assume that $R = \begin{bmatrix} 1 & 0 \end{bmatrix}^\top$ and $S = \begin{bmatrix} 0 & 1 \end{bmatrix}$. It can be easily verified that

$$F_Q(z) = \frac{Q_{21}}{z^3 - Q_{11}z^2 - Q_{22}z + Q_{11}Q_{22} - Q_{12}Q_{21}},$$

where Q_{11}, Q_{12}, Q_{21} and Q_{22} are the entries of the interconnection matrix

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}.$$

We assume that $Q_{21} \neq 0$ such that $F_Q(z)$ is nonzero. Suppose that $F_Q(z) = F_{\bar{Q}}(z)$ for some interconnection matrix $\bar{Q}$. By comparing the numerators of F_Q and $F_{\bar{Q}}$ we see that $Q_{21} = \bar{Q}_{21}$. Moreover, by comparing the coefficients corresponding to z^2 and z in the denominator, we obtain $Q_{11} = \bar{Q}_{11}$ and $Q_{22} = \bar{Q}_{22}$. Finally, by comparing constant terms in the denominator, we see that $Q_{12} = \bar{Q}_{12}$. Hence, $Q = \bar{Q}$ and we conclude that the topology of (7.3) is identifiable. However, S does not have full column rank and R does not have full row rank.

7.4 TOPOLOGY IDENTIFICATION

In this section, we focus on the problem of topology identification, as formulated in Problem 7.2. The proposed solution consists of two steps: first identify the Markov parameters of the networked system (7.3), and then extract the matrix Q. There are several ways of computing the Markov parameters on the basis of input/output data, we will summarize some of them in the next section.

7.4.1 Identification of Markov parameters

Consider a general linear system of the form

$$x(t+1) = Ax(t) + Bu(t) \tag{7.16}$$

$$y(t) = Cx(t) + Du(t), \tag{7.17}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the input and $y \in \mathbb{R}^p$ the output. In this section we recap how one can identify the Markov parameters $D, CB, CAB, \dots, CA^rB$ for $r \in \mathbb{N}$, using measurements of the input and output of (7.16)-(7.17). For a given signal $f(t)$ with $t = 0, 1, \dots, T-1$, we define the Hankel matrix of depth k as

$$\mathcal{H}_k(f) := \begin{bmatrix} f(0) & f(1) & \cdots & f(T-k) \\ f(1) & f(2) & \cdots & f(T-k+1) \\ \vdots & \vdots & & \vdots \\ f(k-1) & f(k) & \cdots & f(T-1) \end{bmatrix}.$$

Recall from Definition 2.1 that the signal $f(0), f(1), \dots, f(T-1)$ is said to be *persistently exciting* of order k if $\mathcal{H}_k(f)$ has full row rank. Now suppose that we measure T samples of the input $u(t)$ and output $y(t)$ of (7.16)-(7.17) for $t = 0, 1, \dots, T-1$. We rearrange these measurements in Hankel matrices of depth $n + r + 1$. Moreover, we partition

$$\mathcal{H}_{n+r+1}(u) = \begin{bmatrix} U_p \\ U_f \end{bmatrix}, \quad \mathcal{H}_{n+r+1}(y) = \begin{bmatrix} Y_p \\ Y_f \end{bmatrix},$$

where U_p and Y_p contain the first n row blocks of $\mathcal{H}_{n+r+1}(u)$ and $\mathcal{H}_{n+r+1}(y)$, respectively. The following result from [125, Prop. 4] shows how the Markov parameters can be obtained from data.

Theorem 7.5. Let (7.16) be controllable and assume that $u(0), u(1), \dots, u(T-1)$ is persistently exciting of order $2n + r + 1$. There exists a matrix $G \in \mathbb{R}^{(T-n-r)\times m}$ such that

$$\begin{bmatrix} U_p \\ Y_p \\ U_f \end{bmatrix} G = \begin{bmatrix} 0 \\ 0 \\ \operatorname{col}(I, 0) \end{bmatrix}.$$

Moreover, the Markov parameters are given as $Y_f G = \operatorname{col}(D, CB, CAB, \dots, CA^rB)$.

Theorem 7.5 shows how the Markov parameters of the system can be obtained from measured input/output data. The input should be designed in such a way that it is persistently exciting, special cases of such inputs have been discussed in [227]. For $u(0), u(1), \dots, u(T-1)$ to be persistently exciting of order $2n + r + 1$ a number of samples $T \geqslant (m+1)(2n + r + 1) - 1$ is necessary. In fact, there are input functions that achieve persistency of excitation of this order exactly for $T = (m+1)(2n + r + 1) - 1$. A refinement of Theorem 7.5 is possible using the notion of weaving trajectories [128], which reduces the order of excitation to $2n + 1$. More generally, one can extend the notion of persistency of excitation to an arbitrary concatenation of multiple trajectories [220] (see also Chapter 2). This is useful in situations where single experiments are individually not sufficiently informative.

Remark 7.5. In addition to the deterministic setting of Theorem 7.5, there are approaches to identify the Markov parameters of systems with disturbances, i.e., systems of the form

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + w(t) \\ y(t) &= Cx(t) + Du(t) + v(t), \end{aligned}$$

where v and w are zero mean, white vector sequences. In particular, the paper [161] studies the identification of the system's Markov parameters from finite data, and provides statistical guarantees for the quality of estimation.

7.4.2 Topology identification

Subsequently, we will turn to the problem of identifying the topology of (7.3) from the network's Markov parameters. As in Theorem 7.1, we will assume that S has full column rank. In fact, to lighten the notation, we will simply assume $S = I$, even though all results can be stated for general matrices S having full column rank. Under the latter assumption, the Markov parameters of (7.3) are given by

$$M_\ell(Q) := C(A+BQC)^\ell BR.$$

Whenever the dependence of $M_\ell(Q)$ on Q is clear, we simply write M_ℓ. It is not immediately clear how to obtain Q from the Markov parameters since M_ℓ depends on the ℓ-th power of $A + BQC$. The following lemma will be helpful since it implies that M_ℓ can essentially be viewed as an affine function in Q and lower order Markov parameters.

Lemma 7.2. We have that

$$M_\ell = CA^\ell BR + \sum_{i=0}^{\ell-1} CA^i BQM_{\ell-i-1}.$$

Proof. First, we claim that for square matrices D_1 and D_2 of the same dimensions, we have

$$(D_1 + D_2)^\ell = D_1^\ell + \sum_{i=0}^{\ell-1} D_1^i D_2 (D_1 + D_2)^{\ell-i-1} \tag{7.18}$$

for all $\ell = 1, 2, \dots$. It is straightforward to prove this claim by induction. Indeed, for $\ell = 1$, (7.18) holds. If (7.18) holds for $\ell \geqslant 1$ then

$$\begin{aligned} (D_1 + D_2)^{\ell+1} &= D_1^\ell (D_1 + D_2) + \sum_{i=0}^{\ell-1} D_1^i D_2 (D_1 + D_2)^{\ell-i} \\ &= D_1^{\ell+1} + \sum_{i=0}^{\ell} D_1^i D_2 (D_1 + D_2)^{\ell-i}, \end{aligned}$$

proving the claim. Subsequently, by substitution of $D_1 = A$ and $D_2 = BQC$ into (7.18), we obtain

$$(A+BQC)^\ell = A^\ell + \sum_{i=0}^{\ell-1} A^i BQC(A+BQC)^{\ell-i-1}.$$

Finally, the lemma follows by pre- and postmultiplication by C and BR, respectively. □

Using Lemma 7.2, we can come up with a system of linear equations in the unknown interconnection matrix Q. To see this, let us denote $K_\ell := M_\ell - CA^\ell BR$. Moreover, define the Toeplitz matrix L by

$$L := \begin{bmatrix} CB & 0 & \cdots & 0 \\ CAB & CB & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{r-1}B & CA^{r-2}B & \cdots & CB \end{bmatrix},$$

where $r \geqslant 2n-1$. We apply Lemma 7.2 for $\ell = 1, \dots, r$ to obtain

$$\begin{bmatrix} K_1 \\ K_2 \\ \vdots \\ K_r \end{bmatrix} = L(I \otimes Q) \begin{bmatrix} M_0 \\ M_1 \\ \vdots \\ M_{r-1} \end{bmatrix}. \tag{7.19}$$

Next, let L_i denote the $(i+1)$-th column block of L and define the matrix $K := \operatorname{col}(K_1, K_2, \dots, K_r)$. We can then write (7.19) in a more compact form as

$$K = \sum_{i=0}^{r-1} L_i Q M_i, \tag{7.20}$$

which reveals that Q is a solution to a *generalized Sylvester equation*. Topology identification thus boils down to i) identifying the network's Markov parameters, ii) constructing the matrices K, L_i and M_i for $i = 0, \dots, r-1$ and iii) solving the Sylvester equation. We summarize this procedure in the following theorem.

Theorem 7.6. Consider the networked system (7.3) with $S = I$. Let the Markov parameters of (7.3) be M_i for $i = 0, 1, \dots, r \geqslant 2n-1$. Let the matrices K and L_i be as before. If the topology of (7.3) is identifiable then the interconnection matrix Q is the unique solution to the generalized Sylvester equation

$$K = \sum_{i=0}^{r-1} L_i \mathbf{Q} M_i \tag{7.21}$$

in the unknown $\mathbf{Q}$.

Proof. Note that the interconnection matrix Q is a solution to (7.21) by construction. Suppose that $\bar{Q}$ is also a solution to (7.21). We want to prove that $Q = \bar{Q}$. Since Q and $\bar{Q}$ are both solutions to (7.21), we have

$$\sum_{i=0}^{\ell-1} CA^i BQM_{\ell-i-1}(Q) = \sum_{i=0}^{\ell-1} CA^i B\bar{Q}M_{\ell-i-1}(Q) \tag{7.22}$$

for $\ell = 1, 2, \dots, r$. Here we have written the dependence of $M_{\ell-i-1}$ on Q explicitly, to distinguish between Q and $\bar{Q}$. By Lemma 7.2 we have

$$M_\ell(Q) = CA^\ell BR + \sum_{i=0}^{\ell-1} CA^i BQM_{\ell-i-1}(Q) \tag{7.23}$$

$$M_\ell(\bar{Q}) = CA^\ell BR + \sum_{i=0}^{\ell-1} CA^i B\bar{Q}M_{\ell-i-1}(\bar{Q}). \tag{7.24}$$

Clearly, $M_0(Q) = CBR = M_0(\bar{Q})$. In fact, we claim that $M_k(Q) = M_k(\bar{Q})$ for all $k = 0, 1, \dots, r$. Suppose on the contrary that there exists an integer s such that $0 < s \leqslant r$ and $M_s(Q) \neq M_s(\bar{Q})$. We assume without loss of generality that s is the smallest integer for which this is the case. Then $M_k(Q) = M_k(\bar{Q})$ for all $k = 0, 1, \dots, s-1$. By combining (7.22) and (7.23) we obtain

$$M_s(Q) = CA^s BR + \sum_{i=0}^{s-1} CA^i B\bar{Q}M_{s-i-1}(Q). \tag{7.25}$$

By hypothesis $M_k(Q) = M_k(\bar{Q})$ for all $k = 0, 1, \dots, s-1$, which yields

$$M_s(Q) = CA^s BR + \sum_{i=0}^{s-1} CA^i B\bar{Q}M_{s-i-1}(\bar{Q}) = M_s(\bar{Q}),$$

using (7.24). This is a contradiction and we conclude that $M_k(Q) = M_k(\bar{Q})$ for all $k = 0, 1, \dots, r$. Since $r \geqslant 2n-1$ it follows from the Cayley-Hamilton theorem that $M_k(Q) = M_k(\bar{Q})$ for all $k \in \mathbb{N}$. Thus, $F_Q(z) = F_{\bar{Q}}(z)$. Finally, as the topology of (7.3) is identifiable, we conclude that $Q = \bar{Q}$. This completes the proof. □

7.4.3 Solving the generalized Sylvester equation

In the previous section, we saw that the generalized Sylvester equation (7.21) plays a central role in our topology identification approach. In this section, we discuss methods to solve this equation. One simple approach to the problem is to vectorize $\mathbf{Q}$ and write (7.21) as the system of linear equations

$$\sum_{i=0}^{r-1} \left(M_i^\top \otimes L_i \right) \operatorname{vec}(\mathbf{Q}) = \operatorname{vec}(K) \tag{7.26}$$

in the unknown $\operatorname{vec}(\mathbf{Q})$ of dimension

$$\left(\sum_{i=1}^{N} m_i \right) \left(\sum_{i=1}^{N} p_i \right).$$

However, a drawback of this approach is that the dimension of $\text{vec}(\mathbf{Q})$ is quadratic in the number of nodes N. This means that for large networks, solving (7.26) is costly from a computational point of view.

For the "ordinary" Sylvester equation of the form

$$L_0\mathbf{Q} + \mathbf{Q}M_1 = K,$$

there are well-known solution methods that avoid vectorization[3]. The general idea is to transform the matrices L_0 and M_1 to a suitable form so that the Sylvester equation is easier to solve. A classic approach is the *Bartels-Stewart* method [11] that transforms L_0 and M_1 to real Schur form by means of two orthogonal similarity transformations. The resulting equivalent Sylvester equation is then simply solved by backward substitution. A Hessenberg-Schur variant of this algorithm was proposed in [69]. The approach was also extended to be able to deal with the more general equation

$$L_0\mathbf{Q}M_0 + L_1\mathbf{Q}M_1 = K,$$

using QZ-decompositions [69, Sec. 7]. The problem with all of these transformation methods is that they rely on the fact that the Sylvester equation consists of exactly two $\mathbf{Q}$-dependent terms, i.e., $r = 1$. Therefore, it does not seem possible to extend such methods to solve generalized Sylvester equations of the form (7.21) for $r > 1$, see also the discussion in [216, Sec. 2].

Nonetheless, we can improve upon the basic approach of vectorization (7.26) by noting that the matrices A, B and C have a special structure. Indeed, recall from (7.2) that these matrices are block diagonal. This allows us to write down a Sylvester equation for each row block of $\mathbf{Q}$. Let $\mathbf{Q}^{(j)}$ denote the j-th block row of $\mathbf{Q}$ for $j \in \mathcal{V}$. Then it is straightforward to show that (7.21) is equivalent to

$$K^{(j)} = \sum_{i=0}^{r-1} L_i^{(j)}\mathbf{Q}^{(j)}M_i \tag{7.27}$$

for all $j \in \mathcal{V}$, where $L_i^{(j)}$ is the $(i+1)$-th column block of the matrix $L^{(j)}$, given by

$$L^{(j)} := \begin{bmatrix} C_jB_j & 0 & \cdots & 0 \\ C_jA_jB_j & C_jB_j & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C_jA_j^{r-1}B_j & C_jA_j^{r-2}B_j & \cdots & C_jB_j \end{bmatrix},$$

and $K^{(j)} := \text{col}(K_1^{(j)}, K_2^{(j)}, \dots, K_r^{(j)})$ with $K_\ell^{(j)}$ the j-th row block of K_ℓ. The importance of (7.27) lies in the fact that each row block of Q can be obtained independently, which significantly reduces the dimensions of the involved matrices. In fact, (7.27) is equivalent to the linear system of equations

$$\sum_{i=0}^{r-1}\left(M_i^\top \otimes L_i^{(j)}\right)\text{vec}\left(\mathbf{Q}^{(j)}\right) = \text{vec}\left(K^{(j)}\right) \tag{7.28}$$

[3] It is typically assumed that the matrices L_0 and M_1 are square [11,69].

in the unknown vec $\left(\mathbf{Q}^{(j)}\right)$ of dimension $m_j \left(\sum_{i=1}^{N} p_i\right)$. Note that the unknown is linear in the number of nodes, assuming that m_j and p_i are small in comparison to N.

7.4.4 Robustness analysis

In the case that the Markov parameters $M_0, M_1, \ldots, M_r$ are identified exactly, we can reconstruct the topology by solving the generalized Sylvester equation (7.21), or equivalently, the system of linear equations (7.26). Now suppose that our estimates of the Markov parameters are inexact, and we have access to

$$\hat{M}_\ell := M_\ell + \Delta_\ell, \quad \ell = 1, 2, \ldots, r \tag{7.29}$$

where the real matrices Δ_ℓ represent the perturbations. Accordingly, we define $\hat{K}_\ell := \hat{M}_\ell - CA^\ell BR = K_\ell + \Delta_\ell$. Let $\Delta := \operatorname{col}(\Delta_1, \Delta_2, \ldots, \Delta_r)$. In this case it is natural to look for an approximate (least squares) solution $\operatorname{vec}(\hat{Q})$ that solves

$$\min_{\operatorname{vec}(\hat{\mathbf{Q}})} \| \sum_{i=0}^{r-1} \left(\hat{M}_i^\top \otimes L_i\right) \operatorname{vec}(\hat{\mathbf{Q}}) - \operatorname{vec}(\hat{K}) \|. \tag{7.30}$$

An obvious question is how the solution $\hat{Q}$ is related to the true interconnection matrix Q. The following lemma provides a bound on the infinity norm of $\operatorname{vec}(\hat{Q}) - \operatorname{vec}(Q)$. In what follows, we will make use of the constant

$$\alpha := \| \left(\sum_{i=0}^{r-1} \left(\hat{M}_i^\top \otimes L_i\right) \right)^\dagger \|_\infty,$$

where $X^\dagger$ denotes the Moore-Penrose inverse of X.

Lemma 7.3. Consider the network (7.3) with $S = I$ and suppose that its topology be identifiable. Assume that the solution $\hat{Q}$ to (7.30) is unique. Then we have that

$$\| \operatorname{vec}(\hat{Q}) - \operatorname{vec}(Q) \|_\infty$$

is upper bounded by

$$\alpha \left(\| \operatorname{vec}(\Delta) \|_\infty + \| \sum_{i=0}^{r-1} (\Delta_i^\top \otimes L_i) \|_\infty \| \operatorname{vec}(Q) \|_\infty \right). \tag{7.31}$$

Note that the bound (7.31) tends to zero as $\Delta_0, \Delta_1, \ldots, \Delta_r$ tend to zero, so $\hat{Q}$ is a good approximation of Q for small perturbations. An overestimate of (7.31) can be obtained if some prior knowledge is available. In particular, note that α is readily computable from the estimated Markov parameters (7.29). The first two norms in (7.31) can be upper bounded if a bound on $\|\Delta_i\|_\infty$ is given. Identification error bounds on the Markov parameters are derived, e.g., in [161]. Finally, to estimate $\| \operatorname{vec}(Q) \|_\infty$ one requires a bound on the largest network weight, i.e., an upper bound on the largest (in magnitude) entry of Q. The upper bound (7.31) is useful in the case that the nonzero weights of the network are lower bounded in magnitude by some known positive scalar γ, an assumption that is common in the literature on consensus networks, cf. [108, Sec. 3]. Indeed, in this case we can can exactly identify the graph structure $\mathcal{G}$ from noisy Markov parameters if

$$\alpha\left(\|\operatorname{vec}(\Delta)\|_\infty + \|\sum_{i=0}^{r-1}(\Delta_i^\top \otimes L_i)\|_\infty \|\operatorname{vec}(Q)\|_\infty\right) < \frac{1}{2}\gamma,$$

since identified entries smaller than $\frac{1}{2}\gamma$ are necessarily zero. We will further illustrate this point in Example 7.3.

Proof. We make use of the shorthand notation

$$E := \sum_{i=0}^{r-1}\left(\Delta_i^\top \otimes L_i\right),\; A_E := \sum_{i=0}^{r-1}\left(\hat{M}_i^\top \otimes L_i\right).$$

The hypothesis that $\hat{Q}$ is unique is equivalent to A_E having full column rank. By using (7.26) and the relation $\hat{M}_i = M_i + \Delta_i$, we get

$$A_E^\top A_E \operatorname{vec}(Q) = A_E^\top(\operatorname{vec}(K) + E\operatorname{vec}(Q)).$$

Therefore, $\operatorname{vec}(Q) = A_E^\dagger(\operatorname{vec}(K) + E\operatorname{vec}(Q))$. Further, $\operatorname{vec}(\hat{Q}) = A_E^\dagger \operatorname{vec}(\hat{K}) = A_E^\dagger \operatorname{vec}(K+\Delta)$. This yields

$$\operatorname{vec}(\hat{Q}) - \operatorname{vec}(Q) = A_E^\dagger(\operatorname{vec}(\Delta) - E\operatorname{vec}(Q)).$$

Finally, taking infinity norms yields the upper bound (7.31). This completes the proof. □

Example 7.3. Consider the networked system in Example 7.1. We consider the situation in which only the first node of the network is externally excited. We already know by the discussion in Example 7.1 that the topology of the system is identifiable. Here, our aim is to reconstruct the topology on the basis of the noisy Markov parameters (7.29), where $r = 40$. The perturbations are drawn randomly from a normal distribution using the Matlab command `randn`, and scaled such that $\|\Delta_i^\top\|_\infty \leqslant 10^{-5}$ for all i. Since Δ_i is a vector, this also implies that $\|\Delta_i\|_\infty \leqslant 10^{-5}$. In this example, we assume that the weights of the network (i.e., the entries of Q) have magnitudes between $\frac{1}{2}$ and 1.

We identify the matrix $\hat{Q}$ by solving (7.30). To get an idea of the quality of estimation, we want to find a bound on (7.31). First, we compute $\alpha = 464.7040$. By the assumptions on the perturbations and network weights, we obtain the bounds $\|\Delta\|_\infty \leqslant 10^{-5}$ and $\|\operatorname{vec}(Q)\|_\infty \leqslant 1$. Moreover,

$$\begin{aligned}\|\sum_{i=0}^{r-1}(\Delta_i^\top \otimes L_i)\|_\infty &\leqslant \sum_{i=0}^{r-1}\|\Delta_i^\top\|_\infty \|L_i\|_\infty \\ &\leqslant 4.0000 \times 10^{-4},\end{aligned}$$

where we have used [106, Thm. 8 & p. 413] to bound the Kronecker product. Combining the previous bounds, we conclude that (7.31) is less then or equal to 0.1883. Since $0.1883 \leqslant 0.25$ we can round all entries of $\hat{Q}$ that are less than 0.25 to zero, since the corresponding entries in Q are necessarily zero. The resulting zero/nonzero structure of $\hat{Q}$ can be captured by a graph $\hat{\mathcal{G}}$ that we display in Figure 7.2. Clearly, the structure of $\hat{\mathcal{G}}$ is identical to the graph defined in Example 7.1, and the weights of $\hat{\mathcal{G}}$ are close to the weights of $\mathcal{G}$. Next, we repeat the experiment for larger perturbations, i.e., for $\|\Delta_i\|_\infty$ and $\|\Delta_i^\top\|_\infty$ bounded by 0.01. We identify $\hat{Q}$ and use the same rounding strategy as before to obtain a graph $\hat{\mathcal{G}}$ in Figure 7.3. Note that $\hat{\mathcal{G}}$ resembles the original network structure $\mathcal{G}$. In fact, all links are identified correctly, except for $(7,8)$ and the spurious link $(4,8)$. In this case, the bound (7.31) equals 49.9997, illustrating the fact that (7.31) can be conservative.

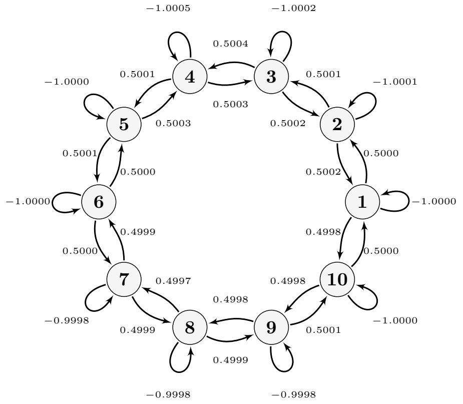

Figure 7.2: $\hat{\mathcal{G}}$ for $\|\Delta_i\|_\infty \leqslant 10^{-5}$.

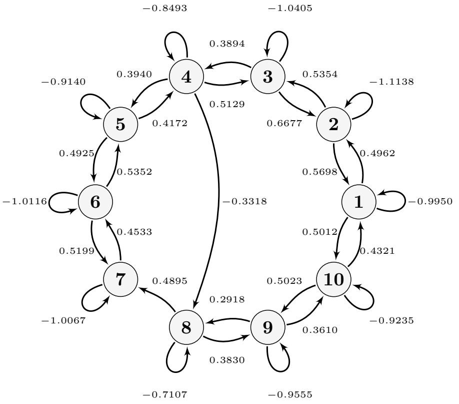

Figure 7.3: $\hat{\mathcal{G}}$ for $\|\Delta_i\|_\infty \leqslant 10^{-2}$.

7.5 CONCLUSIONS

In this chapter we have studied the problem of topology identification of heterogeneous networks of linear systems. First, we have provided necessary and sufficient conditions for topological identifiability. These conditions were stated in terms of the constant kernel of certain network-related transfer matrices. We have also seen that homogeneous SISO networks enjoy quite special identifiability properties that do not extend to the heterogeneous case. Subsequently, we have turned our attention to the topology identification problem. The idea of the identification approach was to solve a generalized Sylvester equation involving the network's Markov parameters to obtain the network topology. One of the attractive features of the approach is that the structure of the networked system can be exploited so that each row block of the interconnection matrix can be obtained individually.

The generalized Sylvester equation (7.21) plays an important role in our identification approach. Numerical solution methods are less well-developed for this equation than they are for the standard Sylvester equation [11, 69]. Hence, it would be of interest to further develop numerical methods for Sylvester equations of the form (7.21). We note that a Krylov subspace method has already been developed in [22]. Another direction for future work is to study topological identifiability with prior information on the interconnection matrix. For example, from physical principles it may be known that Q is Laplacian. Such prior knowledge could be exploited to weaken the identifiability conditions in Theorems 7.1, 7.2 and 7.4.

8 TOPOLOGY RECONSTRUCTION OF AUTONOMOUS NETWORKS

In this chapter we continue studying the problem of topology identification, but for networks of single-integrators without external inputs. In this case, excitation has to be secured through the initial conditions of the network. We will provide conditions under which unique reconstruction is possible, and also develop methods for topology identification itself. We will see that the more specialized setup of this chapter allows for a more particular reconstruction technique, via the solutions to certain Lyapunov equations.

8.1 INTRODUCTION

Networks of dynamical systems appear in many contexts, including biological networks [213], water distribution networks [49] and (wireless) sensor networks [118].

The overall behavior of a dynamical network is greatly influenced by its network structure (also called network topology). For instance, in the case of consensus networks, the dynamical network reaches *consensus* if and only if the network graph is connected [158]. Unfortunately, the interconnection structure of dynamical networks is often unavailable. For instance, in the case of wireless sensor networks [118] the locations of sensors, and hence, communication links between sensors is not always known. Other examples of dynamical networks with unknown network topologies are encountered in biology, for instance in neural networks [213] and genetic networks [97].

Consequently, the problem of *network reconstruction* is studied in the literature. The aim of network reconstruction (also called topology identification) is to find the network structure and weights of a dynamical network, using measurements obtained from the network. To this end, most papers assume that the states of the network nodes can be measured. The literature on network reconstruction methods can roughly be divided into two parts, namely methods for *stochastic* and *deterministic* dynamical networks.

Methods for stochastic network dynamics include *inverse covariance estimation* [79], [145] and methods based on *power spectral analysis* [190]. Moreover, network reconstruction based on *compressive sensing* [184] has been investigated. Furthermore, the authors of [130] consider network reconstruction using *Wiener filtering*.

Apart from methods for stochastic networks, network reconstruction for deterministic network dynamics has been considered. In the paper [153] the concept of *node-knockout* is introduced, and a network reconstruction method based on this concept is discussed. The paper [70] considers the problem of reconstructing a network topology from a transfer matrix of the network. Conditions are investigated under which the network structure can be uniquely determined. Furthermore, the paper [242] considers network reconstruction using a so-called *response network*.

In this chapter, we consider network reconstruction for *deterministic* networks of linear dynamical systems. In contrast to papers studying network reconstruction for specific network dynamics such as consensus dynamics [153] and adjacency dynamics [57], we consider network reconstruction for *general* linear network dynamics described by state matrices contained in the so-called *qualitative class* [87]. It is our aim to infer the unknown network topology of such dynamical networks, from state measurements obtained from the network.

The contributions of this chapter are threefold. Firstly, we rigorously define what we mean by *solvability* of the network reconstruction problem for dynamical networks. Loosely speaking, we say that the network reconstruction problem is solvable if the measurements obtained from a network correspond only with the network under consideration (and not with any other dynamical network). Secondly, we provide necessary and sufficient conditions under which the network reconstruction problem is solvable. Thirdly, we provide a framework for network reconstruction of dynamical networks, using constrained Lyapunov equations. We will show that our framework can be used to establish algorithms to infer network topologies for a variety of network dynamics, including Laplacian and adjacency dynamics. An attractive feature of our approach is that the conditions under which our algorithms reconstruct the network structure are not restrictive. In other words, we show that our algorithms return the correct network structure if and only if the network reconstruction problem is solvable.

Although the chapter mainly focuses on continuous-time network dynamics, we also show how our reconstruction algorithms can be applied to discrete-time systems, and to systems with sampled measurements.

The organization of this chapter is as follows. First, in Section 8.2, we introduce preliminaries and notation. Subsequently, we give a formal problem statement in Section 8.3. In Section 8.4 we discuss necessary and sufficient conditions for the solvability of the network reconstruction problem. Section 8.5 provides our network reconstruction algorithms. We consider an illustrative example in Section 8.6. Finally, Section 8.7 contains our conclusions.

8.2 PRELIMINARIES

We denote by $\mathbb{S}^n$ the set of $n \times n$ symmetric real matrices. For a given set S, the *power set* 2^S is the set of all subsets of S. Let X and Y be nonempty sets. If for each $x \in X$, there exists a set $F(x) \subseteq Y$, we say F is a *set-valued map* from X to Y, and we denote $F : X \to 2^Y$. The *image* of a set-valued map $F : X \to 2^Y$ is defined as $\operatorname{im} F := \{y \in Y \mid \exists x \in X \text{ such that } y \in F(x)\}$.

8.2.1 Systems theory

Consider the linear time-invariant system

$$\begin{aligned} \dot{x}(t) &= Ax(t) \\ y(t) &= Cx(t), \end{aligned} \tag{8.1}$$

where $x \in \mathbb{R}^n$ is the state, $y \in \mathbb{R}^p$ is the output, and the real matrices A and C are of suitable dimensions. We denote the *unobservable subspace* of system (8.1) by $\langle \ker C \mid A \rangle$, i.e.,

$$\langle \ker C \mid A \rangle := \bigcap_{i=0}^{n-1} \ker \left(CA^i \right).$$

The subspace $\langle \ker C \mid A \rangle$ is A-invariant, that is, $A\langle \ker C \mid A \rangle \subseteq \langle \ker C \mid A \rangle$. Furthermore, system (8.1) is *observable* if and only if $\langle \ker C \mid A \rangle = \{0\}$ (see, e.g., Chapter 3 of [208]). If system (8.1) is observable, we say the pair (C, A) is observable.

8.2.2 Graph theory

All graphs considered in this chapter are simple, i.e., without self-loops and with at most one edge between any pair of vertices. We denote the set of simple, undirected graphs of n nodes by $\mathcal{G}^n$. Consider a graph $G \in \mathcal{G}^n$, with vertex set $V = \{1, 2, \ldots, n\}$ and edge set E. The set of neighbours $\mathcal{N}_i$ of vertex $i \in V$ is defined as $\mathcal{N}_i := \{j \in V \mid (i, j) \in E\}$.

We will now define various families of matrices associated with graphs in $\mathcal{G}^n$. To this end, we first define the set-valued map $\mathcal{Q} : \mathcal{G}^n \to 2^{\mathbb{S}^n}$ as

$$\mathcal{Q}(G) := \{X \in \mathbb{S}^n \mid \text{for all } i \neq j,\ X_{ij} \neq 0 \iff (i, j) \in E\}.$$

The set of matrices $\mathcal{Q}(G)$ is called the *qualitative class* of the graph $G \in \mathcal{G}^n$ [87]. The qualitative class has recently been studied in the context of structural controllability of dynamical networks [142], [219]. Note that each matrix $X \in \mathcal{Q}(G)$ carries the graph structure of G, in the sense that X contains nonzero off-diagonal entries in exactly the same positions corresponding to the edges in G. Furthermore, note that the diagonal elements of matrices in the qualitative class are unrestricted. Hence, examples of matrices in $\mathcal{Q}(G)$ include the well-known (weighted) *adjacency* and *Laplacian* matrices, which are defined next. Define the set-valued map $\mathcal{A} : \mathcal{G}^n \to 2^{\mathbb{S}^n}$ as

$$\mathcal{A}(G) := \{A \in \mathcal{Q}(G) \mid \text{for all } i, j,\ A_{ij} \geqslant 0 \text{ and } A_{ii} = 0\}.$$

Matrices in $\mathcal{A}(G)$ are called adjacency matrices associated with the graph G. Subsequently, define $\mathcal{L} : \mathcal{G}^n \to 2^{\mathbb{S}^n}$ as

$$\mathcal{L}(G) := \{L \in \mathcal{Q}(G) \mid L\mathbb{1} = 0 \text{ and for all } i \neq j, L_{ij} \leqslant 0\}.$$

Matrices in the set $\mathcal{L}(G)$ are called Laplacian matrices of G. A Laplacian matrix $L \in \mathcal{L}(G)$ is said to be *unweighted* if $L_{ij} \in \{0, -1\}$ for all $i \neq j$. Similarly, an adjacency matrix $A \in \mathcal{A}(G)$ is called unweighted if $A_{ij} \in \{0, 1\}$ for all i, j.

8.2.3 Consensus dynamics

Consider a graph $G \in \mathcal{G}^n$, with vertex set $V = \{1, 2, \ldots, n\}$ and edge set E. With each vertex $i \in V$, we associate a linear dynamical system $\dot{x}_i(t) = u_i(t)$, where $x_i \in \mathbb{R}$ is the state of node i, and $u_i \in \mathbb{R}$ is its control input. Suppose that each node $i \in V$ applies the control input

$$u_i(t) = -\sum_{j \in \mathcal{N}_i} a_{ij}(x_i(t) - x_j(t)),$$

where $a_{ij} = a_{ji} > 0$ for all $i \in V$ and $j \in \mathcal{N}_i$. Then, the dynamics of the overall system can be written as

$$\dot{x}(t) = -Lx(t), \tag{8.2}$$

where $x = \operatorname{col}(x_1, x_2, \ldots, x_n)$, and $L \in \mathcal{L}(G)$ is a Laplacian matrix. We refer to system (8.2) as a *consensus network*. Consensus networks have been studied extensively in the literature, see, e.g., [158] and the references therein.

8.3 PROBLEM FORMULATION

In this section we define the network reconstruction problem. We consider a linear time-invariant network system, with nodes satisfying single-integrator dynamics. We assume that the state matrix of the system (and hence, the network topology) is not directly available. Moreover, we suppose that the state vector of the system is available for measurement during a time interval $[0, T]$. It is our goal to find conditions on the system under which the exact state matrix can be reconstructed from such measurements. Moreover, if the state matrix can be reconstructed, we want to develop algorithms to infer the state matrix from measurements.

We will now make these problems more precise. Since we want to consider network reconstruction for general network dynamics (instead of specific consensus or adjacency dynamics), we consider any set-valued map $\mathcal{K} : \mathcal{G}^n \to 2^{\mathbb{S}^n}$ such that

$$\varnothing \neq \mathcal{K}(G) \subseteq \mathcal{Q}(G) \tag{8.3}$$

for all $G \in \mathcal{G}^n$. The map $\mathcal{K}$ is specified by the available information on the *type* of network. For example, if we know that we deal with a consensus network, we have $\mathcal{K} = -\mathcal{L}$. On the other hand, if no additional information on the communication weights (such as sign constraints) is known, we let $\mathcal{K} = \mathcal{Q}$. With this in mind, we consider the system

$$\begin{aligned} \dot{x}(t) &= Xx(t) \\ x(0) &= x_0, \end{aligned} \tag{8.4}$$

where $x \in \mathbb{R}^n$ is the state, and $X \in \operatorname{im} \mathcal{K}$ (i.e., $X \in \mathcal{K}(G)$ for some network graph $G \in \mathcal{G}^n$). In what follows, we denote the state trajectory of (8.4) by $x_{x_0}(\cdot)$, where the subscript indicates dependence on the initial condition x_0. We assume that X is unknown, but the state trajectory of (8.4) can be measured during the time-interval $[0, T]$, where $T > 0$. The problem of *network reconstruction* concerns finding the matrix X (and thereby, the graph G), using the state measurements $x_{x_0}(t)$ for $t \in [0, T]$. Of course, this is only possible if the state trajectory $x_{x_0}(\cdot)$ of (8.4) is not a solution to the differential equation $\dot{x}(t) = \bar{X}x(t)$ for some other admissible state matrix $\bar{X} \neq X$. Indeed, if this were the case, the state measurements could correspond to a network described either by X or $\bar{X}$, and we would not be able to distinguish between the two. This leads to the following definition.

Definition 8.1. Consider system (8.4), and denote its state trajectory by $x_{x_0}(\cdot)$. We say that the network reconstruction problem is *solvable* for system (8.4) if for all $\bar{X} \in \operatorname{im} \mathcal{K}$ such that $x_{x_0}(\cdot)$ is a solution to

$$\dot{x}(t) = \bar{X}x(t) \text{ for } t \in [0, T], \tag{8.5}$$

we have $\bar{X} = X$. In the case that the network reconstruction problem is solvable for system (8.4), we say that the network reconstruction problem is solvable for $(x_0, X, \mathcal{K})$.

Remark 8.1. As the state variables of system (8.4) are sums of exponential functions of t, they are *real analytic* functions of t. It is well-known that if two real analytic functions are equal on a non-degenerate interval, they are equal on their whole domain (see, e.g., Corollary 1.2.5 of [104]). Consequently, the state vector $x_{x_0}(\cdot)$ of system (8.4) satisfies (8.5) for $t \in [0, T]$ if and only if $x_{x_0}(\cdot)$ satisfies (8.5) for all $t \geqslant 0$. Therefore, Definition 8.1 can be equivalently stated for $t \geqslant 0$ instead of $t \in [0, T]$.

In this chapter we are interested in conditions on x_0, X, and $\mathcal{K}$ under which the network reconstruction problem is solvable for $(x_0, X, \mathcal{K})$. More explicitly, we have the following problem.

Problem 8.1. Consider system (8.4). Provide necessary and sufficient conditions on x_0, X, and $\mathcal{K}$ under which the network reconstruction problem is solvable for system (8.4).

In addition to Problem 8.1, we are interested in solving the network reconstruction problem itself. This is stated in the following problem.

Problem 8.2. Consider system (8.4), and denote its state vector by $x_{x_0}(\cdot)$. Suppose that $x_{x_0}(t)$ is available for measurement for $t \in [0, T]$, and that the network reconstruction problem is solvable for (8.4). Provide a method to compute the matrix X.

Remark 8.2. Note that we assume that the states of all nodes in the network can be measured. This assumption is *necessary* in the sense that the network reconstruction problem is not solvable (in the case of $\mathcal{Q}(G)$) if we can only measure a part of the state vector. To see this, suppose that we only have access to a p-dimensional output vector $y(t) = Cx(t)$, where

$$C = \begin{bmatrix} I & 0 \end{bmatrix} \in \mathbb{R}^{p \times n}.$$

We claim that for each $X \in \mathcal{Q}(G)$ and $x_0 \in \mathbb{R}^n$ there exists a graph $\bar{G}$, a matrix $\bar{X} \in \mathcal{Q}(\bar{G}) \setminus \{X\}$ and a vector $\bar{x}_0 \in \mathbb{R}^n$ such that

$$Ce^{Xt}x_0 = Ce^{\bar{X}t}\bar{x}_0. \tag{8.6}$$

That is, we cannot distinguish between X and $\bar{X}$ on the basis of output measurements. To see that this claim is true, we write X as

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix},$$

where the partitioning of X is compatible with the one of C. Now we distinguish two cases. First suppose that $X_{21} \neq 0$. Clearly, there exists a vector $z \in \mathbb{R}^{n-p}$ such that $z^\top X_{21} \neq 0$ and $z^\top z = 1$. Define

$$S := \begin{bmatrix} I & 0 \\ 0 & S_{22} \end{bmatrix},$$

where $S_{22} := I - 2zz^\top = S_{22}^{-1}$. Then let $\bar{X} := SXS$ and $\bar{x}_0 := Sx_0$. It is not difficult to see that $\bar{X} \neq X$ and (8.6) is satisfied for this choice of $\bar{X}$ and $\bar{x}_0$. Secondly, consider the case that $X_{21} = 0$. Then $X_{12} = 0$. We can choose $\bar{X}$ as

$$\bar{X} := \begin{bmatrix} X_{11} & 0 \\ 0 & \bar{X}_{22} \end{bmatrix},$$

where $\bar{X}_{22} \neq X_{22}$. In this case, it can be shown that $\bar{X}$ and $\bar{x}_0 := x_0$ satisfy (8.6). Hence, for the network reconstruction problem to be solvable it is necessary to measure all nodes.

8.4 SOLVABILITY OF THE RECONSTRUCTION PROBLEM

In this section we state our main results regarding Problem 8.1. That is, we provide conditions on x_0, X, and $\mathcal{K}$ under which the network reconstruction problem is solvable. Firstly, in Section 8.4.1 we provide necessary and sufficient conditions for the solvability of the network reconstruction problem in the general case that $\mathcal{K}$ is any mapping satisfying (8.3). Later, we consider the special cases in which $\mathcal{K} = \mathcal{Q}$ (Section 8.4.2), and the cases in which $\mathcal{K} = -\mathcal{L}$ or $\mathcal{K} = \mathcal{A}$ (Section 8.4.3).

8.4.1 Solvability for general $\mathcal{K}$

Let $G \in \mathcal{G}^n$ be a graph, and let the mapping $\mathcal{K}$ be as in (8.3). Recall that we consider the dynamical network described by system (8.4). As a preliminary result, we give conditions under which the state trajectory $x_{x_0}(\cdot)$ of system (8.4) is also the solution to the system

$$\begin{aligned}\dot{x}(t) &= \bar{X}x(t)\\ x(0) &= x_0,\end{aligned} \tag{8.7}$$

where $\bar{X} \in \mathcal{K}(\bar{G})$ for some graph $\bar{G} \in \mathcal{G}^n$. This result is given in the following proposition.

Proposition 8.1. Consider systems (8.4) and (8.7), and let $x_{x_0}(\cdot)$ be the state trajectory of (8.4). The trajectory $x_{x_0}(\cdot)$ is also the solution to system (8.7) if and only if $x_0 \in \left\langle \ker(\bar{X} - X) \mid X \right\rangle$.

Proof. Suppose that the state trajectory $x_{x_0}(\cdot)$ of (8.4) is also the solution to system (8.7). This means that $x_{x_0}(\cdot)$ is the solution to both the differential equation

$$\dot{x}(t) = Xx(t), \tag{8.8}$$

and the differential equation

$$\dot{x}(t) = Xx(t) + (\bar{X} - X)x(t). \tag{8.9}$$

In particular, by substitution of $t = 0$, this implies that x_0 is contained in $\ker(\bar{X} - X)$. Moreover, by taking the i-th time-derivative of (8.8) and (8.9), we find that $x_0 \in \ker(\bar{X} - X)X^i$ for $i = 1, 2, \ldots, n-1$. Consequently, we obtain $x_0 \in \left\langle \ker(\bar{X} - X) \mid X \right\rangle$.
Conversely, suppose that x_0 satisfies $x_0 \in \left\langle \ker(\bar{X} - X) \mid X \right\rangle$. By X-invariance of $\left\langle \ker(\bar{X} - X) \mid X \right\rangle$, this implies that the state trajectory $x_{x_0}(\cdot)$ of system (8.4) satisfies $x_{x_0}(t) \in \left\langle \ker(\bar{X} - X) \mid X \right\rangle$ for all $t \geqslant 0$. Specifically, we have that $x_{x_0}(t) \in \ker(\bar{X} - X)$ for all $t \geqslant 0$. We conclude that $x_{x_0}(\cdot)$ is the solution to Equation (8.9), and consequently, to Equation (8.7). □

Remark 8.3. Note that a condition equivalent to the one given in Proposition 8.1 can be stated in terms of the *common eigenspaces* of X and $\bar{X}$. Such a condition was previously proven by Battistelli *et al.* [14], [13] in the case that X and $\bar{X}$ are Laplacian matrices.

By combining Proposition 8.1 and the fact that the state variables of (8.4) are real analytic functions in t (see Remark 8.1), we obtain Theorem 8.1. This theorem states a necessary and sufficient condition under which the network reconstruction problem is solvable for $(x_0, X, \mathcal{K})$.

Theorem 8.1. Let $G \in \mathcal{G}^n$ be a graph, and let the mapping $\mathcal{K}$ be as in (8.3). Moreover, consider a matrix $X \in \mathcal{K}(G)$ and a vector $x_0 \in \mathbb{R}^n$. The network reconstruction problem is solvable for $(x_0, X, \mathcal{K})$ if and only if for all $\bar{X} \in \operatorname{im} \mathcal{K} \setminus \{X\}$, we have

$$x_0 \notin \left\langle \ker(\bar{X} - X) \mid X \right\rangle.$$

Although Theorem 8.1 gives a general necessary and sufficient condition for network reconstruction, it is not directly clear how to verify this condition. Especially since X is assumed to be unknown, it seems difficult to check that $x_0 \notin \left\langle \ker(\bar{X} - X) \mid X \right\rangle$. In fact, we will show in Section 8.5 that the condition of Theorem 8.1 can be checked using only the measurements $x_{x_0}(t)$ for $t \in [0, T]$.

Note that the condition of Theorem 8.1 is not only given in terms of x_0 and X, but also in terms of all other matrices $\bar{X} \in \operatorname{im} \mathcal{K}$. In the following theorem, we provide a simple *sufficient* condition for the solvability of the network reconstruction problem, which is stated in terms of x_0 and X.

Theorem 8.2. Let $G \in \mathcal{G}^n$ be a graph, and let the mapping $\mathcal{K}$ be as in (8.3). Moreover, consider a matrix $X \in \mathcal{K}(G)$ and a vector $x_0 \in \mathbb{R}^n$. The network reconstruction problem is solvable for $(x_0, X, \mathcal{K})$ if the pair $(x_0^\top, X)$ is observable.

Proof. Suppose that $(x_0^\top, X)$ is observable, and assume $x_0 \in \left\langle \ker(\bar{X} - X) \mid X \right\rangle$ for some matrix $\bar{X} \in \operatorname{im} \mathcal{K}$. We want to show that $\bar{X} = X$. Note that by hypothesis, we have $x_0 \in \ker(\bar{X} - X)X^i$, for $i = 0, 1, \dots, n-1$. As a consequence, we obtain the equalities

$$\bar{X} X^i x_0 = X^{i+1} x_0, \tag{8.10}$$

for $i = 0, 1, \dots, n-1$. It is not difficult to see that by induction, Equation (8.10) implies that

$$X^i x_0 = \bar{X}^i x_0, \tag{8.11}$$

for $i = 1, 2, \dots, n$. In other words, the matrix $X \begin{bmatrix} x_0 & X x_0 & \dots & X^{n-1} x_0 \end{bmatrix}$ is equal to

$$\bar{X} \begin{bmatrix} x_0 & \bar{X} x_0 & \dots & \bar{X}^{n-1} x_0 \end{bmatrix}. \tag{8.12}$$

Since $(x_0^\top, X)$ is observable and $X = X^\top$, the matrix $\begin{bmatrix} x_0 & X x_0 & \dots & X^{n-1} x_0 \end{bmatrix}$ is invertible. This allows us to conclude that

$$X = \bar{X} \begin{bmatrix} x_0 & \bar{X} x_0 & \dots & \bar{X}^{n-1} x_0 \end{bmatrix} \begin{bmatrix} x_0 & X x_0 & \dots & X^{n-1} x_0 \end{bmatrix}^{-1}.$$

However, by (8.11), this implies that $X = \bar{X}$. Consequently, for all $\bar{X} \in \operatorname{im} \mathcal{K} \setminus \{X\}$ we have $x_0 \notin \left\langle \ker(\bar{X} - X) \mid X \right\rangle$. Finally, we conclude by Theorem 8.1 that the network reconstruction problem is solvable for $(x_0, X, \mathcal{K})$. □

In the next section, we show that for $\mathcal{K} = \mathcal{Q}$, the observability condition of Theorem 8.2 is necessary *and* sufficient. However, in general, the observability condition is not necessary. In particular, this will be shown for consensus networks in Section 8.4.3.

8.4.2 Solvability for $\mathcal{K} = \mathcal{Q}$

In this subsection, we consider the case that $\mathcal{K} = \mathcal{Q}$. This case corresponds to the situation where we do not have any additional information (such as sign constraints) on the entries of the state matrix X. To be precise, we consider system (8.4), where $X \in \mathcal{Q}(G)$ for some network graph $G \in \mathcal{G}^n$. We will see that the solvability of the network reconstruction problem for $(x_0, X, \mathcal{Q})$ is in fact equivalent to the observability of the pair $(x_0^\top, X)$. This is stated in the following theorem.

Theorem 8.3. Consider a graph $G \in \mathcal{G}^n$, let $X \in \mathcal{Q}(G)$, and let $x_0 \in \mathbb{R}^n$. The network reconstruction problem is solvable for $(x_0, X, \mathcal{Q})$ if and only if the pair $(x_0^\top, X)$ is observable.

Proof. Sufficiency follows immediately from Theorem 8.2 by taking $\mathcal{K} = \mathcal{Q}$. Hence, assume that the pair $(x_0^\top, X)$ is unobservable. We want to show that the network reconstruction problem is not solvable for $(x_0, X, \mathcal{Q})$. To do so, we will construct a matrix $\bar{X} \neq X$ such that $x_0 \in \left\langle \ker(\bar{X} - X) \mid X \right\rangle$.
Let $v \in \mathbb{R}^n$ be a nonzero vector such that

$$v^\top \begin{bmatrix} x_0 & Xx_0 & \dots & X^{n-1}x_0 \end{bmatrix} = 0. \tag{8.13}$$

Subsequently, define the matrix $\bar{X} := X + vv^\top$. By definition of v, we obtain $\bar{X}^i x_0 = X^i x_0$, for $i = 1, 2, \dots, n$. Consequently, $x_0 \in \left\langle \ker(\bar{X} - X) \mid X \right\rangle$. It remains to be shown that $\bar{X} \in \operatorname{im} \mathcal{Q}$, i.e., $\bar{X} \in \mathcal{Q}(\bar{G})$ for some $\bar{G} \in \mathcal{G}^n$. Define the simple undirected graph $\bar{G} = (V, E)$, where $V := \{1, 2, \dots, n\}$, and for distinct $i, j \in V$, we have $(i, j) \in E$ if and only if $\bar{X}_{ij} \neq 0$. By definition of the qualitative class $\mathcal{Q}(\bar{G})$, we obtain $\bar{X} \in \mathcal{Q}(\bar{G})$. We conclude that the network reconstruction problem is not solvable for $(x_0, X, \mathcal{Q})$. □

8.4.3 Solvability for $\mathcal{K} = -\mathcal{L}$ and $\mathcal{K} = \mathcal{A}$

In what follows, we consider solvability of the network reconstruction problem for consensus and adjacency networks. We will start with consensus networks. That is, we consider the system

$$\begin{aligned} \dot{x}(t) &= -Lx(t) \\ x(0) &= x_0, \end{aligned} \tag{8.14}$$

where $x \in \mathbb{R}^n$ is the state and $L \in \mathcal{L}(G)$ denotes the Laplacian matrix of a graph $G \in \mathcal{G}^n$. In this section we show by means of an example that observability of $(x_0^\top, -L)$ is not necessary for the solvability of the network reconstruction problem for $(x_0, -L, -\mathcal{L})$. In Section 8.5 we will use this fact to establish an algorithm for network reconstruction of consensus networks, that does not require observability of the pair $(x_0^\top, -L)$. Consider the Laplacian matrix

$$L = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix},$$

corresponding to the star graph G depicted in Figure 8.1.

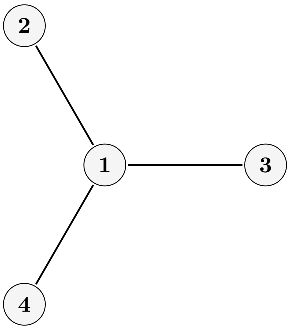

Figure 8.1: Star graph G.

Moreover, consider the initial condition $x_0 \in \mathbb{R}^4$ given by $x_0 = \text{col}(1,0,3,1)$. We claim that the network reconstruction problem is solvable for $(x_0, -L, -\mathcal{L})$, even though the pair $(x_0^\top, -L)$ is unobservable. Indeed, it can be verified that the unobservable subspace of $(x_0^\top, -L)$ is $\langle \ker x_0^\top \mid -L \rangle = \operatorname{im} v$, where the vector v is defined as $v := \text{col}(0,2,1,-3)$. This implies that $(x_0^\top, -L)$ is unobservable. To prove that the network reconstruction problem is solvable for $(x_0, -L, -\mathcal{L})$, consider a Laplacian matrix $\bar{L} \in \operatorname{im} \mathcal{L}$ such that $x_0 \in \langle \ker (L - \bar{L}) \mid -L \rangle$. We obtain

$$(L - \bar{L}) \begin{bmatrix} x_0 & -Lx_0 & \dots & (-L)^{n-1}x_0 \end{bmatrix} = 0. \tag{8.15}$$

In other words, the columns of the matrix $D := L - \bar{L}$ are contained in the unobservable subspace of $(x_0^\top, -L)$. Since D is symmetric and $\langle \ker x_0^\top \mid -L \rangle = \operatorname{im} v$, we find

$$D = \alpha \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 4 & 2 & -6 \\ 0 & 2 & 1 & -3 \\ 0 & -6 & -3 & 9 \end{bmatrix}, \tag{8.16}$$

for some $\alpha \in \mathbb{R}$. If $\alpha \neq 0$, the entries D_{32} and D_{42} of the matrix D have opposite sign. Since we have $L_{32} = L_{42} = 0$, we conclude from the relation $\bar{L} = L - D$ that $\bar{L}_{32}$ and $\bar{L}_{42}$ have opposite sign. However, this is a contradiction as $\bar{L}$ is a Laplacian matrix. Therefore, we conclude that $\alpha = 0$, and hence, $D = 0$. Consequently, we obtain $L = \bar{L}$. By Theorem 8.1, we conclude that the network reconstruction problem is solvable for $(x_0, -L, -\mathcal{L})$. Thus, we have shown that observability of the pair $(x_0^\top, L)$ is not necessary for the solvability of the network reconstruction problem for $(x_0, -L, -\mathcal{L})$.

It can be shown that $\operatorname{im} v$ also equals $\langle \ker x_0^\top \mid A \rangle$, where $A \in \mathcal{A}(G)$ denotes the unweighted *adjacency* matrix associated with the star graph G depicted in Figure 8.1. Then, using the exact same reasoning as before, we conclude that the pair $(x_0^\top, A)$ is unobservable, but the network reconstruction problem is solvable for $(x_0, A, \mathcal{A})$. In other words, observability of $(x_0^\top, A)$ is not necessary for the solvability of the network reconstruction problem for $(x_0, A, \mathcal{A})$.

8.5 THE NETWORK RECONSTRUCTION PROBLEM

In this section, we provide a solution to Problem 8.2. That is, given measurements generated by an unknown network, we establish algorithms to infer the network topology. Similar to the setup of Section 8.4, we start with the most general case in which $\mathcal{K}$ is any mapping satisfying (8.3). For this case, we obtain a general methodology to infer $X \in \mathcal{K}(G)$ from measurements. Subsequently, we provide specific algorithms for network reconstruction in the case that $\mathcal{K} = \mathcal{Q}$ (Section 8.5.2), and in the case of consensus and adjacency networks (Section 8.5.3).

8.5.1 Network reconstruction for general $\mathcal{K}$

Recall that we consider the system (8.4), where the matrix X and graph G are unknown, but the state vector $x_{x_0}(\cdot)$ of (8.4) can be measured during the time interval $[0, T]$. In this section, we establish a method to infer the matrix X and graph G using the vector $x_{x_0}(t)$ for $t \in [0, T]$. Firstly, define the matrix

$$P := \int_0^T x_{x_0}(t) x_{x_0}(t)^\top dt = \int_0^T e^{Xt} x_0 x_0^\top e^{Xt} dt. \tag{8.17}$$

Note that P can be computed from the measurements $x_{x_0}(t)$ for $t \in [0, T]$. The unknown matrix X is a solution to a *Lyapunov equation* involving the matrix P. Indeed, we have

$$\begin{aligned} XP + PX &= \int_0^T \left(X e^{Xt} x_0 x_0^\top e^{Xt} + e^{Xt} x_0 x_0^\top e^{Xt} X \right) dt \\ &= \int_0^T \frac{d}{dt} \left(e^{Xt} x_0 x_0^\top e^{Xt} \right) dt \\ &= x_T x_T^\top - x_0 x_0^\top, \end{aligned} \tag{8.18}$$

where $x_T := x_{x_0}(T) = e^{XT} x_0$. In other words, X satisfies the Lyapunov equation

$$XP + PX = Q, \tag{8.19}$$

where Q is defined as $Q := x_T x_T^\top - x_0 x_0^\top$. Note that we can compute the matrix Q from the measurements $x_{x_0}(t)$ at time $t = 0$ and time $t = T$. Therefore, if the matrix $S = X$ is the *unique* solution to the Lyapunov equation

$$SP + PS = Q, \tag{8.20}$$

we can find X (and therefore G), by solving (8.20) for S. However, it turns out that in general it is not necessary for network reconstruction that the Lyapunov equation (8.20) has a unique solution S. In fact, we only need a unique solution S in the image of $\mathcal{K}$. That is, the Lyapunov equation (8.20) may have many solutions, but if only one of these solutions is contained in $\operatorname{im} \mathcal{K}$, we can solve the network reconstruction problem for $(x_0, X, \mathcal{K})$. This is stated more formally in the following theorem.

Theorem 8.4. Let $G \in \mathcal{G}^n$, and let the mapping $\mathcal{K}$ be as in (8.3). Moreover, consider $X \in \mathcal{K}(G)$, $x_0 \in \mathbb{R}^n$, and let P and Q be as defined in (8.17) and (8.19) respectively. The network reconstruction problem is solvable for $(x_0, X, \mathcal{K})$ if and only if there exists a unique matrix S satisfying

$$SP + PS = Q, \quad S \in \operatorname{im} \mathcal{K}. \tag{8.21}$$

Moreover, under this condition, we have $S = X$.

Before we can prove Theorem 8.4, we need the following proposition, which states that $\ker P$ equals the unobservable subspace of the pair $(x_0^\top, X)$.

Proposition 8.2. Let P, x_0 and X be as in (8.17). Then we have that $\ker P = \langle \ker x_0^\top \mid X \rangle$.

Proof. Let $v \in \ker P$. We compute

$$v^\top P v = \int_0^\top \left(x_0^\top e^{Xt} v \right)^2 dt = 0, \tag{8.22}$$

from which we obtain $x_0^\top e^{Xt} v = 0$ for all $t \in [0, T]$. Since $x_0^\top e^{Xt} v$ is a real analytic function, we see that $x_0^\top e^{Xt} v = 0$ for all $t \geqslant 0$ (cf. Remark 8.1). This implies that $v \in \langle \ker x_0^\top \mid X \rangle$.

Conversely, suppose that $v \in \langle \ker x_0^\top \mid X \rangle$. This implies that $x_0^\top e^{Xt} v = 0$ for all $t \geqslant 0$. We compute

$$Pv = \int_0^T e^{Xt} x_0 x_0^\top e^{Xt} v \, dt = 0. \tag{8.23}$$

In other words, we obtain $v \in \ker P$. We conclude that $\ker P = \langle \ker x_0^\top \mid X \rangle$, which completes the proof. □

Proof of Theorem 8.4. To prove the "if" part, suppose that the network reconstruction problem is not solvable for $(x_0, X, \mathcal{K})$. We want to prove that the solution to (8.21) is not unique. By hypothesis, there exists a matrix $\bar{X} \in \operatorname{im} \mathcal{K} \setminus \{X\}$ such that $e^{Xt} x_0 = e^{\bar{X}t} x_0$ for all $t \in [0, T]$. We can repeat the discussion of Equation (8.18) for $\bar{X}$, to show that $\bar{X}$ also solves the Lyapunov equation (8.21). Consequently, we conclude that there exists no unique solution S satisfying (8.21).

Conversely, to prove the “only if" part, suppose that there exists no unique solution to (8.21). Note that $S = X$ is always a solution to (8.21) by Equation (8.19). This implies that there exists a matrix $\bar{X} \neq X$ satisfying (8.21). Consequently, $\bar{X} \in \operatorname{im} \mathcal{K}$, and

$$(\bar{X} - X)P + P(\bar{X} - X) = 0. \tag{8.24}$$

Since P is symmetric positive semidefinite, there exists an orthogonal matrix $U \in \mathbb{R}^{n\times n}$ such that $P = U\Lambda U^\top$, where

$$\Lambda = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix},$$

with D a positive definite diagonal matrix. We define the matrix $\hat{X} := U^\top(\bar{X} - X)U$. It follows from (8.24) that $\hat{X}$ satisfies the Lyapunov equation $\hat{X}\Lambda + \Lambda\hat{X} = 0$. Next, we partition $\hat{X}$ as

$$\hat{X} = \begin{bmatrix} \hat{X}_{11} & \hat{X}_{12} \\ \hat{X}_{21} & \hat{X}_{22} \end{bmatrix},$$

where the partitioning of $\hat{X}$ is compatible with the one of Λ. Then, we rewrite $\hat{X}\Lambda + \Lambda\hat{X} = 0$ as

$$\begin{bmatrix} \hat{X}_{11}D + D\hat{X}_{11} & D\hat{X}_{12} \\ \hat{X}_{21}D & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Since D is nonsingular, $\hat{X}_{12} = 0$. Moreover, since D and $-D$ do not have common eigenvalues, the Lyapunov equation $\hat{X}_{11}D + D\hat{X}_{11} = 0$ has a unique solution given by $\hat{X}_{11} = 0$ (cf. Theorem 2.5.10 of [12]). This means that $\Lambda\hat{X} = 0$. Therefore, $P(\bar{X} - X) = 0$. By Proposition 8.2 we have $x_0^\top X^i(\bar{X} - X) = 0$ for $i = 0, 1, \ldots, n-1$. By exploiting symmetry, we obtain $x_0 \in \langle \ker(\bar{X} - X) \mid X \rangle$. We conclude by Theorem 8.1 that the network reconstruction problem is not solvable for $(x_0, X, \mathcal{K})$.

Finally, as we have shown in Equation (8.19) that $X \in \operatorname{im} \mathcal{K}$ is always a solution to the Lyapunov equation $SP + PS = Q$, it is immediate that $S = X$ if there exists a unique solution S to (8.21). □

Theorem 8.4 provides a general framework for network reconstruction. Indeed, suppose that the network reconstruction problem is solvable for $(x_0, X, \mathcal{K})$. We can compute the matrices P and Q from the state measurements $x_{x_0}(t)$ for $t \in [0, T]$. Then, network reconstruction boils down to computing the unique solution S to the constrained Lyapunov equation (8.21). In the subsequent sections, we will show how this can be done for several types of network dynamics.

8.5.2 Network reconstruction for $\mathcal{K} = \mathcal{Q}$

In this section, we consider network reconstruction in the case that $\mathcal{K}$ is equal to $\mathcal{Q}$. Based on Theorem 8.4 we will derive an algorithm to identify the unknown matrix $X \in \mathcal{Q}(G)$ using state measurements taken from the network.

Recall from Theorem 8.4 that the network reconstruction problem is solvable for $(x_0, X, \mathcal{Q})$ if and only if there exists a unique matrix S satisfying $SP + PS = Q$ and $S \in \operatorname{im} \mathcal{Q}$. Note that $\operatorname{im} \mathcal{Q}$ is equal to $\mathbb{S}^n$, the set of $n \times n$ symmetric matrices. In other words, if the network reconstruction problem is solvable for $(x_0, X, \mathcal{Q})$, the solution to the problem can be found by computing the unique symmetric solution to the Lyapunov equation $SP + PS = Q$. It is not difficult to see that there exists a unique *symmetric solution* to $SP + PS = Q$ if and only if there exists a unique *solution* to $SP + PS = Q$. This yields the following corollary of Theorem 8.4.

Corollary 8.1. Let $G \in \mathcal{G}^n$ be a graph, and let $X \in \mathcal{Q}(G)$. Moreover, consider a vector $x_0 \in \mathbb{R}^n$, and let P and Q be as defined in (8.17) and (8.19) respectively. The network reconstruction problem is solvable for $(x_0, X, \mathcal{Q})$ if and only if the Lyapunov equation $SP + PS = Q$ admits a unique solution S. Under this condition, we have $S = X$.

Based on Corollary 8.1, we establish Algorithm 1, which infers the state matrix X and graph G from measurements. Recall from Theorem 8.3 that the network reconstruction problem is solvable for $(x_0, X, \mathcal{Q})$ if and only if $(x_0^\top, X)$ is observable. Of course, we can not directly check observability of $(x_0^\top, X)$ since X is unknown. However, we can in fact check observability of the pair $(x_0^\top, X)$ using the matrix P. Indeed, by Proposition 8.2, $(x_0^\top, X)$ is observable if and only if the matrix P is nonsingular. This condition is similar to a *persistency of excitation* condition, found in the literature on adaptive systems, cf. Section 3.4.3 of [119].

Algorithm 1 Network reconstruction for $(x_0, X, \mathcal{Q})$

```
Input: Measurements $x_{x_0}(t)$ for $t \in [0, T]$;
Output: Matrix $X$ or “No unique solution exists";
1: Compute the matrix $P = \int_0^T x_{x_0}(t) x_{x_0}(t)^\top dt$;
2: if rank $P < n$ then
3:     return “No unique solution exists";
4: else
5:     Compute the matrix $Q = x_0 x_0^\top - x_T x_T^\top$;
6:     Solve $SP + PS = Q$ with respect to $S$;
7:     return $X = S$;
8: end if
```

A classic method to solve the Lyapunov equation in Step 6 of Algorithm 1 is the *Bartels-Stewart algorithm* [11]. In addition, much effort has been made to develop methods for solving large-scale Lyapunov equations [77], [195]. Typically, such methods use the Galerkin projection of the Lyapunov equation onto a lower-dimensional Krylov subspace [195]. The resulting reduced problem is then solved by means of standard schemes for (small) Lyapunov equations. Using these techniques, it is possible to efficiently solve large-scale ($n > 10000$) Lyapunov equations [195].

Remark 8.4. In theory, the correctness of Theorem 8.4, Corollary 8.1, and Algorithm 1 is independent of the exact choice of time $T > 0$. However, choosing small T results in a matrix P with high condition number, and hence numerical rank computation (as in line 2 of Algorithm 1) becomes inaccurate. Consequently, in practice the value of T should be sufficiently large.

Remark 8.5. Even though the focus of this chapter is on continuous-time systems, we remark that Algorithm 1 can also be applied for network reconstruction of *discrete-time* networks of the form

$$\begin{aligned} z(k+1) &= Mz(k) \text{ for } k \in \mathbb{N} \\ z(0) &= z_0, \end{aligned} \tag{8.25}$$

where $z \in \mathbb{R}^n$ and $M \in \operatorname{im} \mathcal{Q}$. In this case, we assume that we can measure the state $z(k)$, for $k = 0, 1, \ldots, m$, where $m \geqslant n$. From these measurements, we compute

$$P := \sum_{k=0}^{m-1} z(k)z(k)^\top, \quad Q := \sum_{k=0}^{m-1} z(k+1)z(k)^\top + z(k)z(k+1)^\top.$$

Similar to the continuous-time case, the matrix P is nonsingular if and only if $(z_0^\top, M)$ is observable. Under this condition, we can reconstruct M by computing the unique solution to the Lyapunov equation $MP + PM = Q$.

The above approach can also be used for the continuous-time network (8.4) in the case that we cannot measure the state trajectory $x_{x_0}(\cdot)$ during a *time interval*, but only have access to *sampled measurements*. Indeed, suppose that we can measure $x_{x_0}(k\tau)$ for $k = 0, 1, \ldots, m$, where $\tau > 0$ is some sampling period. We can then use the framework for discrete-time systems on $z(k) := x_{x_0}(k\tau)$ to reconstruct the matrix $M = e^{X\tau}$. Subsequently, we can reconstruct X by computing the (unique) matrix logarithm of $e^{X\tau}$.

8.5.3 Network reconstruction for $\mathcal{K} = -\mathcal{L}$ and $\mathcal{K} = \mathcal{A}$

Although Algorithm 1 is applicable to general network dynamics described by state matrices $X \in \mathcal{Q}(G)$, the observability condition guaranteeing uniqueness of the solution to (8.20) can be quite restrictive if the *type* of network is a priori known. We have already seen in Section 8.4.3 that observability of the pair $(x_0^\top, X)$ is not necessary for the solvability of the network reconstruction problem for adjacency or consensus networks. Therefore, in this section we focus on network reconstruction for $(x_0, -L, -\mathcal{L})$ and $(x_0, A, \mathcal{A})$.

Recall from Theorem 8.4 that the network reconstruction problem is solvable for $(x_0, -L, -\mathcal{L})$ if and only if there exists a unique matrix S satisfying $SP + PS = Q$ and $S \in -\operatorname{im} \mathcal{L}$. Based on the definition of $\mathcal{L}$ (see Section 8.2.2), we find the following corollary of Theorem 8.4.

Corollary 8.2. Let $G \in \mathcal{G}^n$ be a graph, and let $L \in \mathcal{L}(G)$. Moreover, consider a vector $x_0 \in \mathbb{R}^n$, and let P and Q be as defined in (8.17) and (8.19) respectively.

The network reconstruction problem is solvable for $(x_0, -L, -\mathcal{L})$ if and only if there exists a unique solution S to

$$SP + PS = Q, \quad S \in \mathbb{S}^n, \quad S\mathbb{1} = 0, \quad S_{ij} \geqslant 0 \text{ for } i \neq j. \tag{8.26}$$

Moreover, under this condition, we have $S = -L$.

The constraint $S_{ij} \geqslant 0$ for $i \neq j$ can be stated as a *linear matrix inequality* (LMI) in the matrix variable S. Indeed, $S_{ij} \geqslant 0$ is equivalent to $e_i^\top S e_j \geqslant 0$, where e_k denotes the k-th column of the $n \times n$ identity matrix. Consequently, by Corollary 8.2, network reconstruction for $(x_0, -L, -\mathcal{L})$ boils down to finding the matrix S satisfying linear matrix equations and linear matrix inequalities, given by (8.26). We can deduce a corollary similar to Corollary 8.2 for the class $\mathcal{A}(G)$. In this case, the restrictions on the elements of S are $S_{ii} = 0$ and $S_{ij} \geqslant 0$ for all $i \in V$ and all $j \neq i$.

8.6 ILLUSTRATIVE EXAMPLE

In this section we illustrate the developed theory by considering an example of a sensor network. Specifically, consider a graph $G = (V, E)$ consisting of 100 sensor nodes, monitoring a region of 1 km × 1 km (see Figure 8.2). It is assumed that the sensors are linked using a so-called *geometric link model* [166]. This means that there is a connection between two nodes in the network if and only if the distance between the two nodes is less than a certain threshold, set to be equal to 135 m in this example. It is assumed that the sensors run consensus dynamics, that is, the dynamics of the network is given by $\dot{x}(t) = -Lx(t)$, where $x \in \mathbb{R}^{100}$, and $L \in \mathcal{L}(G)$ is the unweighted Laplacian associated with G. The components of the initial condition $x_0 \in \mathbb{R}^{100}$ were selected randomly within $[0, 10]$. Moreover, for this example, measurements were used over the time-interval $[0, 10]$, i.e., $T = 10$. We compute the matrices P and Q, and solve (8.26) using Yalmip with Sedumi as an LMI solver. The resulting identified Laplacian matrix is denoted by L_r. The relative and maximum element-wise errors between the identified Laplacian L_r and original Laplacian L are very small. Specifically, we obtain

$$\frac{\|L_r - L\|}{\|L_r\|} = 1.56 \cdot 10^{-8}, \quad \max_{i,j \in V} |L_{ij} - (L_r)_{ij}| = 2.21 \cdot 10^{-7},$$

where $\|\cdot\|$ denotes the induced 2-norm.

8.7 CONCLUSIONS

In this chapter, we have considered the problem of network reconstruction for networks of linear dynamical systems. In contrast to papers studying network reconstruction for specific network dynamics such as consensus dynamics [153]

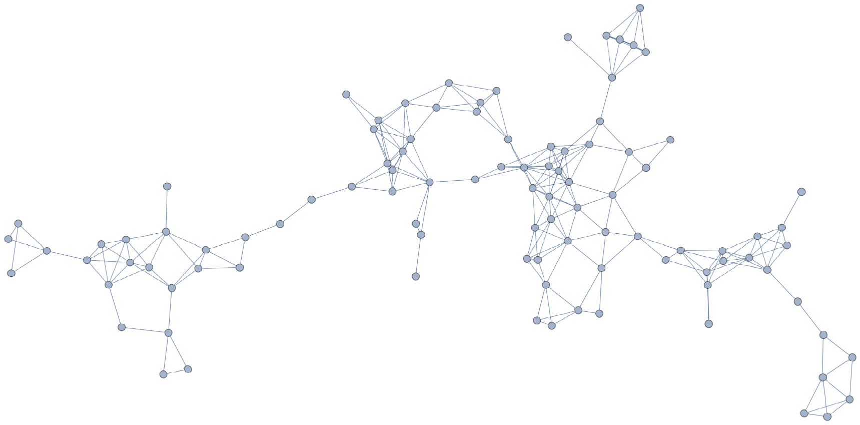

Figure 8.2: Graph G of the sensor network.

and adjacency dynamics [57], we considered network reconstruction for general linear network dynamics described by state matrices contained in the qualitative class. We formulated what is meant by solvability of the network reconstruction problem. Subsequently, we provided necessary and sufficient conditions under which the network reconstruction problem is solvable. Using constrained Lyapunov equations, we established a general framework for network reconstruction of networks of dynamical systems. We have shown that this framework can be used for a variety of network types, including consensus and adjacency networks. Finally, we have illustrated the theory by reconstructing the network topology of a sensor network.

9 NETWORK IDENTIFIABILITY AND GRAPH SIMPLIFICATION

In the preceeding two chapters we have aimed at identifying networks with an unknown interconnection structure. However, for certain networks, such as water distribution networks, the interconnection structure may be readily available. Therefore, in this chapter we focus on the identifiability of network models with *known* structure. We will focus on a notion of global identifiability of the model set, which allows us to fully characterize identifiability in terms of the locations of the measured nodes and the graph structure underlying the network. We will see that the assumption of known network structure is beneficial in the sense that network identifiability can typically be guaranteed with relatively few measured nodes.

9.1 INTRODUCTION

Networks of dynamical systems appear in a variety of domains, including power systems, robotic networks, and aerospace systems [137]. In this chapter, we consider a dynamical network model in which the relations between node signals are modelled by proper transfer functions. Such network models have received much attention in recent years, see e.g. [46, 81, 214, 223, 235].

The interconnection structure of a dynamical network can be represented by a directed graph, where vertices (or nodes) represent scalar signals, and edges correspond to transfer functions connecting different node signals. We will assume that the underlying graph (i.e., the topology) of the dynamical network is *known*. We remark that the related problem of *topology identification* has also been studied, see e.g. [70, 130, 154, 190, 226, 246].

We are interested in conditions for identifiability of dynamical networks. Identifiability is a fundamental property of a model set that guarantees that a unique (network) model can be identified, given informative data. Thus identifiability can be thought of as a prerequisite for identification: if identifiability does not hold then it is impossible to uniquely determine a network model, irrespective of the particular identification method and the experimental conditions.

In the literature, several methods have been proposed for network identification [46, 76, 214, 236], these methods all exploit the structure of the network. For instance, a prediction error method was considered in [236], where consistency and minimum variance properties were proven under the assumption that the network is identifiable, the disturbances are filtered white noise, and the inputs are persistently exciting and uncorrelated with the disturbances. Another work [76] considers subspace identification of networks with a path graph topology. As we will see, the structure of the network plays a fundamental role also with respect to the question of identifiability.

We follow the setup of [81], where all network nodes can be externally excited, but only a subset of nodes can be measured. Within this setup, we are interested in two identifiability problems. Firstly, we want to find conditions under which the transfer functions from a given node to its out-neighbours are identifiable. Secondly, we wonder under which conditions the transfer functions of all edges in the network are identifiable. In particular, our aim is to find *graph-theoretic* conditions for the above problems, that is, conditions in terms of the network structure and the locations of measured nodes. Such conditions based on the network topology are desirable since they give insight on the types of network structures that allow unique identification, and in addition may aid in the *selection* of measured nodes. Graph-theoretic methods have also been succesfully applied to assess other system-theoretic properties like structural controllability [30, 96, 113, 142, 219] and fault detection [49, 95, 174].

Identifiability of dynamical networks is an active research area, see e.g. [3, 80, 81, 152, 223, 224, 234, 235] and the references therein. The papers that are most closely related to the work presented here are [152], [224], [81], and [223], in which identifiability is also considered from *graph-theoretic* perspective. In [152] and [224], sufficient graph-theoretic conditions for identifiability have been presented for a class of *state-space* systems.

In [81], graph-theoretic conditions have been established for *generic* identifiability. That is, conditions were given under which transfer functions in the network can be identified for *"almost all"* network matrices associated with the graph. The authors of [81] show that generic identifiability is equivalent to the existence of certain *vertex-disjoint paths*, which yields elegant conditions for generic identifiability.

Inspired by the work in [81], we are interested in graph-theoretic conditions for a stronger notion, namely identifiability *for all* network matrices associated with the graph, a notion often referred to as *global identifiability*. This problem is motivated by the fact that, although generic identifiability guarantees identifiability for almost all network matrices, there are meaningful examples of network matrices that are not contained in this set of almost all network matrices. As a consequence, a situation may arise in which the system under consideration is not identifiable, even though the conditions for generic identifiability are satisfied. For an example of such a situation, we refer to Section 9.3. On the other hand, if the conditions derived in this chapter are satisfied, then it is guaranteed that the network is identifiable *for all* network matrices associated with the graph. The contributions of this chapter are the following.

1. We introduce the so-called *graph simplification process*. Based on this process, we provide necessary and sufficient conditions for the left-invertibility of certain network-related transfer matrices.

2. Using the fact that identifiability is characterized by the left-invertibility of transfer matrices [81], [223], we provide necessary and sufficient graph-

theoretic conditions for identifiability based on graph simplification. We also show that these conditions can be verified by polynomial time algorithms.

3. We compare our results with the sufficient topological conditions for identifiability based on constrained vertex-disjoint paths [223]. In particular, we show that the results obtained in this chapter generalize those in [223].

This chapter is organized as follows. In Section 9.2 we discuss the preliminaries that are used throughout this chapter. Subsequently, in Section 9.3 we state and motivate the problem. Next, in Section 9.4 we recall rank conditions for identifiability. Sections 9.5 and 9.6 contain our main results. In Section 9.5 we introduce the graph simplification process and show its relation to the left-invertibility of transfer matrices. Subsequently, in Section 9.6 we provide graph-theoretic conditions for identifiability. Our main results are compared to previous work in Section 9.7. Finally, Section 9.8 contains our conclusions.

9.2 PRELIMINARIES

9.2.1 Rational functions and rational matrices

Consider a scalar variable z and a rational function $f(z) = \frac{p(z)}{q(z)}$, where $p(z)$ and $q(z)$ are real polynomials and q is nonzero. The function f is *proper* if the degree of $p(z)$ is less than or equal to the degree of $q(z)$. We say f is *strictly proper* if the degree of $p(z)$ is less than the degree of $q(z)$. An $m \times n$ matrix $A(z)$ is called *rational* if its entries are rational functions in the variable z. In addition, $A(z)$ is *proper* if its entries are proper rational functions. We omit the argument z whenever the dependency of A on z is clear from the context. The *normal rank* of $A(z)$ is defined as $\max_{\lambda \in \mathbb{C}} \operatorname{rank} A(\lambda)$ and denoted by $\operatorname{rank} A(z)$, with slight abuse of notation. We say $A(z)$ is *left-invertible* if $\operatorname{rank} A(z) = n$. We denote the (i, j)-th entry of a matrix A by A_{ij}. Moreover, the j-th column of A is given by $A_{\bullet j}$. More generally, let $\mathcal{M} \subseteq \{1, 2, \dots, m\}$ and $\mathcal{N} \subseteq \{1, 2, \dots, n\}$. Then, $A_{\mathcal{M},\mathcal{N}}$ denotes the submatrix of A containing the rows of A indexed by $\mathcal{M}$ and the columns of A indexed by $\mathcal{N}$. Next, consider the case that A is square, i.e., $m = n$. The *determinant* of A is denoted by $\det A$, while the *adjugate* of A is denoted by $\operatorname{adj} A$. A *principal submatrix* of A is a submatrix $A_{\mathcal{M},\mathcal{M}}$, where $\mathcal{M} \subseteq \{1, 2, \dots, m\}$. The determinant of $A_{\mathcal{M},\mathcal{M}}$ is called a *principal minor* of A.

9.2.2 Graph theory

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph, with vertex (or node) set $\mathcal{V} = \{1, 2, \dots, n\}$ and edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. The graphs considered in this chapter are *simple*, i.e., without self-loops and with at most one edge from one node to another. Consider an edge $(i, j) \in \mathcal{E}$. Then (i, j) is called an *outgoing* edge of node $i \in \mathcal{V}$ and j is called an *out-neighbour* of $i \in \mathcal{V}$. The set of out-neighbours of i is denoted by

$\mathcal{N}_i^+$. Similarly, (i,j) is called an *incoming* edge of $j \in \mathcal{V}$ and node i is called an *in-neighbour* of j. The set of in-neighbours of node j is denoted by $\mathcal{N}_j^-$. For any subset $\mathcal{S} = \{v_1, v_2, \ldots, v_s\} \subseteq \mathcal{V}$ we define the $s \times n$ matrix $P(\mathcal{V};\mathcal{S})$ as $P_{ij} := 1$ if $j = v_i$, and $P_{ij} := 0$ otherwise. The complement of $\mathcal{S}$ in $\mathcal{V}$ is defined as $\mathcal{S}^c := \mathcal{V} \setminus \mathcal{S}$. Moreover, the cardinality of $\mathcal{S}$ is denoted by $|\mathcal{S}|$. A *path* $\mathcal{P}$ is a set of edges in $\mathcal{G}$ of the form $\mathcal{P} = \{(v_i, v_{i+1}) \mid i = 1, 2, \ldots, k\} \subseteq \mathcal{E}$, where the vertices $v_1, v_2, \ldots, v_{k+1}$ are *distinct*. The vertex v_1 is called a *starting node* of $\mathcal{P}$, while v_{k+1} is the *end node*. The cardinality of $\mathcal{P}$ is called the *length* of the path. A collection of paths $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_l$ is called *vertex-disjoint* if the paths have no vertex in common, that is, if for all distinct $i, j \in \{1, 2, \ldots, l\}$, we have that

$$(u_i, w_i) \in \mathcal{P}_i, (u_j, w_j) \in \mathcal{P}_j \implies u_i, w_i, u_j \text{ and } w_j \text{ are distinct.}$$

Let $\mathcal{U}, \mathcal{W} \subseteq \mathcal{V}$ be disjoint. We say there exists a path *from* $\mathcal{U}$ *to* $\mathcal{W}$ if there exist vertices $u \in \mathcal{U}$ and $w \in \mathcal{W}$ such that there exists a path in $\mathcal{G}$ with starting node u and end node w. Similarly, we say there are m vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ if there exist m vertex-disjoint paths[1] in $\mathcal{G}$ with starting nodes in $\mathcal{U}$ and end nodes in $\mathcal{W}$. In the case that $\mathcal{U} \cap \mathcal{W} \neq \varnothing$, we say there exist m vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ if there are $\max\{0, m - |\mathcal{U} \cap \mathcal{W}|\}$ vertex-disjoint paths from $\mathcal{U} \setminus \mathcal{W}$ to $\mathcal{W} \setminus \mathcal{U}$. Roughly speaking, this means that we count paths of “length zero" from every node in $\mathcal{U} \cap \mathcal{W}$ to itself.

9.3 PROBLEM STATEMENT AND MOTIVATION

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a simple directed graph with vertex set $\mathcal{V} = \{1, 2, \ldots, n\}$ and edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. We associate with each node $i \in \mathcal{V}$ a scalar *node signal* w_i, an external *excitation signal* r_i and a *disturbance signal* v_i. Then, we consider the dynamics

$$w_i = \sum_{j \in \mathcal{N}_i^-} G_{ij} w_j + r_i + v_i,$$

where $G_{ij}(z)$ is a scalar transfer function. By concatenation of the node signals, excitation signals and disturbance signals, we can write the dynamics of all nodes compactly as $w = Gw + r + v$, where w, r, and v are n-dimensional vectors and $G(z)$ is a $n \times n$ rational matrix. In addition, we consider a measured output vector y of dimension p that consists of the node signals of a subset $\mathcal{C} \subseteq \mathcal{V}$ of so-called *measured nodes*. By defining an associated binary matrix C as $C := P(\mathcal{V}, \mathcal{C})$, we can write this output as $y = Cw$. Finally, by combining the equations for w and y, we obtain the networked system

$$\begin{aligned} w &= Gw + r + v \\ y &= Cw. \end{aligned} \tag{9.1}$$

We call the matrix $G(z)$ the *network matrix* and assume that it satisfies the following properties:

[1] Such sets of vertex-disjoint paths have been studied in detail in [151], where they were called linkings.

(P1) For all $i, j \in \mathcal{V}$, the entry $G_{ij}(z)$ is a proper rational (transfer) function.

(P2) The function $G_{ij}(z)$ is nonzero if and only if $(j,i) \in \mathcal{E}$. A matrix $G(z)$ that satisfies this property is said to be *consistent* with the graph $\mathcal{G}$.

(P3) Every principal minor of $\lim_{z\to\infty}(I - G(z))$ is nonzero. This implies that the network model (9.1) is *well-posed* in the sense of Definition 2.11 of [46].

A network matrix $G(z)$ satisfying Properties P1, P2, and P3 is called *admissible*. The set of all admissible network matrices is denoted by $\mathcal{A}(\mathcal{G})$.

For the development of this chapter, it is important to distinguish between the following two concepts:

- *Identifiability*: this is a fundamental property of the set of models of the form (9.1) that captures under what conditions identification is conceptually possible. If this property is not satisfied, one cannot uniquely identify the dynamics, no matter which identification algorithm is used. Identifiability does not involve any use of data.

- *Identification*: this involves the development of numerical algorithms for identifying the system dynamics from data. If identifiability holds then identification can be successfully performed in different ways under hypotheses on the noise and the informativity of the data [114].

This chapter focuses on characterizations of *identifiability*. To explain what identifiability means in a network context, we first write (9.1) in input/output form as

$$y = C(I - G)^{-1}r + \bar{v},$$

where $\bar{v} := C(I - G)^{-1}v$. It is well-known that the transfer matrix $C(I - G(z))^{-1}$ from r to y can be obtained from $\{r, y\}$-data, under suitable assumptions on r and v [114]. The question of network identifiability is then the following: which entries of $G(z)$ can be uniquely reconstructed from $C(I - G(z))^{-1}$? In this chapter we restrict our attention to the identifiability of the transfer functions outgoing a given node i (i.e., identifiability of a column of $G(z)$), and to the identifiability of the entire matrix $G(z)$. A standing assumption in our work is that we *know* the graph structure $\mathcal{G}$ underlying the dynamical network.

In recent work [81], [15] the problem of identifiability has been considered from *generic* viewpoint. Graph-theoretic conditions were given under which certain entries of $G(z)$ can be uniquely reconstructed from $C(I - G(z))^{-1}$ *for almost all* network matrices G consistent with the graph. For a formal definition of generic identifiability we refer to Definition 1 of [81]. Here, we will informally illustrate the approach of [81]. We will use the shorthand notation $T(z; G) := (I - G(z))^{-1}$. This means that the transfer matrix from r to y equals CT.

Example 9.1. Consider the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in Figure 9.1. We assume that the node signals of nodes 4 and 5 can be measured, that is, $\mathcal{C} = \{4, 5\}$. Suppose that we want to identify the transfer functions from node 1 to its out-neighbours, i.e., the transfer functions $G_{21}(z)$ and $G_{31}(z)$. According to Corollary 5.1 of [81], this is possible if and only if there exist two vertex-disjoint paths from $\mathcal{N}_1^+$ to $\mathcal{C}$. Note that this is the case in this example, since the edges $(2,4)$ and $(3,5)$ are two vertex-disjoint paths. To see why we can generically identify the transfer functions G_{21} and G_{31}, we compute CT as:

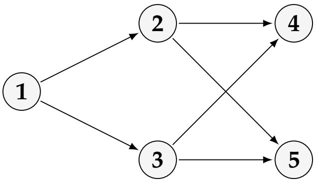

Figure 9.1: Graph used in Example 9.1.

$$CT = \begin{bmatrix} G_{42}G_{21} + G_{43}G_{31} & G_{42} & G_{43} & 1 & 0 \\ G_{52}G_{21} + G_{53}G_{31} & G_{52} & G_{53} & 0 & 1 \end{bmatrix},$$

where we omit the argument z. Clearly, we can uniquely obtain the transfer functions G_{42}, G_{43}, G_{52}, and G_{53} from CT. Moreover, the transfer matrices G_{21} and G_{31} satisfy

$$\begin{bmatrix} G_{42} & G_{43} \\ G_{52} & G_{53} \end{bmatrix} \begin{bmatrix} G_{21} \\ G_{31} \end{bmatrix} = \begin{bmatrix} T_{41} \\ T_{51} \end{bmatrix}. \tag{9.2}$$

Equation (9.2) has a unique solution in the unknowns G_{21} and G_{31} if $G_{42}G_{53} - G_{43}G_{52} \neq 0$, which means that we can identify G_{21} and G_{31} for “almost all" G consistent with $\mathcal{G}$.

As mentioned before, the approach based on vertex-disjoint paths [81] gives necessary and sufficient conditions for *generic* identifiability. This implies that for some network matrices G, it might be impossible to identify the transfer functions, even though the path-based conditions are satisfied. For instance, in Example 9.1 we cannot identify G_{21} and G_{31} if the network matrix G is such that $G_{42} = G_{43} = G_{52} = G_{53}$. Nonetheless, scenarios in which some (or all) transfer functions in the network are equal occur frequently, for example in the study of undirected (electrical) networks [53], in unweighted consensus networks [158], and in the study of Cartesian products of graphs [29]. Therefore, instead of generic identifiability, we are interested in graph-theoretic conditions that guarantee identifiability *for all* admissible network matrices. Such a problem might seem like a simple extension of the work on generic identifiability [81]. However, to analyze *strong structural* network properties (for all network matrices), we typically need completely different graph-theoretic tools than the ones used in the analysis of *generic* network properties. For instance, in the literature on *controllability*, weak structural controllability is related to maximal matchings [113], while strong structural controllability is related to zero forcing sets [142] and constrained matchings [29]. To make the problem of this chapter more precise, we state a few definitions. First, we are interested in conditions under which all transfer functions from a node i to its out-neighbours $\mathcal{N}_i^+$ are identifiable (for any admissible network matrix $G \in \mathcal{A}(\mathcal{G})$). If this is the case, we say $(i, \mathcal{N}_i^+)$ is *globally identifiable*, or simply $(i, \mathcal{N}_i^+)$ is identifiable for short.

Definition 9.1. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let $i \in \mathcal{V}$ and $\mathcal{C} \subseteq \mathcal{V}$. Moreover, define $C = P(\mathcal{V}, \mathcal{C})$. We say $(i, \mathcal{N}_i^+)$ is *(globally) identifiable* from $\mathcal{C}$ if the implication

$$CT(z; G) = CT(z; \bar{G}) \implies G_{\bullet i}(z) = \bar{G}_{\bullet i}(z)$$

holds for all $G(z), \bar{G}(z) \in \mathcal{A}(\mathcal{G})$.

In addition, we are interested in conditions under which the *entire* network matrix G can be identified. If this is the case, we say the graph $\mathcal{G}$ is (globally) identifiable.

Definition 9.2. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let $\mathcal{C} \subseteq \mathcal{V}$ and $C = P(\mathcal{V}, \mathcal{C})$. We say $\mathcal{G}$ is *(globally) identifiable* from $\mathcal{C}$ if the implication

$$CT(z; G) = CT(z; \bar{G}) \implies G(z) = \bar{G}(z)$$

holds for all $G(z), \bar{G}(z) \in \mathcal{A}(\mathcal{G})$.

The main goals of this chapter are to find graph-theoretic conditions for identifiability of $(i, \mathcal{N}_i^+)$ and $\mathcal{G}$.

Problem 9.1. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and measured nodes $\mathcal{C} \subseteq \mathcal{V}$. Provide necessary and sufficient graph-theoretic conditions under which, respectively, $(i, \mathcal{N}_i^+)$ and $\mathcal{G}$ are identifiable from $\mathcal{C}$.

Graph-theoretic conditions for global identifiability are attractive for two reasons. First, such conditions will give insight on the types of graph structures that allow identification. Secondly, they allow us to select measured nodes guaranteeing identifiability *before* collecting data. To deal with Problem 9.1, we make use of rank conditions for identifiability which we will recall in Section 9.4. To verify such rank conditions, we introduce a novel graph-theoretic concept called the *graph simplification process* in Section 9.5.

9.4 RANK CONDITIONS FOR IDENTIFIABILITY

First, we review some of the conditions for identifiability in terms of the normal rank of transfer matrices. For the proofs of all results in this section, we refer to [223]. Recall from Section 9.2 that $T_{\mathcal{C},\mathcal{N}_i^+}(z; G)$ denotes the submatrix of T formed by taking the rows of T indexed by $\mathcal{C}$ and the columns of T corresponding to $\mathcal{N}_i^+$. This means that $T_{\mathcal{C},\mathcal{N}_i^+}(z; G)$ is a submatrix of the transfer matrix $CT(z; G)$ from r to y, obtained by selecting the columns corresponding to $\mathcal{N}_i^+$. The following lemma (Lemma 5 of [223]) asserts that identifiability of $(i, \mathcal{N}_i^+)$ is equivalent to a rank condition on the matrix $T_{\mathcal{C},\mathcal{N}_i^+}(z; G)$.

Lemma 9.1. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, let $i \in \mathcal{V}$, and $\mathcal{C} \subseteq \mathcal{V}$. Then, $(i, \mathcal{N}_i^+)$ is identifiable from $\mathcal{C}$ if and only if $\operatorname{rank} T_{\mathcal{C},\mathcal{N}_i^+}(z; G) = |\mathcal{N}_i^+|$ for all $G(z) \in \mathcal{A}(\mathcal{G})$.

As an immediate consequence of Lemma 9.1, we find conditions for the identifiability of $\mathcal{G}$ based on the normal rank of transfer matrices. This is stated in the following corollary.

Corollary 9.1. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let $\mathcal{C} \subseteq \mathcal{V}$. Then, $\mathcal{G}$ is identifiable from $\mathcal{C}$ if and only if $\operatorname{rank} T_{\mathcal{C},\mathcal{N}_i^+}(z; G) = |\mathcal{N}_i^+|$ for all $i \in \mathcal{V}$ and all $G(z) \in \mathcal{A}(\mathcal{G})$.

Although Lemma 9.1 and Corollary 9.1 give necessary and sufficient conditions for the identifiability of respectively $(i, \mathcal{N}_i^+)$ and $\mathcal{G}$, these conditions are limited since there is no obvious method to *check* left-invertibility of $T_{\mathcal{C},\mathcal{N}_i^+}(z; G)$ for an infinite number of matrices G. Therefore, one of the main results of this chapter will be graph-theoretic conditions for the left-invertibility of $T_{\mathcal{W},\mathcal{U}}(z; G)$, where $\mathcal{U}, \mathcal{W} \subseteq \mathcal{V}$ are any two subsets of vertices. These conditions will be introduced in the next section.

9.5 THE GRAPH SIMPLIFICATION PROCESS

In this section we provide necessary and sufficient conditions for left-invertibility of $T_{\mathcal{W},\mathcal{U}}(z; G)$ for all $G(z) \in \mathcal{A}(\mathcal{G})$, where $\mathcal{U}, \mathcal{W} \subseteq \mathcal{V}$. Loosely speaking, the idea is to simplify the graph $\mathcal{G}$ and nodes $\mathcal{W}$ in such a way that checking left-invertibility becomes easy. To give the reader some intuition for the approach, we start with the following basic lemma, which asserts that $T_{\mathcal{W},\mathcal{U}}(z; G)$ is left-invertible if $\mathcal{U} \subseteq \mathcal{W}$.

Lemma 9.2. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let $\mathcal{U}, \mathcal{W} \subseteq \mathcal{V}$. If $\mathcal{U} \subseteq \mathcal{W}$ then $\operatorname{rank} T_{\mathcal{W},\mathcal{U}}(z; G) = |\mathcal{U}|$ for all $G(z) \in \mathcal{A}(\mathcal{G})$.

The proof of Lemma 9.2 is postponed to Section 9.9.1. The condition $\mathcal{U} \subseteq \mathcal{W}$ considered in Lemma 9.2 is clearly not necessary for left-invertibility. One can show this using the example $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2\}$, $\mathcal{E} = \{(1, 2)\}$, and the subsets $\mathcal{U}$ and $\mathcal{W}$ are chosen as $\mathcal{U} = \{1\}$ and $\mathcal{W} = \{2\}$. However, the *main idea* of the graph simplification process is to simplify $\mathcal{G}$ and to “move" the nodes in $\mathcal{W}$ closer to the nodes in $\mathcal{U}$ such that the condition $\mathcal{U} \subseteq \mathcal{W}$ possibly holds *after* applying these operations. Of course, we cannot blindly modify the graph $\mathcal{G}$ since this would affect the left-invertibility of $T_{\mathcal{W},\mathcal{U}}(z; G)$. Instead, we will now state two lemmas in which we consider two different operations on $\mathcal{G}$ and $\mathcal{W}$ that *preserve* left-invertibility of $T_{\mathcal{W},\mathcal{U}}(z; G)$. We emphasize that the graph operations are introduced for analysis purposes only. Indeed, since the condition of Lemma 9.2 is simple to check, the graph operations should be seen as a *tool* to check left-invertibility of the transfer matrix of a given *fixed* graph $\mathcal{G}$. First, we state

Lemma 9.3 which asserts that left-invertibility of $T_{\mathcal{W},\mathcal{U}}(z;G)$ is unaffected by the removal of the outgoing edges of $\mathcal{W}$.

Lemma 9.3. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let $\mathcal{U}, \mathcal{W} \subseteq \mathcal{V}$. Moreover, let $\bar{\mathcal{G}} = (\mathcal{V}, \bar{\mathcal{E}})$ be the graph obtained from $\mathcal{G}$ by removing all outgoing edges of the nodes in $\mathcal{W}$. Then $\operatorname{rank} T_{\mathcal{W},\mathcal{U}}(z;G) = |\mathcal{U}|$ for all $G(z) \in \mathcal{A}(\mathcal{G})$ if and only if $\operatorname{rank} T_{\mathcal{W},\mathcal{U}}(z;\bar{G}) = |\mathcal{U}|$ for all $\bar{G}(z) \in \mathcal{A}(\bar{\mathcal{G}})$.

Proof. Let $G(z) \in \mathcal{A}(\mathcal{G})$. Relabel the nodes in $\mathcal{V}$ such that

$$G = \begin{bmatrix} G_{\mathcal{R},\mathcal{R}} & G_{\mathcal{R},\mathcal{W}} \\ G_{\mathcal{W},\mathcal{R}} & G_{\mathcal{W},\mathcal{W}} \end{bmatrix}, \tag{9.3}$$

where $\mathcal{R} := \mathcal{V} \setminus \mathcal{W}$ and the argument z has been omitted. Define the matrix $\bar{G}$ as

$$\bar{G} = \begin{bmatrix} G_{\mathcal{R},\mathcal{R}} & 0 \\ G_{\mathcal{W},\mathcal{R}} & 0 \end{bmatrix}. \tag{9.4}$$

The matrix $\bar{G}$ is an admissible matrix consistent with $\bar{\mathcal{G}}$, i.e., $\bar{G} \in \mathcal{A}(\bar{\mathcal{G}})$. To see this, note that $\bar{G}$ satisfies Property P1. Moreover, since all outgoing edges of nodes in $\mathcal{W}$ are removed in the graph $\bar{\mathcal{G}}$, the matrix $\bar{G}$ is consistent with $\bar{\mathcal{G}}$. Hence, $\bar{G}$ satisfies property P2. Finally, to see that $\bar{G}$ satisfies Property P3, note that any principal minor of

$$\lim_{z\to\infty} \begin{bmatrix} I - G_{\mathcal{R},\mathcal{R}}(z) & 0 \\ -G_{\mathcal{W},\mathcal{R}}(z) & I \end{bmatrix} \tag{9.5}$$

is either 1 or equal to a principal minor of $\lim_{z\to\infty}(I - G_{\mathcal{R},\mathcal{R}}(z))$, which is nonzero by the assumption that G is admissible. We conclude that $\bar{G} \in \mathcal{A}(\bar{\mathcal{G}})$. Next, by Proposition 2.8.7 of [19], the inverse of $I - G$ can be written as

$$T = (I-G)^{-1} = \begin{bmatrix} * & * \\ S(G)G_{\mathcal{W},\mathcal{R}}(I - G_{\mathcal{R},\mathcal{R}})^{-1} & S(G) \end{bmatrix},$$

where $S(G) := (I - G_{\mathcal{W},\mathcal{W}} - G_{\mathcal{W},\mathcal{R}}(I - G_{\mathcal{R},\mathcal{R}})^{-1}G_{\mathcal{R},\mathcal{W}}))^{-1}$ denotes the inverse Schur complement of $I - G$. Using the same formula to compute the inverse of $I - \bar{G}$, we find

$$\bar{T} := (I - \bar{G})^{-1} = \begin{bmatrix} * & * \\ G_{\mathcal{W},\mathcal{R}}(I - G_{\mathcal{R},\mathcal{R}})^{-1} & I \end{bmatrix}.$$

The above expressions for T and $\bar{T}$ imply that

$$T_{\mathcal{W},\mathcal{U}} = S(G)\bar{T}_{\mathcal{W},\mathcal{U}},$$

and because $S(G)$ has full normal rank, we obtain

$$\operatorname{rank} T_{\mathcal{W},\mathcal{U}} = \operatorname{rank} \bar{T}_{\mathcal{W},\mathcal{U}}. \tag{9.6}$$

Next, we use (9.6) to prove the lemma. First, to prove the “if" statement, suppose that $\operatorname{rank} T_{\mathcal{W},\mathcal{U}}(z;\bar{G}) = |\mathcal{U}|$ for all matrices $\bar{G} \in \mathcal{A}(\bar{\mathcal{G}})$. Let $G \in \mathcal{A}(\mathcal{G})$. Using G, construct the matrix $\bar{G} \in \mathcal{A}(\bar{\mathcal{G}})$ in (9.4). By hypothesis, $\operatorname{rank} T_{\mathcal{W},\mathcal{U}}(z;\bar{G}) = |\mathcal{U}|$ and therefore we conclude from (9.6) that $\operatorname{rank} T_{\mathcal{W},\mathcal{U}}(z;G) = |\mathcal{U}|$.

Subsequently, to prove the "only if" statement, suppose that $\operatorname{rank} T_{\mathcal{W},\mathcal{U}}(z;G) = |\mathcal{U}|$ for all $G(z) \in \mathcal{A}(\mathcal{G})$. Consider any matrix $\bar{G}(z) \in \mathcal{A}(\bar{\mathcal{G}})$ and note that $\bar{G}$ can be written in the form (9.4). Next, we choose the matrices $G_{\mathcal{R},\mathcal{W}}$ and $G_{\mathcal{W},\mathcal{W}}$ such that the matrix G in (9.3) is consistent with the graph $\mathcal{G}$, and such that the nonzero entries of $G_{\mathcal{R},\mathcal{W}}$ and $G_{\mathcal{W},\mathcal{W}}$ are *strictly proper* rational functions. This means that G readily satisfies Properties P1 and P2 (see Section 9.3). In fact, G also satisfies P3. Indeed, since $\lim_{z\to\infty}(I - G(z))$ is given by (9.5), it follows that every principal minor of $\lim_{z\to\infty}(I - G(z))$ is either 1 or equal to a principal minor of $\lim_{z\to\infty}(I - G_{\mathcal{R},\mathcal{R}})$, which is nonzero by the hypothesis that $\bar{G}(z) \in \mathcal{A}(\bar{\mathcal{G}})$. We conclude that G satisfies Properties P1, P2, and P3, equivalently, $G \in \mathcal{A}(\mathcal{G})$. By hypothesis, $\operatorname{rank} T_{\mathcal{W},\mathcal{U}}(z;G) = |\mathcal{U}|$ and consequently, by (9.6) we conclude that $\operatorname{rank} T_{\mathcal{W},\mathcal{U}}(z;\bar{G}) = |\mathcal{U}|$. This proves the lemma. □

Remark 9.1. In similar fashion as in the proof of Lemma 9.3, we can prove that all *incoming* edges of nodes in $\mathcal{U}$ can be removed without affecting the left-invertibility of $T_{\mathcal{W},\mathcal{U}}(z;G)$.

Inspired by Lemma 9.3, we wonder what type of operations we can further perform on the graph $\mathcal{G}$ and nodes $\mathcal{W}$ without affecting left-invertibility of $T_{\mathcal{W},\mathcal{U}}(z;G)$. In what follows we will show that under suitable conditions it is possible to "move" the nodes in $\mathcal{W}$ closer to the nodes in $\mathcal{U}$. Here the notion of *reachability* in graphs will play an important role. For a subset $\mathcal{U} \subseteq \mathcal{V}$ and a node $j \in \mathcal{V} \setminus \mathcal{U}$, we say j is *reachable* from $\mathcal{U}$ if there exists at least one path from $\mathcal{U}$ to j. By convention, if $j \in \mathcal{U}$ then j is reachable from $\mathcal{U}$. In the following lemma, we will show that the rank of $T_{\mathcal{W},\mathcal{U}}(z;G)$ is unaffected if we replace a node $k \in \mathcal{W} \setminus \mathcal{U}$ by j, provided that j is the *only* in-neighbour of k that is reachable from $\mathcal{U}$.

Lemma 9.4. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let $\mathcal{U}, \mathcal{W} \subseteq \mathcal{V}$. Suppose that $k \in \mathcal{W} \setminus \mathcal{U}$ has exactly one in-neighbour $j \in \mathcal{N}_k^-$ that is reachable from $\mathcal{U}$. Then for all $G(z) \in \mathcal{A}(\mathcal{G})$, we have

$$\operatorname{rank} T_{\mathcal{W},\mathcal{U}}(z;G) = \operatorname{rank} T_{\bar{\mathcal{W}},\mathcal{U}}(z;G),$$

where $\bar{\mathcal{W}} := (\mathcal{W} \setminus \{k\}) \cup \{j\}$.

Remark 9.2. We emphasize that Lemma 9.4 does not require node k to have exactly one in-neighbour. In general, node k may have multiple in-neighbours, but if exactly one of such neighbours is reachable from $\mathcal{U}$, we can apply Lemma 9.4. The intuition of Lemma 9.4 is as follows: under the assumptions, all information from the nodes in $\mathcal{U}$ enters node k *via* node j. Therefore, choosing node k or node j as a node in $\mathcal{W}$ does not make any difference. An interesting special case is obtained when *both* nodes j and k are contained in $\mathcal{W}$. In this case, we obtain $\bar{\mathcal{W}} = \mathcal{W} \setminus \{k\}$, that is, node k can be removed from $\mathcal{W}$ without affecting the rank of $T_{\mathcal{W},\mathcal{U}}(z;G)$.

Proof of Lemma 9.4. By Lemma 9.3, we can assume without loss of generality that the nodes in $\mathcal{W}$ have no outgoing edges. Let $G(z) \in \mathcal{A}(\mathcal{G})$. In what follows we omit the dependence of G on z and the dependence of $T(z;G)$ on both z and G. Consider a vertex $v \in \mathcal{U}$. Note that

$$(I - G)T = I \tag{9.7a}$$

$$\sum_{l=1}^{n} (I - G)_{kl} T_{lv} = 0, \tag{9.7b}$$

where $n := |\mathcal{V}|$ and (9.7b) follows from the fact that $k \in \mathcal{W} \setminus \mathcal{U}$ and $v \in \mathcal{U}$ are *distinct*. Equation (9.7b) implies that

$$T_{kv} = \sum_{l \in \mathcal{N}_k^-} G_{kl} T_{lv}. \tag{9.8}$$

Note that $j \in \mathcal{N}_k^-$, but possibly $\mathcal{N}_k^-$ contains other vertices. We will now prove that for all these other vertices, the corresponding transfer function T_{lv} equals zero. That is, $T_{lv} = 0$ for all $l \in \mathcal{N}_k^- \setminus \{j\}$. To see this, we first observe that there does not exist a path in $\mathcal{G}$ from v to $l \in \mathcal{N}_k^- \setminus \{j\}$. Indeed, suppose that there is a path $\mathcal{P}$ from v to l. Then this path cannot contain the edge (j,k), since node $k \in \mathcal{W} \setminus \mathcal{U}$ does not have any outgoing edges. This implies that there exists a path $\mathcal{P} \cup (l,k)$ from v to k via node l. This is a contradiction since by hypothesis j is the only in-neighbour of k that is reachable from $\mathcal{U}$. Therefore, we conclude that there does not exist a path from v to l. By Lemma 3 of [214] we conclude that $T_{lv} = 0$. This means that (9.8) can be simplified as

$$T_{kv} = G_{kj} T_{jv}.$$

Since $v \in \mathcal{U}$ is arbitrary, it follows that

$$T_{k,\mathcal{U}} = G_{kj} T_{j,\mathcal{U}}.$$

As $G_{kj} \neq 0$, we conclude that

$$\operatorname{rank} T_{\mathcal{W},\mathcal{U}} = \operatorname{rank} T_{\bar{\mathcal{W}},\mathcal{U}},$$

where $\bar{\mathcal{W}} := (\mathcal{W} \setminus \{k\}) \cup \{j\}$. This proves the lemma. □

From Lemma 9.3 and Lemma 9.4, we see that (i) we can always remove the outgoing edges of nodes in $\mathcal{W}$ and (ii) we can move nodes in $\mathcal{W}$ closer to $\mathcal{U}$ under suitable conditions. Of course, since both operations do not affect left-invertibility of $T_{\mathcal{W},\mathcal{U}}$, we can also apply these operations multiple times consecutively. Therefore, we introduce the following process to simplify the graph $\mathcal{G}$ and move the nodes in $\mathcal{W}$. The idea of this process is to apply the above operations to the graph until no more changes are possible.

Graph simplification process:
Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph and let $\mathcal{U}, \mathcal{W} \subseteq \mathcal{V}$. Consider the following two operations on the graph $\mathcal{G}$ and nodes $\mathcal{W}$.

1. Remove all outgoing edges of nodes in $\mathcal{W}$ from $\mathcal{G}$.
2. If $k \in \mathcal{W} \setminus \mathcal{U}$ has exactly one in-neighbour $j \in \mathcal{N}_k^-$ that is reachable from $\mathcal{U}$, replace k by j in $\mathcal{W}$.

Consecutively apply operations 1 and 2 on the graph $\mathcal{G}$ and nodes $\mathcal{W}$ until no more changes are possible.

Clearly, the graph simplification process terminates after a finite number of applications of operations 1 and 2. Indeed, operation 1 can only be applied once in a row, and a node in $\mathcal{W} \setminus \mathcal{U}$ can be “moved" at most $|\mathcal{V}| - 1$ times which means that operation 2 can be applied only a finite number of times. In fact, it is attractive to apply the operations 1 and 2 in alternating fashion since the process will then terminate within $|\mathcal{V}|$ operations of both types. This is due to the fact that if the outgoing edges of a node $j \in \mathcal{V}$ are removed, then we cannot apply operation 2 to replace a node k by j. A graph obtained by applying the graph simplification process to $\mathcal{G}$ is called a *derived graph*, which we denote by $\mathcal{D}(\mathcal{G})$. Similarly, we call a vertex set obtained by applying the graph simplification process to $\mathcal{W}$ a *derived vertex set*, denoted by $\mathcal{D}(\mathcal{W})$. To stress the fact that $\mathcal{D}(\mathcal{G})$ and $\mathcal{D}(\mathcal{W})$ do not only depend on the graph $\mathcal{G}$ and set $\mathcal{W}$, but *also* on the set $\mathcal{U}$, we say that $\mathcal{D}(\mathcal{G})$ and $\mathcal{D}(\mathcal{W})$ are a derived graph of $\mathcal{G}$ and derived vertex set of $\mathcal{W}$ *with respect to* the set $\mathcal{U}$. We emphasize that derived graphs and derived vertex sets are not necessarily unique. In general, the derived graph and derived vertex set that are obtained from the graph simplification process depend on the *order* in which the operations 1 and 2 are applied, and on the order in which operation 2 is applied to the nodes in $\mathcal{W}$. However, it turns out that the non-uniqueness of derived graphs and derived vertex sets is not a problem for the application (left-invertibility) we have in mind. In fact, we will show in Theorem 9.1 that *any* derived graph and derived vertex set will lead to the same conclusions about left-invertibility.

Remark 9.3. In step 2 of the graph simplification process, we have to decide whether there exists a node $k \in \mathcal{W} \setminus \mathcal{U}$ that has exactly one in-neighbour $j \in \mathcal{N}_k^-$ which is reachable from $\mathcal{U}$. Therefore, we want to find which in-neighbours of k are reachable from $\mathcal{U}$. One of the ways to do this, is to use Dijkstra’s single source shortest path (SSSP) algorithm [51], [160]. This algorithm computes the shortest paths (i.e., paths of minimum length) from a given source node s to every other node in the graph, and returns an “infinite" distance for each node which is not reachable from s. If we apply the SSSP algorithm to each node in $\mathcal{U}$, we obtain all nodes in $\mathcal{V}$ that are reachable from $\mathcal{U}$. Dijkstra’s SSSP algorithm has time complexity $\mathcal{O}(n + e)$, where $n = |\mathcal{V}|$ and $e = |\mathcal{E}|$ [160], and therefore we can find all nodes reachable from $\mathcal{U}$ in time complexity $\mathcal{O}(un + ue)$, where $u = |\mathcal{U}|$. Once we know the nodes in $\mathcal{V}$ that are reachable from $\mathcal{U}$, we can simply check whether there exists exactly one $j \in \mathcal{N}_k^-$ that is reachable from $\mathcal{U}$. In particular, this shows that the graph simplification process can be implemented in polynomial time since both operations 1 and 2 can be implemented in polynomial time, and the graph simplification process executes at most n operations of type 1 and 2 (if applied in this order).

Example 9.2. Consider the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in Figure 9.2 and define $\mathcal{U} := \{2\}$ and $\mathcal{W} := \{5, 6\}$. The goal of this example is to apply the graph simplification process to obtain a derived graph and derived vertex set. After this simplification, it will be easy to check left-invertibility of $T_{\mathcal{W},\mathcal{U}}(z; G)$.

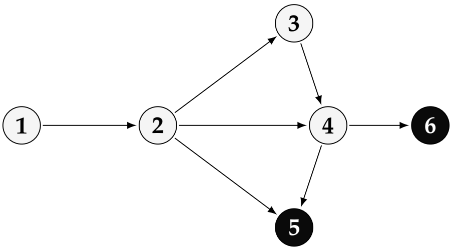

Figure 9.2: Graph $\mathcal{G}$ with nodes $\mathcal{W}$ colored black.

First, note that both nodes 5 and 6 do not have outgoing edges, so at the moment we cannot apply operation 1. However, we observe that node 6 has exactly one in-neighbour (node 4) that is reachable from $\mathcal{U}$. Consequently, we can replace node 6 by node 4 in $\mathcal{W}$ (see Figure 9.3).

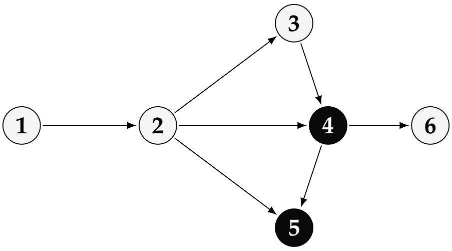

Figure 9.3: Graph with nodes $\mathcal{W}$, obtained by applying operation 2 to node 6.

To follow up, we see that node 4 has outgoing edges, which we can remove by applying operation 1, see Figure 9.4.

Subsequently, node 5 has exactly one in-neighbour that is (trivially) reachable from $\mathcal{U}$. Therefore, we replace vertex 5 by 2 in $\mathcal{W}$. Next, we can remove all outgoing edges of node 2 using operation 1. These result of these two operations is depicted in Figure 9.5.

Note that nodes 2 and 4 do not have any outgoing edges. Moreover, the in-neighbour 3 of node 4 is not reachable from node 2, so we cannot use operation 2 to node 4. In addition, operation 2 cannot be applied to node 2 since 2 $\in$

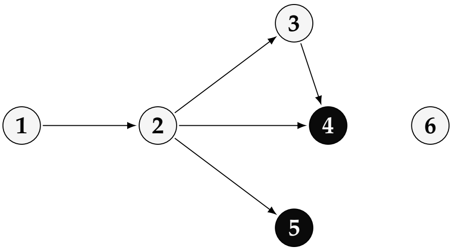

Figure 9.4: Graph with nodes $\mathcal{W}$, obtained by applying operation 1 to node 4.

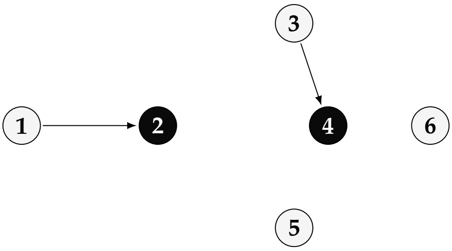

Figure 9.5: Derived graph $\mathcal{D}(\mathcal{G})$ with derived vertex set $\mathcal{D}(\mathcal{W})$ (in black), obtained by applying operation 2 to node 5 and operation 1 to node 2.

$\mathcal{U}$. Therefore, the graph simplification process terminates. We conclude that the graph $\mathcal{D}(\mathcal{G})$ in Figure 9.5 is a *derived graph* of $\mathcal{G}$, whereas the vertex set $\mathcal{D}(\mathcal{W}) = \{2, 4\}$ is a *derived vertex set* of $\mathcal{W}$ (with respect to $\mathcal{U}$). This example shows the strength of the graph simplification process in the following way: since $\mathcal{U} \subseteq \mathcal{D}(\mathcal{W})$, we conclude by Lemma 9.2 that $T_{\mathcal{D}(\mathcal{W}),\mathcal{U}}(z; G)$ is left-invertible for all $G(z) \in \mathcal{A}(\mathcal{D}(\mathcal{G}))$. However, by Lemma 9.3 and Lemma 9.4, we immediately see that $T_{\mathcal{W},\mathcal{U}}(z; G)$ is left-invertible for all $G(z) \in \mathcal{A}(\mathcal{G})$. This suggests that the graph simplification process is a promising tool to study left-invertibility of transfer matrices (and hence, to study identifiability of dynamical networks).

To summarize, we have seen that it is possible to remove the outgoing edges of nodes in $\mathcal{W}$ and to move the nodes in $\mathcal{W}$ closer to $\mathcal{U}$ if certain conditions are satisfied. Since left-invertibility is preserved by both operations due to Lemmas 9.3 and 9.4, we see that left-invertibility of $T_{\mathcal{W},\mathcal{U}}(z; G)$ for all $G(z) \in \mathcal{A}(\mathcal{G})$ is equivalent to the left-invertibility of $T_{\mathcal{D}(\mathcal{W}),\mathcal{U}}(z; G)$ for all $G(z) \in \mathcal{A}(\mathcal{D}(\mathcal{G}))$. Using Lemma 9.2, this shows that the condition $\mathcal{U} \subseteq \mathcal{D}(\mathcal{W})$ is *sufficient* for the left-invertibility of $T_{\mathcal{W},\mathcal{U}}(z; G)$. Remarkably, the condition $\mathcal{U} \subseteq \mathcal{D}(\mathcal{W})$ turns out to be also *necessary* for left-invertibility of $T_{\mathcal{W},\mathcal{U}}(z; G)$. This is stated more formally in the following theorem, which is one of the main results of this chapter.

Theorem 9.1. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let $\mathcal{U}, \mathcal{W} \subseteq \mathcal{V}$. Let $\mathcal{D}(\mathcal{W})$ be any derived vertex set of $\mathcal{W}$ with respect to $\mathcal{U}$. Then $\operatorname{rank} T_{\mathcal{W},\mathcal{U}}(z; G) = |\mathcal{U}|$ for all matrices $G(z) \in \mathcal{A}(\mathcal{G})$ if and only if $\mathcal{U} \subseteq \mathcal{D}(\mathcal{W})$.

Before we prove Theorem 9.1, we need some auxiliary results. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, let $n = |\mathcal{V}|$, $s = |\mathcal{E}|$, and index the edges as $\mathcal{E} = \{e_1, e_2, \dots, e_s\}$. We associate with each edge $e \in \mathcal{E}$ an indeterminate g_e. Moreover, we define the s-dimensional vector

$$\mathsf{g} := \begin{bmatrix} \mathsf{g}_{e_1} & \mathsf{g}_{e_2} & \cdots & \mathsf{g}_{e_s} \end{bmatrix}^\top,$$

which we call the *indeterminate vector* of $\mathcal{G}$. Next, we define the $n \times n$ matrix G as

$$\mathsf{G}_{ji} = \begin{cases} \mathsf{g}_{e_k} & \text{if } e_k = (i,j) \text{ for some } k \\ 0 & \text{otherwise.} \end{cases}$$

We emphasize that not all entries of G are indeterminates, but some are fixed zeros. Note that we write G in sans-serif font, to clearly distinguish between G and a *fixed* rational matrix $G(z)$. It is clear that the determinants of square submatrices of $I - \mathsf{G}$ are real polynomials in the indeterminate entries of G, i.e., in the indeterminate vector g. Hence, the entries of the adjugate of $I - \mathsf{G}$ are real polynomials in g. We state the following basic lemma, which gives conditions under which an entry of $\operatorname{adj}(I - \mathsf{G})$ is a *nonzero* polynomial.

Lemma 9.5. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let $i, j \in \mathcal{V}$. Let g and G be the indeterminate vector and matrix of $\mathcal{G}$, respectively, and define $\mathsf{A} := \operatorname{adj}(I - \mathsf{G})$. Then A_{ji} is a nonzero polynomial in g if and only if there exists a path from i to j.

Lemma 9.5 follows from Proposition 5.1 of [81]. Next, we state the following basic result on polynomials.

Proposition 9.1. Consider k nonzero real polynomials $p_i(x)$, where $i = 1, 2, ..., k$ and $x = (x_1, x_2, \dots, x_n)$. There exists an $\bar{x} \in \mathbb{R}^n$ such that $p_i(\bar{x}) \neq 0$ for all $i = 1, 2, ..., k$.

Remark 9.4. Without loss of generality, we can assume that $\bar{x}$ in Proposition 9.1 has only nonzero coordinates. Indeed, by continuity, if $p_i(\bar{x}) \neq 0$ for $i = 1, 2, ..., k$, there exists an open ball $B(\bar{x})$ around $\bar{x}$ in which $p_i(x) \neq 0$ for all $i = 1, 2, ..., k$ and all $x \in B(\bar{x})$. Clearly, this open ball contains a point with only nonzero coordinates.

Finally, we require a proposition on rational matrices.

Proposition 9.2. Let $A(z)$ be an $m \times n$ rational matrix and assume that each row of $A(z)$ contains at least one nonzero entry. There exists a vector $b \in \mathbb{R}^n$ such that each entry of $A(z)b$ is a nonzero rational function.

The proof of Proposition 9.2 follows simply from induction on the number of rows of $A(z)$ and is therefore omitted. With these results in place, we are ready to prove Theorem 9.1.

Proof of Theorem 9.1. Let $\mathcal{D}(\mathcal{G})$ and $\mathcal{D}(\mathcal{W})$ be a derived graph and derived vertex set with respect to $\mathcal{U}$ obtained from the graph simplification process. To prove the "if" statement, suppose that $\mathcal{U} \subseteq \mathcal{D}(\mathcal{W})$. By Corollary 9.2 we find that $\operatorname{rank} T_{\mathcal{D}(\mathcal{W}),\mathcal{U}}(z;G) = |\mathcal{U}|$ for all $G(z) \in \mathcal{A}(\mathcal{D}(\mathcal{G}))$. By consecutive application of Lemmas 9.3 and 9.4, we conclude that $\operatorname{rank} T_{\mathcal{W},\mathcal{U}}(z;G) = |\mathcal{U}|$ for all $G(z) \in \mathcal{A}(\mathcal{G})$.

Conversely, to prove the "only if" statement, suppose that $\mathcal{U} \not\subseteq \mathcal{D}(\mathcal{W})$. We want to show that

$$\operatorname{rank} T_{\mathcal{D}(\mathcal{W}),\mathcal{U}}(z;G) < |\mathcal{U}| \text{ for some } G(z) \in \mathcal{A}(\mathcal{D}(\mathcal{G})).$$

Since $\mathcal{U} \not\subseteq \mathcal{D}(\mathcal{W})$, the set $\bar{\mathcal{U}} := \mathcal{U} \setminus \mathcal{D}(\mathcal{W})$ is nonempty. Furthermore, as $\mathcal{D}(\mathcal{G})$ and $\mathcal{D}(\mathcal{W})$ result from the graph simplification process, it is clear that nodes in $\mathcal{D}(\mathcal{W})$ do not have outgoing edges. In addition, each node in the set $\bar{\mathcal{W}} := \mathcal{D}(\mathcal{W}) \setminus \mathcal{U}$ has either *zero* or *at least two* in-neighbours that are reachable from $\mathcal{U}$. As nodes in $\mathcal{D}(\mathcal{W}) \cap \mathcal{U}$ have no outgoing edges, this means that each node in $\bar{\mathcal{W}}$ has either zero or at least two in-neighbours that are reachable *from* $\bar{\mathcal{U}}$. Finally, we assume that the nodes in $\mathcal{U}$ do not have any incoming edges, which is without loss of generality by Remark 9.1.

The idea of the proof is to show that $T_{\mathcal{D}(\mathcal{W}),\bar{\mathcal{U}}}(z;G)b = 0$, for some to-be-determined network matrix $G(z) \in \mathcal{A}(\mathcal{D}(\mathcal{G}))$ and nonzero vector b. Hence, $\operatorname{rank} T_{\mathcal{D}(\mathcal{W}),\bar{\mathcal{U}}}(z;G) < |\bar{\mathcal{U}}|$ and since $T_{\mathcal{D}(\mathcal{W}),\bar{\mathcal{U}}}$ is a submatrix of $T_{\mathcal{D}(\mathcal{W}),\mathcal{U}}$, it will then immediately follow that $\operatorname{rank} T_{\mathcal{D}(\mathcal{W}),\mathcal{U}}(z;G) < |\mathcal{U}|$.

We investigate a row $T_{w,\bar{\mathcal{U}}}(z;G)$ of the transfer matrix $T_{\mathcal{D}(\mathcal{W}),\bar{\mathcal{U}}}(z;G)$ and we distinguish two cases, namely the case that $w \in \mathcal{D}(\mathcal{W}) \cap \mathcal{U}$ and the case that $w \in \bar{\mathcal{W}}$. First, suppose that $w \in \mathcal{D}(\mathcal{W}) \cap \mathcal{U}$. This implies that $w \in \mathcal{U}$. Recall that the nodes in $\mathcal{U}$ do not have any incoming edges. Consequently, there are no paths from v to w for any $v \in \bar{\mathcal{U}}$. We conclude from Lemma 3 of [214] that $T_{wv}(z;G) = 0$ for all $G(z) \in \mathcal{A}(\mathcal{D}(\mathcal{G}))$. Therefore, $T_{w,\bar{\mathcal{U}}}(z;G) = 0$ for all $G(z) \in \mathcal{A}(\mathcal{D}(\mathcal{G}))$. Obviously, this implies that $T_{w,\bar{\mathcal{U}}}(z;G)b = 0$ for all $G(z) \in \mathcal{A}(\mathcal{D}(\mathcal{G}))$ and all real vectors b.

Next, we consider the second case in which $w \in \bar{\mathcal{W}}$. Let G denote the indeterminate matrix of $\mathcal{D}(\mathcal{G})$. In addition, define $\mathsf{A} := \operatorname{adj}(I - \mathsf{G})$. Then, we have

$$(I - \mathsf{G})\mathsf{A} = \det(I - \mathsf{G})I \tag{9.9a}$$

$$(I - \mathsf{G})_{\bar{\mathcal{W}},\mathcal{V}}\mathsf{A}_{\mathcal{V},\bar{\mathcal{U}}} = 0, \tag{9.9b}$$

where (9.9b) follows from the fact that $\bar{\mathcal{U}}$ and $\bar{\mathcal{W}}$ are disjoint. Recall that nodes in $\bar{\mathcal{W}}$ do not have any outgoing edges, and therefore $(I - \mathsf{G})_{\bar{\mathcal{W}},\bar{\mathcal{W}}} = I$. This means that we can rewrite (9.9b) as

$$\mathsf{A}_{\bar{\mathcal{W}},\bar{\mathcal{U}}} = \mathsf{G}_{\bar{\mathcal{W}},\bar{\mathcal{W}}^c}\mathsf{A}_{\bar{\mathcal{W}}^c,\bar{\mathcal{U}}}, \tag{9.10}$$

where we recall that $\bar{\mathcal{W}}^c := \mathcal{V} \setminus \bar{\mathcal{W}}$. Note that for $j \in \bar{\mathcal{W}}^c$, the column $\mathsf{G}_{\bar{\mathcal{W}},j}$ is equal to 0 if j is not an in-neighbour of any node in $\bar{\mathcal{W}}$. In addition, for any $j \in \bar{\mathcal{W}}^c$, the row $\mathsf{A}_{j,\bar{\mathcal{U}}}$ equals 0 if there is no path from $\bar{\mathcal{U}}$ to j (by Lemma 9.5). Therefore, we can rewrite (9.10) as

$$\mathsf{A}_{\bar{\mathcal{W}},\bar{\mathcal{U}}} = \mathsf{G}_{\bar{\mathcal{W}},\mathcal{N}}\mathsf{A}_{\mathcal{N},\bar{\mathcal{U}}}, \tag{9.11}$$

where $\mathcal{N} \subseteq \bar{\mathcal{W}}^c$ is characterized by the following property: we have $j \in \mathcal{N}$ if and only if j is an in-neighbour of a node in $\bar{\mathcal{W}}$ and there is a path from $\bar{\mathcal{U}}$ to j. By definition of the adjugate, the entries of $\mathsf{A}_{\mathcal{N},\bar{\mathcal{U}}}$ are polynomials in the indeterminate entries of G. We claim that the indeterminate entries of $\mathsf{G}_{\bar{\mathcal{W}},\mathcal{N}}$ do not appear in any entry of $\mathsf{A}_{\mathcal{N},\bar{\mathcal{U}}}$, that is, $\mathsf{A}_{\mathcal{N},\bar{\mathcal{U}}}$ is *independent* of the indeterminate entries of $\mathsf{G}_{\bar{\mathcal{W}},\mathcal{N}}$. For the sake of clarity, we postpone the proof of this claim to the end. For now, we assume that $\mathsf{A}_{\mathcal{N},\bar{\mathcal{U}}}$ is independent of the indeterminate entries of $\mathsf{G}_{\bar{\mathcal{W}},\mathcal{N}}$.

By definition, there is a path from $\bar{\mathcal{U}}$ to each node in $\mathcal{N}$. Let $\mathcal{N} = \{n_1, n_2, \dots, n_r\}$, where $r = |\mathcal{N}|$. Then, for each node $n_i \in \mathcal{N}$, there exists a node $u_i \in \bar{\mathcal{U}}$ such that A_{n_i,u_i} is a nonzero polynomial in the indeterminate entries of G (by Lemma 9.5). We emphasize that u_i and u_j are not necessarily distinct. We focus on the r nonzero polynomials

$$\mathsf{A}_{n_1,u_1}, \mathsf{A}_{n_2,u_2}, \dots, \mathsf{A}_{n_r,u_r}. \tag{9.12}$$

The idea is to apply Proposition 9.1 and Remark 9.4 to these r polynomials. By Remark 9.4, we can substitute nonzero real numbers for the indeterminate entries of G such that all r polynomials (9.12) evaluate to nonzero real numbers. Since the polynomials (9.12) are independent of the indeterminate entries of $\mathsf{G}_{\bar{\mathcal{W}},\mathcal{N}}$, we do not have to fix the entries of $\mathsf{G}_{\bar{\mathcal{W}},\mathcal{N}}$. In addition, it is possible to substitute *strictly proper functions* in z for the indeterminate entries of G (except for entries of $\mathsf{G}_{\bar{\mathcal{W}},\mathcal{N}}$) such that the polynomials (9.12) evaluate to *nonzero* rational functions. Indeed, one can simply choose all indeterminate entries of G as nonzero real numbers as before, and then divide all of these real numbers by z.

To summarize the progress so far, we have substituted strictly proper functions for the indeterminate entries of G (except for the entries of $\mathsf{G}_{\bar{\mathcal{W}},\mathcal{N}}$) such that the polynomials (9.12) evaluate to *nonzero* rational functions. Note that this implies that the matrix $\mathsf{A}_{\mathcal{N},\bar{\mathcal{U}}}$ evaluates to a rational matrix, which we denote by $A_{\mathcal{N},\bar{\mathcal{U}}}(z)$ from now on. Since each row of $A_{\mathcal{N},\bar{\mathcal{U}}}(z)$ contains a nonzero rational function, by Proposition 9.2 there exists a nonzero real vector b such that $A_{\mathcal{N},\bar{\mathcal{U}}}(z)b$ has only *nonzero* rational entries.

Subsequently, we will choose the indeterminate entries of $\mathsf{G}_{\bar{\mathcal{W}},\mathcal{N}}$ such that $\mathsf{G}_{\bar{\mathcal{W}},\mathcal{N}} A_{\mathcal{N},\bar{\mathcal{U}}}(z)b = 0$. Recall that the nodes in $\bar{\mathcal{W}}$ either have zero or at least two in-neighbours from the set $\mathcal{N}$. If a node $w \in \bar{\mathcal{W}}$ has no in-neighbours, then $\mathsf{G}_{w,\mathcal{N}} = 0$, and therefore clearly $\mathsf{G}_{w,\mathcal{N}} A_{\mathcal{N},\bar{\mathcal{U}}}(z)b = 0$. If a node $w \in \bar{\mathcal{W}}$ has at least two in-neighbours, say $n_1, n_2, \dots, n_p \in \mathcal{N}$, then we substitute *strictly proper* functions for the indeterminate entries $\mathsf{G}_{w,n_1}, \mathsf{G}_{w,n_2}, \dots, \mathsf{G}_{w,n_p}$ so that $\mathsf{G}_{w,\mathcal{N}} A_{\mathcal{N},\bar{\mathcal{U}}}(z)b = 0$. Note that this is possible since the vector $A_{\mathcal{N},\bar{\mathcal{U}}}(z)b$ has only *nonzero* rational entries. To conclude, we have substituted strictly proper functions for the indeterminate entries of G which yields a matrix which we denote by $G(z)$. The adjugate of $I - G(z)$ is denoted by $A(z) = \operatorname{adj}(I - G(z))$. We have shown that $G_{\bar{\mathcal{W}},\mathcal{N}}(z) A_{\mathcal{N},\bar{\mathcal{U}}}(z)b = 0$. By (9.11), this yields $A_{\bar{\mathcal{W}},\bar{\mathcal{U}}}(z)b = 0$. Note that $\det(I - G(z))$ is nonzero since all nonzero entries of G are strictly proper functions. Therefore,

$$T(z; G) = \frac{1}{\det(I - G(z))} A(z),$$

from which we find that $T_{\bar{\mathcal{W}},\bar{\mathcal{U}}}(z;G)b = 0$. Consequently, $T_{\mathcal{D}(\mathcal{W}),\bar{\mathcal{U}}}(z;G)b = 0$, and $\operatorname{rank} T_{\mathcal{D}(\mathcal{W}),\bar{\mathcal{U}}}(z;G) < |\bar{\mathcal{U}}|$. Therefore, we conclude that $\operatorname{rank} T_{\mathcal{D}(\mathcal{W}),\mathcal{U}}(z;G) < |\mathcal{U}|$. We still have to show that $G(z)$ is admissible, i.e., $G(z) \in \mathcal{A}(\mathcal{D}(\mathcal{G}))$. Since the indeterminate matrix G is consistent with the graph $\mathcal{D}(\mathcal{G})$ and we substituted (nonzero) strictly proper functions for each indeterminate entry of G, the matrix $G(z)$ readily satisfies Properties P1 and P2. In addition, since all nonzero entries of $G(z)$ are strictly proper, we obtain $\lim_{z\to\infty} I - G(z) = I$, and hence, $G(z)$ also satisfies Property P3. We conclude that $\operatorname{rank} T_{\mathcal{D}(\mathcal{W}),\mathcal{U}}(z;G) < |\mathcal{U}|$ for some $G(z) \in \mathcal{A}(\mathcal{D}(\mathcal{G}))$. Finally, by consecutive application of Lemmas 9.3 and 9.4, we conclude that $\operatorname{rank} T_{\mathcal{W},\mathcal{U}}(z;G) < |\mathcal{U}|$ for some $G(z) \in \mathcal{A}(\mathcal{G})$.

Recall that we have so far assumed that $\mathsf{A}_{\mathcal{N},\bar{\mathcal{U}}}$ is *independent* of the indeterminate entries of $\mathsf{G}_{\bar{\mathcal{W}},\mathcal{N}}$. It remains to be shown that this is true. To this end, label the nodes in $\mathcal{V}$ such that G can be written as

$$\mathsf{G} = \begin{bmatrix} \mathsf{G}_{\bar{\mathcal{W}}^c,\bar{\mathcal{W}}^c} & \mathsf{G}_{\bar{\mathcal{W}}^c,\bar{\mathcal{W}}} \\ \mathsf{G}_{\bar{\mathcal{W}},\bar{\mathcal{W}}^c} & \mathsf{G}_{\bar{\mathcal{W}},\bar{\mathcal{W}}} \end{bmatrix} \tag{9.13a}$$

$$= \begin{bmatrix} \mathsf{G}_{\bar{\mathcal{W}}^c,\bar{\mathcal{W}}^c} & 0 \\ \mathsf{G}_{\bar{\mathcal{W}},\bar{\mathcal{W}}^c} & 0 \end{bmatrix}, \tag{9.13b}$$

where (9.13b) follows from the fact that nodes in $\bar{\mathcal{W}}$ have no outgoing edges. This implies that

$$I - \mathsf{G} = \begin{bmatrix} I - \mathsf{G}_{\bar{\mathcal{W}}^c,\bar{\mathcal{W}}^c} & 0 \\ -\mathsf{G}_{\bar{\mathcal{W}},\bar{\mathcal{W}}^c} & I \end{bmatrix},$$

and therefore

$$\mathsf{A} = \operatorname{adj}(I - \mathsf{G}) = \begin{bmatrix} \operatorname{adj}(I - \mathsf{G}_{\bar{\mathcal{W}}^c,\bar{\mathcal{W}}^c}) & 0 \\ * & * \end{bmatrix}. \tag{9.14}$$

Since the entries of $\mathsf{G}_{\bar{\mathcal{W}}^c,\bar{\mathcal{W}}^c}$ are independent of the indeterminate entries of $\mathsf{G}_{\bar{\mathcal{W}},\bar{\mathcal{W}}^c}$, we conclude from (9.14) that the matrix $\mathsf{A}_{\bar{\mathcal{W}}^c,\bar{\mathcal{W}}^c} = \operatorname{adj}(I - \mathsf{G}_{\bar{\mathcal{W}}^c,\bar{\mathcal{W}}^c})$ is independent of the indeterminate entries of $\mathsf{G}_{\bar{\mathcal{W}},\bar{\mathcal{W}}^c}$. Now, to prove the claim, note that $\bar{\mathcal{U}}$ and $\bar{\mathcal{W}}$ are disjoint by definition, and therefore $\bar{\mathcal{U}} \subseteq \bar{\mathcal{W}}^c$. In addition, we have $\mathcal{N} \subseteq \bar{\mathcal{W}}^c$. Therefore, the matrix $\mathsf{A}_{\mathcal{N},\bar{\mathcal{U}}}$ is a *submatrix* of $\mathsf{A}_{\bar{\mathcal{W}}^c,\bar{\mathcal{W}}^c}$. Furthermore, we see that $\mathsf{G}_{\bar{\mathcal{W}},\mathcal{N}}$ is a submatrix of $\mathsf{G}_{\bar{\mathcal{W}},\bar{\mathcal{W}}^c}$ by using the fact that $\mathcal{N} \subseteq \bar{\mathcal{W}}^c$. We conclude that the entries of $\mathsf{A}_{\mathcal{N},\bar{\mathcal{U}}}$ are independent of the indeterminate entries of $\mathsf{G}_{\bar{\mathcal{W}},\mathcal{N}}$. This proves the theorem. □

9.6 IDENTIFIABILITY AND GRAPH SIMPLIFICATION

In this section we use Theorem 9.1 to provide solutions to the identifiability problems introduced in Section 9.3. Specifically, the following theorem follows from Theorem 9.1 and Lemma 9.1 and states necessary and sufficient graph-theoretic conditions for identifiability of $(i, \mathcal{N}_i^+)$.

Theorem 9.2. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, let $i \in \mathcal{V}$ and $\mathcal{C} \subseteq \mathcal{V}$. Moreover, let $\mathcal{D}(\mathcal{C})$ be any derived vertex set of $\mathcal{C}$ with respect to $\mathcal{N}_i^+$. Then $(i, \mathcal{N}_i^+)$ is identifiable from $\mathcal{C}$ in $\mathcal{G}$ if and only if $\mathcal{N}_i^+ \subseteq \mathcal{D}(\mathcal{C})$.

Example 9.3. Consider the graph in Figure 9.2. We wonder whether $(1, \mathcal{N}_1^+)$ is identifiable. Note that we have $\mathcal{N}_1^+ = \{2\}$. The set of measured nodes is $\mathcal{C} = \{5, 6\}$. As shown in Example 9.2, a derived vertex set of $\mathcal{C}$ with respect to $\mathcal{N}_1^+$ is given by $\mathcal{D}(\mathcal{C}) = \{2, 4\}$. Since $\{2\} \subseteq \mathcal{D}(\mathcal{C})$, we conclude by Theorem 9.2 that $(1, \mathcal{N}_1^+)$ is identifiable. In other words, we can uniquely reconstruct $G_{21}(z)$ from the transfer matrix $CT(z; G)$. This approach shows the strength of our approach. Indeed, note that to check identifiability, we do not have to verify Definition 9.1 directly. Also, we do not have to compute $CT(z; G) = C(I - G(z))^{-1}$ and verify its rank for all $G(z) \in \mathcal{A}(\mathcal{G})$, which is required to check the condition of Lemma 9.1.

By definition of the graph simplification process, we have that $|\mathcal{D}(\mathcal{C})| \leqslant |\mathcal{C}|$. Hence, it follows from Theorem 9.2 that identifiability of $(i, \mathcal{N}_i^+)$ implies that the number of measured nodes is greater or equal to the number of out-neighbours of node i.

Corollary 9.2. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, let $i \in \mathcal{V}$ and $\mathcal{C} \subseteq \mathcal{V}$. If $(i, \mathcal{N}_i^+)$ is identifiable from $\mathcal{C}$ in $\mathcal{G}$ then $|\mathcal{N}_i^+| \leqslant |\mathcal{C}|$.

The next result gives necessary and sufficient graph-theoretic conditions under which the entire graph $\mathcal{G}$ is identifiable. This result is a corollary of Theorem 9.2 but is stated as a theorem due to its importance.

Theorem 9.3. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let $\mathcal{C} \subseteq \mathcal{V}$. Then $\mathcal{G}$ is identifiable from $\mathcal{C}$ if and only if for all $i \in \mathcal{V}$, we have $\mathcal{N}_i^+ \subseteq \mathcal{D}(\mathcal{C})$, where $\mathcal{D}(\mathcal{C})$ is any derived vertex set of $\mathcal{C}$ with respect to $\mathcal{N}_i^+$.

We emphasize that the derived set $\mathcal{D}(\mathcal{C})$ of $\mathcal{C}$ depends on the choice of neighbour set $\mathcal{N}_i^+$, and hence, for each node $i \in \mathcal{V}$ we have to compute the derived set of $\mathcal{C}$ with respect to $\mathcal{N}_i^+$.

9.7 CONSTRAINED VERTEX-DISJOINT PATHS

In the previous section we established necessary and sufficient graph-theoretic conditions for the identifiability of respectively $(i, \mathcal{N}_i^+)$ and $\mathcal{G}$. The purpose of the current section is to compare these results to the ones based on so-called *constrained vertex-disjoint paths* [223]. We first recall the definition in what follows.

Definition 9.3. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph. Consider a set of m vertex-disjoint paths in $\mathcal{G}$ with starting nodes $\bar{\mathcal{U}} \subseteq \mathcal{V}$ and end nodes $\bar{\mathcal{W}} \subseteq \mathcal{V}$. We say that the set of vertex-disjoint paths is *constrained* if it is the *only* set of m vertex-disjoint paths from $\bar{\mathcal{U}}$ to $\bar{\mathcal{W}}$.

Next, let $\mathcal{U}, \mathcal{W} \subseteq \mathcal{V}$ be disjoint subsets of vertices. We say that there exists a constrained set of m vertex-disjoint paths *from $\mathcal{U}$ to $\mathcal{W}$* if there exists a constrained set of m vertex-disjoint paths in $\mathcal{G}$ with starting nodes $\bar{\mathcal{U}} \subseteq \mathcal{U}$ and end nodes $\bar{\mathcal{W}} \subseteq \mathcal{W}$. In the case that $\mathcal{U} \cap \mathcal{W} \neq \varnothing$, we say that there is a constrained set of m vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ if there exists a constrained set of $\max\{0, m - |\mathcal{U} \cap \mathcal{W}|\}$ vertex-disjoint paths from $\mathcal{U} \setminus \mathcal{W}$ to $\mathcal{W} \setminus \mathcal{U}$.

Remark 9.5. Note that for a set of m vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ to be constrained, we do not require the existence of a unique set of m vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$. In fact, we only require the existence of a unique set of vertex-disjoint paths between the *starting nodes* $\bar{\mathcal{U}}$ of the paths and the *end nodes* $\bar{\mathcal{W}}$. We will illustrate the definition of constrained vertex-disjoint paths in Example 9.4.

Remark 9.6. The notion of constrained vertex-disjoint paths is strongly related to the notion of *constrained matchings* in bipartite graphs [82]. In fact, a constrained matching can be seen as a special case of a constrained set of vertex-disjoint paths where all paths are of length one.

Example 9.4. Consider the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in Figure 9.6. Moreover, consider the subsets of vertices $\mathcal{U} := \{2, 3\}$ and $\mathcal{W} := \{6, 7, 8\}$. Clearly, the paths $\{(2, 4), (4, 6)\}$ and $\{(3, 5), (5, 7)\}$ form a set of two vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$. In fact, this set of vertex-disjoint paths is *constrained* since there does not exist another set of two vertex-disjoint paths from $\bar{\mathcal{U}} = \{2, 3\}$ to $\bar{\mathcal{W}} = \{6, 7\}$. Therefore, there exists a constrained set of two vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$. Note that there are also other sets of vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$. For example, the paths $\{(2, 4), (4, 7)\}$ and $\{(3, 5), (5, 8)\}$ also form a set of two vertex-disjoint paths. However, this set of vertex-disjoint paths is *not* constrained. To see this, note that we have another set of vertex-disjoint paths from $\bar{\mathcal{U}} = \{2, 3\}$ to $\bar{\mathcal{W}} = \{7, 8\}$, namely the set consisting of the paths $\{(2, 4), (4, 8)\}$ and $\{(3, 5), (5, 7)\}$.

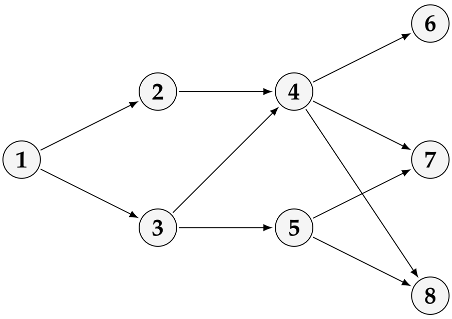

Figure 9.6: Graph used in Example 9.4.

In the following theorem, we recall the main result presented in [223], which relates the notion of constrained vertex-disjoint paths and identifiability of $(i, \mathcal{N}_i^+)$.

Theorem 9.4. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, let $i \in \mathcal{V}$ and $\mathcal{C} \subseteq \mathcal{V}$. If there exists a constrained set of $|\mathcal{N}_i^+|$ vertex-disjoint paths from $\mathcal{N}_i^+$ to $\mathcal{C}$ then $(i, \mathcal{N}_i^+)$ is identifiable from $\mathcal{C}$.

The proof of Theorem 9.4 can be found in [223] (see Theorem 13). A natural question to ask is whether the condition given in Theorem 9.4 is also *necessary* for identifiability. It turns out that this is not the case, as demonstrated next.

Example 9.5. In this example, we revisit the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in Figure 9.2. Suppose that we are interested in the identifiability of $(1, \mathcal{N}_1^+)$, i.e., in the identifiability of the transfer function corresponding to the edge $(1, 2)$. The set of measured nodes is given by $\mathcal{C} = \{5, 6\}$. The purpose of this example is to show that Theorem 9.4 is not necessary, i.e., we have to show that $(1, \mathcal{N}_1^+)$ is identifiable even though there does not exist a constrained set of one (vertex-disjoint) path from $\mathcal{N}_1^+$ to $\mathcal{C}$.

Note that $\mathcal{N}_1^+ = \{2\}$ and that there are three different paths from 2 to 5. In addition, there are two different paths from node 2 to node 6. This implies that there does not exist a constrained set of one (vertex-disjoint) path from $\mathcal{N}_1^+$ to $\mathcal{C}$. Nonetheless, we can show that $(1, \mathcal{N}_1^+)$ is identifiable. The easiest way to show this is by noting that we already proved in Example 9.2 that $\mathcal{N}_1^+ \subseteq \mathcal{D}(\mathcal{C})$, where $\mathcal{D}(\mathcal{C})$ is a derived vertex set of $\mathcal{C}$. Hence, by Theorem 9.2 we conclude that $(1, \mathcal{N}_1^+)$ is identifiable. However, to gain a bit more insight we will prove identifiability of $(1, \mathcal{N}_1^+)$ by inspection of the transfer matrix $T_{\mathcal{C},\mathcal{N}_1^+}(z; G)$. For any $G(z) \in \mathcal{A}(\mathcal{G})$, we obtain

$$T_{\mathcal{C},\mathcal{N}_1^+} = \begin{bmatrix} G_{52} + G_{54}(G_{42} + G_{43}G_{32}) \\ G_{64}(G_{42} + G_{43}G_{32}) \end{bmatrix}, \tag{9.15}$$

where we omitted the argument z. If $G_{42} + G_{43}G_{32} \neq 0$ then $G_{64}(G_{42} + G_{43}G_{32}) \neq 0$ and therefore $\operatorname{rank} T_{\mathcal{C},\mathcal{N}_1^+} = 1$. If $G_{42} + G_{43}G_{32} = 0$, we see that $G_{52} + G_{54}(G_{42} + G_{43}G_{32}) = G_{52} \neq 0$ so also in this case $\operatorname{rank} T_{\mathcal{C},\mathcal{N}_1^+} = 1$. We conclude that $\operatorname{rank} T_{\mathcal{C},\mathcal{N}_1^+} = 1$ for all admissible network matrices, which means that $(1, \mathcal{N}_1^+)$ is identifiable by Lemma 9.1.

Example 9.5 also gives some intuition for the fact that Theorem 9.4 is not necessary for identifiability. Indeed, the condition based on constrained vertex-disjoint paths guarantees that a *square* submatrix of $T_{\mathcal{C},\mathcal{N}_i^+}(z; G)$ is invertible *for all* admissible G, where the columns and rows of this submatrix are indexed by the starting nodes and end nodes of the paths, respectively [223]. However, as can be seen from (9.15), the matrix $T_{\mathcal{C},\mathcal{N}_i^+}(z; G)$ might be left-invertible for all admissible G, even though there *does not exist* a square $|\mathcal{N}_i^+| \times |\mathcal{N}_i^+|$ submatrix of $T_{\mathcal{C},\mathcal{N}_i^+}(z; G)$ that is invertible for all admissible G. In general, the particular square submatrix of $T_{\mathcal{C},\mathcal{N}_i^+}(z; G)$ that is invertible *depends* on the network matrix G. Interestingly, we can use the general theory developed in this chapter to show that the condition of Theorem 9.4 is necessary and sufficient in the special case that $T_{\mathcal{C},\mathcal{N}_i^+}(z; G)$ is *square itself* (a proof is given in Section 9.9.2).

Theorem 9.5. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Let $i \in \mathcal{V}$ and $\mathcal{C} \subseteq \mathcal{V}$ be such that $|\mathcal{C}| = |\mathcal{N}_i^+|$. Then, $(i, \mathcal{N}_i^+)$ is identifiable if and only if there exists a constrained set of $|\mathcal{N}_i^+|$ vertex-disjoint paths from $\mathcal{N}_i^+$ to $\mathcal{C}$.

The main message of this section is that the conditions in terms of constrained vertex-disjoint paths [223] are only necessary and sufficient in the special case that $|\mathcal{N}_i^+| = |\mathcal{C}|$. This case is quite particular, especially if one is interested in identifiability of the entire network. In the latter situation, Theorem 9.5 can only be applied if the number of out-neighbours of *each node* is equal to the number of measured nodes, which is very restrictive. Therefore, we conclude that the necessary and sufficient conditions for identifiability based on graph simplification are much more general. Additional advantages of the conditions based on the graph simplification process are that they are conceptually simpler, and appealing from computational point of view, cf. Remark 9.3.

9.8 CONCLUSIONS

In this chapter we have considered the problem of identifiability of dynamical networks for which interactions between nodes are modelled by transfer functions. We have been interested in graph-theoretic conditions for two identifiability problems. First, we wanted to find conditions under which the transfer functions of all outgoing edges of a given node are identifiable. Secondly, we have been interested in conditions under which all transfer functions in the network are identifiable. It is known that these problems are equivalent to the left-invertibility of certain transfer matrices *for all* networked matrices associated with the graph [81], [223]. However, the downside of such rank conditions is that it is not clear how to *check* the rank of a transfer matrix for an *infinite* number of network matrices.

Therefore, as our first contribution, we have provided a necessary and sufficient graph-theoretic condition under which a transfer matrix has full column rank *for all* network matrices. To this end, we have introduced a new concept called the *graph simplification process*. The idea of this process is to apply simplifying operations to the graph, after which left-invertibility can be verified by simply checking a set inclusion. Based on the graph simplification process, we have given necessary and sufficient conditions for identifiability. Notably, we have shown that our conditions can be verified by polynomial time algorithms. Finally, we have shown that our results generalize existing sufficient conditions based on constrained vertex-disjoint paths [223].

It is interesting to observe that our topological conditions for global identifiability are quite different from the path-based conditions for *generic* identifiability [81]. This is analogous to the *controllability* literature, where it was shown that weak structural controllability can be characterized in terms of maximal matchings [113], while strong structural controllability was characterized using a (different) graph-theoretic concept called zero forcing [142].

9.9 SOME PROOFS

9.9.1 Proof of Lemma 9.2

Proof of Lemma 9.2. Without loss of generality, we assume that the nodes in $\mathcal{W}$ do not have outgoing edges (see Lemma 9.3). Since $\mathcal{U} \subseteq \mathcal{W}$, the nodes in $\mathcal{U}$ do not have outgoing edges. We now relabel the nodes in $\mathcal{G}$ such that $G(z) \in \mathcal{A}(\mathcal{G})$ can be written as

$$G = \begin{bmatrix} G_{\mathcal{U},\mathcal{U}} & G_{\mathcal{U},\mathcal{U}^c} \\ G_{\mathcal{U}^c,\mathcal{U}} & G_{\mathcal{U}^c,\mathcal{U}^c} \end{bmatrix} = \begin{bmatrix} 0 & G_{\mathcal{U},\mathcal{U}^c} \\ 0 & G_{\mathcal{U}^c,\mathcal{U}^c} \end{bmatrix},$$

where we omitted the argument z, and where the zeros are present due to the fact that nodes in $\mathcal{U}$ do not have outgoing edges. Consequently, we obtain

$$T = (I - G)^{-1} = \begin{bmatrix} I & -G_{\mathcal{U},\mathcal{U}^c} \\ 0 & I - G_{\mathcal{U}^c,\mathcal{U}^c} \end{bmatrix}^{-1} = \begin{bmatrix} I & * \\ 0 & * \end{bmatrix},$$

and therefore, $T_{\mathcal{U},\mathcal{U}} = I$. Hence, $T_{\mathcal{U},\mathcal{U}}$ has full rank for all $G(z) \in \mathcal{A}(\mathcal{G})$ and we conclude that $T_{\mathcal{W},\mathcal{U}}$ has rank $|\mathcal{U}|$ for all $G(z) \in \mathcal{A}(\mathcal{G})$. □

9.9.2 Proof of Theorem 9.5

To prove Theorem 9.5, we will first state two lemmas. Under the assumption that $|\mathcal{U}| = |\mathcal{W}|$, the following lemma asserts that the existence of a set of constrained vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ is preserved by operation 1.

Lemma 9.6. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph and consider $\mathcal{U}, \mathcal{W} \subseteq \mathcal{V}$ such that $|\mathcal{U}| = |\mathcal{W}|$. Moreover, let $\bar{\mathcal{G}} = (\mathcal{V}, \bar{\mathcal{E}})$ be the graph obtained from $\mathcal{G}$ by removing all outgoing edges of the nodes in $\mathcal{W}$. There exists a constrained set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ in $\mathcal{G}$ if and only if there exists a constrained set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ in $\bar{\mathcal{G}}$.

Proof. The lemma follows from the following important observation: if $|\mathcal{U}| = |\mathcal{W}|$, then a set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ *does not contain* any outgoing edge of a node in $\mathcal{W}$. Indeed, if a path $\mathcal{P}$ from $\mathcal{U}$ to $\mathcal{W}$ in such a set of vertex-disjoint paths contains an edge (w, v), where $w \in \mathcal{W}$ and $v \in \mathcal{V}$, then the path $\mathcal{P}$ contains at least two vertices in $\mathcal{W}$ (namely w and the end node). This means that $\mathcal{P}$ is contained in a set of at most $|\mathcal{U}| - 1$ vertex disjoint paths from $\mathcal{U}$ to $\mathcal{W}$. However, this is a contradiction since we assumed that $\mathcal{P}$ was contained in a set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$.

Next, we prove the "if" statement. Suppose that there exists a constrained set $\mathcal{S}$ of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ in $\bar{\mathcal{G}}$. Then $\mathcal{S}$ is also a set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ in $\mathcal{G}$. We want to prove that $\mathcal{S}$ is constrained (in the graph $\mathcal{G}$). Therefore, suppose on the contrary that there exists another set $\bar{\mathcal{S}}$ of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ in $\mathcal{G}$. By the above discussion, we know that no path in $\bar{\mathcal{S}}$ contains an outgoing edge of a node in $\mathcal{W}$. Therefore, $\bar{\mathcal{S}}$ is a set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ in $\bar{\mathcal{G}}$. As such, we conclude that

$\bar{\mathcal{S}} = \mathcal{S}$. In other words, $\mathcal{S}$ is a constrained set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ in $\mathcal{G}$.

Conversely, to prove the "only if" statement, suppose that there exists a constrained set $\mathcal{S}$ of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ in $\mathcal{G}$. Again, by the previous discussion we know that no path in $\mathcal{S}$ contains an outgoing edge of a node in $\mathcal{W}$. Therefore, $\mathcal{S}$ is also a constrained set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ in $\bar{\mathcal{G}}$. This proves the lemma. □

The following lemma relates the existence of a constrained set of vertex-disjoint paths and the *second* graph operation.

Lemma 9.7. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let $\mathcal{U}, \mathcal{W} \subseteq \mathcal{V}$. Suppose that $k \in \mathcal{W} \setminus \mathcal{U}$ has exactly one in-neighbour $j \in \mathcal{N}_k^-$ that is reachable from $\mathcal{U}$. Then there exists a constrained set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ if and only if there exists a constrained set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\bar{\mathcal{W}}$ with $\bar{\mathcal{W}} := (\mathcal{W} \setminus \{k\}) \cup \{j\}$.

Proof. We will first show that $\mathcal{S}$ is a set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ if and only if $\bar{\mathcal{S}}$ is a set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\bar{\mathcal{W}}$, where $\bar{\mathcal{S}}$ will be specified.

Suppose that $\mathcal{S}$ is a set of $|\mathcal{U}|$ vertex disjoint paths from $\mathcal{U}$ to $\mathcal{W}$. Consider the path $\mathcal{P} \in \mathcal{S}$ that goes from $\mathcal{U}$ to k. Since $j \in \mathcal{N}_k^-$ is the only in-neighbour of k that is reachable from $\mathcal{U}$, we obtain $(j,k) \in \mathcal{P}$. This means that $\bar{\mathcal{P}} := \mathcal{P} \setminus (j,k)$ is a path from $\mathcal{U}$ to j. Clearly, $\bar{\mathcal{S}} := (\mathcal{S} \setminus \mathcal{P}) \cup \bar{\mathcal{P}}$ is a set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\bar{\mathcal{W}}$.

Conversely, suppose that $\bar{\mathcal{S}}$ is a set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\bar{\mathcal{W}}$. Consider the path $\bar{\mathcal{P}} \in \bar{\mathcal{S}}$ that goes from $\mathcal{U}$ to $j \in \bar{\mathcal{W}}$. Since $j \in \mathcal{N}_k^-$ is the only in-neighbour of k that is reachable from $\mathcal{U}$, the path $\bar{\mathcal{P}}$ does not pass through the vertex k. Consequently, $\mathcal{P} := \bar{\mathcal{P}} \cup (j,k)$ is a path from $\mathcal{U}$ to k. Again using the fact that j is the only in-neighbour of k that is reachable from $\mathcal{U}$, we see that no path in $\bar{\mathcal{S}}$ passes through the vertex k. This implies that $\mathcal{S} := (\bar{\mathcal{S}} \setminus \bar{\mathcal{P}}) \cup \mathcal{P}$ is a set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$.

To conclude, we have shown that $\mathcal{S}$ is a set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ if and only if $\bar{\mathcal{S}}$ is a set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\bar{\mathcal{W}}$, where the set $\bar{\mathcal{S}}$ is defined as $\bar{\mathcal{S}} := (\mathcal{S} \setminus \mathcal{P}) \cup \bar{\mathcal{P}}$. This implies that $\mathcal{S}$ is a *constrained* set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\mathcal{W}$ if and only if $\bar{\mathcal{S}}$ is a *constrained* set of $|\mathcal{U}|$ vertex-disjoint paths from $\mathcal{U}$ to $\bar{\mathcal{W}}$. □

Proof of Theorem 9.5. The "if" statement follows from Theorem 9.4. To prove the "only if" part, suppose that $(i, \mathcal{N}_i^+)$ is identifiable. By Theorem 9.2, $\mathcal{N}_i^+ \subseteq \mathcal{D}(\mathcal{C})$, where $\mathcal{D}(\mathcal{C})$ is a derived vertex set of $\mathcal{C}$ with respect to $\mathcal{N}_i^+$. In fact, we obtain $\mathcal{N}_i^+ = \mathcal{D}(\mathcal{C})$ as $|\mathcal{N}_i^+| = |\mathcal{C}|$. Let $\mathcal{D}(\mathcal{G})$ denote the associated derived graph of $\mathcal{G}$. By definition, there exists a constrained set of $|\mathcal{N}_i^+|$ vertex-disjoint paths from $\mathcal{N}_i^+$ to $\mathcal{D}(\mathcal{C})$ in $\mathcal{D}(\mathcal{G})$ (see Section 9.7). By consecutive application of Lemmas 9.6 and 9.7, we conclude that there exists a constrained set of $|\mathcal{N}_i^+|$ vertex-disjoint paths from $\mathcal{N}_i^+$ to $\mathcal{C}$ in the graph $\mathcal{G}$. □

10 NETWORK IDENTIFIABILITY OF UNDIRECTED NETWORKS

In this chapter we continue our work on identifiability of networks with known graph structure. We will, however, focus on a different class of *undirected* networks described by state-space models. For this (more specific) class of networks we are able to derive stronger results. In fact, we prove that identifiability can generally be secured by both measuring and exciting a subset of nodes in the network.

10.1 INTRODUCTION

Networks of dynamical systems appear in multiple contexts, including power networks, sensor networks, and robotic networks, cf. [137, Sec. 1]. It is natural to describe such networks by a graph, where nodes correspond with dynamical subsystems, and edges represent interaction between different systems. Often, the graph structure of dynamical networks is not directly available. For instance, in neuroscience, the interactions between brain areas are typically unknown [213]. Other examples of networks with unknown interconnection structure include genetic networks [97] and wireless sensor networks [118].

Consequently, the problem of *network reconstruction* is considered in the literature. Network reconstruction is quite a broad concept, and there exist multiple variants of this problem. For example, the goal in [190], [130] is to reconstruct the Boolean structure of the network (i.e., the locations of the edges). Moreover, simultaneous identification of the graph structure *and* the network weights has been considered in [79], [223], [249]. Typically, the conditions under which the network structure is uniquely identifiable are rather strong, and it is often assumed that the states of all nodes in the network can be measured [130], [79], [223], [249]. In fact, it has been shown [163] that measuring all network nodes is *necessary* for network reconstruction of dynamical networks (described by a class of state-space systems).

In this chapter, we consider undirected dynamical networks described by state-space systems. In contrast to the above discussed papers, we assume that the graph structure is *known*, but the state matrix of the network is *unavailable*. Such a situation arises, for example, in electrical or power networks in which the locations of links are typically known, but link weights require identification. Our goal is to find graph-theoretic conditions under which the state matrix of the network can be uniquely identified.

Graph-theoretic conditions have previously been used to assess other system-theoretic properties such as structural controllability [142], [219], fault detection

[174], [49], and parameter-independent stability [102]. Conditions based on the graph structure are often desirable since they avoid potential numerical issues associated with more traditional linear algebra tests. In addition, graph-theoretic conditions provide insight in the types of networks having certain system-theoretic properties, and can aid the selection of input/output nodes [167].

The papers that are most closely related to the work in this chapter are [152] and [15]. Nabavi *et al.* [152] consider weighted, undirected consensus networks with a single input. They assume that the graph structure is known, and aim to identify the weights in the network. A sensor placement algorithm is presented, which selects a set of sensor nodes on the basis of the graph structure. It is shown that this set of sensor nodes is sufficient to guarantee weight identifiability. Bazanella *et al.* [15] consider a network model where interactions between nodes are modeled by proper transfer functions (see also [214], [235]). Also in this chapter, the graph structure is assumed to be known, and the goal is to find conditions under which the transfer functions can be identified. Under the assumption that all nodes are externally excited, necessary and sufficient graph-theoretic conditions are presented under which all transfer functions can be (generically) identified.

Note that the above papers make explicit assumptions on the number of input or output nodes. Indeed, in [152] there is a single input node, all nodes are input nodes in [15], and all nodes are measured in [235]. In contrast to these papers, the current chapter considers graph-theoretic conditions for identifiability of dynamical networks where the sets of input and output nodes can be any two (known) subsets of the vertex set. Our main contribution consists of a graph coloring condition for identifiability of dynamical networks with single-integrator node dynamics. Specifically, we prove a relation between identifiability and so-called *zero forcing sets* [87] (see also [142], [219], [207]). As our second result, we show how our framework can be used to assess identifiability of dynamical networks with general, higher-order node dynamics.

The organization of this chapter is as follows. First, in Section 10.2 we introduce the notation and preliminaries used throughout the chapter. Subsequently, in Section 10.3 we state the problem. Section 10.4 contains our main results, and Section 10.5 treats an extension to higher-order dynamics. Finally, our conclusions are stated in Section 10.6.

10.2 PRELIMINARIES

10.2.1 Graph theory

All graphs considered in this chapter are simple, that is, without self-loops and with at most one edge between any pair of vertices. Let $G = (V, E)$ be an undirected graph, where $V = \{1, 2, \ldots, n\}$ is the set of *vertices* (or nodes), and $E \subseteq V \times V$ denotes the set of *edges*. A node $j \in V$ is said to be a *neighbour* of $i \in V$

if $(i, j) \in E$. An *induced subgraph* $G_S = (V_S, E_S)$ of G is a graph with the properties that $V_S \subseteq V$, $E_S \subseteq E$ and for each $i, j \in V_S$ we have $(i, j) \in E_S$ if and only if $(i, j) \in E$. For any subset of nodes $V' = \{v_1, v_2, \ldots, v_r\} \subseteq V$ we define the $n \times r$ matrix $P(V; V')$ as $P_{ij} := 1$ if $i = v_j$ and $P_{ij} := 0$ otherwise, where P_{ij} denotes the (i, j)-th entry of P. We will now define two families of matrices associated with the graph G. Firstly, we define the *qualitative class* $\mathcal{Q}(G)$ as [87]

$$\mathcal{Q}(G) := \{X \in \mathbb{S}^n \mid \text{for } i \neq j,\, X_{ji} \neq 0 \iff (i, j) \in E\}.$$

The off-diagonal entries of matrices in $\mathcal{Q}(G)$ carry the graph structure of G in the sense that X_{ji} is nonzero if and only if there exists an edge (i, j) in the graph G. Note that the diagonal elements of matrices in $\mathcal{Q}(G)$ are free, and hence, both Laplacian and adjacency matrices associated with G are contained in $\mathcal{Q}(G)$ (see [142]). In this chapter, we focus on a subclass of $\mathcal{Q}(G)$, namely the class of matrices with *non-negative* off-diagonal entries. This class is denoted by $\mathcal{Q}_p(G)$, and defined as

$$\mathcal{Q}_p(G) := \{X \in \mathcal{Q}(G) \mid \text{for } i \neq j,\, X_{ji} \neq 0 \implies X_{ji} > 0\}.$$

Note that (weighted) adjacency and negated Laplacian matrices are members of the class $\mathcal{Q}_p(G)$.

10.2.2 Zero forcing sets

In this section we review the notion of zero forcing. Let $G = (V, E)$ be an undirected graph with vertices colored either black or white. The *color-change rule* is defined in the following way. If $u \in V$ is a black vertex and exactly one neighbour $v \in V$ of u is white, then change the color of v to black [87]. When the color-change rule is applied to u to change the color of v, we say *u forces v*, and write $u \to v$. Given a coloring of G, that is, given a set $Z \subseteq V$ containing black vertices only, and a set $V \setminus Z$ consisting of only white vertices, the *derived set* $D(Z)$ is the set of black vertices obtained by applying the color-change rule until no more changes are possible [87]. A *zero forcing set* for G is a subset of vertices $Z \subseteq V$ such that if initially the vertices in Z are colored black and the remaining vertices are colored white, then $D(Z) = V$. Finally, a zero forcing set $Z \subseteq V$ is called a *minimum zero forcing set* if for any zero forcing set Y in G we have $|Y| \geqslant |Z|$.

10.2.3 Dynamical networks

Consider an undirected graph $G = (V, E)$. Let $V_I \subseteq V$ be the set of so-called *input nodes*, and let $V_O \subseteq V$ be the set of *output nodes*, with cardinalities $|V_I| = m$ and $|V_O| = p$, respectively. Associated with G, V_I, and V_O, we consider the dynamical system

$$\dot{x}(t) = Xx(t) + Mu(t) \tag{10.1a}$$

$$y(t) = Nx(t), \tag{10.1b}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the input, and $y \in \mathbb{R}^p$ is the output. Furthermore, $X \in \mathcal{Q}_p(G)$ and the matrices M and N are indexed by V_I and V_O in the sense that

$$M = P(V; V_I), \text{ and } N = P^\top(V; V_O). \tag{10.2}$$

We use the shorthand notation (X, M, N) to denote the dynamical system (10.1). The transfer matrix of (10.1) is given by $T(s) := N(sI - X)^{-1}M$.

Remark 10.1. Note that in this chapter, we focus on dynamical networks (10.1), where the state matrix X is contained in $\mathcal{Q}_p(G)$. This implies that X is symmetric and the off-diagonal elements of X are non-negative. Dynamical networks of this form appear, for example, in consensus problems [158], and in the study of resistive-capacitive electrical networks, cf. [53, Sec. VB]. In addition, as we will see, the constraints on the matrix X are also attractive from identification point of view in the sense that we can often identify X with relatively small sets of input and output nodes. This is in contrast to the case of identifiability of matrices that do not satisfy symmetry and/or sign constraints. This is explained in more detail in Remark 10.3.

10.2.4 Network identifiability

In this section, we define the notion of network identifiability. It is well-known that the transfer matrix from u to y of system (10.1) can be identified from measurements of $u(t)$ and $y(t)$ if the input function u is sufficiently rich [114]. Then, the question is whether we can uniquely reconstruct the state matrix X from the transfer matrix $T(s)$. Specifically, since we assume that the matrix X is *unknown*, we are interested in conditions under which X can be reconstructed from $T(s)$ *for all* matrices $X \in \mathcal{Q}_p(G)$. This is known as *global identifiability* (see, e.g., [73]). To be precise, we have the following definition.

Definition 10.1. Consider an undirected graph $G = (V, E)$ with input nodes $V_I \subseteq V$ and output nodes $V_O \subseteq V$. Define M and N as in (10.2). We say $(G; V_I; V_O)$ is *identifiable* if for all matrices $X, \bar{X} \in \mathcal{Q}_p(G)$ the following implication holds:

$$N(sI - X)^{-1}M = N(sI - \bar{X})^{-1}M \implies X = \bar{X}. \tag{10.3}$$

Note that identifiability of $(G; V_I; V_O)$ is a property of the graph and the input/output nodes only, and not of the particular state matrix $X \in \mathcal{Q}_p(G)$.

Observation 10.1. The implication (10.3) that appears in Definition 10.1 can be equivalently stated as

$$NX^kM = N\bar{X}^kM \text{ for all } k \in \mathbb{N} \implies X = \bar{X}.$$

The matrices NX^kM for $k \in \mathbb{N}$ are often referred to as the *Markov parameters* of (X, M, N).

In addition to identifiability of $(G; V_I; V_O)$, we are interested in a more general property, namely identifiability of an *induced subgraph* of G. This is defined as follows.

Definition 10.2. Consider an undirected graph $G = (V, E)$ with input nodes $V_I \subseteq V$ and output nodes $V_O \subseteq V$, and let G_S be an induced subgraph of G. Define M and N as in (10.2). We say $(G_S; V_I; V_O)$ is *identifiable* if for all matrices $X, \bar{X} \in \mathcal{Q}_p(G)$ the following implication holds:

$$N(sI - X)^{-1}M = N(sI - \bar{X})^{-1}M \implies X_S = \bar{X}_S,$$

where $X_S, \bar{X}_S \in \mathcal{Q}_p(G_S)$ are the principal submatrices of X and $\bar{X}$ corresponding to the nodes in G_S.

Note that identifiability of $(G; V_I; V_O)$ is a special case of identifiability of $(G_S; V_I; V_O)$, where the subgraph G_S is simply equal to G.

10.3 PROBLEM STATEMENT

Let $G = (V, E)$ be an undirected graph with input nodes $V_I \subseteq V$ and output nodes $V_O \subseteq V$, and consider the associated dynamical system (10.1). Throughout this chapter, we assume G, V_I, and V_O to be *known*. We want to investigate which principal submatrices of X can be identified from input/output data (for all $X \in \mathcal{Q}_p(G)$). In other words, we want to find out for which induced subgraphs G_S of G, the triple $(G_S; V_I; V_O)$ is identifiable. In particular, we are interested in conditions under which $(G; V_I; V_O)$ is identifiable. Note that it is not straightforward to check the condition for identifiability in Definitions 10.1 and 10.2. Indeed, these definitions requires the computation and comparison of an infinite number of transfer matrices (for all $X, \bar{X} \in \mathcal{Q}_p(G)$). Instead, in this chapter we want to establish a condition for identifiability of $(G_S; V_I; V_O)$ in terms of zero forcing sets. Such a graph-based condition has the potential of being more efficient to check than the condition of Definition 10.2. In addition, graph-theoretic conditions have the advantage of avoiding possible numerical errors in the linear algebra computations appearing in Definition 10.2. Explicitly, the considered problem in this chapter is as follows.

Problem 10.1. Consider an undirected graph $G = (V, E)$ with input nodes $V_I \subseteq V$ and ouput nodes $V_O \subseteq V$, and let G_S be an induced subgraph of G. Provide graph-theoretic conditions under which $(G_S; V_I; V_O)$ is identifiable.

10.4 MAIN RESULTS

In this section, we state our main results. First, we establish a lemma which will be used to prove our main contributions (Theorems 10.1 and 10.2). The following lemma considers the case that $V_I = V_O$, and asserts that identifiability of $(G_S; U; U)$ is invariant under the color-change rule.

Lemma 10.1. Let G_S be an induced subgraph of the undirected graph $G = (V, E)$, and let $U \subseteq V$. Suppose that $u \to v$, where $u \in U$ and $v \in V \setminus U$. Then $(G_S; U; U)$ is identifiable if and only if $(G_S; U \cup \{v\}; U \cup \{v\})$ is identifiable.

Proof. The "only if" part of the statement follows directly from the fact that identifiability is preserved under the addition of input and output nodes. Therefore, in what follows, we focus on proving the "if" part. Suppose that $(G_S; U \cup \{v\}; U \cup \{v\})$ is identifiable. Let $\bar{M} := P(V; U \cup \{v\})$ denote the associated input matrix, and let $\bar{N} := \bar{M}^\top$ be the output matrix. In addition, let $M := P(V; U)$ and $N := M^\top$. The idea of this proof is as follows. For any $X \in \mathcal{Q}_p(G)$, we will show that the Markov parameters $\bar{N} X^k \bar{M}$ for $k \in \mathbb{N}$ can be obtained from the Markov parameters

$$NX^k M \text{ for } k \in \mathbb{N}. \tag{10.4}$$

Then, we will show that this implies that $(G_S; U; U)$ is identifiable. In particular, due to the overlap in the Markov parameters of $(\bar{N}, X, \bar{M})$ and (N, X, M), we only need to show that $(X^k)_{vw} = (X^k)_{wv}$ and $(X^k)_{vv}$ can be obtained from (10.4) for all $k \in \mathbb{N}$ and all $w \in U$. We start by showing that X_{uv} can be obtained from (10.4). To this end, we define $V_u := \{u\} \cup \{j \in V \mid (u, j) \in E\}$ and compute

$$\begin{aligned}(X^2)_{uu} &= \sum_{z \in V_u} X_{uz} X_{zu} \\ &= X_{uv}^2 + \sum_{z \in V_u \setminus \{v\}} X_{uz} X_{zu}.\end{aligned}$$

By hypothesis, $u \to v$ and hence $V_u \setminus \{v\} \subseteq U$. This implies that $X_{uv}^2 = (X^2)_{uu} - \sum_{z \in V_u \setminus \{v\}} X_{uz} X_{zu}$ can be obtained from the Markov parameters (10.4). As $X \in \mathcal{Q}_p(G)$, we have $X_{uv} > 0$ and therefore also X_{uv} can be obtained from (10.4).

Next, we prove that $(X^k)_{vw}$ can be obtained from (10.4) for any $k \in \mathbb{N}$ and any $w \in U$. To this end, we write

$$(X^{k+1})_{uw} = X_{uv}(X^k)_{vw} + \sum_{z \in V_u \setminus \{v\}} X_{uz}(X^k)_{zw}.$$

Since $V_u \setminus \{v\} \subseteq U$, and X_{uv} can be obtained from (10.4), this shows that we can find $(X^k)_{vw}$ from the Markov parameters (10.4) using the formula

$$(X^k)_{vw} = \frac{1}{X_{uv}} \left((X^{k+1})_{uw} - \sum_{z \in V_u \setminus \{v\}} X_{uz}(X^k)_{zw} \right).$$

Finally, we have to show that $(X^k)_{vv}$ can be obtained from (10.4) for any $k \in \mathbb{N}$. To do so, we compute

$$\begin{aligned}(X^{k+2})_{uu} &= \sum_{i,j \in V_u} X_{ui}(X^k)_{ij} X_{ju} \\ &= X_{uv}^2 (X^k)_{vv} + \sum_{\substack{i,j \in V_u \\ \{i,j\} \neq \{v\}}} X_{ui}(X^k)_{ij} X_{ju}.\end{aligned}$$

Note that $(X^{k+2})_{uu}$ appears as an entry of one of the Markov parameters (10.4). Furthermore, we have already established that X_{uv} can be obtained from (10.4). If $i = v$, then $X_{ui} = X_{uv}$, and we obtain X_{ui} from (10.4). Otherwise, $i \in V_u \setminus \{v\}$, and X_{ui} already appears as an entry of one of the Markov parameters (10.4). We can repeat the exact same argument for X_{ju}, to show that it can be obtained from (10.4). Finally, consider the term $(X^k)_{ij}$ for i and j not both equal to v. If $i, j \in V_u \setminus \{v\}$, then $i, j \in U$ and $(X^k)_{ij}$ appears directly as an entry of a Markov parameter in (10.4). Next, if $i = v$, then $j \in U$ and we have already proven that $(X^k)_{vj}$ can be obtained from (10.4). By symmetry, this also holds for the entry $(X^k)_{iv}$, where $i \in U$. This shows that $(X^k)_{vv}$ can be found using the Markov parameters (10.4) via the fomula

$$(X^k)_{vv} = \frac{1}{X_{uv}^2} \left((X^{k+2})_{uu} - \sum_{\substack{i,j \in V_u \\ \{i,j\} \neq \{v\}}} X_{ui}(X^k)_{ij} X_{ju} \right).$$

Now, by hypothesis, for any $X, \bar{X} \in \mathcal{Q}_p(G)$ the following implication holds:

$$\bar{N} X^k \bar{M} = \bar{N} \bar{X}^k \bar{M} \text{ for all } k \in \mathbb{N} \implies X_S = \bar{X}_S, \tag{10.5}$$

where X_S and $\bar{X}_S$ are the principal submatrices of respectively X and $\bar{X}$ corresponding to the nodes in G_S. Suppose that $NX^kM = N\bar{X}^kM$ for all $k \in \mathbb{N}$. By the above formulae for $(X^k)_{vv}$ and $(X^k)_{vw}$ (and for $(\bar{X}^k)_{vv}$, $(\bar{X}^k)_{vw}$), we conclude that $\bar{N}X^k\bar{M} = \bar{N}\bar{X}^k\bar{M}$ for all $k \in \mathbb{N}$, and consequently $X_S = \bar{X}_S$ by (10.5). Therefore, $(G_S; U; U)$ is identifiable. □

Based on the previous lemma, we state the following theorem, which is one of the main results of this chapter. Loosely speaking, it states that we can identify the principal submatrix of X corresponding to the *derived set* (cf. Section 10.2.2) of $V_I \cap V_O$.

Theorem 10.1. Let $G_S = (V_S, E_S)$ be an induced subgraph of the undirected graph $G = (V, E)$, and let $V_I, V_O \subseteq V$. Define $W := V_I \cap V_O$ and let $D(W)$ be the derived set of W in G. If $V_S \subseteq D(W)$ then $(G_S; V_I; V_O)$ is identifiable.

Proof. Let G_W denote the induced subgraph of G with vertex set $D(W)$. Note that $(G_W; D(W); D(W))$ is trivially identifiable. By consecutive application of Lemma

10.1, we find that $(G_W; W; W)$ is identifiable. By hypothesis, G_S is a subgraph of G_W and hence $(G_S; W; W)$ is identifiable. Finally, note that $W \subseteq V_I$ and $W \subseteq V_O$. Since identifiability is invariant under the addition of input/output nodes, we conclude that $(G_S; V_I; V_O)$ is identifiable. □

As a particular case of Theorem 10.1, we find that $(G; V_I; V_O)$ is identifiable if $D(W) = V$, that is, if W is a zero forcing set in the graph G. This is the topic of the following theorem.

Theorem 10.2. Let $G = (V, E)$ be an undirected graph and let $V_I, V_O \subseteq V$. If $V_I \cap V_O$ is a zero forcing set in G then $(G; V_I; V_O)$ is identifiable.

Remark 10.2. For a graph $G = (V, E)$, checking whether a given subset is a zero forcing set in G can be done in time complexity $\mathcal{O}(n^2)$, where $n = |V|$ (cf. [207]). Consequently, checking the condition of Theorem 10.2 is still feasible for large-scale graphs. Although the focus of this chapter is on the *analysis* of identifiability, we remark that Theorem 10.2 can also be used in the *design* of sets V_I and V_O that ensure identifiability of $(G; V_I; V_O)$. Specifically, input and output sets with small cardinality are obtained by choosing $V_I = V_O$ as a *minimum zero forcing set* in G. Minimum zero forcing sets are known for several types of graphs including path, cycle, and complete graphs, and for the class of tree graphs (see Section IV-B of [142]). Finding a minimum zero forcing set in general graphs is NP-hard [2]. However, there also exist heuristic algorithms for finding (minimum) zero forcing sets. For instance, it can be shown that for any graph G, it is possible to find a zero forcing set of cardinality $n - \operatorname{diam}(G)$, where $\operatorname{diam}(G)$ denotes the diameter of G.

Remark 10.3. It is interesting to remark that Theorem 10.2 implies that for many networks it is sufficient to excite and measure only a fraction of nodes (see, for instance, Example 10.1). This is in contrast with the case of identifiability of dynamical networks with unknown graph structure, for which it was shown that all nodes need to be measured [163]. Apart from the fact that we assume that the graph G is known, the rather mild conditions of Theorem 10.2 are also due to the fact that we consider *undirected* graphs with state matrices that satisfy *sign constraints*. In fact, in the case of *directed* graphs it can be shown that the condition $V_I \cup V_O = V$ is *necessary* for identifiability, i.e., each node of the graph needs to be an input or output node (or both). To see this, let G_d be a directed graph, and define $\mathcal{Q}_p(G_d)$ analogous to the definition for undirected graphs (Section 10.2.1), with the distinction that $X \in \mathcal{Q}_p(G_d)$ is not necessarily symmetric. Assume that $V_I \cup V_O \neq V$. We partition $X \in \mathcal{Q}_p(G_d)$, and pick a nonsingular $S \in \mathbb{R}^{n \times n}$ as

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}, \quad S = \begin{bmatrix} I & 0 \\ 0 & \epsilon I \end{bmatrix},$$

where the row block $\begin{bmatrix} X_{21} & X_{22} \end{bmatrix}$ corresponds to the nodes in $V \setminus (V_I \cup V_O)$. The partition of S is compatible with the one of X, and ϵ is a positive real number, not equal to 1. If X_{12} and X_{21} are not both zero matrices, then $\bar{X} := S^{-1} X S$ is contained in $\mathcal{Q}_p(G_d)$ and $\bar{X} \neq X$, but (X, M, N) and $(\bar{X}, M, N)$ have the same Markov parameters. That is, $(G; V_I; V_O)$ is not identifiable. If both X_{12} and X_{21} are zero, then the Markov parameters of (X, M, N) are independent of X_{22}, hence, $(G; V_I; V_O)$ is also not identifiable. Therefore, for directed graphs the condition $V_I \cup V_O = V$ is necessary for identifiability. The above discussion also implies that $V_I \cup V_O = V$ is necessary for identifiability of *undirected graphs* for which $X \in \mathcal{Q}(G)$ (i.e., for which X does not necessarily satisfy the sign constraints). Indeed, this can be shown by the same arguments as above, using $\epsilon = -1$. We conclude that the conditions for identifiability become much more restrictive once we remove either the assumption on sign constraints or the assumption that the graph is undirected.

Example 10.1. In this example, we illustrate Theorem 10.2. Consider the tree graph $G = (V, E)$ of Figure 10.1. The input set V_I and output set V_O have been designed in such a way that $V_I = V_O$ is a minimum zero forcing set in G. In fact, the nodes of V_I have been chosen as initial nodes of paths in a so-called *minimal*

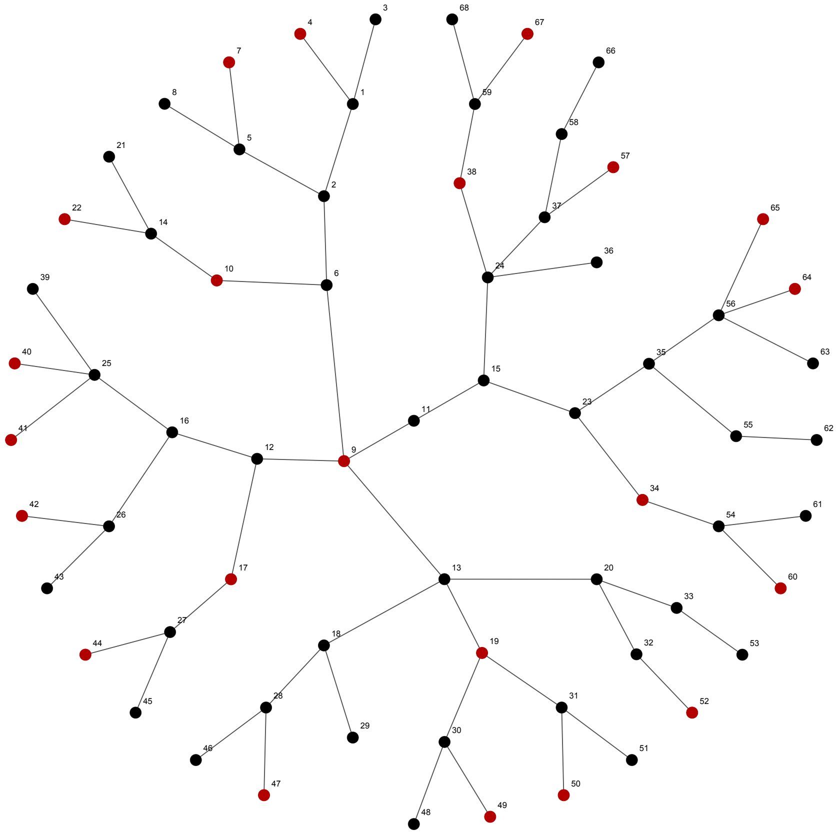

Figure 10.1: Tree graph G with input and output set $V_I = V_O = \{4, 7, 9, 10, 17, 19, 22, 34, 38, 40, 41, 42, 44, 47, 49, 50, 52, 57, 60, 64, 65, 67\}$.

path cover of G [142], and therefore, V_I is a minimum zero forcing set in G by Proposition IV.12 of [142]. Hence $V_I \cap V_O$ is a zero forcing set, and therefore, by Theorem 10.2 we conclude that $(G; V_I; V_O)$ is identifiable.

Example 10.2. It is important to note that the condition in Theorem 10.2 is not necessary for identifiability. This is shown in the following example. Consider a graph $G = (V, E)$, where $V = \{1,2,3\}$, and $E = \{(1,2),(2,1),(2,3),(3,2)\}$, and let $V_I = \{2\}$ and $V_O = V$. A straightforward calculation shows that any matrix $X \in \mathcal{Q}_p(G)$ can be identified from the Markov parameters NXM and NX^2M. This shows that $(G; V_I; V_O)$ is identifiable. However, note that $V_I \cap V_O = \{2\}$ is not a zero forcing set in G.

10.5 HIGHER-ORDER NODE DYNAMICS

The purpose of this section is to generalize the results of Section 10.4 to the case of higher-order node dynamics. Suppose that node $i \in V$ has the associated dynamics

$$\dot{x}_i(t) = \begin{cases} Ax_i(t) + Bu_i(t) + Fz_i(t) & \text{if } i \in V_I \\ Ax_i(t) + Fz_i(t) & \text{otherwise} \end{cases},$$

where $x_i \in \mathbb{R}^q$ is the state of node i, $u_i \in \mathbb{R}^r$ is the input (only applied to nodes in V_I), and $z_i \in \mathbb{R}^s$ describes the coupling between the nodes. The real matrices A, B, and F are of appropriate dimensions. In addition, we associate with each node $i \in V_O$ the output equation

$$y_i(t) = Cx_i(t).$$

The coupling variable z_i is chosen as

$$z_i(t) = \sum_{j=1}^{n} X_{ij} K x_j(t),$$

where $K \in \mathbb{R}^{s \times q}$, $X_{ii} \in \mathbb{R}$, $X_{ij} = X_{ji}$, and for $i \neq j$, $X_{ij} \geqslant 0$ and $X_{ij} > 0$ if and only if $(i,j) \in E$. We define $x := \operatorname{col}(x_1, x_2, \ldots, x_n)$, $u := \operatorname{col}(u_{i_1}, u_{i_2}, \ldots, u_{i_m})$, and $y := \operatorname{col}(y_{j_1}, y_{j_2}, \ldots, y_{j_p})$, where $i_k \in V_I$ and $j_l \in V_O$ for all $k = 1,2,\ldots,m$ and $l = 1,2,\ldots,p$. Then, the dynamics of the entire network is described by the system

$$\dot{x}(t) = (I \otimes A + X \otimes FK)x(t) + (M \otimes B)u(t) \tag{10.6a}$$
$$y(t) = (N \otimes C)x(t), \tag{10.6b}$$

where the (i,j)-th entry of the matrix $X \in \mathcal{Q}_p(G)$ is equal to X_{ij}, and the matrices M and N are defined in (10.2). Dynamics of the form (10.6) arise, for example, when synchronizing networks of linear oscillators [185]. In what follows, we use the notation $X_e := I \otimes A + X \otimes FK$, $M_e := M \otimes B$, and $N_e := N \otimes C$.

We assume that the matrices A, B, C, F and K are known, and we are interested in conditions under which we can identify an induced subgraph G_S of G. To make this precise, we say $(G_S; V_I; V_O)$ is *identifiable with respect to* (10.6) if for all $X, \bar{X} \in \mathcal{Q}_p(G)$ the following implication holds:

$$N_e(sI - X_e)^{-1}M_e = N_e(sI - \bar{X}_e)^{-1}M_e \implies X_S = \bar{X}_S,$$

where $X_e := I \otimes A + X \otimes FK$, $\bar{X}_e := I \otimes A + \bar{X} \otimes FK$, and the matrices $X_S, \bar{X}_S \in \mathcal{Q}_p(G)$ are the principal submatrices of X and $\bar{X}$ corresponding to the nodes of G_S. The following theorem states conditions for identifiability of $(G_S; V_I; V_O)$ for the case of general network dynamics.

Theorem 10.3. Let $G_S = (V_S, E_S)$ be an induced subgraph of the undirected graph $G = (V, E)$, and let $V_I, V_O \subseteq V$. Define $W := V_I \cap V_O$ and let $D(W)$ be the derived set of W in G. Then $(G_S; V_I; V_O)$ is identifiable with respect to (10.6) if $V_S \subseteq D(W)$ and $C(FK)^k B \neq 0$ for all $k \in \mathbb{N}$.

Proof. Consider two matrices $X, \bar{X} \in \mathcal{Q}_p(G)$ and define $X_e := I \otimes A + X \otimes FK$ and $\bar{X}_e := I \otimes A + \bar{X} \otimes FK$. Moreover, let $M_e := M \otimes B$, and $N_e := N \otimes C$. Suppose that $N_e X_e^k M_e = N_e \bar{X}_e^k M_e$ for all $k \in \mathbb{N}$. We want to prove by induction that $NX^kM = N\bar{X}^kM$ for all $k \in \mathbb{N}$. For $k = 1$, the equation $N_e X_e M_e = N_e \bar{X}_e M_e$ implies

$$(N \otimes C)(I \otimes A + X \otimes FK - I \otimes A - \bar{X} \otimes FK)(M \otimes B) = 0,$$

and hence $N(X - \bar{X})M \otimes CFKB = 0$. By assumption, $CFKB \neq 0$, and therefore $NXM = N\bar{X}M$. Next, suppose that $NX^iM = N\bar{X}^iM$ for all $i = 1, \dots, k$. The aim is to prove that $NX^{k+1}M = N\bar{X}^{k+1}M$. Note that we obtain

$$N_e X_e^{k+1} M_e = NX^{k+1}M \otimes C(FK)^{k+1}B + \sum_{i=0}^{k} NX^iM \otimes R_i,$$

where R_i is a matrix that depends on A, B, C, F and K only. Completely analogously, an expression for $N_e \bar{X}_e^{k+1} M_e$ can be derived. By the induction hypothesis, $NX^iM = N\bar{X}^iM$ for $i = 1, \dots, k$, and therefore $N_e X_e^{k+1} M_e = N_e \bar{X}_e^{k+1} M_e$ implies

$$(NX^{k+1}M - N\bar{X}^{k+1}M) \otimes C(FK)^{k+1}B = 0.$$

Since $C(FK)^{k+1}B \neq 0$, we find $NX^{k+1}M = N\bar{X}^{k+1}M$. Therefore, $NX^kM = N\bar{X}^kM$ for all $k \in \mathbb{N}$. However, since $V_S \subseteq D(W)$ we find $X = \bar{X}$ by Theorem 10.1. Hence $(G_S; V_I; V_O)$ is identifiable with respect to (10.6). □

10.6 CONCLUSIONS

In this chapter we have considered the problem of identifiability of undirected dynamical networks. Specifically, we have assumed that the graph structure of the network is known, and we were interested in graph-theoretic conditions under which (a submatrix of) the network's state matrix can be identified. To this end, we have used a graph coloring rule called zero forcing. We have shown that a principal submatrix of the state matrix can be identified if the intersection of input and output nodes can color all nodes corresponding to the rows and columns of the submatrix. In particular, the entire state matrix can be identified if the intersection of input and output nodes constitutes a so-called zero forcing set in the graph. Checking whether a given set of nodes is a zero forcing set can be done in $\mathcal{O}(n^2)$, where n is the number of nodes in the network [207]. We emphasize that the results we have presented here only treat the *identifiability* of dynamical networks, and we have not discussed any network reconstruction algorithms, like in [190], [130], [79], [223]. However, if the conditions of Theorem 10.2 are satisfied, then the state matrix of the network can be identified using any suitable method, given sufficiently rich data.

11 A UNIFYING FRAMEWORK FOR STRUCTURAL CONTROLLABILITY

So far, we have studied notions of network identifiability. We saw that global identifiability of a model set is a *structural* property: it can be characterized in terms of the network graph and the locations of excited and measured nodes. In this chapter, we focus on a different structural property, namely *strong structural controllability*. As with identifiability, we will see that structural controllability can be characterized in graph-theoretic terms, via two graphs associated to the considered linear system.

11.1 INTRODUCTION

Controllability is a fundamental concept in systems and control. For linear time-invariant systems of the form

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{11.1}$$

controllability can be be verified using the Kalman rank test or the Hautus test [208]. Often, the exact values of the entries in the matrices A and B are not known, but the underlying interconnection structure between the input and state variables is known exactly.

In order to formalize this, Mayeda and Yamada have introduced a framework in which, instead of a fixed pair of real matrices, only the *zero/nonzero* structure of A and B is given [133]. This means that each entry of these matrices is known to be either a *fixed zero* or an *arbitrary nonzero* real number. Given this zero/nonzero structure, they then study controllability of the family of systems for which the state and input matrices have this zero/nonzero structure. In this setup, this family of systems is called *strongly structurally controllable* if all members of the family are controllable in the classical sense [133].

Most of the existing literature up to now has considered strong structural controllability under the above rather restrictive assumption that for each of the entries of the system matrices there are only *two possibilities*: it is either a fixed zero, or an arbitrary nonzero value [8, 29, 93, 133, 135, 157, 175, 207]. There are, however, many scenarios in which, in addition to these two possibilities, there is a third possibility, namely, that a given entry is not a fixed zero or nonzero, but can take *any real value*. In such a scenario, it is not possible to represent the system using a zero/nonzero structure, but a third possibility needs to be taken into account. To illustrate this, consider the following example.

Example 11.1. The electrical circuit in Figure 11.1 consists of a resistor, two capacitors, an inductor, an independent voltage source, an independent current

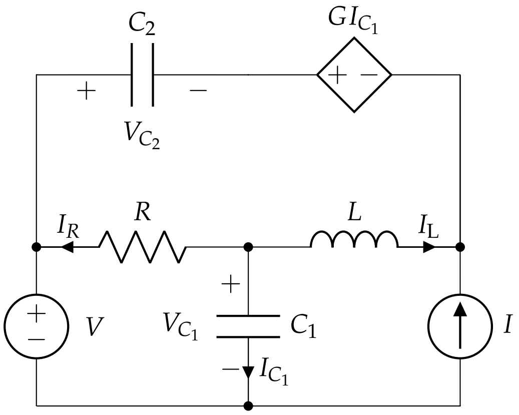

Figure 11.1: Example of electrical circuit.

source and a current controlled voltage source. Assume that the parameters R, C_1, C_2 and L are positive but not known exactly. We denote the current through R, L, and C_1 by I_R, I_L, and I_{C_1}, respectively, and the voltage across C_1 and C_2 by V_{C_1} and V_{C_2}, respectively. The current controlled voltage source is represented by GI_{C_1} with gain G assumed to be positive. Define the state vector as $x = [V_{C_1} \ V_{C_2} \ I_L]^T$ and the input as $u = [V \ I]^T$. By Kirchhoff's current and voltage laws, the circuit is represented by a linear time-invariant system (11.1) with

$$A = \begin{bmatrix} -\frac{1}{RC_1} & 0 & -\frac{1}{C_1} \\ 0 & 0 & -\frac{1}{C_2} \\ \frac{R-G}{RL} & \frac{1}{L} & -\frac{G}{L} \end{bmatrix}, \quad B = \begin{bmatrix} \frac{1}{RC_1} & 0 \\ 0 & -\frac{1}{C_2} \\ \frac{G-R}{RL} & 0 \end{bmatrix}. \tag{11.2}$$

Recall that the parameters $R, C_1, C_2, L > 0$ are not known exactly. This means that the matrices in (11.2) are not known exactly, but we do know that they have the following structure. Firstly, some entries are *fixed zeros*. Secondly, some of the entries are always *nonzero*, for instance, the entry with value $-\frac{1}{RC_1}$. The third type of entries, those with value $\frac{R-G}{RL}$ and $\frac{G-R}{RL}$, can be either zero (if $R = G$) or nonzero. Since the system matrices in this example do not have a zero/nonzero structure, the existing tests for strong structural controllability [8, 29, 93, 133, 157, 175, 207] are not applicable.

A similar problem as in Example 11.1 appears in the context of linear networked systems. Strong structural controllability of such systems has been well-studied [29, 142, 147, 168, 207]. In the setup of these references, the weights on the edges of the network graph are unknown, while the network graph itself is known. Under the assumption that the edge weights are arbitrary but nonzero, linear networked systems can thus be regarded as systems with a given zero/nonzero structure. This zero/nonzero structure is determined by the network graph, in the sense that nonzero entries in the system matrices correspond to edges in the network graph. However, often even exact knowledge of the network graph is not available, in the sense that it is unknown whether certain edges in the graph exist or not. This issue of missing knowledge appears, for example, in social networks [103], the world wide web [237], biological networks [34,74] and ecological systems [105]. Another cause for uncertainty about the network graph might be malicious attacks and unintentional failures. This issue is encountered in transportation networks [107], sensor networks [98] and gas networks [27].

To conclude, both in the context of modeling physical systems, as well as in representing networked systems, capturing the system simply by a zero/nonzero structure is not always possible, and a more general system structure is required. The papers [142,146,147,149,170,219] study classes[1] of zero/nonzero/arbitrary patterns in the context of strong structural controllability. However, necessary and sufficient conditions for strong structural controllability of general zero/nonzero/arbitrary patterns have not yet been established. The goal of this chapter is to provide such general necessary and sufficient conditions. In particular, our main contributions are the following:

1. We extend the notion of zero/nonzero structure to a more general type of *zero/nonzero/arbitrary structure*, and formalize this structure in terms of suitable pattern matrices.

2. We establish necessary and sufficient conditions for strong structural controllability for families of systems with a given zero/nonzero/arbitrary structure. These conditions are of an algebraic nature and can be verified by a rank test on two pattern matrices.

3. We provide a graph-theoretic condition for a given pattern matrix to have full row rank. This condition can be verified using a new *color change rule*, that will be defined in this chapter.

4. We establish a graph-theoretic test for strong structural controllability for the new families of structured systems.

5. Finally, we relate our results to those existing in the literature by showing how existing results can be recovered from those we present in this chapter. We find that seemingly incomparable results of [207] and [142] follow from our main results, which reveals an overarching theory. For these reasons, this chapter can be seen as a unifying approach to strong structural controllability of linear time-invariant systems.

We conclude this section by giving a brief account of research lines that are related to strong structural controllability but that will not be pursued in this chapter. The concept of weak structural controllability was introduced by Lin in [110] and has been studied extensively, see [37,38,52,110,113,194,201]. Another, more recent, line of work focuses on structural controllability of systems for which

[1] In [142,146,147,149,219], a special structure where only the diagonal entries of the state matrix are arbitrary entries (typically arising from a network context) were studied. In [170], the authors call zero/nonzero/arbitrary structure a “selective structure".

there are *dependencies* among the arbitrary entries of the system matrices [94, 112]. An important special case of dependencies among parameters arises when the state matrix is constrained to be symmetric, which was considered in [134,136,147]. The problem of *minimal input selection* for controllability has also been well-studied, see, e.g., [159, 167, 200, 211]. Strong structural controllability was also studied for *time-varying systems* in [177], and conditions for controllability were established for both discrete-time and continuous-time systems. Finally, weak and strong structural *targeted* controllability have been investigated in [109] and [141, 219], respectively.

The outline of the rest of the chapter is as follows. In Section 11.2, we present some preliminaries. Next, in Section 11.3, we formulate the main problem treated in this chapter. Then, in Section 11.4 we state our main results. Section 11.5 contains a comparison of our results with previous work. In Section 11.6 we state proofs of the main results. Finally, in Section 11.7 we formulate our conclusions.

11.2 PRELIMINARIES

In this chapter, we will use so-called pattern matrices. By a pattern matrix we mean a matrix with entries in the set of symbols $\{0, *, ?\}$. These symbols will be given a meaning in the sequel.

The set of all $p \times q$ pattern matrices will be denoted by $\{0, *, ?\}^{p\times q}$. For a given $p \times q$ pattern matrix $\mathcal{M}$, we define the *pattern class* of $\mathcal{M}$ as

$$\begin{aligned}\mathcal{P}(\mathcal{M}) := \{M \in \mathbb{R}^{p\times q} \,| M_{ij} = 0 \text{ if } \mathcal{M}_{ij} = 0,\\ M_{ij} \neq 0 \text{ if } \mathcal{M}_{ij} = *\}.\end{aligned}$$

This means that for a matrix $M \in \mathcal{P}(\mathcal{M})$, the entry M_{ij} is either (i) *zero* if $\mathcal{M}_{ij} = 0$, (ii) *nonzero* if $\mathcal{M}_{ij} = *$, or (iii) *arbitrary* (zero or nonzero) if $\mathcal{M}_{ij} = ?$.

11.3 PROBLEM FORMULATION

Let $\mathcal{A} \in \{0, *, ?\}^{n\times n}$ and $\mathcal{B} \in \{0, *, ?\}^{n\times m}$ be pattern matrices. Consider the linear dynamical system

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad (11.3)$$

where the system matrix A is in $\mathcal{P}(\mathcal{A})$ and the input matrix B is in $\mathcal{P}(\mathcal{B})$, and where $x \in \mathbb{R}^n$ is the state and $u \in \mathbb{R}^m$ is the input.

We will call the family of systems (11.3) a *structured system*. To simplify the notation, we denote this structured system by the pair of pattern matrices $(\mathcal{A}, \mathcal{B})$.

Example 11.2. Consider the electrical circuit discussed in Example 11.1. Recall that this was modelled as the state space system (11.2) in which the entries of the system matrix and input matrix were either fixed zeros, strictly nonzero or undetermined. This can be represented as a structured system $(\mathcal{A}, \mathcal{B})$ with pattern matrices

$$\mathcal{A} = \begin{bmatrix} * & 0 & * \\ 0 & 0 & * \\ ? & * & * \end{bmatrix} \text{ and } \mathcal{B} = \begin{bmatrix} * & 0 \\ 0 & * \\ ? & 0 \end{bmatrix}. \tag{11.4}$$

In this chapter we will study structural controllability of structured systems. In particular, we will focus on strong structural controllability.

Definition 11.1. The system $(\mathcal{A}, \mathcal{B})$ is called *strongly structurally controllable* if the pair (A, B) is controllable for all $A \in \mathcal{P}(\mathcal{A})$ and $B \in \mathcal{P}(\mathcal{B})$.

The problem that we will investigate in the present chapter is stated as follows.

Problem 11.1. Given two pattern matrices $\mathcal{A} \in \{0, *, ?\}^{n \times n}$ and $\mathcal{B} \in \{0, *, ?\}^{n \times m}$, provide necessary and sufficient conditions under which $(\mathcal{A}, \mathcal{B})$ is strongly structurally controllable.

In the remainder of this chapter, we will simply call the structured system $(\mathcal{A}, \mathcal{B})$ *controllable* if it is strongly structurally controllable.

Remark 11.1. In addition to strong structural controllability, *weak structural controllability* has also been studied extensively. This concept was introduced by Lin in [110]. Instead of requiring *all* systems in a family associated with a given structured system to be controllable, weak structural controllability only asks for the existence of at least one controllable member of that family, see [52, 110, 194]. In these references, conditions were established for weak structural controllability of structured systems in which the pattern matrices only contain 0 or ? entries. The question then arises: is it possible to generalize the results from [52, 110, 194] to structured systems in the context of our chapter, with more general pattern matrices $\mathcal{A} \in \{0, *, ?\}^{n \times n}$ and $\mathcal{B} \in \{0, *, ?\}^{n \times m}$. Indeed, it turns out that the results in [52, 110, 194] can immediately be applied to assess weak structural controllability of our more general structured systems. To show this, for given pattern matrices $\mathcal{A} \in \{0, *, ?\}^{n \times n}$ and $\mathcal{B} \in \{0, *, ?\}^{n \times m}$ we define two new pattern matrices $\mathcal{A}' \in \{0, ?\}^{n \times n}$ and $\mathcal{B}' \in \{0, ?\}^{n \times m}$ as follows: $\mathcal{A}'_{ij} = 0 \iff \mathcal{A}_{ij} = 0$ and $\mathcal{B}'_{ij} = 0 \iff \mathcal{B}_{ij} = 0$. The new structured system $(\mathcal{A}', \mathcal{B}')$ is now a structured system of the form studied in [52, 110, 194]. Using the fact that weak structural controllability is a generic property [194], it can then be shown that weak structural controllability of $(\mathcal{A}', \mathcal{B}')$ is equivalent to that of $(\mathcal{A}, \mathcal{B})$. In other words, weak structural controllability of general $(\mathcal{A}, \mathcal{B})$ can be verified using the conditions established in previous work [52, 110, 194].

11.4 MAIN RESULTS

In this section, the main results of this chapter will be stated. Firstly, we will establish an algebraic condition for controllability of a given structured system.

This condition states that controllability of a structured system is equivalent to full rank conditions on two pattern matrices associated with the system. Secondly, a graph-theoretic condition for a given pattern matrix to have full row rank will be given in terms of a so-called *color change rule*. Finally, based on the above algebraic condition and graph-theoretic condition, we will establish a graph-theoretic condition for controllability of a structured system.

Our first main result is a rank test for controllability of a structured system. In the sequel, we say that a pattern matrix $\mathcal{M}$ *has full row rank* if every matrix $M \in \mathcal{P}(\mathcal{M})$ has full row rank.

Theorem 11.1. The system $(\mathcal{A}, \mathcal{B})$ is controllable if and only if the following two conditions hold:

1. The pattern matrix $\begin{bmatrix}\mathcal{A} & \mathcal{B}\end{bmatrix}$ has full row rank.

2. The pattern matrix $\begin{bmatrix}\bar{\mathcal{A}} & \mathcal{B}\end{bmatrix}$ has full row rank where $\bar{\mathcal{A}}$ is the pattern matrix obtained from $\mathcal{A}$ by modifying the diagonal entries of $\mathcal{A}$ as follows:

$$\bar{\mathcal{A}}_{ii} := \begin{cases} * & \text{if } \mathcal{A}_{ii} = 0, \\ ? & \text{otherwise.} \end{cases} \tag{11.5}$$

We note here that the two rank conditions in Theorem 11.1 are independent, in the sense that one does not imply the other in general. To show that the first rank condition does not imply the second, consider the pattern matrices $\mathcal{A}$, the corresponding $\bar{\mathcal{A}}$, and $\mathcal{B}$ given by

$$\mathcal{A} = \begin{bmatrix} * & * \\ 0 & 0 \end{bmatrix}, \bar{\mathcal{A}} = \begin{bmatrix} ? & * \\ 0 & * \end{bmatrix} \text{ and } \mathcal{B} = \begin{bmatrix} * \\ * \end{bmatrix}.$$

It is evident that the pattern matrix $\begin{bmatrix}\mathcal{A} & \mathcal{B}\end{bmatrix}$ has full row rank. However, with

$$\bar{A} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \in \mathcal{P}(\bar{\mathcal{A}}) \text{ and } B = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathcal{P}(\mathcal{B}),$$

the matrix $\begin{bmatrix}\bar{A} & B\end{bmatrix}$ does not have full row rank. To show that the second condition does not imply the first one, consider the pattern matrix $\mathcal{A}$, the corresponding $\bar{\mathcal{A}}$, and $\mathcal{B}$ given by

$$\mathcal{A} = \begin{bmatrix} ? & 0 \\ * & 0 \end{bmatrix}, \bar{\mathcal{A}} = \begin{bmatrix} ? & 0 \\ * & * \end{bmatrix} \text{ and } \mathcal{B} = \begin{bmatrix} * \\ * \end{bmatrix}.$$

Obviously, the pattern matrix $\begin{bmatrix}\bar{\mathcal{A}} & \mathcal{B}\end{bmatrix}$ has full row rank. However, for the choice

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \in \mathcal{P}(\mathcal{A}) \text{ and } B = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathcal{P}(\mathcal{B}),$$

we see that $\begin{bmatrix}A & B\end{bmatrix}$ does not have full row rank.

Next, we discuss a noteworthy special case in which the first rank condition in Theorem 11.1 is implied by the second one. Indeed, if none of the diagonal entries of $\mathcal{A}$ is zero, it follows from (11.5) that $\mathcal{P}(\mathcal{A}) \subseteq \mathcal{P}(\bar{\mathcal{A}})$. Hence, we obtain the following corollary to Theorem 11.1.

Corollary 11.1. Suppose that none of the diagonal entries of $\mathcal{A}$ is zero. Let $\bar{\mathcal{A}}$ be as defined in (11.5). The system $(\mathcal{A}, \mathcal{B})$ is controllable if and only if $\begin{bmatrix}\bar{\mathcal{A}} & \mathcal{B}\end{bmatrix}$ has full row rank.

Note that both $\begin{bmatrix}\mathcal{A} & \mathcal{B}\end{bmatrix}$ and $\begin{bmatrix}\bar{\mathcal{A}} & \mathcal{B}\end{bmatrix}$ appearing in Theorem 11.1 are $n \times (n+m)$ pattern matrices. Next, we will develop a graph-theoretic test for checking whether a given pattern matrix has full rank. To do so, we first need to introduce some terminology. Let $\mathcal{M} \in \{0, *, ?\}^{p \times q}$ be a pattern matrix with $p \leqslant q$. We associate a directed graph $G(\mathcal{M}) = (V, E)$ with $\mathcal{M}$ as follows. Take as node set $V = \{1, 2, \ldots, q\}$ and define the edge set $E \subseteq V \times V$ such that $(j, i) \in E$ if and only if $\mathcal{M}_{ij} = *$ or $\mathcal{M}_{ij} = ?$. If $(i, j) \in E$, then we call j an *out-neighbor* of i. Also, in order to distinguish between $*$ and $?$ entries in $\mathcal{M}$, we define two subsets E_* and $E_?$ of the edge set E as follows: $(j, i) \in E_*$ if and only if $\mathcal{M}_{ij} = *$ and $(j, i) \in E_?$ if and only if $\mathcal{M}_{ij} = ?$. Then, obviously, $E = E_* \cup E_?$ and $E_* \cap E_? = \varnothing$. To visualize this, we use solid and dashed arrows to represent edges in E_* and $E_?$, respectively.

Example 11.3. As an example, consider the pattern matrix $\mathcal{M}$ given by

$$\mathcal{M} = \begin{bmatrix} 0 & 0 & * & 0 & 0 \\ 0 & * & * & ? & * \\ * & 0 & ? & 0 & 0 \\ 0 & * & 0 & 0 & ? \end{bmatrix}.$$

The associated directed graph $G(\mathcal{M})$ is then given in Figure 11.2.

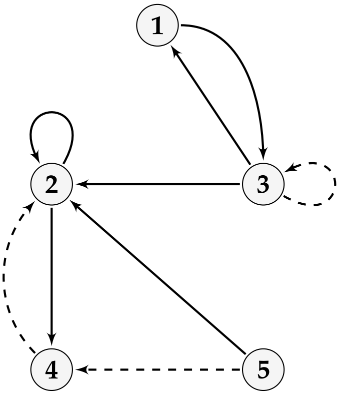

Figure 11.2: The graph $G(\mathcal{M})$ associated with $\mathcal{M}$.

Next, we introduce the notion of *colorability* for $G(\mathcal{M})$:

1. Initially, color all nodes of $G(\mathcal{M})$ white.
2. If a node i has exactly one white out-neighbor j and $(i, j) \in E_*$, we change the color of j to black.

3. Repeat step 2 until no more color changes are possible.

The graph $G(\mathcal{M})$ is called *colorable* if the nodes $1, 2, \ldots, p$ are colored black following the procedure above. Note that the remaining nodes $p+1, \ldots, q$ can never be colored black since they have no incoming edges. We refer to step 2 in the above procedure as the *color change rule*. Similar color change rules have appeared in the literature before (see e.g. [87, 142, 207]). Unlike some of these rules, node i in step 2 does not need to be black in order to change the color of a neighboring node.

Example 11.4. Consider the pattern matrix $\mathcal{M}$ given by

$$\mathcal{M} = \begin{bmatrix} * & 0 & 0 & 0 & * & 0 \\ 0 & ? & 0 & * & 0 & * \\ * & 0 & 0 & * & 0 & 0 \\ 0 & ? & * & * & 0 & 0 \end{bmatrix}.$$

The directed graph $G(\mathcal{M})$ associated with $\mathcal{M}$ is depicted in Figure 11.3. By repeated application of the color change rule as shown in Figure 11.4 to 11.6, we obtain the derived set $\mathcal{D} = \{1, 2, 3, 4\}$. Hence, $G(\mathcal{M})$ is colorable.

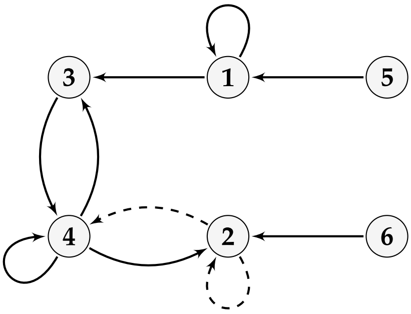

Figure 11.3: The graph $G(\mathcal{M})$.

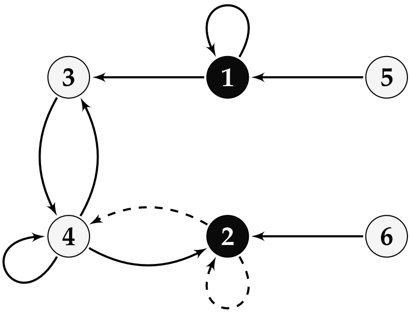

Figure 11.4: Node 5 colors 1 and 6 colors 2.

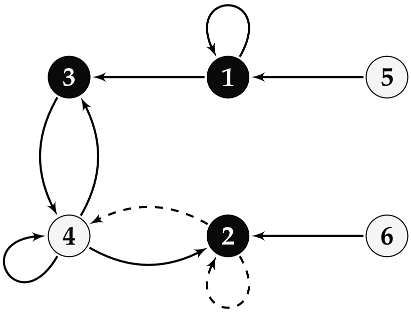

Figure 11.5: Node 1 colors 3.

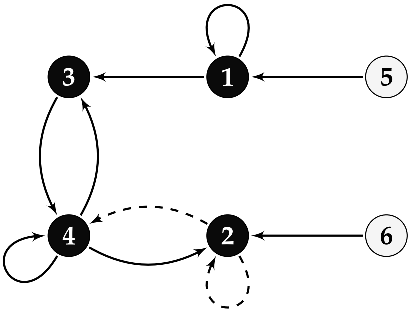

Figure 11.6: Node 3 colors 4.

The following theorem now provides a necessary and sufficient graph-theoretic condition for a given pattern matrix to have full row rank.

Theorem 11.2. Let $\mathcal{M} \in \{0, *, ?\}^{p \times q}$ be a pattern matrix with $p \leqslant q$. Then, $\mathcal{M}$ has full row rank if and only if $G(\mathcal{M})$ is colorable.

It is clear from the definition of the color change rule that colorability of a given graph can be checked in polynomial time.

Finally, based on the rank test in Theorem 11.1 and the result in Theorem 11.2, the following necessary and sufficient graph-theoretic condition for controllability of a given structured system is obtained.

Theorem 11.3. Let $\mathcal{A} \in \{0, *, ?\}^{n \times n}$ and $\mathcal{B} \in \{0, *, ?\}^{n \times m}$ be pattern matrices. Also, let $\bar{\mathcal{A}}$ be obtained from $\mathcal{A}$ by modifying the diagonal entries of $\mathcal{A}$ as follows:

$$\bar{\mathcal{A}}_{ii} := \begin{cases} * & \text{if } \mathcal{A}_{ii} = 0, \\ ? & \text{otherwise.} \end{cases} \tag{11.6}$$

Then, the structured system $(\mathcal{A}, \mathcal{B})$ is controllable if and only if both $G(\begin{bmatrix}\mathcal{A} & \mathcal{B}\end{bmatrix})$ and $G(\begin{bmatrix}\bar{\mathcal{A}} & \mathcal{B}\end{bmatrix})$ are colorable.

As an example, we study controllability of the electrical circuit discussed in Example 11.1.

Example 11.5. According to Example 11.2, the electrical circuit depicted in Figure 11.1 can be modelled as a structured system of the form (11.3). For this example, we have

$$\mathcal{A} = \begin{bmatrix} * & 0 & * \\ 0 & 0 & * \\ ? & * & * \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} * & 0 \\ 0 & * \\ ? & 0 \end{bmatrix}, \quad \text{and} \quad \bar{\mathcal{A}} = \begin{bmatrix} ? & 0 & * \\ 0 & * & * \\ ? & * & ? \end{bmatrix}.$$

The graphs $G(\begin{bmatrix}\mathcal{A} & \mathcal{B}\end{bmatrix})$ and $G(\begin{bmatrix}\bar{\mathcal{A}} & \mathcal{B}\end{bmatrix})$ are depicted in Figure 11.7 and Figure 11.8, respectively. Both graphs are colorable. Indeed, node 5 colors 2, node 2 colors 3, and finally 3 colors 1 in both graphs. Therefore, the system $(\mathcal{A}, \mathcal{B})$ is controllable by Theorem 11.3.

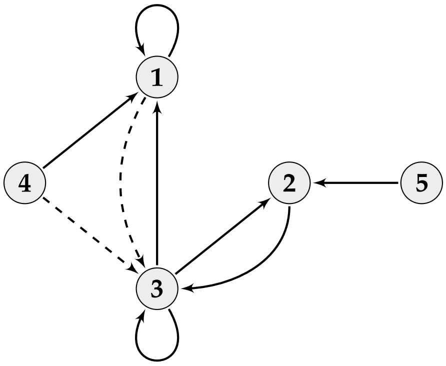

Figure 11.7: The graph $G([\mathcal{A}\ \mathcal{B}])$.

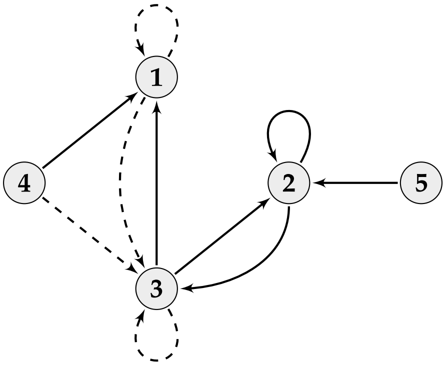

Figure 11.8: The graph $G([\bar{\mathcal{A}}\ \mathcal{B}])$.

By applying Theorem 11.3 to the special case discussed in Corollary 11.1, we obtain the following.

Corollary 11.2. Suppose that none of the diagonal entries of $\mathcal{A}$ is zero. Let $\bar{\mathcal{A}}$ be defined as in (11.6). Then, the system $(\mathcal{A}, \mathcal{B})$ is controllable if and only if $G([\bar{\mathcal{A}}\ \mathcal{B}])$ is colorable.

To conclude this section, the results we have obtained for controllability lead to an interesting observation in the context of structural stabilizability. We say that a structured system $(\mathcal{A}, \mathcal{B})$ is *stabilizable* if the pair (A, B) is stabilizable for all $A \in \mathcal{P}(\mathcal{A})$ and $B \in \mathcal{P}(\mathcal{B})$.

Theorem 11.4. The system $(\mathcal{A}, \mathcal{B})$ is stabilizable if and only if it is controllable.

11.5 DISCUSSION OF EXISTING RESULTS

In this section, we compare our results with those existing in the literature. We focus on the most relevant related work [8, 29, 93, 133, 142, 157, 175, 207]. The structured systems studied in these references are all special cases of those we study in this chapter. In Table 11.1, we summarize the different pattern matrices $\mathcal{A}$ and $\mathcal{B}$ studied in these references. We also include the type of conditions that were developed, i.e., either graph-theoretic, algebraic or both. Note that the references [29, 142, 207] study controllability in a network context, where the pattern matrix $\mathcal{B}$ has a particular structure in the sense that each column has exactly one $*$-entry, and each row has at most one $*$-entry. Additionally, the paper [142] considers a particular class of systems where the diagonal entries of $\mathcal{A}$ are all ? and none of the off-diagonal entries is ?. In the following two subsections, we elaborate on the existing graph-theoretic conditions and algebraic conditions, respectively. In both sections, we also compare these results to the present work.

Ref.	$\mathcal{A}$	$\mathcal{B}$	Conditions	
			GTC	AC
[133]	$\{0, *\}^{n \times n}$	$\{0, *\}^{n \times 1}$	✓	–
[157]			–	✓
[8]		$\{0, *\}^{n \times m}$	✓	–
[175]			✓	✓
[93]			✓	✓
[29]		particular $\{0, *\}^{n \times m}$	–	✓
[207]			–	✓
[142]	particular $\{0, *, ?\}^{n \times n}$		✓	✓

Table 11.1: An overview of previous work. Graph-theoretic conditions are abbreviated by "GTC" and algebraic conditions by "AC".

11.5.1 Graph-theoretic conditions

The graph-theoretic conditions provided in [133, Thm. 1] for the single-input case ($m = 1$) and extended to the multi-input case in [8, Satz 3] are based on the graph $G = (V, E)$ associated with a pattern matrix $\begin{bmatrix}\mathcal{A} & \mathcal{B}\end{bmatrix}$ where $\mathcal{A} \in \{0, *\}^{n \times n}$ and $\mathcal{B} \in \{0, *\}^{n \times m}$. Note that $V = \{1, 2, \ldots, n + m\}$ in this case. The graph-theoretic characterization in [8, Satz 3] (or in [133, Thm. 1] if $m = 1$) consists of three conditions. The first one requires checking the so-called accessibility of each node in $\{1, 2, \ldots, n\}$ from the nodes in $\{n + 1, n + 2, \ldots, n + m\}$. The remaining two conditions require checking certain relations for *all* subsets of $\{1, 2, \ldots, n\}$. As such, the computational complexity of checking these conditions is *at least* exponential in n. Note that, in contrast, the computational complexity of checking the colorability conditions of our Theorem 11.3 is polynomial in n.

The paper [133] provides another set of graph-theoretic conditions, stated, more specifically, in [133, Thm. 2] (only for the case $m = 1$). As argued in [133, p. 135], this theorem performs better than [133, Thm. 1] for sparse graphs. Essentially, the conditions given in [133, Thm. 2] require checking the existence of a unique serial buds cactus as well as nonexistence of certain cycles within the graph G. How these conditions can be checked in an algorithmic manner is not clear, whereas the colorability conditions given in Theorem 11.3 can be checked by a simple algorithm.

On top of the advantages of computational complexity, the conditions provided in Theorem 11.3 are more attractive because of their conceptual simplicity. Indeed, colorability is a simpler and more intuitive notion than those appearing in the results of [133] and [8].

Yet another graph-theoretical characterization is provided in [93, Thm. 5]. In order to verify the conditions of [93, Thm. 5], one needs to check whether a unique spanning cycle family with certain properties exists in $\binom{n+m}{n}$ directed graphs obtained from the pattern matrices $\mathcal{A}$ and $\mathcal{B}$. Needless to say, checking the conditions of Theorem 11.3 is much easier than checking these conditions.

Also in the context of networked systems, graph-theoretic conditions for strong structural controllability have been obtained (see e.g. [29, 142, 207]). To elaborate further on the relationship between the work on networked systems and our work, we first need to explain the framework of the papers [29, 142, 207]. The starting point of these papers is a directed graph $H = (W, F)$ where $W = \{1, 2, \ldots, n\}$ denotes the node set and F the edge set. The graphs considered in [29, 207] are so-called loop graphs, that are graphs which are allowed to contain self-loops, whereas [142] does not allow self-loops. Apart from the graph H, these papers consider a subset of the node set W, the so-called leader set, say $W_L = \{w_1, w_2, \ldots, w_m\}$. Based on the graph H and W_L, [29, 142, 207] introduce systems of the form (11.3) where the pattern matrix $\mathcal{B}$ is defined by

$$\mathcal{B}_{ij} = \begin{cases} * & \text{if } i = w_j \\ 0 & \text{otherwise} \end{cases} \tag{11.7}$$

for $i \in \{1, 2, \ldots, n\}$, $j \in \{1, 2, \ldots, m\}$. In [29] and [207] the pattern matrix $\mathcal{A}$ is defined by

$$\mathcal{A}_{ij} = \begin{cases} * & \text{if } (j, i) \in F \\ 0 & \text{otherwise} \end{cases} \tag{11.8}$$

whereas in [142] the pattern matrix $\mathcal{A}$ is defined by

$$\mathcal{A}_{ij} = \begin{cases} * & \text{if } (j, i) \in F \\ ? & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{11.9}$$

for $i, j \in \{1, 2, \ldots, n\}$.

In [29], the authors first define two bipartite graphs obtained from the pattern matrices $\mathcal{A}$ and $\mathcal{B}$. Then, they show in [29, Thm. 5] that $(\mathcal{A}, \mathcal{B})$ is strongly structurally controllable if and only if there exist so-called constrained matchings with certain properties in these bipartite graphs. Later, in [207, Thm. 5.4] an equivalence between the existence of constrained matchings and so-called zero forcing sets for loop graphs was established. To explain this in more detail, we need to introduce the notion of zero forcing that was originally studied in the context of minimal rank problems (see e.g. [87]).

Let $H = (W, F)$ be a directed loop graph and $S \subseteq W$. Color all nodes in S black and the others white.

If a node i (of any color) has exactly one white out-neighbor j, we change the color of j to black and write $i \rightarrow j$. If all the nodes in W can be colored black by repeated application of this color change rule, we say that S is a *loopy zero forcing set* for H. Given a loopy zero forcing set, we can list the color changes in the order in which they were performed to color all nodes black. This list is called a chronological list of color changes.

In order to quote [207, Thm. 5.5], we need two more definitions. Define $W_{\text{loop}} \subseteq W$ to be the subset of all nodes with self-loops and let H^* be the graph obtained from H by placing a self-loop at every node.

Theorem 11.5. [207, Thm. 5.5] Let H be a directed loop graph and W_L be a leader set. Consider the pattern matrices defined in (11.7) and (11.8). Then, the structured system $(\mathcal{A}, \mathcal{B})$ is controllable if and only if the following conditions hold:

1. W_L is a loopy zero forcing set for H.

2. W_L is a loopy zero forcing set for H^* for which there is a chronological list of color changes that does not contain a color change of the form $i \to i$ with $i \in W_{\text{loop}}$.

A result similar to this theorem was obtained in [142] for controllability of pattern matrices defined by (11.7) and (11.9) that are obtained from a graph H *without* self-loops. However, in order to deal with this class of pattern matrices, [142] introduces a slightly different notion of zero forcing to be defined below.

Let $H = (W, F)$ be a directed graph without self-loops and $S \subseteq W$. Color all nodes in S black and the others white. If a *black* node i has exactly one white out-neighbor j, we change the color of j to black. If all the nodes in W can be colored black by repeated application of this color change rule, we say that S is a *ordinary zero forcing set* for H.

We now state the graph-theoretic characterization of controllability established in [142].

Theorem 11.6. [142, Thm. IV.4] Let H be a directed graph without self-loops and W_L be a leader set. Consider the pattern matrices given by (11.7) and (11.9). Then, the structured system $(\mathcal{A}, \mathcal{B})$ is controllable if and only if W_L is an ordinary zero forcing set for H.

Even though Theorems 11.5 and 11.6 present conditions that are similar in nature, it is not possible to compare these results immediately as they deal with two different and non-overlapping system classes. Indeed, the pattern matrices considered in [207] (given by (11.8)) *do not contain* any ? entries whereas those studied in [142] (given by (11.9)) contain *only* ? entries on their diagonals.

Next, we will show that the conditions of Theorem 11.3 are equivalent to those of Theorems 11.5 and 11.6 if specialized to the corresponding pattern matrices. This will shed light on the relationship between these results based on the different zero forcing notions.

We start with Theorem 11.5. According to our color change rule, the nodes belonging to W_L will be colored black in both $G([\mathcal{A}\ \mathcal{B}])$ and $G([\bar{\mathcal{A}}\ \mathcal{B}])$ because $\mathcal{B}$ is a pattern matrix with structure defined by (11.7). Since $\mathcal{A}$ does not contain ? entries, $G([\mathcal{A}\ \mathcal{B}])$ is colorable if and only if W_L is a loopy zero forcing set for $G(\mathcal{A})$. By noting that $H = G(\mathcal{A})$, we see that the first condition in Theorem 11.3 is equivalent to that of Theorem 11.5. Now, let the pattern matrix $\mathcal{A}^*$ be such that $H^* = G(\mathcal{A}^*)$. Since $W_{\text{loop}} = \{i \mid \bar{\mathcal{A}}_{ii} = ?\}$, we see that $G([\bar{\mathcal{A}}\ \mathcal{B}])$ is colorable if and only if the second condition of Theorem 11.5 holds. Thus, the second condition of Theorem 11.3 is equivalent to that of Theorem 11.5.

Now, we turn attention to Theorem 11.6. It follows from (11.6) and (11.9) that $\bar{\mathcal{A}} = \mathcal{A}$, i.e., graphs $G(\begin{bmatrix}\bar{\mathcal{A}} & \mathcal{B}\end{bmatrix})$ and $G(\begin{bmatrix}\mathcal{A} & \mathcal{B}\end{bmatrix})$ are the same. As in the discussion above, the nodes belonging to W_L will be colored black in $G(\begin{bmatrix}\bar{\mathcal{A}} & \mathcal{B}\end{bmatrix})$ because $\mathcal{B}$ is a pattern matrix with structure defined by (11.7). According to our color change rule, a white node can never color any other white node in $G(\begin{bmatrix}\bar{\mathcal{A}} & \mathcal{B}\end{bmatrix})$ since $(i,i) \in E_?$ for every node i of $G(\bar{\mathcal{A}})$. This means that $G(\begin{bmatrix}\bar{\mathcal{A}} & \mathcal{B}\end{bmatrix})$ is colorable if and only if W_L is an ordinary zero forcing set for $G(\bar{\mathcal{A}})$. By noting that $H = G(\mathcal{A}) = G(\bar{\mathcal{A}})$, we see that the conditions in Theorem 11.3 are equivalent to the single condition of Theorem 11.6.

11.5.2 Algebraic conditions

In this subsection, we will compare our rank tests for strong structural controllability with those provided in [29, 142, 175]. More precisely, we will show that the rank tests in Theorem 11.1 reduce to those in [29, 142, 175] for the corresponding special cases of pattern matrices.

An algebraic condition for controllability of $(\mathcal{A}, \mathcal{B})$ was provided in [175, Thm. 2] for $\mathcal{A} \in \{0,*\}^{n\times n}$ and $\mathcal{B} \in \{0,*\}^{n\times m}$. Later, these conditions were reformulated in [29, Thm. 3]. These conditions rely on a matrix property that will be defined next for pattern matrices that may also contain ? entries.

Definition 11.2. Consider a pattern matrix $\mathcal{M} \in \{0,*,?\}^{p\times q}$ with $p \leqslant q$. The matrix $\mathcal{M}$ is said to be of Form III if there exist two permutation matrices P_1 and P_2 such that

$$P_1\mathcal{M}P_2 = \begin{bmatrix} \otimes & \cdots & \otimes & * & 0 & \cdots & 0 \\ \vdots & & \vdots & \ddots & \ddots & \ddots & \vdots \\ \otimes & \cdots & \otimes & \cdots & \otimes & * & 0 \\ \otimes & \cdots & \otimes & \cdots & \otimes & \otimes & * \end{bmatrix}, \tag{11.10}$$

where the symbol $\otimes$ indicates an entry that can be either 0, $*$ or ?.

The above-mentioned algebraic conditions are stated next.

Theorem 11.7. [29, Thm. 3] Let $\mathcal{A} \in \{0,*\}^{n\times n}$ and $\mathcal{B} \in \{0,*\}^{n\times m}$ be two pattern matrices. Also, let $\mathcal{A}^*$ be the pattern matrix obtained from $\mathcal{A}$ by replacing all diagonal entries by $*$. The system $(\mathcal{A}, \mathcal{B})$ is controllable if and only if the following two conditions hold:

1. The matrix $\begin{bmatrix}\mathcal{A} & \mathcal{B}\end{bmatrix}$ is of Form III.

2. The matrix $\begin{bmatrix}\mathcal{A}^*\mathcal{B}\end{bmatrix}$ is of Form III with the additional property that $*$ entries appearing in (11.10) do not originate from diagonal elements in $\mathcal{A}$ that are $*$ entries.

It can be shown that our algebraic conditions in Theorem 11.1 are equivalent to those in Theorem 11.7 for the special case of pattern matrices that only contain 0 and $*$ entries. Recall that it follows from Theorem 11.1 that $(\mathcal{A}, \mathcal{B})$ is controllable if and only if both $\left[\mathcal{A}\ \mathcal{B}\right]$ and $\left[\bar{\mathcal{A}}\ \mathcal{B}\right]$ have full row rank, where $\bar{\mathcal{A}}$ is given in (11.6). To relate our algebraic conditions with the ones in Theorem 11.7, we need the following lemma.

Lemma 11.1. Let $\mathcal{M} \in \{0, *, ?\}^{p \times q}$ with $p \leqslant q$. Then, $\mathcal{M}$ has full row rank if and only if $\mathcal{M}$ is of Form III.

From Lemma 11.1 it immediately follows that $\left[\mathcal{A}\ \mathcal{B}\right]$ has full row rank if and only if $\left[\mathcal{A}\ \mathcal{B}\right]$ is of Form III. Hence, the first condition of Theorem 11.1 is equivalent to that of Theorem 11.7. We will now also show that $\left[\bar{\mathcal{A}}\ \mathcal{B}\right]$ has full row rank if and only if the second condition of Theorem 11.7 holds. From Lemma 11.1, we have that $\left[\bar{\mathcal{A}}\ \mathcal{B}\right]$ has full row rank if and only if $\left[\bar{\mathcal{A}}\ \mathcal{B}\right]$ is of Form III. By definition of $\bar{\mathcal{A}}$ and $\mathcal{A}^*$, it follows that $\bar{\mathcal{A}}_{ij} = \mathcal{A}^*_{ij}$ for all $i \neq j$. If $\mathcal{A}_{ii} = 0$ then both $\bar{\mathcal{A}}_{ii} = *$ and $\mathcal{A}^*_{ii} = *$. On the other hand, if $\mathcal{A}_{ii} = *$ then $\bar{\mathcal{A}}_{ii} = ?$ and $\mathcal{A}^*_{ii} = *$. To sum up, $\bar{\mathcal{A}}_{ij} \neq \mathcal{A}^*_{ij}$ if and only if $i = j$ and $\mathcal{A}_{ii} = *$. In other words, all entries of $\bar{\mathcal{A}}$ and $\mathcal{A}^*$ are *the same*, except for those that correspond to the diagonal elements of $\mathcal{A}$ that are $*$ entries. Hence, there exist two permutation matrices P_1 and P_2 such that all entries of the matrices $P_1 \left[\bar{\mathcal{A}}\ \mathcal{B}\right] P_2$ and $P_1 \left[\mathcal{A}^*\ \mathcal{B}\right] P_2$ are the same, except those that originate from diagonal elements of $\mathcal{A}$ that are $*$ entries. This implies that $\left[\bar{\mathcal{A}}\ \mathcal{B}\right]$ is of Form III if and only if $\left[\mathcal{A}^*\ \mathcal{B}\right]$ is of Form III with the additional property that the $*$ entries in (11.10) do not originate from diagonal elements in $\mathcal{A}$ that are $*$ entries. In other words, the second conditions of Theorem 11.1 and 11.7 are equivalent. Since also the first conditions in these theorems are equivalent, we conclude that the algebraic conditions in Theorem 11.1 are equivalent to those in Theorem 11.7 for the special case in which $\mathcal{A} \in \{0, *\}^{n \times n}$ and $\mathcal{B} \in \{0, *\}^{n \times m}$.

A different algebraic condition was introduced in [142] for systems defined on simple directed graphs. The pattern matrices of such systems can be represented by $\mathcal{A}$ and $\mathcal{B}$ given by (11.9) and (11.7), respectively. The algebraic condition referred to above is then stated as follows.

Theorem 11.8. [142, Lem. IV.1] Consider the pattern matrices $\mathcal{A}$ and $\mathcal{B}$ given by (11.9) and (11.7), respectively. Then, $(\mathcal{A}, \mathcal{B})$ is controllable if and only if $\left[\mathcal{A}\ \mathcal{B}\right]$ has full row rank.

In order to see that this theorem follows from Corollary 11.1, note that $\mathcal{A} = \bar{\mathcal{A}}$ since all diagonal entries of $\mathcal{A}$ are ?'s.

11.6 PROOFS

11.6.1 Proof of Theorem 11.1

Proof. To prove the "only if" part, assume that $(\mathcal{A}, \mathcal{B})$ is controllable. By the Hautus test [208, Thm. 3.13] and the definition of strong structural controllability, it follows that $\left[A - \lambda I\ \ B\right]$ has full row rank for all $(A, B) \in \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{B})$ and all $\lambda \in \mathbb{C}$. By substitution of $\lambda = 0$ we conclude that condition 1 is satisfied.

To prove that condition 2 also holds, suppose that $x^T \begin{bmatrix} \bar{A} & B \end{bmatrix} = 0$ for some pair $(\bar{A}, B) \in \mathcal{P}(\bar{\mathcal{A}}) \times \mathcal{P}(\mathcal{B})$ and $x \in \mathbb{R}^n$. We want to prove that $x = 0$. Let $\alpha \in \mathbb{R}$ be a nonzero real number such that $\alpha \notin \{\bar{A}_{ii} \mid i \text{ is such that } \mathcal{A}_{ii} = *\}$. Then, define a nonsingular diagonal matrix $X \in \mathbb{R}^{n \times n}$ as

$$X_{ii} = \begin{cases} 1 & \text{if } \bar{\mathcal{A}}_{ii} = ? \\ \alpha / \bar{A}_{ii} & \text{if } \bar{\mathcal{A}}_{ii} = *. \end{cases}$$

It is clear that $\bar{A}X \in \mathcal{P}(\bar{\mathcal{A}})$ and $x^T \begin{bmatrix} \bar{A}X & B \end{bmatrix} = 0$. Furthermore, by the choice of α and X we obtain $\hat{A} := \bar{A}X - \alpha I \in \mathcal{P}(\mathcal{A})$. By assumption, $\begin{bmatrix} \hat{A} + \alpha I & B \end{bmatrix}$ has full row rank (by substitution of $\lambda = -\alpha$). In other words, $\begin{bmatrix} \bar{A}X & B \end{bmatrix}$ has full row rank and therefore $x = 0$. We conclude that condition 2 is satisfied.

To prove the "if" part, assume that conditions 1 and 2 are satisfied. Suppose that $z^H \begin{bmatrix} A - \lambda I & B \end{bmatrix} = 0$ for some $(A, B) \in \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{B})$ and $(\lambda, z) \in \mathbb{C} \times \mathbb{C}^n$, and z^H denotes the conjugate transpose of z. We want to prove that $z = 0$. Note that if $\lambda = 0$, it readily follows that $z = 0$ by condition 1. Therefore, it remains to be shown that $z = 0$ if $\lambda \neq 0$. To this end, write $z = \xi + j\eta$, where $\xi, \eta \in \mathbb{R}^n$ and j denotes the imaginary unit. Next, let $\alpha \in \mathbb{R}$ be a nonzero real number such that

$$\alpha \notin \left\{ -\frac{\xi_i}{\eta_i} \mid \eta_i \neq 0 \right\} \cup \left\{ -\frac{(\xi^T A)_i}{(\eta^T A)_i} \mid (\eta^T A)_i \neq 0 \right\}.$$

We define $x := \xi + \alpha\eta$. Now, we claim that

(a) $x_i = 0$ if and only if $z_i = 0$.

(b) $x_i = 0$ if and only if $(x^T A)_i = 0$.

Note that (a) follows directly from the definition of x and the choice of α. To prove the "only if" part of (b), suppose that $x_i = 0$. By (a), this implies that $z_i = 0$. Since $z^H A = \lambda z^H$, we see that $(z^H A)_i = 0$. Equivalently, $((\xi^T - j\eta^T)A)_i = 0$. Therefore, both $(\xi^T A)_i = 0$ and $(\eta^T B)_i = 0$. We conclude that $(x^T A)_i = ((\xi^T + \alpha\eta^T)A)_i = 0$.

To prove the "if" part of (b), suppose that $(x^T A)_i = 0$. This means that $((\xi^T + \alpha\eta^T)A)_i = 0$. Equivalently, $(\xi^T A)_i + \alpha(\eta^T A)_i = 0$. By the choice of α, this implies that $(\xi^T A)_i = (\eta^T A)_i = 0$. We conclude that $(z^H A)_i = 0$. Recall that $z^H A = \lambda z^H$, where λ was assumed to be *nonzero*. This implies that $z_i = 0$. Again, using (a) we conclude that $x_i = 0$. This proves (b).

Next, we define the diagonal matrix $X' \in \mathbb{R}^{n \times n}$ as

$$X'_{ii} = \begin{cases} 1 & \text{if } x_i = 0 \\ \frac{(x^T A)_i}{x_i} & \text{otherwise.} \end{cases}$$

We know that X' is nonsingular by (b). By definition of X' we have $x^T A = x^T X'$. Furthermore, as $z^H B = 0$ we obtain $\xi^T B = \eta^T B = 0$ and therefore $x^T B = 0$. Hence $x^T \begin{bmatrix} A - X' & B \end{bmatrix} = 0$. Since X' is nonsingular, it follows that $A - X' \in \mathcal{P}(\bar{\mathcal{A}})$. By condition 2, this means that $x = 0$. Finally, we conclude that $z = 0$ using (a). □

11.6.2 Proof of Theorem 11.2

To prove Theorem 11.2, we need the following auxiliary result.

Lemma 11.2. Let $\mathcal{M} \in \{0, *, ?\}^{p \times q}$ be a pattern matrix with $p \leqslant q$. Consider the directed graph $G(\mathcal{M})$. Suppose that each node is colored white or black. Let $D \in \mathbb{R}^{p \times p}$ be the diagonal matrix defined by

$$D_{kk} = \begin{cases} 1 & \text{if node } k \text{ is black,} \\ 0 & \text{otherwise.} \end{cases}$$

Suppose further that $j \in \{1, 2, \ldots, p\}$ is a node for which there exists a node $i \in \{1, 2, \ldots, p\}$, possibly identical to j, such that j is the only white out-neighbor of i and $(i, j) \in E_*$. Then for all $M \in \mathcal{P}(\mathcal{M})$ we have that $\begin{bmatrix} M & D \end{bmatrix}$ has full row rank if and only if $\begin{bmatrix} M & D + e_j e_j^T \end{bmatrix}$ has full row rank where e_j denotes the jth column of I.

Proof. The "only if" part is trivial. To prove the "if" part, suppose that $M \in \mathcal{P}(\mathcal{M})$ and $\begin{bmatrix} M & D + e_j e_j^T \end{bmatrix}$ has full row rank. Let $z \in \mathbb{R}^p$ be such that $z^T \begin{bmatrix} M & D \end{bmatrix} = 0$. Our aim is to show that $z_j = 0$. Indeed, if z_j is zero then $z^T \begin{bmatrix} M & D + e_j e_j^T \end{bmatrix} = z^T \begin{bmatrix} M & D \end{bmatrix} = 0$ and hence z must be zero. This would prove that $\begin{bmatrix} M & D \end{bmatrix}$ has full row rank. We will distinguish two cases: $i = j$ and $i \neq j$. Suppose first that $i = j$. Let $\beta, \omega \subseteq \{1, 2, \ldots, p\}$ be defined as the index sets $\beta = \{k \mid k \neq j \text{ and } k \text{ is black}\}$ and $\omega = \{\ell \mid \ell \neq j \text{ and } \ell \text{ is white}\}$. In the sequel, to simplify the notations, for a given vector $z \in \mathbb{R}^p$ and a given index set $\alpha \subseteq \{1, 2, \ldots, p\}$, we define $z_\alpha := \{x \in \mathbb{R}^{|\alpha|} \mid x_i = z_{\alpha(i)}, i \in \{1, 2, \ldots, |\alpha|\}\}$, where $|\alpha|$ is the cardinality of α. From $z^T M = 0$, we get

$$z_j M_{jj} + z_\beta^T M_{\beta j} + z_\omega^T M_{\omega j} = 0. \tag{11.11}$$

Since j is the only white out-neighbor of itself, we must have that M_{jj} is nonzero and that $M_{\omega j}$ is a zero vector. Moreover, it follows from $z^T D = 0$ that z_β must a zero vector. Therefore, (11.11) implies that z_j must be zero.

Next, suppose that $i \neq j$. Let $\beta, \omega \subseteq \{1, 2, \ldots, p\}$ be defined as the index sets $\beta = \{k \mid k \neq i,\, k \neq j, \text{ and } k \text{ is black}\}$ and $\omega = \{\ell \mid \ell \neq i,\, \ell \neq j, \text{ and } \ell \text{ is white}\}$. From $z^T M = 0$, we now get

$$z_i M_{ii} + z_j M_{ji} + z_\beta^T M_{\beta i} + z_\omega^T M_{\omega i} = 0. \tag{11.12}$$

Since j is the only white out-neighbor of i, we must have that M_{ji} is nonzero and that

vector. Therefore, (11.12) implies that

$$z_i M_{ii} + z_j M_{ji} = 0. \tag{11.13}$$

Now, we distinguish two cases: i is black and i is white. If i is black, then we have that z_i is zero because $z^T D = 0$. Therefore, (11.13) implies that $z_j = 0$ as desired.

Finally, if i is white, then we have that $M_{ii} = 0$ for otherwise i would have two white out-neighbors. Again, (11.13) implies that z_j is zero. This completes the proof. □

Now, we can give the proof of Theorem 11.2.

Proof of Theorem 11.2. To prove the "if" part, suppose that $G(\mathcal{M})$ is colorable. Let $M \in \mathcal{P}(\mathcal{M})$. By repeated application of Lemma 11.2, it follows that M has full row rank if and only if $\begin{bmatrix} M & I \end{bmatrix}$ has full row rank, which is obviously true. Therefore, we conclude that M has full row rank.

To prove the "only if" part, suppose that $\mathcal{M}$ has full row rank but $G(\mathcal{M})$ is not colorable. Let C be the set of nodes that are colored black by repeated application of the color change rule until no more color changes are possible. Then, C is a *strict* subset of $\{1, 2, \dots, p\}$. Thus, possibly after reordering the nodes, we can partition $\mathcal{M}$ as

$$\mathcal{M} = \begin{bmatrix} \mathcal{M}_1 \\ \mathcal{M}_2 \end{bmatrix},$$

where the rows of the matrix $\mathcal{M}_1$ correspond to the nodes in C and the matrix $\mathcal{M}_2$ correspond to the remaining white nodes. Note that $C = \varnothing$ means that $\mathcal{M}_2 = \mathcal{M}$ and $\mathcal{M}_1$ is absent. Since no more color changes are possible, there is no column of $\mathcal{M}_2$ that has exactly one $*$ entry while all other entries are 0. Therefore, for any column of $\mathcal{M}_2$, we have one of the following three cases:

a. All entries are 0.

b. There exists exactly one $?$ entry while all other entries are 0.

c. At least two entries belong to the set $\{*, ?\}$.

Consequently, there exists a matrix $M_2 \in \mathcal{P}(\mathcal{M}_2)$ such that its column sums are zero, that is $\mathbb{1}^T M_2 = 0$, where $\mathbb{1}$ denotes the vector of ones of appropriate size. Take any $M_1 \in \mathcal{P}(\mathcal{M}_1)$. Then

$$M = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} \in \mathcal{P}(\begin{bmatrix} \mathcal{M}_1 \\ \mathcal{M}_2 \end{bmatrix}) = \mathcal{P}(\mathcal{M})$$

satisfies $\begin{bmatrix} 0^T & \mathbb{1}^T \end{bmatrix} \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = 0$. Hence, M does not have full row rank and we have reached a contradiction. □

11.6.3 Proof of Theorem 11.3

Proof. By Theorems 11.1 and 11.2, we have that $\begin{bmatrix} \mathcal{A} & \mathcal{B} \end{bmatrix}$ is controllable if and only if $G(\begin{bmatrix} \mathcal{A} & \mathcal{B} \end{bmatrix})$ and $G(\begin{bmatrix} \bar{\mathcal{A}} & \mathcal{B} \end{bmatrix})$ are colorable. □

11.6.4 Proof of Theorem 11.4

Proof. The "if" part is evident. Therefore, it is enough to prove the "only if" part. Suppose that the system $(\mathcal{A}, \mathcal{B})$ is stabilizable. Let $(A, B) \in \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{B})$. Then, (A, B) is stabilizable. Note that $-A \in \mathcal{P}(\mathcal{A})$. Therefore, both (A, B) and $(-A, B)$ are stabilizable. It follows from the Hautus test for stabilizability (see e.g. [208, Thm. 3.32]) that (A, B) is controllable. Consequently, the system $(\mathcal{A}, \mathcal{B})$ is controllable. □

11.6.5 Proof of Lemma 11.1

Proof. Since the "if" part is evident, it remains to prove the "only if" part. Suppose that $\mathcal{M}$ has full row rank. From Theorem 11.2, it follows that $G(\mathcal{M})$ is colorable. In particular, there exist $i \in \{1, 2, \ldots, q\}$ and $j \in \{1, 2, \ldots, p\}$ such that $\mathcal{M}_{ji} = *$ and $\mathcal{M}_{ki} = 0$ for all $k \neq j$. Therefore, we can find permutation matrices P_1' and P_2' such that

$$P_1' \mathcal{M} P_2' = \left[\begin{array}{c|c} \mathcal{M}' & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline \otimes \cdots \otimes & * \end{array}\right]$$

where the symbol $\otimes$ indicates an entry that can be either 0, $*$ or $?$. Note that M has full row rank for all $M \in \mathcal{P}(\mathcal{M})$ if and only if M' has full row rank for all $M \in \mathcal{P}(\mathcal{M}')$. Therefore, repeated application of the argument above results in permutation matrices P_1 and P_2 such that (11.10) holds. □

11.7 CONCLUSIONS

In most of the existing literature on strong structural controllability of structured systems, a zero/nonzero structure of the system matrices is assumed to be given. However, in many physical systems or linear networked systems, apart from fixed zero entries and nonzero entries we need to allow a third kind of entries, namely those that can take arbitrary (zero or nonzero) values. To deal with this, we have extended the notion of zero/nonzero structure to what we have called zero/nonzero/arbitrary structure. We have formalized this more general class of structured systems using pattern matrices containing fixed zero, arbitrary nonzero and arbitrary entries. In this setup, we have established necessary and sufficient algebraic conditions for strong structural controllability of these systems in terms of full rank tests on two associated pattern matrices. Moreover, a necessary and sufficient graph-theoretic condition for a given pattern matrix to have full row rank has been given in terms of a new color change rule. We have then established a graph-theoretic test for strong structural controllability of the new class of structured systems. Finally, we have shown how our results generalize previous work. We have also shown that some existing results [142, 207] that are seemingly incomparable to ours, can be put in our framework, thus unveiling an overarching theory.

In addition to strong structural controllability, weak structural controllability and strong structural stabilizability of zero/nonzero/arbitrary structured systems has been analyzed. We have shown that weak structural controllability of our structured systems can be checked using tests that already exist in the literature. We have also shown that a structured system with zero/nonzero/arbitrary structure is strongly structurally stabilizable if and only if it is strongly structurally controllable.

It would be interesting to adopt our new framework of structured systems to other problem areas in systems and control, such as network identifiability [223] or fault detection and isolation [174]. This is left as a possibility for future research.

12 PROPERTIES OF PATTERN MATRICES

In this chapter we take a closer look at properties of pattern matrices. As in the previous chapter, we consider pattern matrices with three types of entries: zero, nonzero and arbitrary. We will introduce addition and multiplication of such pattern matrices. Thereafter, we will investigate the pattern classes of matrices that are either the sum or product of two pattern matrices. These results are then applied to three structural problems, namely strong structural input-state observability, output controllability, and controllability of differential algebraic equations. In each of these problems we will see that addition and multiplication of pattern matrices plays an important role.

12.1 INTRODUCTION

Often, the exact parameters of a dynamical system are unavailable. Nonetheless, in many scenarios we do know something about the *structure* of these parameters, for example that some of them are nonzero. It is useful to capture such prior knowledge by a so-called pattern matrix. A pattern matrix is simply an array of symbols, where each of the symbols represents some prior information. In this chapter[1] we focus on pattern matrices having three types of entries: "0", "$*$" and "?". Naturally, 0 captures the prior knowledge that a parameter is zero, while $*$ represents nonzero real numbers and ? represents arbitrary (zero or nonzero) real numbers. Given a pattern matrix, its *pattern class* is the set of all real matrices having the same zero/nonzero/arbitrary structure as the pattern matrix.

Sometimes we can conclude that a property of a dynamical system holds *for all* system matrices in a pattern class. Such properties are referred to as *strong structural* properties. Strong structural properties are thus independent of the particular numerical parameters of the system, and instead depend on the configuration of symbols in the involved pattern matrices. These structural properties are valuable since they allow us to ascertain system-theoretic properties of a system even though its parameters are uncertain.

The purpose of the chapter is to take a closer look at properties of pattern matrices. In particular, we will introduce pattern matrix addition and multiplication. We then investigate the pattern class of matrices that are either the sum or product of two pattern matrices. We will conclude that the pattern class of the sum of two pattern matrices is equal to the sum of the pattern classes of

[1] We note that other types of pattern matrices have also been studied, see for instance the work on sign patterns [78, 148].

each of the matrices. However, a similar equality generally does not hold for multiplication. Nonetheless, we can show that the product of two pattern classes of pattern matrices is *contained* in the pattern class of the product of these pattern matrices.

Next, we follow up by applying our results on pattern matrix addition and multiplication to strong structural system properties. In particular, we focus on three strong structural properties, namely controllability of differential algebraic equations (DAE's), input-state observability, and output controllability. Strong structural controllability of DAE's was studied in [170], and the different notion of weak (or generic) structural controllability was also considered in [176, 244]. Our approach differs from [170] in the sense that we relate strong structural controllability to full rank properties of certain (sums of) pattern matrices.

Strong and weak structural input-state observability were studied in a network setting in [72]. Here, a zero/nonzero pattern was considered and input-state observability was characterized in terms of a graph underlying this pattern. In this chapter, we consider more general zero/nonzero/arbitrary pattern matrices and characterize input-state observability algebraically, via full rank properties of sums of pattern matrices. Our results can also be verified in a graph-theoretic manner using a so-called color change rule [96].

Finally, strong structural output controllability has been considered in [141, 219] and its weak structural version was also studied in [63, 109, 144]. Both [141] and [219] study the problem in a network context, where the input and output matrices have a particular structure. In this context, output controllability is often referred to as *targeted controllability*. In these papers, also the state matrix has particular structure in the sense that its pattern has only arbitrary diagonal entries and zero/nonzero off-diagonal entries. On top of this, the paper [219] considers a subclass of the pattern class, called the *distance-information preserving* subclass. In this chapter, we study strong structural output controllability in the more general setting where the structure of the system matrices is any zero/nonzero/arbitrary pattern. We will see that the multiplication of pattern matrices plays in important role in the characterization of strong structural output controllability.

The organization of this chapter is as follows. In Section 12.2 we introduce addition and multiplication of pattern matrices. Subsequently, in Section 12.3 we treat applications. In particular, we consider strong structural controllability of DAE's in Section 12.3.1, input-state observability in Section 12.3.2 and output controllability in Section 12.3.3. Finally, we provide our conclusions in Section 12.4.

12.2 PATTERN MATRICES

In this section, we will review the concept of pattern matrix. A particular type of pattern matrices was introduced in [96] in order to formalize the idea of matrices whose entries are not known precisely but are known to be zero, nonzero or arbitrary real numbers. More precisely, a *pattern matrix* is a matrix with entries from the set of symbols $\{0, *, ?\}$. Here $*$ represents nonzero real numbers and ? represents arbitrary real numbers, as is made precise in the following definition.

Definition 12.1. Let $\mathcal{A} \in \{0, *, ?\}^{m \times n}$. The set

$$\mathcal{P}(\mathcal{A}) = \left\{ A \in \mathbb{R}^{m \times n} \;\middle|\; \begin{matrix} A_{ij} = 0 \text{ if } \mathcal{A}_{ij} = 0 \\ A_{ij} \neq 0 \text{ if } \mathcal{A}_{ij} = * \end{matrix} \right\}$$

is called the *pattern class* of $\mathcal{A}$.

We can define properties of pattern matrices in terms of the properties of the real matrices in their pattern classes. For example, we say that a pattern matrix $\mathcal{A}$ has full rank if A has full rank for all $A \in \mathcal{P}(\mathcal{A})$. Rank properties will be crucial in the applications to structured systems as most system-theoretic properties are characterized in terms of full rank conditions. Fortunately, conditions under which a pattern matrix has full row rank exist and can be checked algorithmically (see Theorem 11 and Lemma 21 of [96]).

In practice, we will be working with several "unknown" matrices that belong to the pattern classes of some known pattern matrices. This will naturally lead to expressions involving sums and products. In order to understand the results of such expression, we will define a sensible way of adding and multiplying pattern matrices. Here sensible means that the result of adding and multiplying pattern matrices gives us some useful information on the result of adding and multiplying matrices belonging to their pattern classes.

To this end, we will define addition for a pair of pattern matrices in such a way that the sum of any pair of real matrices belonging to their pattern class is contained in the pattern class of the sum of the pattern matrices. We know that the sum of zero and any real number is just the number itself, while the sum of two nonzero real numbers can be any real number. Motivated by this, we define addition for the set $\{0, *, ?\}$ as shown in Table 12.1.

Table 12.1: Addition for the set $\{0, *, ?\}$.

$+$	0	$*$	$?$
0	0	$*$	$?$
$*$	$*$	$?$	$?$
$?$	$?$	$?$	$?$

Then addition for pattern matrices is defined element-wise.

Definition 12.2. Let $\mathcal{A}, \mathcal{B} \in \{0, *, ?\}^{m \times n}$. Their sum $\mathcal{A} + \mathcal{B} \in \{0, *, ?\}^{m \times n}$ is defined as

$$(\mathcal{A} + \mathcal{B})_{ij} = \mathcal{A}_{ij} + \mathcal{B}_{ij}$$

for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots n\}$.

By definition, if $\mathcal{A}$ and $\mathcal{B}$ are pattern matrices of the same dimensions, then $\mathcal{P}(\mathcal{A}+\mathcal{B}) \supseteq \mathcal{P}(\mathcal{A}) + \mathcal{P}(\mathcal{B})$, where

$$\mathcal{P}(\mathcal{A}) + \mathcal{P}(\mathcal{B}) = \{A + B \mid A \in \mathcal{P}(\mathcal{A}),\ B \in \mathcal{P}(\mathcal{B})\}$$

is the Minowski sum of sets. It turns out that the converse is true as well.

Proposition 12.1. If $\mathcal{A}$ and $\mathcal{B}$ are pattern matrices of the same dimensions, then $\mathcal{P}(\mathcal{A}+\mathcal{B}) = \mathcal{P}(\mathcal{A}) + \mathcal{P}(\mathcal{B})$.

Proof. The inclusion $\mathcal{P}(\mathcal{A}+\mathcal{B}) \supseteq \mathcal{P}(\mathcal{A}) + \mathcal{P}(\mathcal{B})$ follows from the definition of addition. For the other inclusion, let $C \in \mathcal{P}(\mathcal{A}+\mathcal{B})$ and consider an entry C_{ij}. The goal is to show that there exist entries $A_{ij} \in \mathcal{P}(\mathcal{A}_{ij})$ and $B_{ij} \in \mathcal{P}(\mathcal{B}_{ij})$ such that $C_{ij} = A_{ij} + B_{ij}$. We will consider the cases $C_{ij} = 0$ and $C_{ij} \neq 0$ separately.

Suppose that $C_{ij} = 0$. Then either $(\mathcal{A}+\mathcal{B})_{ij} = 0$ or $(\mathcal{A}+\mathcal{B})_{ij} = ?$. In the former, we must have that $\mathcal{A}_{ij} = 0$ and $\mathcal{B}_{ij} = 0$, hence $A_{ij} = 0$ and $B_{ij} = 0$ would work. In the latter, there are several possibilities whose solutions are listed below.

$\mathcal{A}_{ij}$	$\mathcal{B}_{ij}$	A_{ij}	B_{ij}
$*, ?$	$*, ?$	1	-1
0	$?$	0	0
$?$	0	0	0

Suppose that $C_{ij} \neq 0$. Then either $(\mathcal{A}+\mathcal{B})_{ij} = *$ or $(\mathcal{A}+\mathcal{B})_{ij} = ?$. In the former, exactly one of $\mathcal{A}_{ij}$ and $\mathcal{B}_{ij}$ is $*$ and the other one is 0, hence we can pick either $A_{ij} = C_{ij}$ and $B_{ij} = 0$, or $A_{ij} = 0$ and $B_{ij} = C_{ij}$. In the latter, there are several possibilities whose solutions are listed below.

$\mathcal{A}_{ij}$	$\mathcal{B}_{ij}$	A_{ij}	B_{ij}
$*, ?$	$*, ?$	$\frac{C_{ij}}{2}$	$\frac{C_{ij}}{2}$
0	$?$	0	C_{ij}
$?$	0	C_{ij}	0

The element C_{ij} was chosen arbitrarily, hence we can always find matrices $A \in \mathcal{P}(\mathcal{A})$ and $B \in \mathcal{P}(\mathcal{B})$ such that $A + B = C$ and thus $\mathcal{P}(\mathcal{A}+\mathcal{B}) \subseteq \mathcal{P}(\mathcal{A}) + \mathcal{P}(\mathcal{B})$. □

In the same vein, we now turn to the definition of multiplication for pattern matrices. Note that the product of zero and any real number is just zero, while the product of two nonzero real numbers is always a nonzero real number. This motivates the definition of multiplication for the set $\{0, *, ?\}$ shown in Table 12.2. Then we can define pattern matrix multiplication in the usual way.

Definition 12.3. Let $\mathcal{A} \in \{0, *, ?\}^{m \times p}$ and $\mathcal{B} \in \{0, *, ?\}^{p \times n}$. Their product $\mathcal{A}\mathcal{B} \in \{0, *, ?\}^{m \times n}$ is defined as

$$(\mathcal{A}\mathcal{B})_{ij} = \sum_{k=1}^{p} \mathcal{A}_{ik}\mathcal{B}_{kj}$$

for all $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$.

Table 12.2: Multiplication for the set $\{0, *, ?\}$.

$\cdot$	0	$*$	$?$
0	0	0	0
$*$	0	$*$	$?$
$?$	0	$?$	$?$

By definition, if $\mathcal{A}$ and $\mathcal{B}$ are of appropriate dimensions, then $\mathcal{P}(\mathcal{A}\mathcal{B}) \supseteq \mathcal{P}(\mathcal{A})\mathcal{P}(\mathcal{B})$, where

$$\mathcal{P}(\mathcal{A})\mathcal{P}(\mathcal{B}) = \{AB \mid A \in \mathcal{A},\ B \in \mathcal{B}\}.$$

Unfortunately, the converse is generally not true. When multiplying matrices with at least two rows or columns, we typically create dependencies between the entries of the product. These dependencies cannot be captured by the operations with pattern matrices.

Example 12.1. Consider the pattern vectors

$$\mathcal{A} = \begin{bmatrix} * \\ * \end{bmatrix} \quad \text{and} \quad \mathcal{B} = \begin{bmatrix} * & * \end{bmatrix}.$$

Their product is easily computed as

$$\mathcal{A}\mathcal{B} = \begin{bmatrix} * & * \\ * & * \end{bmatrix},$$

whose pattern class contains the matrix

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \in \mathcal{P}(\mathcal{A}\mathcal{B}).$$

Note that the latter is a matrix of rank 2, thus it cannot be written as the outer product of two vectors. In other words, the fact that the columns (or rows) of AB, where $A \in \mathcal{P}(\mathcal{A})$ and $B \in \mathcal{P}(\mathcal{B})$, are linearly dependent cannot be inferred from the product $\mathcal{A}\mathcal{B}$.

Although the equality $\mathcal{P}(\mathcal{A}\mathcal{B}) = \mathcal{P}(\mathcal{A})\mathcal{P}(\mathcal{B})$ does not hold in general, there are special cases of $\mathcal{A}$ and $\mathcal{B}$ for which equality holds. A notable special case is when either $\mathcal{A}$ or $\mathcal{B}$ is the identity pattern matrix $\mathcal{I}$ of appropriate dimensions, defined as

$$\mathcal{I} := \begin{bmatrix} * & 0 & \cdots & 0 \\ 0 & * & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & * \end{bmatrix}.$$

Indeed, suppose that $\mathcal{A} = \mathcal{I}$. It is not difficult to see that $\mathcal{I}\mathcal{B} = \mathcal{B}$ and $\mathcal{P}(\mathcal{I})\mathcal{P}(\mathcal{B}) = \mathcal{P}(\mathcal{B})$, and thus, $\mathcal{P}(\mathcal{I}\mathcal{B}) = \mathcal{P}(\mathcal{B}) = \mathcal{P}(\mathcal{I})\mathcal{P}(\mathcal{B})$. In the case that $\mathcal{B} = \mathcal{I}$ we can prove the equality in an analogous way.

Finally, we will consider the following lemma that will come in handy when considering the applications to structured systems in the next section.

Lemma 12.1. Let $\mathcal{A}, \mathcal{B} \in \{0, *, ?\}^{m \times n}$. Then $A - \lambda B$ has full rank for all $A \in \mathcal{P}(\mathcal{A})$, $B \in \mathcal{P}(\mathcal{B})$ and nonzero $\lambda \in \mathbb{C}$ if and only if $\mathcal{A} + \mathcal{B}$ has full rank.

Proof. Suppose that $A - \lambda B$ has full rank for all $A \in \mathcal{P}(\mathcal{A})$, $B \in \mathcal{P}(\mathcal{B})$ and nonzero $\lambda \in \mathbb{C}$. Fixing $\lambda = -1$ shows that $A + B$ has full rank for all $A \in \mathcal{P}(\mathcal{A})$, $B \in \mathcal{P}(\mathcal{B})$. But this is equivalent to C having full rank for all $C \in \mathcal{P}(\mathcal{A}) + \mathcal{P}(\mathcal{B})$, hence $\mathcal{A} + \mathcal{B}$ has full rank due to Proposition 12.1.

Conversely, suppose that $\mathcal{A} + \mathcal{B}$ has full rank. We will only treat the case where $m \leqslant n$ since the case where $n > m$ follows the same reasoning after transposing $\mathcal{A}$ and $\mathcal{B}$. With this in mind, let $z \in \mathbb{C}^m$ be such that $z^*A - \lambda z^*B = 0$. The goal is to show that z must be the zero vector. Write $z = x + iy$, where $x, y \in \mathbb{R}^m$ and i denotes the imaginary unit, and consider $\hat{z} = x + \alpha y$ with $\alpha \in \mathbb{R}$ such that

$$\alpha \notin \{\frac{x_k}{y_k} \mid y_k \neq 0, k = 1, 2, \ldots, m\}, \tag{12.1}$$

$$\alpha \notin \{\frac{(x^\top A)_k}{(y^\top A)_k} \mid (y^\top A)_k \neq 0, k = 1, 2, \ldots, m\}, \tag{12.2}$$

$$\alpha \notin \{\frac{(x^\top B)_k}{(y^\top B)_k} \mid (y^\top B)_k \neq 0, k = 1, 2, \ldots, m\}. \tag{12.3}$$

Note that (12.1) implies that $z_k = 0$ if and only if $\hat{z}_k = 0$. Similarly, (12.2) and (12.3) imply that $(z^*A)_k = 0$ if and only if $(\hat{z}^\top A)_k = 0$, and $(z^*B)_k = 0$ if and only if $(\hat{z}^\top B)_k = 0$. Furthermore, since $\lambda \neq 0$ and $z^*A = \lambda z^*B$, we have that $(z^*A)_k = 0$ if and only if $(z^*B)_k = 0$, hence $(\hat{z}^\top A)_k = 0$ if and only if $(\hat{z}^\top B)_k = 0$. Therefore, the diagonal matrix $\Delta \in \mathbb{R}^{n \times n}$ defined as

$$\Delta_{kk} = \begin{cases} 1 & \text{if } (\hat{z}^\top B)_k = 0, \\ \dfrac{(\hat{z}^\top A)_k}{(\hat{z}^\top B)_k} & \text{otherwise,} \end{cases}$$

is a member of $\mathcal{P}(\mathcal{I})$. Moreover, we have $\hat{z}^\top A = \hat{z}^\top B\Delta$. Since $\mathcal{P}(\mathcal{B})\mathcal{P}(\mathcal{I}) = \mathcal{P}(\mathcal{B})$ it holds that $-B\Delta \in \mathcal{P}(\mathcal{B})$. Therefore, $A - B\Delta \in \mathcal{P}(\mathcal{A} + \mathcal{B})$ due to Proposition 12.1. Then $A - B\Delta$ has full row rank, which implies that $\hat{z} = 0$ and thus $z = 0$ because of (12.1). This proves the lemma. □

12.3 APPLICATIONS

In this section, we will show how $\{0, *, ?\}$ pattern matrices can be used to characterize properties of structured systems. This has already been done for strong structural controllability in [96]. There it was shown that a structured system is strongly structurally controllable if and only if a pair of pattern matrices has full row rank. As the latter can be checked algorithmically, this provides a way to verify strong structural controllability for a given structured system. Here, we will extend the work of [96] by studying strong structural controllability of systems of differential algebraic equations. In addition, we will treat the problems of input-state observability and output controllability. As we will see, addition and multiplication of pattern matrices will play an important role in each of these three problems.

12.3.1 Controllability of linear DAE's

In this subsection, we will extend the results on strong structural controllability in [96] to systems of linear differential-algebraic equations (DAE's). Let (E, A, B) denote the system

$$E\dot{x}(t) = Ax(t) + Bu(t), \tag{12.4}$$

where $t \geqslant 0$, $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the input, $E \in \mathbb{R}^{n\times n}$, $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. The system (E, A, B) is called *regular* if $sE - A$ is invertible as a rational matrix in s. Typically, the matrix E is singular, which puts algebraic constraints on the state. This leads to (E, A, B) having special features that are not found in systems described by linear differential equations only. Consequently, there are different kinds of controllability notions defined for (E, A, B), some of which make sense only in the presence of algebraic constraints. We will not go into the analysis of DAE systems, and will instead focus on a particular definition of controllability and its characterization, as presented in [43]. To this end, let $x(t; x_0, u)$ denote the state trajectory at time $t \geqslant 0$ for the initial condition $x(0) = x_0 \in \mathbb{R}^n$ and input u.

Definition 12.4. The regular system (E, A, B) is controllable if for any $T > 0$, $x_0 \in \mathbb{R}^n$ and $x_1 \in \mathbb{R}^n$, there exists an input function[2] $u \in C_p^{h-1}$ such that $x(T; x_0, u) = x_1$.

Then we have the following characterization of controllability for (E, A, B).

Theorem 12.1. [43, Thm. 2-2.1] The regular system (E, A, B) is controllable if and only if

$$\operatorname{rank}\begin{bmatrix} E & B \end{bmatrix} = \operatorname{rank}\begin{bmatrix} A - \lambda E & B \end{bmatrix} = n \tag{12.5}$$

for all $\lambda \in \mathbb{C}$.

Now, suppose that E, A and B are not known precisely but are known to belong to the pattern classes of some known pattern matrices. In other words, we know that $E \in \mathcal{P}(\mathcal{E})$, $A \in \mathcal{P}(\mathcal{A})$ and $B \in \mathcal{P}(\mathcal{B})$ for given pattern matrices $\mathcal{E} \in \{0, *, ?\}^{n\times n}$, $\mathcal{A} \in \{0, *, ?\}^{n\times n}$ and $\mathcal{B} \in \{0, *, ?\}^{n\times m}$. This naturally leads to a family of systems as E, A and B range over the respective pattern classes. This family is completely characterized by $\mathcal{E}$, $\mathcal{A}$ and $\mathcal{B}$, hence we denote it by $(\mathcal{E}, \mathcal{A}, \mathcal{B})$ and call it a *structured system*.

[2] The input is assumed to be in C_p^{h-1}, the set of $(h-1)$-times piecewise continuously differentiable functions. Here h denotes the index of the DAE, see Chapter 1 of [43].

We are interested in conditions under which all regular $(E, A, B) \in \mathcal{P}(\mathcal{E}) \times \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{B})$ are controllable. This motivates the following definition.

Definition 12.5. The structured system $(\mathcal{E}, \mathcal{A}, \mathcal{B})$ is *regularly strongly structurally controllable* if all regular systems $(E, A, B) \in \mathcal{P}(\mathcal{E}) \times \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{B})$ are controllable.

Making use of the results from Section 12.2, we now have the following theorem that provides a sufficient condition for regular strong structural controllability.

Theorem 12.2. The rank conditions (12.5) hold for all $(E, A, B) \in \mathcal{P}(\mathcal{E}) \times \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{B})$ if and only if the pattern matrices

$$\begin{bmatrix} \mathcal{E} & \mathcal{B} \end{bmatrix}, \quad \begin{bmatrix} \mathcal{A} & \mathcal{B} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathcal{A} + \mathcal{E} & \mathcal{B} \end{bmatrix}$$

have full row rank. Moreover, if these pattern matrices have full row rank then $(\mathcal{E}, \mathcal{A}, \mathcal{B})$ is regularly strongly structurally controllable.

Proof. Note that the rank conditions (12.5) hold for all $(E, A, B) \in \mathcal{P}(\mathcal{E}) \times \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{B})$ if and only if

$$\operatorname{rank} \begin{bmatrix} E & B \end{bmatrix} = \operatorname{rank} \begin{bmatrix} A & B \end{bmatrix} = \operatorname{rank} \begin{bmatrix} A - \lambda E & B \end{bmatrix} = n$$

for all nonzero $\lambda \in \mathbb{C}$, $E \in \mathcal{P}(\mathcal{E})$, $A \in \mathcal{P}(\mathcal{A})$ and $B \in \mathcal{P}(\mathcal{B})$. The result then follows from Lemma 12.1. □

Remark 12.1. In the special case that $\mathcal{E} = \mathcal{I}$, all systems $(E, A, B) \in \mathcal{P}(\mathcal{I}) \times \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{B})$ are regular. In this case, we can also write (12.4) as

$$\dot{x} = E^{-1}Ax + E^{-1}Bu$$

for all $E \in \mathcal{P}(\mathcal{I})$, $A \in \mathcal{P}(\mathcal{A})$ and $B \in \mathcal{P}(\mathcal{B})$. Clearly, $E^{-1} \in \mathcal{P}(\mathcal{I})$ for all $E \in \mathcal{P}(\mathcal{I})$. Since $\mathcal{P}(I)\mathcal{P}(\mathcal{A}) = \mathcal{P}(\mathcal{A})$ and $\mathcal{P}(I)\mathcal{P}(\mathcal{B}) = \mathcal{P}(\mathcal{B})$, we see that regular strong structural controllability of $(\mathcal{I}, \mathcal{A}, \mathcal{B})$ is equivalent to strong structural controllability of $(\mathcal{A}, \mathcal{B})$, i.e., to controllability of (A, B) for all $A \in \mathcal{P}(\mathcal{A})$ and $B \in \mathcal{P}(\mathcal{B})$. In fact, in the special case $\mathcal{E} = \mathcal{I}$, the conditions of Theorem 12.2 coincide with the conditions for strong structural controllability given in [96, Thm. 7]. To see this, note that $\begin{bmatrix} \mathcal{I} & \mathcal{B} \end{bmatrix}$ has full row rank for any $\mathcal{B}$. In addition, the matrix $\bar{\mathcal{A}} := \mathcal{A} + \mathcal{I}$ is the pattern matrix obtained from $\mathcal{A}$ by changing the diagonal entries of $\mathcal{A}$ to

$$\bar{\mathcal{A}}_{kk} = \begin{cases} * & \text{if } \mathcal{A}_{kk} = 0, \\ ? & \text{otherwise.} \end{cases}$$

A such, Theorem 12.2 requires $\begin{bmatrix} \mathcal{A} & \mathcal{B} \end{bmatrix}$ and $\begin{bmatrix} \bar{\mathcal{A}} & \mathcal{B} \end{bmatrix}$ to have full row rank, which are exactly the two conditions of [96, Thm. 7]. These conditions are, in fact, necessary and sufficient for controllability of $(\mathcal{A}, \mathcal{B})$ [96]. The lack of necessity in the characterization of regular strong structural controllability (Theorem 12.2) stems from the fact that in general not all $(E, A, B) \in \mathcal{P}(\mathcal{E}) \times \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{B})$ are regular.

12.3.2 Input-state observability

In this section, we will use the techniques developed in the analysis of strong structural controllability to another property, namely, input-state observability. Let (A, B, C, D) denote the system

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{12.6}$$

$$y(t) = Cx(t) + Du(t), \tag{12.7}$$

where $t \geqslant 0$ represents time, $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the input, $y(t) \in \mathbb{R}^p$ is the output, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$ and $D \in \mathbb{R}^{p\times m}$. For a given initial condition $x(0) = x_0 \in \mathbb{R}^n$ and input function u, we denote the corresponding output trajectory at time $t \geqslant 0$ by $y(t; x_0, u)$. Consider the following definition.

Definition 12.6. The system (A, B, C, D) is *input-state observable* if $y(t; x_1, u_1) = y(t; x_2, u_2)$ for all $t \geqslant 0$ implies that $x_1 = x_2$ and $u_1(t) = u_2(t)$ for all $t \geqslant 0$.

In other words, a system (A, B, C, D) is input-state observable if different initial states and input functions can be distinguished on the basis of the output of the system. Conditions under which this is the case are provided in the following theorem.

Theorem 12.3. [191, Thm. 3.3] The system (A, B, C, D) is input-state observable if and only if

$$\operatorname{rank} \begin{bmatrix} A - \lambda I & B \\ C & D \end{bmatrix} = n + m$$

for all $\lambda \in \mathbb{C}$.

As before, instead of considering a single system (A, B, C, D), we consider the family of systems where $A \in \mathcal{P}(\mathcal{A})$, $B \in \mathcal{P}(\mathcal{B})$, $C \in \mathcal{P}(\mathcal{C})$ and $D \in \mathcal{P}(\mathcal{D})$ for given pattern matrices $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$ of appropriate dimensions. We denote this family by $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ and refer to it as a *structured system*. We are interested in finding necessary and sufficient conditions under which (A, B, C, D) is guaranteed to be input-state observable for all $A \in \mathcal{P}(\mathcal{A})$, $B \in \mathcal{P}(\mathcal{B})$, $C \in \mathcal{P}(\mathcal{C})$ and $D \in \mathcal{P}(\mathcal{D})$.

Definition 12.7. The system $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ is strongly structurally input-state observable if (A, B, C, D) is input-state observable for all $A \in \mathcal{P}(\mathcal{A})$, $B \in \mathcal{P}(\mathcal{B})$, $C \in \mathcal{P}(\mathcal{C})$ and $D \in \mathcal{P}(\mathcal{D})$.

In view of Theorem 12.3 and the results presented so far, the following characterization of strong structural input-state observability follows naturally.

Theorem 12.4. The system $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ is strongly structurally input-state observable if and only if the pattern matrices

$$\begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathcal{A} + \mathcal{I} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix}$$

have full column rank.

Proof. We claim that $(\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D})$ is strongly structurally input-state observable if and only if

$$\operatorname{rank}\begin{bmatrix} A-\lambda\Delta & B \\ C & D \end{bmatrix} = n+m \tag{12.8}$$

for all $\lambda \in \mathbb{C}$, $\Delta \in \mathcal{P}(\mathcal{I})$, $A \in \mathcal{P}(\mathcal{A})$, $B \in \mathcal{P}(\mathcal{B})$, $C \in \mathcal{P}(\mathcal{C})$ and $D \in \mathcal{P}(\mathcal{D})$. Indeed, (12.8) holds if and only if

$$\operatorname{rank}\begin{bmatrix} \Delta^{-1}A-\lambda I & \Delta^{-1}B \\ C & D \end{bmatrix} = n+m$$

and $\Delta^{-1}A \in \mathcal{P}(\mathcal{A})$, $\Delta^{-1}B \in \mathcal{P}(\mathcal{B})$ if and only if $A \in \mathcal{P}(\mathcal{A})$, $B \in \mathcal{P}(\mathcal{B})$. Therefore, $(\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D})$ is strongly structurally input-state observable if and only if

$$\operatorname{rank}\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \operatorname{rank}\begin{bmatrix} A & B \\ C & D \end{bmatrix} - \lambda\begin{bmatrix} \Delta & 0 \\ 0 & 0 \end{bmatrix} = n+m$$

for all $\lambda \neq 0$, $\Delta \in \mathcal{P}(\mathcal{I})$, $A \in \mathcal{P}(\mathcal{A})$, $B \in \mathcal{P}(\mathcal{B})$, $C \in \mathcal{P}(\mathcal{C})$ and $D \in \mathcal{P}(\mathcal{D})$, which is equivalent to

$$\begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathcal{A}+\mathcal{I} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix}$$

having full column rank by Lemma 12.1. This proves the theorem. □

12.3.3 Output controllability

In this section, we will show how pattern matrix multiplication and its properties can be used to characterize strong structural output controllability. To this end, consider the system (A,B,C,D) as defined in Section 12.3.2.

Definition 12.8. The system (A,B,C,D) is *output controllable* if for any $x_0 \in \mathbb{R}^n$ and $y_1 \in \mathbb{R}^p$, there exist a time $T>0$ and an input u such that $y(T;x_0,u)=y_1$.

The following is a well-known characterization of output controllability of (A,B,C,D), c.f., [208, Ex. 3.22].

Theorem 12.5. The system (A,B,C,D) is output controllable if and only if

$$\operatorname{rank}\begin{bmatrix} D & CB & CAB & \cdots & CA^{n-1}B \end{bmatrix} = p.$$

Now, consider the structured system $(\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D})$ as defined in Section 12.3.2.

Definition 12.9. The system $(\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D})$ is *strongly structurally output controllable* if (A,B,C,D) is output controllable for all $A \in \mathcal{P}(\mathcal{A})$, $B \in \mathcal{P}(\mathcal{B})$, $C \in \mathcal{P}(\mathcal{C})$ and $D \in \mathcal{P}(\mathcal{D})$.

We are interested in conditions under which $(\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D})$ is strongly structurally output controllable. Note that the condition for output controllability of (A,B,C,D) involves products of system matrices, unlike the conditions for controllability of (E, A, B) or input-state observability of (A, B, C, D). This suggest that we need to consider products of pattern matrices when investigating strong structural output controllability of $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$. Unfortunately, since products of pattern matrices do not share the same favourable property as sums, i.e., $\mathcal{P}(\mathcal{A}\mathcal{B}) \neq \mathcal{P}(\mathcal{A})\mathcal{P}(\mathcal{B})$, we cannot easily derive necessary and sufficient conditions. Nevertheless, we state and prove the following sufficient condition.

Theorem 12.6. The system $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ is strongly structurally output controllable if the pattern matrix

$$\begin{bmatrix}\mathcal{D} \ \mathcal{C}\mathcal{B} \ \mathcal{C}\mathcal{A}\mathcal{B} \ \cdots \ \mathcal{C}\mathcal{A}^{n-1}\mathcal{B}\end{bmatrix}$$

has full row rank.

Proof. Let $A \in \mathcal{P}(\mathcal{A})$, $B \in \mathcal{P}(\mathcal{B})$, $C \in \mathcal{P}(\mathcal{C})$ and $D \in \mathcal{P}(\mathcal{D})$. Recall that $\mathcal{P}(\mathcal{C})\mathcal{P}(\mathcal{B}) \subseteq \mathcal{P}(\mathcal{C}\mathcal{B})$, that is, $CB \in \mathcal{P}(\mathcal{C}\mathcal{B})$ for all $C \in \mathcal{P}(\mathcal{C})$ and $B \in \mathcal{P}(\mathcal{B})$. Similarly, we find that

$$\mathcal{P}(\mathcal{C})\mathcal{P}(\mathcal{A})\mathcal{P}(\mathcal{B}) \subseteq \mathcal{P}(\mathcal{C}\mathcal{A})\mathcal{P}(\mathcal{B}) \subseteq \mathcal{P}(\mathcal{C}\mathcal{A}\mathcal{B}),$$

and, more generally, that $\mathcal{P}(\mathcal{C})\mathcal{P}(\mathcal{A})^k\mathcal{P}(\mathcal{B}) \subseteq \mathcal{P}(\mathcal{C}\mathcal{A}^k\mathcal{B})$ for all positive integers k. In other words, we have that

$$\begin{bmatrix}D \ CB \ CAB \ \cdots \ CA^{n-1}B\end{bmatrix} \subseteq \mathcal{P}(\begin{bmatrix}\mathcal{D} \ \mathcal{C}\mathcal{B} \ \mathcal{C}\mathcal{A}\mathcal{B} \ \cdots \ \mathcal{C}\mathcal{A}^{n-1}\mathcal{B}\end{bmatrix}).$$

Hence (A, B, C, D) is output controllable. As A, B, C and D were chosen arbitrarily, it follows that $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ is strongly structurally output controllable. □

12.4 CONCLUSION

In this chapter we have studied addition and multiplication of pattern matrices having a zero/nonzero/arbitrary structure. We have seen that addition of such pattern matrices preserves the pattern class, while multiplication of pattern matrices enlarges the pattern class in general. We have applied these results to assess strong structural properties of linear systems. In particular, we have derived conditions for strong structural input-state observability and output controllability of linear systems. We have also studied controllability of differential algebraic equations. In each of the three problems we saw that addition and/or multiplication of pattern matrices plays an important role. We are confident that our results on pattern matrices can also be applied to other system-theoretic properties such as fault detection and isolation [95].

13 A DISTANCE-BASED APPROACH TO TARGET CONTROLLABILITY

In this chapter we continue our work on strong structural output controllability. This time, we will consider the problem in a more specific (network) setting. Here, the input and output matrices have a particular structure reflecting the fact that inputs are applied to a subset of network nodes and outputs consist of a subset of nodal states. In this setting, output controllability is often referred to as targeted controllability. We will additionally focus on a smaller class of so-called distance-information preserving state matrices. This allows us to come up with stronger results for targeted controllability. We will provide both a sufficient and a necessary condition for targeted controllability. In addition, we will propose a method to select inputs that guarantee targeted controllability.

13.1 INTRODUCTION

During the last two decades, networks of dynamical agents have been extensively studied. It is customary to represent the infrastructure of such networks by a graph, where nodes are identified with agents and arcs correspond to the communication between agents. In the study of controllability of networks, two types of nodes are distinguished: leaders, which are influenced by external input, and followers whose dynamics are completely determined by the behaviour of their neighbours. Network controllability comprises the ability to drive the states of all nodes of the network to any desired state, by applying appropriate input to the leaders.

Motivated by model uncertainties, the notion of structural controllability of linear control systems described by the pair (A, B) was introduced by Lin [110]. Here the entries of the matrices A and B are either fixed zeros or free parameters. In this framework, weak structural controllability requires almost all realizations of (A, B) to be controllable. That is, for almost all parameter settings of the entries of A and B, the pair (A, B) is controllable. Lin provided a graph-theoretic condition under which (A, B) is weakly structurally controllable in the single-input case. Many papers followed [110], amongst others we name [67] and [194] in which extensions to multiple leaders are given, and the article [133], that introduces strong structural controllability, which requires all realizations of (A, B) to be controllable.

In recent years, structural controllability gained much attention in the study of networks of dynamical agents [24], [29], [113], [142], [207]. With a given network graph, a family of linear control systems is associated, where the structure of the state matrix of each system depends on the network topology, and the input matrix is determined by the leader set. In this framework, a network is said to be weakly (strongly) structurally controllable if almost all (all) systems associated with the network are controllable. The graph-theoretic results obtained in classical papers [110], [133] lend themselves excellently to the study of structural controllability of networks. A topological condition for (weak) structural controllability of networks is given in terms of maximum matchings in [113], while strong structural controllability is fully characterized in terms of zero forcing sets in [142], and in terms of constrained matchings in [29].

However, in large-scale networks with high vertex degrees, a substantial amount of nodes must be chosen as leader to achieve full control in the strong sense, which is often infeasible. Furthermore, in some applications full control over the network is unnecessary. Hence, we are interested in controlling a subset of agents, called target nodes. This specific form of output control is known under the name target control [63], [141]. Potential applications of target control within the areas of biology, chemical engineering and economic networks are identified in [63].

A network is said to be strongly targeted controllable if all systems in the family associated with the network graph are targeted controllable. In this chapter we consider strong targeted controllability for the class of state matrices called distance-information preserving matrices. The adjacency matrix and symmetric, indegree and outdegree Laplacian matrices are examples of distance-information preserving matrices. As these matrices are often used to describe network dynamics (see, e.g., [54], [68], [171], [205], [247]), distance-information preserving matrices form an important class of matrices associated with network graphs.

Our main results are threefold. Firstly, we provide a sufficient topological condition for strong targeted controllability of networks, that generalizes the results of [141] for the class of distance-information preserving matrices. Specifically, the results of [141] are restricted to target nodes having distance one with respect to the so-called derived set of the leaders. However, our result is applicable to target nodes that have arbitrary distance with respect to the leaders. Our sufficient topological condition can be understood as a "k-walk theory" [63] for strong targeted controllability. However, we remark that the k-walk theory for (weak) targeted controllability established in Theorem 2 of [63] is only applicable to single-input directed tree networks. On the other hand, our result is applicable to arbitrary directed networks with multiple leaders.

Secondly, noting that our proposed sufficient condition for target control is not a one-to-one correspondence, we establish a necessary graph-theoretic condition for strong targeted controllability.

Finally, we consider the minimum leader selection problem in the context of strong targeted controllability of networks. Recently, leader selection (and in general, actuator placement) has received much attention in the literature. Minimum actuator placement in the context of controllability is studied in [159], [165], [200] and [211]. Also, minimum actuator placement for reachability problems was studied [210]. Moreover, the selection of minimum input sets achieving (strong) structural controllability has been considered in [29], [113], [167] and [168]. It was shown [29] that determining minimum input sets for strong structural controllability is NP-hard in general. In the context of (weak) targeted controllability, a leader selection method has been given in [63]. However, an algorithm for leader selection for *strong* targeted controllability is still missing. We first prove that there exists no polynomial-time algorithm to determine minimum leader sets achieving strong target control (assuming P $\neq$ NP). Subsequently, we provide a heuristic two-phase leader selection algorithm consisting of a binary linear programming phase and a greedy approach to obtain leader sets achieving strong target control.

The chapter is organized as follows: in Section 13.2, we introduce preliminaries and notation. Subsequently, the problem is stated in Section 13.3. Our main results are presented in Section 13.4. To illustrate the main results, an example is given in Section 13.5. Finally, Section 13.6 contains our conclusions.

13.2 PRELIMINARIES

Consider a directed graph $G = (V, E)$, where V is a set of n vertices, and E is the set of directed arcs. Throughout this chapter, all graphs are assumed to be simple and without self-loops.

We define the *distance* $d(u, v)$ between two vertices $u, v \in V$ as the length of the shortest path from u to v. If there does not exist a path in the graph G from vertex u to v, the distance $d(u, v)$ is defined as infinite. Moreover, the distance from a vertex to itself is equal to zero.

For a nonempty subset $S \subseteq V$ and a vertex $j \in V$, the distance from S to j is defined as

$$d(S, j) := \min_{i \in S} d(i, j). \tag{13.1}$$

A directed graph $G = (V, E)$ is called *bipartite* if there exist disjoint sets of vertices V^- and V^+ such that $V = V^- \cup V^+$ and $(u, v) \in E$ only if $u \in V^-$ and $v \in V^+$. We denote bipartite graphs by $G = (V^-, V^+, E)$, to indicate the partition of the vertex set.

13.2.1 Qualitative class and pattern class

The *qualitative class* of a directed graph G is a family of matrices associated with the graph. Each of the matrices of this class contains a nonzero element in position i, j if and only if there is an arc (j, i) in G, for $i \neq j$. More explicitly, the qualitative class $\mathcal{Q}(G)$ of a graph G is given by

$$\mathcal{Q}(G) = \{X \in \mathbb{R}^{n \times n} \mid \text{ for } i \neq j,\ X_{ij} \neq 0 \iff (j, i) \in E\}.$$

Note that the diagonal elements of a matrix $X \in \mathcal{Q}(G)$ do not depend on the structure of G, these are “free elements" in the sense that they can be either zero or nonzero.

Next, we look at a different class of matrices associated with a bipartite graph $G = (V^-, V^+, E)$, where the vertex sets V^- and V^+ are given by

$$\begin{aligned} V^- &= \{r_1, r_2, ..., r_s\} \\ V^+ &= \{q_1, q_2, ..., q_t\}. \end{aligned} \tag{13.2}$$

The *pattern class* $\mathcal{P}(G)$ of the bipartite graph G, with vertex sets V^- and V^+ given by (13.2), is defined as

$$\mathcal{P}(G) = \{M \in \mathbb{R}^{t \times s} \mid M_{ij} \neq 0 \iff (r_j, q_i) \in E\}. \tag{13.3}$$

Note that the cardinalities of V^- and V^+ can differ, hence the matrices in the pattern class $\mathcal{P}(G)$ are not necessarily square.

13.2.2 Subclass of distance-information preserving matrices

In this subsection we investigate properties of the powers of matrices belonging to the qualitative class $\mathcal{Q}(G)$. The relevance of these properties will become apparent later on, when we provide a graph-theoretic condition for targeted controllability of systems defined on graphs.

We first provide the following lemma, which states that if the distance between two nodes is greater than k, the corresponding element in X^k is zero.

Lemma 13.1. Consider a directed graph $G = (V, E)$, two distinct vertices $i, j \in V$, a matrix $X \in \mathcal{Q}(G)$ and a positive integer k. If $d(j, i) > k$, then $(X^k)_{ij} = 0$.

Proof. The proof follows easily by induction on k, and is therefore omitted. □

Subsequently, we consider the class of matrices for which $(X^k)_{ij}$ is nonzero if the distance $d(j, i)$ is exactly equal to k. Such matrices are called distance-information preserving, more precisely:

Definition 13.1. Consider a directed graph $G = (V, E)$. A matrix $X \in \mathcal{Q}(G)$ is called *distance-information preserving* if for any two distinct vertices $i, j \in V$ we have that $d(j, i) = k$ implies $(X^k)_{ij} \neq 0$.

Although the distance-information preserving property does not hold for all matrices $X \in \mathcal{Q}(G)$, it does hold for the adjacency and Laplacian matrices [174]. Because these matrices are often used to describe network dynamics, distance-information preserving matrices form an important subclass of $\mathcal{Q}(G)$, which from now on will be denoted by $\mathcal{Q}_d(G)$.

13.2.3 Zero forcing sets

In this section we review the notion of zero forcing. The reason for this is the correspondence between zero forcing sets and the sets of leaders rendering a system defined on a graph controllable. More on this will follow in the next subsection.

For now, let $G = (V, E)$ be a directed graph with vertices colored either black or white. The *color-change rule* is defined in the following way: If $u \in V$ is a black vertex and exactly one out-neighbour $v \in V$ of u is white, then change the color of v to black [87].

When the color-change rule is applied to u to change the color of v, we say u *forces* v, and write $u \to v$.

Given a coloring of G, that is, given a set $C \subseteq V$ containing black vertices only, and a set $V \setminus C$ consisting of only white vertices, the *derived set* $D(C)$ is the set of black vertices obtained by applying the color-change rule until no more changes are possible [87].

A *zero forcing set* for G is a subset of vertices $Z \subseteq V$ such that if initially the vertices in Z are colored black and the remaining vertices are colored white, then $D(Z) = V$.

The *zero forcing number* $\rho(G)$ of the graph $G = (V, E)$ is the minimum of $|Z|$ over all zero forcing sets $Z \subseteq V$. Moreover, a zero forcing set $Z \subseteq V$ is called a *minimum zero forcing set* if $|Z|$ equals $\rho(G)$.

Finally, for a given zero forcing set, we can construct the derived set, listing the forces in the order in which they were performed. This list is called a *chronological list of forces*. Note that such a list does not have to be unique.

13.2.4 Output controllability of linear systems

In this section we review the notion of output controllability for linear, time-invariant systems. This notion will become useful in the next section, where we will discuss targeted controllability of systems defined on graphs.

Consider the linear time-invariant system

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{13.4}$$

where $x \in \mathbb{R}^n$ is the state of the system, $u \in \mathbb{R}^m$ is the input and the real matrices A and B are of appropriate dimensions. For a given initial condition $x_0 \in \mathbb{R}^n$ and input function u, we denote the state of (13.4) at time t by $x_u(t, x_0)$. It is well-known that system (13.4) is called *controllable* if for all $x_0, x_1 \in \mathbb{R}^n$, there exists an input function u and a finite time T, such that $x_u(T, x_0) = x_1$. In the case (13.4) is controllable, we say the pair (A, B) is controllable. If in addition to Equation 13.4 we specify an output equation $y(t) = Cx(t)$, we obtain the system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t), \end{aligned} \tag{13.5}$$

where $y \in \mathbb{R}^p$ is the output of the system and $C \in \mathbb{R}^{p \times n}$. For a given initial condition x_0 and input function u, we denote the output of (13.5) at time t by $y_u(t, x_0)$. We are now ready to introduce the notion of output controllability.

Definition 13.2. [208] System (13.5) is called *output controllable* if for all $x_0 \in \mathbb{R}^n$ and $y_1 \in \mathbb{R}^p$, there exists an input function u and finite time T such that $y_u(T, x_0) = y_1$.

In case (13.5) is output controllable, we say (A, B, C) is output controllable.

13.2.5 Targeted controllability of systems defined on graphs

Consider a directed graph $G = (V, E)$, where the vertex set is given by $V = \{1, 2, ..., n\}$. Furthermore, let $V' = \{v_1, v_2, ..., v_r\} \subseteq V$ be a subset. The $n \times r$ matrix $P(V; V')$ is defined by

$$P_{ij} = \begin{cases} 1 \text{ if } i = v_j \\ 0 \text{ otherwise.} \end{cases} \tag{13.6}$$

We now introduce the subset $V_L \subseteq V$ consisting of so-called *leader nodes*, i.e. agents of the network to which an external control input is applied. The remaining nodes $V \setminus V_L$ are called *followers*. We consider finite-dimensional linear time-invariant systems of the form

$$\dot{x}(t) = Xx(t) + Uu(t), \tag{13.7}$$

where $x \in \mathbb{R}^n$ is the state and $u \in \mathbb{R}^m$ is the input of the system. Here $X \in \mathcal{Q}(G)$ and $U = P(V; V_L)$, for some leader set $V_L \subseteq V$. An important notion regarding systems of the form (13.7) is the notion of strong structural controllability.

Definition 13.3. [142] A system of the form (13.7) is called *strongly structurally controllable* if the pair (X, U) is controllable for all $X \in \mathcal{Q}(G)$.

In the case that (13.7) is strongly structurally controllable we say $(G; V_L)$ is controllable, with a slight abuse of terminology. There is a one-to-one correspondence between strong structural controllability and zero forcing sets, as stated in the following theorem.

Theorem 13.1. [142] Let $G = (V, E)$ be a directed graph and let $V_L \subseteq V$ be a leader set. Then $(G; V_L)$ is controllable if and only if V_L is a zero forcing set.

Remark 13.1. In the context of minimum leader selection, it is particularly interesting to compute *minimum* zero forcing sets. Unfortunately, the problem of computing a minimum zero forcing set has been shown to be NP-hard [2]. In fact, the inapproximability of a related problem suggests that it is even hard to *approximate* the minimum *number* of leaders (i.e., the zero forcing number $\rho(G)$). Following Trefois et al. [207], it can be shown that in a directed bipartite graph G with n vertices, we have that $n - \rho(G)$ is equal to the so-called maximum size *constrained matching* in G. It has been shown in [139] that it is hard to approximate the maximum size constrained matching in bipartite graphs within a factor $\mathcal{O}(n^{\frac{1}{3}-\epsilon})$ for any $\epsilon > 0$. This suggests that also the zero forcing number (and thereby, the minimum number of leaders) cannot be approximated in polynomial time within a large factor. The above considerations are for bipartite graphs. However, note that if the zero forcing number is hard to approximate in bipartite graphs, it is certainly hard to approximate in general directed graphs.

In this chapter, we are primarily interested in controlling the states of a subset $V_T \subseteq V$ of nodes, called *target nodes*. We specify an output equation $y(t) = Hx(t)$, which together with (13.7) yields the system

$$\begin{aligned} \dot{x}(t) &= Xx(t) + Uu(t) \\ y(t) &= Hx(t), \end{aligned} \tag{13.8}$$

where $y \in \mathbb{R}^p$ is the output of the system consisting of the states of the target nodes, and $H = P^\top(V; V_T)$. Note that the ability to control the states of all target nodes in V_T is equivalent with the output controllability of system (13.8) [141]. As the output of system (13.8) specifically consists of the states of the target nodes, we say (13.8) is targeted controllable if it is output controllable.

Furthermore, system (13.8) is called strongly targeted controllable if (X, U, H) is targeted controllable for all $X \in \mathcal{Q}(G)$ [141]. In case (13.8) is strongly targeted controllable, we say $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}(G)$. The term "with respect to $\mathcal{Q}(G)$" clarifies the class of state matrices under consideration. This chapter mainly considers strong targeted controllability with respect to $\mathcal{Q}_d(G)$. We conclude this section with well-known conditions for strong targeted controllability. Let $U = P(V; V_L)$ and $H = P^\top(V; V_T)$ be the input and output matrices respectively, and define the reachable subspace $\langle X \mid \operatorname{im} U \rangle = \operatorname{im} U + X \operatorname{im} U + \cdots + X^{n-1} \operatorname{im} U$.

Proposition 13.1. [141] The following statements are equivalent:

1) $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}(G)$
2) $\operatorname{rank} \begin{bmatrix} HU & HXU & \cdots & HX^{n-1}U \end{bmatrix} = p \;\; \forall\, X \in \mathcal{Q}(G)$
3) $H\langle X \mid \operatorname{im} U \rangle = \mathbb{R}^p \;\; \forall\, X \in \mathcal{Q}(G)$
4) $\ker H + \langle X \mid \operatorname{im} U \rangle = \mathbb{R}^n \;\; \forall\, X \in \mathcal{Q}(G)$.

13.3 PROBLEM STATEMENT

Strong targeted controllability with respect to $\mathcal{Q}(G)$ was studied in [141], and a sufficient graph-theoretic condition was provided. Motivated by the fact that $\mathcal{Q}_d(G)$ contains important network-related matrices like the adjacency and Laplacian matrices, we are interested in extending the results of [141] to the class of distance-information preserving matrices $Q_d(G)$. More explicitly, the problem that we will investigate in this chapter is given as follows.

Problem 13.1. Given a directed graph $G = (V, E)$, a leader set $V_L \subseteq V$ and target set $V_T \subseteq V$, provide necessary and sufficient graph-theoretic conditions under which $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}_d(G)$.

The study of such graph-theoretic conditions is motivated by the fact that known rank conditions for strong targeted controllability (Proposition 13.1, condition 2) are computationally infeasible. Indeed, verifying condition 2 of Proposition

3 would require the rank computation of an infinite number of output controllability matrices (one for each $X \in \mathcal{Q}_d(G)$). Furthermore, graph-theoretic conditions for targeted controllability may aid in finding leader selection procedures.

In addition to Problem 1, we are interested in a method to compute leader sets achieving targeted controllability. More precisely:

Problem 13.2. Given a directed graph $G = (V, E)$ and target set $V_T \subseteq V$, compute a leader set $V_L \subseteq V$ of minimum cardinality such that $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}_d(G)$.

13.4 MAIN RESULTS

Our main results are presented in this section. Firstly, in Section 13.4.1, we provide a sufficient graph-theoretic condition for strong targeted controllability with respect to $Q_d(G)$. Subsequently, in Section 13.4.2, we review the notion of sufficient richness of subclasses, and prove that the subclass $\mathcal{Q}_d(G)$ is sufficiently rich. This result allows us to establish a necessary condition for strong targeted controllability, which is presented in Section 13.4.3. Finally, in Section 13.4.4, we show there is no polynomial-time algorithm solving Problem 13.2 (assuming P $\neq$ NP). Therefore, we provide a heuristic leader selection algorithm to determine leader sets achieving targeted controllability.

13.4.1 Sufficient condition for targeted controllability

This section discusses a sufficient graph-theoretic condition for strong targeted controllability. We first introduce some notions that will become useful later on.

Consider a directed graph $G = (V, E)$ with leader set V_L and target set V_T. In this section, we assume all target nodes have finite distance with respect to V_L. This assumption is without loss of generality. Indeed, it is easy to see that $(G; V_L; V_T)$ is not targeted controllable if a target node $v \in V_T$ cannot be reached from any leader.

The derived set of V_L is given by $D(V_L)$. Furthermore, let $V_S \subseteq V \setminus D(V_L)$ be a subset. We partition the set V_S according to the distance of its nodes with respect to $D(V_L)$, that is

$$V_S = V_1 \cup V_2 \cup \cdots \cup V_d, \tag{13.9}$$

where for $j \in V_S$ we have $j \in V_i$ if and only if $d(D(V_L), j) = i$ for $i = 1, 2, ..., d$. Moreover, we define $\check{V}_i$ and $\hat{V}_i$ to be the sets of vertices in V_S of distance respectively less than i and greater than i with respect to $D(V_L)$. More precisely:

$$\begin{aligned} \check{V}_i &:= V_1 \cup ... \cup V_{i-1} \text{ for } i = 2, ..., d \\ \hat{V}_i &:= V_{i+1} \cup ... \cup V_d \text{ for } i = 1, ..., d-1 \end{aligned} \tag{13.10}$$

By convention $\check{V}_1 = \varnothing$ and $\hat{V}_d = \varnothing$. With each of the sets $V_1, V_2, \ldots, V_d$ we associate a bipartite graph $G_i = (D(V_L), V_i, E_i)$, where for $j \in D(V_L)$ and $k \in V_i$ we have $(j,k) \in E_i$ if and only if $d(j,k) = i$ in the network graph G.

Example 13.1. We consider the network graph $G = (V, E)$ as depicted in Figure 13.1. The set of leaders is $V_L = \{1, 2\}$, which implies that $D(V_L) = \{1, 2, 3\}$.

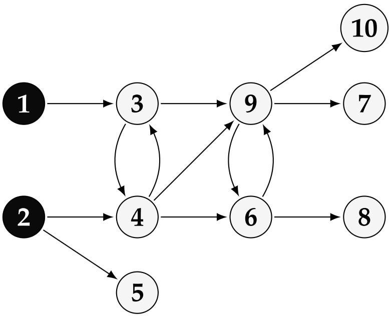

Figure 13.1: Graph G with $V_L = \{1,2\}$.

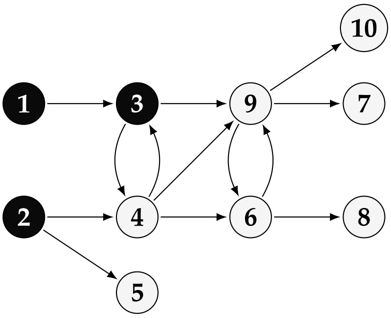

Figure 13.2: $D(V_L) = \{1,2,3\}$.

In this example, we define the subset $V_S \subseteq V \setminus D(V_L)$ as $V_S := \{4,5,6,7,8\}$. Note that V_S can be partitioned according to the distance of its nodes with respect to $D(V_L)$ as $V_S = V_1 \cup V_2 \cup V_3$, where $V_1 = \{4,5\}$, $V_2 = \{6,7\}$ and $V_3 = \{8\}$. The bipartite graphs G_1, G_2 and G_3 are given in Figures 13.3,13.4 and 13.5 respectively.

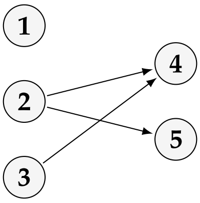

Figure 13.3: Graph G_1.

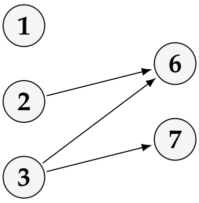

Figure 13.4: Graph G_2.

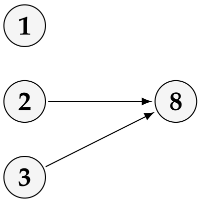

Figure 13.5: Graph G_3.

The main result presented in this section is given in Theorem 13.2. This statement provides a sufficient graph-theoretic condition for targeted controllability of $(G; V_L; V_T)$ with respect to $\mathcal{Q}_d(G)$.

Theorem 13.2. Consider a directed graph $G = (V, E)$, with leader set $V_L \subseteq V$ and target set $V_T \subseteq V$. Let $V_T \setminus D(V_L)$ be partitioned as in (13.9), and assume $D(V_L)$ is a zero forcing set in $G_i = (D(V_L), V_i, E_i)$ for $i = 1, 2, \ldots, d$. Then $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}_d(G)$.

In the special case of a single leader, i.e. $|V_L| = 1$, the condition of Theorem 13.2 can be simplified. In this case, $(G; V_L; V_T)$ is targeted controllable if no pair of target nodes has the same distance with respect to the leader. This is formulated in the following corollary.

Corollary 13.1. Consider a directed graph $G = (V, E)$, with singleton leader set $V_L = \{v\} \subseteq V$ and target set $V_T \subseteq V$. $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}_d(G)$ if $d(v,i) \neq d(v,j)$ for all distinct $i, j \in V_T$.

Note that the condition of Corollary 13.1 is similar to the "k-walk theory" for (weak) targeted controllability established in Theorem 2 of [63]. However, it is worth mentioning that k-walk theory [63] was only proven for directed tree networks with a single leader. On the other hand, Theorem 13.2 establishes a condition for strong targeted controllability that is applicable to general directed networks with multiple leaders.

Furthermore, note that Theorem 13.2 significantly improves the known condition for strong targeted controllability given in [141] for the class $\mathcal{Q}_d(G)$. In Theorem 13.2 target nodes with arbitrary distance with respect to the derived set are allowed, while the main result Theorem VI.6 of [141] is restricted to target nodes of distance one with respect to $D(V_L)$. Before proving Theorem 13.2, we provide an illustrative example and two auxiliary lemmas.

Example 13.2. Once again, consider the network graph depicted in Figure 13.1, with leader set $V_L = \{1, 2\}$ and assume the target set is given by $V_T = \{1, 2, ..., 8\}$. The goal of this example is to prove that $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}_d(G)$.

Note that $V_S := V_T \setminus D(V_L)$ is given by $V_S = \{4, 5, 6, 7, 8\}$, which is partitioned according to (13.9) as $V_S = V_1 \cup V_2 \cup V_3$, where $V_1 = \{4, 5\}$, $V_2 = \{6, 7\}$ and $V_3 = \{8\}$. The graphs G_1, G_2 and G_3 have been computed in Example 13.1. Note that $D(V_L) = \{1, 2, 3\}$ is a zero forcing set in all three graphs (see Figures 13.3, 13.4 and 13.5). We conclude by Theorem 13.2 that $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}_d(G)$.

Lemma 13.2. Consider a directed graph $G = (V, E)$ with leader set $V_L \subseteq V$ and target set $V_T \subseteq V$. Let $\mathcal{Q}_s(G) \subseteq \mathcal{Q}(G)$ be any subclass. Then $(G; V_L; V_T)$ is targeted controllable with respect to $Q_s(G)$ if and only if $(G; D(V_L); V_T)$ is targeted controllable with respect to $Q_s(G)$.

Proof. Let $U = P(V; V_L)$ index the leader set V_L and $W = P(V; D(V_L))$ index the derived set of V_L. Furthermore, let the matrix H be given by $H = P^\top(V; V_T)$. We have that $(G; V_L; V_T)$ is targeted controllable with respect to $Q_s(G)$ if and only if

$$H\langle X \mid \operatorname{im} U\rangle = \mathbb{R}^p \text{ for all } X \in Q_s(G). \tag{13.11}$$

However, as $\langle X \mid \operatorname{im} U\rangle = \langle X \mid \operatorname{im} W\rangle$ for any $X \in \mathcal{Q}(G)$ (see Lemma VI.2 of [141]), (13.11) holds if and only if

$$H\langle X \mid \operatorname{im} W\rangle = \mathbb{R}^p \text{ for all } X \in Q_s(G). \tag{13.12}$$

We conclude that $(G; V_L; V_T)$ is targeted controllable with respect to $Q_s(G)$ if and only if $(G; D(V_L); V_T)$ is targeted controllable with respect to $Q_s(G)$. □

Lemma 13.3. Let $G = (V^-, V^+, E)$ be a bipartite graph and assume V^- is a zero forcing set in G. Then all matrices $M \in \mathcal{P}(G)$ in the pattern class of G have full row rank.

Proof. Note that forces of the form $u \to v$, where $u, v \in V^+$ are not possible, as G is a bipartite graph. Relabel the nodes of V^- and V^+ such that a chronological list of forces is given by $u_i \to v_i$, where $u_i \in V^-$ and $v_i \in V^+$ for $i = 1, 2, ..., |V^+|$. Let $M \in \mathcal{P}(G)$ be a matrix in the pattern class of G. Note that the element M_{ii} is nonzero, as $u_i \to v_i$. Furthermore, M_{ji} is zero for all $j > i$. The latter follows from the fact that u_i would not be able to force v_i if there was an arc $(u_i, v_j) \in E$. We conclude that the columns $1, 2, ..., |V^+|$ of M are linearly independent, hence M has full row rank. □

Proof of Theorem 13.2. Let $D(V_L) = \{1, 2, ..., m\}$, and assume without loss of generality that the matrix U has the form (see Lemma 13.2):

$$U = \begin{bmatrix} I_{m\times m} & 0_{m\times(n-m)} \end{bmatrix}^\top. \tag{13.13}$$

Furthermore, we let $V_S := V_T \setminus D(V_L)$ be given by $\{m+1, m+2, ..., p\}$, where the vertices are ordered in non-decreasing distance with respect to $D(V_L)$. Partition V_S according to the distance of its nodes with respect to $D(V_L)$ as

$$V_S = V_1 \cup V_2 \cup \cdots \cup V_d, \tag{13.14}$$

where for $j \in V_S$ we have $j \in V_i$ if and only if $d(D(V_L), j) = i$ for $i = 1, 2, ..., d$. Finally, assume the target set V_T contains all nodes in the derived set $D(V_L)$. This implies that the matrix H is of the form

$$H = \begin{bmatrix} I_{p\times p} & 0_{p\times(n-p)} \end{bmatrix}. \tag{13.15}$$

Note that by the structure of H and U, the matrix HX^iU is simply the $p \times m$ upper left corner submatrix of X^i. We now claim that HX^iU can be written as follows.

$$HX^iU = \begin{bmatrix} \Lambda_i \\ M_i \\ 0_i \end{bmatrix}, \tag{13.16}$$

where $M_i \in \mathcal{P}(G_i)$ is a $|V_i| \times m$ matrix in the pattern class of G_i, Λ_i is an $(m + |\check{V}_i|) \times m$ matrix containing elements of lesser interest, and 0_i is a zero matrix of dimension $|\hat{V}_i| \times m$.

We proceed as follows: first we prove that the bottom submatrix of (13.16) contains zeros only, secondly we prove that $M_i \in \mathcal{P}(G_i)$. From this, we conclude that equation (13.16) holds.

Note that for $k \in D(V_L)$ and $j \in \hat{V}_i$, we have $d(k, j) > i$ and by Lemma 13.1 it follows that $(X^i)_{jk} = 0$. As $D(V_L) = \{1, 2, ..., m\}$, this means that the bottom $|\hat{V}_i| \times m$ submatrix of HX^iU is a zero matrix.

Subsequently, we want to prove that M_i, the middle block of (13.16), is an element of the pattern class $\in \mathcal{P}(G_i)$. Note that the jth row of M_i corresponds to the element $l := m + |\check{V}_i| + j \in V_i$.

Suppose $(M_i)_{jk} \neq 0$ for a $k \in \{1, 2, ..., m\}$ and $j \in \{1, 2, ..., |V_i|\}$. As M_i is a submatrix of HX^iU, this implies $(HX^iU)_{lk} \neq 0$. Recall that HX^iU is the $p \times m$ upper left corner submatrix of X^i, therefore it holds that $(X^i)_{lk} \neq 0$. Note that for the vertices $k \in D(V_L)$ and $l \in V_i$ we have $d(k, l) \geqslant i$ by the partition of V_S. However, as $(X^i)_{lk} \neq 0$ it follows from Lemma 13.1 that $d(k, l) = i$. Therefore, by the definition of G_i, there is an arc $(k, l) \in E_i$.

Conversely, suppose there is an arc $(k, l) \in E_i$ for $l \in V_i$ and $k \in D(V_L)$. This implies $d(k, l) = i$ in the network graph G. By the distance-information preserving property of X we consequently have $(X^i)_{lk} \neq 0$. We conclude that $(M_i)_{jk} \neq 0$ and hence $M_i \in \mathcal{P}(G_i)$. This implies that equation (13.16) holds, We compute the first dm columns of the output controllability matrix $\left[HU \;\; HXU \;\; HX^2U \;\; \ldots \;\; HX^dU\right]$ as follows:

$$\begin{bmatrix} I & * & * & \ldots & * & * \\ 0 & M_1 & * & \ldots & * & * \\ 0 & 0 & M_2 & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & * & * \\ \vdots & \vdots & \vdots & \ddots & M_{d-1} & * \\ 0 & 0 & 0 & \ldots & 0 & M_d \end{bmatrix}, \tag{13.17}$$

where zeros denote zero matrices and asterisks denote matrices of less interest. As $D(V_L)$ is a zero forcing set in G_i for $i = 1, 2, ..., d$, the matrices $M_1, M_2, ..., M_d$ have full row rank by Lemma 13.3. We conclude that the matrix (13.17) has full row rank, and consequently $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}_d(G)$. □

Note that the condition given in Theorem 13.2 is sufficient, but not necessary. One can verify that the graph $G = (V, E)$ with leader set $V_L = \{1\}$ and target set $V_T = \{2, 3\}$ depicted in Figure 13.6 is an example of a graph for which $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}_d(G)$. However, this graph does not satisfy the conditions of Theorem 13.2.

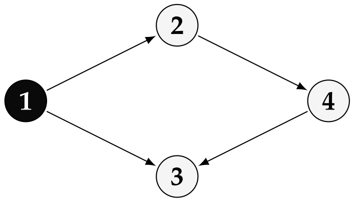

Figure 13.6: Theorem 13.2 not necessary.

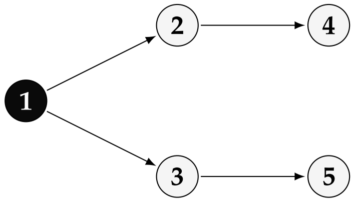

Figure 13.7: Theorem 13.4 not sufficient.

13.4.2 Sufficient richness of $\mathcal{Q}_d(G)$

The notion of sufficient richness of a qualitative subclass was introduced in [142]. We provide an equivalent definition as follows.

Definition 13.4. Let $G = (V, E)$ be a directed graph with leader set $V_L \subseteq V$. A subclass $Q_s(G) \subseteq \mathcal{Q}(G)$ is called *sufficiently rich* if $(G; V_L)$ is controllable with respect to $Q_s(G)$ implies $(G; V_L)$ is controllable with respect to $Q(G)$.

The following geometric characterization of sufficient richness is proven in [142].

Proposition 13.2. A qualitative subclass $Q_s(G) \subseteq \mathcal{Q}(G)$ is sufficiently rich if for all $z \in \mathbb{R}^n$ and $X \in \mathcal{Q}(G)$ satisfying $z^\top X = 0$, there exists an $X' \in Q_s(G)$ such that $z^\top X' = 0$.

The goal of this section is to prove that the qualitative subclass of distance-information preserving matrices is sufficiently rich. This result will be used later on, when we provide a necessary condition for targeted controllability with respect to $\mathcal{Q}_d(G)$. First, however, we state two auxiliary lemmas which will be the building blocks to prove the sufficient richness of $\mathcal{Q}_d(G)$.

Lemma 13.4. Consider q nonzero multivariate polynomials $p_i(x)$, where $i = 1, 2, ..., q$ and $x \in \mathbb{R}^n$. There exists an $\bar{x} \in \mathbb{R}^n$ such that $p_i(\bar{x}) \neq 0$ for $i = 1, 2, ..., q$.

Proof. The proof follows immediately from continuity of polynomials and is omitted. □

Remark 13.2. Without loss of generality, we can assume that the point $\bar{x} \in \mathbb{R}^n$ has only nonzero coordinates. Indeed, if $p_i(\bar{x}) \neq 0$ for $i = 1, 2, ..., q$, there exists an open ball $B(\bar{x})$ around $\bar{x}$ in which $p_i(x) \neq 0$ for $i = 1, 2, ..., q$. Obviously, this open ball contains a point with the aforementioned property.

Lemma 13.5. Let $X \in \mathcal{Q}(G)$ and $D = \text{diag}\,(d_1, d_2, ..., d_n)$ be a matrix with variable diagonal entries. If $d(i, j) = k$ for distinct vertices i and j, then $((XD)^k)_{ji}$ is a nonzero polynomial in the variables $d_1, d_2, ..., d_n$.

Proof. Note that $((XD)^k)_{ji}$ is given by

$$\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_{k-1}=1}^{n} (XD)_{i_1,i}\,(XD)_{i_2,i_1} \cdots (XD)_{j,i_{k-1}},$$

which equals

$$\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_{k-1}=1}^{n} d_i X_{i_1,i}\, d_{i_1} X_{i_2,i_1} \cdots d_{i_{k-1}} X_{j,i_{k-1}}. \tag{13.18}$$

Since the distance $d(i, j)$ is equal to k, there exists at least one path of length k from i to j, which we denote by $(i, i_1), (i_1, i_2), ..., (i_{k-1}, j)$. It follows that the corresponding elements of the matrix X, i.e. the elements $X_{i_1,i}, X_{i_2,i_1}, X_{j,i_{k-1}}$ are nonzero. Therefore, the term

$$d_i X_{i_1,i}\, d_{i_1} X_{i_2,i_1} \cdots d_{i_{k-1}} X_{j,i_{k-1}} \tag{13.19}$$

is nonzero (as a function of $d_i, d_{i_1}, d_{i_2}, ..., d_{i_{k-1}}$). Furthermore, this combination of k diagonal elements is unique in the sense that there does not exist another summand on the right-hand side of (13.18) with exactly the same elements. This implies that the term (13.19) does not vanish (as a polynomial). We conclude that $((XD)^k)_{ji}$ is a nonzero polynomial function in the variables $d_1, d_2, ..., d_n$. □

Theorem 13.3. The subclass $\mathcal{Q}_d(G)$ is sufficiently rich.

Proof. Given a matrix $X \in \mathcal{Q}(G)$, using Lemmas 13.4 and 13.5, we first prove there exists a diagonal matrix $\bar{D}$ with nonzero diagonal components such that $X\bar{D} \in \mathcal{Q}_d(G)$. From this we will conclude $\mathcal{Q}_d(G)$ is sufficiently rich.

Let $D = \text{diag}\,(d_1, d_2, ..., d_n)$ be a matrix with variable diagonal entries. We define $p_{ij} := ((XD)^{d(i,j)})_{ji}$ for distinct $i, j = 1, 2, ..., n$. By Lemma 13.5 we have that $p_{ij}(d_1, d_2, ..., d_n)$ is a nonzero polynomial in the variables $d_1, d_2, ..., d_n$. Moreover, Lemma 13.4 states the existence of nonzero real constants $\bar{d}_1, \bar{d}_2, ..., \bar{d}_n$ such that

$$p_{ij}(\bar{d}_1, \bar{d}_2, ..., \bar{d}_n) \neq 0 \text{ for distinct } i, j = 1, 2..., n. \tag{13.20}$$

Therefore, the choice $\bar{D} = \text{diag}\,(\bar{d}_1, \bar{d}_2, ..., \bar{d}_n)$ implies $X\bar{D} \in \mathcal{Q}_d(G)$. Let $z \in \mathbb{R}^n$ be a vector such that $z^\top X = 0$ for an $X \in \mathcal{Q}(G)$. The choice of $X' = X\bar{D}$ yields a matrix $X' \in \mathcal{Q}_d(G)$ for which $z^\top X' = 0$. By Proposition 13.2 it follows that $\mathcal{Q}_d(G)$ is sufficiently rich. □

13.4.3 Necessary condition for targeted controllability

In addition to the previously established sufficient condition for targeted controllability, we give a necessary graph-theoretic condition for targeted controllability in Theorem 13.4.

Theorem 13.4. Let $G = (V, E)$ be a directed graph with leader set $V_L \subseteq V$ and target set $V_T \subseteq V$. If $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}_d(G)$ then $V_L \cup (V \setminus V_T)$ is a zero forcing set in G.

Proof. Assume without loss of generality that $V_L \cap V_T = \varnothing$. Hence, $V_L \cup (V \setminus V_T) = V \setminus V_T$. We partition the vertex set V into V_L, $V \setminus (V_L \cup V_T)$ and V_T. Without loss, we write $V_L = \{1, 2, \ldots, m\}$ and $V_T = \{n-p+1, n-p+2, \ldots, n\}$. Moreover, we use the shorthand notation $\bar{n} = n - p - m$ to denote the number of nodes that are neither a target node nor a leader node. Accordingly, the input and output matrices $U = P(V; V_L)$ and $H = P^\top(V; V_T)$ satisfy

$$U = \begin{bmatrix} I_{m\times m} & 0_{m\times \bar{n}} & 0_{m\times p} \end{bmatrix}^\top, \tag{13.21}$$

and

$$H = \begin{bmatrix} 0_{p\times m} & 0_{p\times \bar{n}} & I_{p\times p} \end{bmatrix}. \tag{13.22}$$

Note that $\ker H = \operatorname{im} R$, where $R := P(V; (V \setminus V_T))$ is given by

$$R = \begin{bmatrix} I_{m\times m} & 0_{m\times \bar{n}} & 0_{m\times p} \\ 0_{\bar{n}\times m} & I_{\bar{n}\times \bar{n}} & 0_{\bar{n}\times p} \end{bmatrix}^{\top}. \tag{13.23}$$

Since for all $X \in \mathcal{Q}_d(G)$ we have

$$\ker H + \langle X \mid \operatorname{im} U \rangle = \mathbb{R}^n, \tag{13.24}$$

or, equivalently,

$$\operatorname{im} R + \langle X \mid \operatorname{im} U \rangle = \mathbb{R}^n, \tag{13.25}$$

we obtain

$$\langle X \mid \operatorname{im} \begin{bmatrix} U & R \end{bmatrix} \rangle = \mathbb{R}^n. \tag{13.26}$$

As $\operatorname{im} U \subseteq \operatorname{im} R$, (13.26) implies $\langle X \mid \operatorname{im} R \rangle = \mathbb{R}^n$ for all $X \in \mathcal{Q}_d(G)$, or, equivalently, the pair (X, R) is controllable for all $X \in \mathcal{Q}_d(G)$. Furthermore, by sufficient richness of $\mathcal{Q}_d(G)$, it follows that (X, R) is controllable for all $X \in \mathcal{Q}(G)$. We conclude from Theorem 13.1 that $V \setminus V_T$ is a zero forcing set. □

Example 13.3. Consider the directed graph $G = (V, E)$ with leader set $V_L = \{1, 2\}$ and target set $V_T = \{1, 2, ..., 8\}$ as depicted in Figure 13.1. We know from Example 13.2 that $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}_d(G)$. The set $V_L \cup (V \setminus V_T) = \{1, 2, 9, 10\}$ is colored black in Figure 13.8. Indeed, $V_L \cup (V \setminus V_T)$ is a zero forcing set in G. A possible chronological list of forces is: $1 \to 3$, $3 \to 4$, $2 \to 5$, $4 \to 6$, $6 \to 8$ and $9 \to 7$.

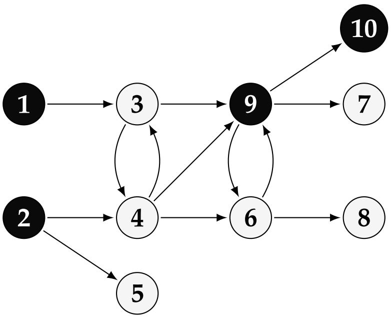

Figure 13.8: Zero forcing set.

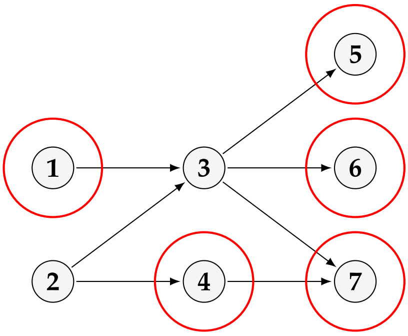

Figure 13.9: Directed graph $G = (V, E)$ with target set $V_T = \{1, 4, 5, 6, 7\}$.

The condition provided in Theorem 13.4 is necessary for targeted controllability, but not sufficient. To prove this fact, consider the directed graph with leader set $V_L = \{1\}$ and target set $V_T = \{4, 5\}$ given in Figure 13.7. It can be shown that $(G; V_L; V_T)$ is not targeted controllable with respect to $\mathcal{Q}_d(G)$, even though $V_L \cup (V \setminus V_T) = \{1, 2, 3\}$ is a zero forcing set.

So far, we have provided a necessary and a sufficient topological condition for targeted controllability. However, given a network graph with target set, it is not clear how to choose leaders achieving target control. Hence, in the following section we focus on a leader selection algorithm.

13.4.4 Leader selection algorithm

We now address Problem 2, as introduced in Section 13.3. That is, given a directed graph $G = (V, E)$ with target set $V_T \subseteq V$, we want to find a leader set $V_L \subseteq V$ of minimum cardinality such that $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}_d(G)$. Such a leader set is called a *minimum leader set*. In general, a graph G with target set V_T can have multiple minimum leader sets. In this section we first prove that there is no polynomial-time algorithm that solves Problem 2 (assuming $\mathrm{P} \neq \mathrm{NP}$). Subsequently, we provide a heuristic algorithm to determine leader sets achieving targeted controllability.

Theorem 13.5. Assuming $\mathrm{P} \neq \mathrm{NP}$, there is no polynomial-time algorithm that solves Problem 2.

Proof. Assume that $\mathrm{P} \neq \mathrm{NP}$. The problem of finding a minimum zero forcing set was proven to be NP-hard in [2], by a reduction from the directed Hamiltonian cycle problem. Consequently, by Theorem 13.1 there is no polynomial-time algorithm to determine a minimum leader set V_L that achieves controllability of $(G; V_L)$ with respect to $\mathcal{Q}(G)$. Recall from Theorem 13.3 that the subclass $\mathcal{Q}_d(G)$ is sufficiently rich. Hence, controllability of $(G; V_L)$ with respect to $\mathcal{Q}(G)$ is equivalent with controllability of $(G; V_L)$ with respect to $\mathcal{Q}_d(G)$. Therefore, there is no polynomial-time algorithm to determine a minimum leader set V_L such that $(G; V_L)$ is controllable with respect to $\mathcal{Q}_d(G)$. Note that "ordinary" controllability of $(G; V_L)$ can be regarded as a special case of targeted controllability of $(G; V_L; V_T)$, where $V_T = V$. We conclude that there is no polynomial-time algorithm solving Problem 2 (assuming $\mathrm{P} \neq \mathrm{NP}$). □

Next, we propose a heuristic approach to compute a (minimum) leader set that achieves targeted controllability. The algorithm consists of two phases. Firstly, we identify a set of nodes in the graph G from which all target nodes can be reached. These nodes are taken as leaders. Secondly, this set of leaders is extended to achieve targeted controllability.

To explain the first phase of the algorithm, we introduce some notation. First of all, we define the notion of *root set*.

Definition 13.5. Consider a directed graph $G = (V, E)$ and a target set $V_T \subseteq V$. A subset $V_R \subseteq V$ is called a *root set* of V_T if for any $v \in V_T$ there exists a vertex $u \in V_R$ such that $d(u, v) < \infty$.

A root set of V_T of minimum cardinality is called a *minimum root set* of V_T. Note that the cardinality of a minimum root set of V_T is a lower bound on the minimum number of leaders rendering $(G; V_L; V_T)$ targeted controllable. Indeed, it is easy to see that if there are no paths from any of the leader nodes to a target node, the graph is not targeted controllable. The first step of the proposed algorithm is to compute the minimum root set of V_T. Let the vertex and target sets be given by $V = \{1, 2, ..., n\}$ and $V_T = \{v_1, v_2, ..., v_p\} \subseteq V$ respectively. Furthermore, define a matrix $A \in \mathbb{R}^{p \times n}$ in the following way. For $j \in V$ and $v_i \in V_T$ let

$$A_{ij} := \begin{cases} 1 \text{ if } d(j, v_i) < \infty, \\ 0 \text{ otherwise.} \end{cases} \tag{13.27}$$

That is, the matrix A contains zeros and ones only, where coefficients with value one indicate the existence of a path between the corresponding vertices.

Remark 13.3. The matrix A can be found using p runs of Dijkstra's algorithm [51], with total computational complexity $\mathcal{O}(pn^2)$. This can be done by transposing the graph $G = (V, E)$, i.e., computing $\bar{G} = (V, \bar{E})$, where $(j, i) \in \bar{E}$ if and only if $(i, j) \in E$. Dijkstra's algorithm can be applied to find the distance from a (target) node to all other nodes. We apply Dijkstra's algorithm in $\bar{G}$ to all target nodes, to find a $p \times n$ distance matrix D, with elements in $\mathbb{N} \cup \{\infty\}$. Here, each element D_{ij} is equal to the distance from node $v_i \in V_T$ to $j \in V$ in the graph $\bar{G}$. As the graph $\bar{G}$ is the transposed of G, we have that D_{ij} equals the distance from $j \in V$ to $v_i \in V_T$ in the original graph G. Consequently, the matrix A is easily obtained from D by changing "∞"-elements in D to 0, and all other elements in D to 1. Dijkstra's algorithm and graph transposition have computational complexity $\mathcal{O}(n^2)$. As we execute Dijkstra's algorithm p times, the total procedure has computational complexity $\mathcal{O}(pn^2)$. The distance matrix D will also become useful in the second phase of the leader selection algorithm.

Now, finding a minimum root set of V_T boils down to finding a binary vector $x \in \mathbb{R}^n$ with minimum number of ones such that $Ax \geqslant \mathbb{1}_p$, where the inequality is defined element-wise and $\mathbb{1}_p$ denotes the p-dimensional vector of all ones. In the vector x, coefficients with value one correspond to elements in the root set of V_T. It is for this reason we can formulate the minimum root set problem as a binary integer linear program

$$\begin{aligned} &\text{minimize } \mathbb{1}_n^\top x \\ &\text{subject to } Ax \geqslant \mathbb{1}_p \\ &\qquad \text{and } x \in \{0, 1\}^n. \end{aligned} \tag{13.28}$$

Linear programs of this form can be solved using software like CPLEX or Matlab.

Remark 13.4. Note that the minimum root set problem (and in general binary integer programming) is NP-hard. This can be shown by constructing an NP-reduction from the well-known NP-hard set-covering problem [33] to the minimum root set problem. Therefore, for large-scale problems it is advisable to apply an approximation algorithm to compute an approximate minimum root set. The greedy algorithm for the set-covering problem [33] can be directly applied to the minimum root set problem. This algorithm has computational complexity

$\mathcal{O}(p^2 n)$, and has an approximation ratio of $\mathcal{O}(\ln(n))$. That is, the approximation algorithm returns a root set containing at most $\mathcal{O}(\ln(n))$ times the optimal number of vertices in the minimum root set. Furthermore, it has been shown in [58] that no polynomial-time algorithm for the set-covering problem can have a better approximation ratio than $\mathcal{O}(\ln(n))$.

In the following example, we illustrate how the minimum root set problem can be regarded as a binary integer linear program.

Example 13.4. Consider the directed graph $G = (V, E)$ with target set $V_T = \{1, 4, 5, 6, 7\}$ depicted in Figure 13.9. The goal of this example is to find a minimum root set for V_T. The matrix A, as defined in (13.27), is given by

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}. \tag{13.29}$$

Note that $x = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}^\top$ satisfies the inequality $Ax \geqslant \mathbb{1}_p$ and the constraint $x \in \{0, 1\}^7$. Furthermore, the vector x minimizes $\mathbb{1}_n^\top x$ under these constraints. This can be seen in the following way: there is no column of A in which all elements equal 1, hence there is no vector x with a single one such that $Ax \geqslant \mathbb{1}_p$ is satisfied. Therefore, x solves the binary integer linear program (13.28), from which we conclude that the choice $V_R = \{1, 4\}$ yields a minimum root set for V_T. Indeed, observe in Figure 13.9 that we can reach all nodes in the target set starting from the nodes 1 and 4. It is worth mentioning that the choice of minimum root set is not unique: the set $\{1, 2\}$ is also a minimum root set for V_T.

In general, the minimum root set V_R of V_T does not guarantee targeted controllability of $(G; V_R; V_T)$ with respect to $\mathcal{Q}_d(G)$. For instance, it can be shown for the graph G and target set V_T of Example 13.4 that the leader set $V_L = \{1, 4\}$ does not render $(G; V_L; V_T)$ targeted controllable with respect to $\mathcal{Q}_d(G)$. Hence, we propose a greedy approach to extend the minimum root set of V_T to a leader set that does achieve targeted controllability.

Recall from Theorem 13.2 that $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}_d(G)$ if $D(V_L)$ is a zero forcing set in the bipartite graphs $G_i = (D(V_L), V_i, E_i)$ for $i = 1, 2, ..., d$, where $V_i \subseteq V_T$ is the set of target nodes having distance i from $D(V_L)$. Given an initial set of leaders V_L, we compute its derived set $D(V_L)$ and verify whether we can force all nodes in the bipartite graphs G_i for $i = 1, 2, ..., d$. Suppose that in the bipartite graph G_k the set V_k cannot be forced by $D(V_L)$ for a $k \in \{1, 2, ..., d\}$. In this case, we choose an additional leader as follows. Let $V_u \subseteq V_k$ be the set of vertices in V_k that can't be forced. Suppose $v_i \in V_u$ is the vertex in V_u from which most target nodes can be reached. Then we choose v_i as additional leader. Consequently, we have extended our leader set V_L to $V_L \cup \{v_i\}$. Note that v_i can be easily found by computing the column sums of the columns in A corresponding to the nodes in V_u.

With the extended leader set we can repeat the procedure, until the leaders render the graph targeted controllable. This idea is captured more formally in the following leader selection algorithm. One should recognize the two phases of leader selection: firstly, a minimum root set is computed. Subsequently, the minimum root set is greedily extended to a leader set achieving targeted controllability.

Algorithm 2 Leader Selection Procedure

```
Input: Directed graph $G = (V, E)$, target set $V_T \subseteq V$;
Output: Leader set $V_L \subseteq V$ achieving target control;
 1: Let $V_L = \varnothing$;
 2: Compute matrix $A$, given in (13.27);
 3: Compute a solution $x$ to the linear program (13.28);
 4: for $i = 1$ to $n$ do
 5:     if $x_i = 1$ then
 6:         $V_L \leftarrow V_L \cup \{i\}$;
 7:     end if
 8: end for
 9: Compute $D(V_L)$;
10: $i \leftarrow 1$;
11: repeat
12:     Compute $V_i$ and $G_i = (D(V_L), V_i, E_i)$;
13:     if $D(V_L)$ forces $V_i$ in $G_i$ then
14:         $i \leftarrow i + 1$;
15:     else
16:         Find unforced $v \in V_i$ reaching the most targets;
17:         $V_L \leftarrow V_L \cup \{v\}$;
18:         Compute $D(V_L)$;
19:         $i \leftarrow 1$;
20:     end if
21: until $d(D(V_L), w) < i$ for all $w \in V_T$;
22: return $V_L$.
```

Remark 13.5. Algorithm 1 can be implemented with computational complexity $\mathcal{O}(p^2n^2)$ if the heuristic set-covering procedure [33] is used to compute an approximate minimum root set (step 3). This can be seen as follows. First, note that the repeat (step 11) runs at most p^2 times. Moreover, every step within the repeat runs in time at most $\mathcal{O}(n^2)$. Indeed, using the distance matrix D (see Remark 13.3), we can compute V_i and $G_i = (D(V_L), V_i, E_i)$ in time $\mathcal{O}(pn)$. Furthermore, the derived sets in steps 13 and 18 can be computed in $\mathcal{O}(n^2)$ time [207]. Finally, step 16 compares the column sums of at most p columns of length p, and can hence be implemented in $\mathcal{O}(p^2)$. As $p \leqslant n$, we find that all steps within the repeat run with computational complexity at most $\mathcal{O}(n^2)$. Consequently, steps 11-21 have computational complexity $\mathcal{O}(p^2n^2)$. Steps 1-10 of the algorithm run in time complexity less than $\mathcal{O}(p^2n^2)$ (see Remarks 13.3 and 13.4). We conclude that Algorithm 1 can be implemented with computational complexity $\mathcal{O}(p^2n^2)$ if the heuristic set-covering procedure is applied to step 3.

Note that Algorithm 1 is a heuristic algorithm, and the quality of its solution with respect to the actual minimum leader set is not known. The problem of finding a minimum leader set achieving target control is more general than the problem of finding a minimum leader set achieving (full) strong structural controllability. Indeed, by the choice $V_T = V$, one can solve the latter problem using the former. Consequently, Remark 13.1 suggests that the minimum number of leaders achieving target control cannot be approximated within a large factor.

However, it is worth mentioning that Algorithm 1 does return optimal results for some specific types of graphs. In the case of a cycle or a complete graph (with target set $V_T = V$), Algorithm 1 returns leader sets of respectively 2 and $n-1$ leaders. This is in agreement with the optimal results found in [142] for cycle and complete graphs.

13.5 ILLUSTRATIVE EXAMPLE

In this section, we illustrate our algorithm using an example. Consider the directed graph given in Figure 13.10, with targets $V_T = \{2, 3, 6, 8, 10, 13, 15, 16, 17, 20\}$. The goal of this example is to compute a leader set V_L such that $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}_d(G)$.

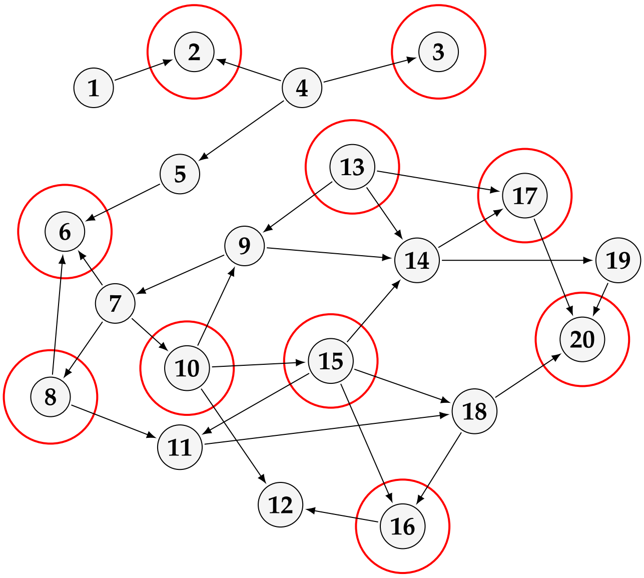

Figure 13.10: Directed graph $G = (V, E)$ with encircled target nodes $V_T = \{2, 3, 6, 8, 10, 13, 15, 16, 17, 20\}$.

The first step of Algorithm 1 is to compute the matrix A, defined in (13.27). For this example, A is given as follows.

$$A = \begin{bmatrix} 1&1&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&1&1&1&1&1&1&1&0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&1&1&1&0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&0&1&1&0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&0&1&1&0&0&1&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&1&1&1&1&1&0&1&0&1&1&0&1&0&0 \\ 0&0&0&0&0&0&1&0&1&1&0&0&1&1&1&0&1&0&0&0 \\ 0&0&0&0&0&0&1&1&1&1&1&0&1&1&1&0&1&1&1&1 \end{bmatrix}$$

Using the Matlab function `intlinprog`, we find the optimal solution

$$x = \begin{bmatrix} 0&0&0&1&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0 \end{bmatrix}^\top$$

to the binary linear program (13.28). Hence, a minimum root set for V_T is given by $\{4, 13\}$. Following Algorithm 1, we define our initial leader set $V_L = \{4, 13\}$. As nodes 4 and 13 both have three out-neighbours, the derived set of V_L is simply given by $D(V_L) = \{4, 13\}$. The next step of the algorithm is to compute the first bipartite graph $G_1 = (D(V_L), V_1, E_1)$, which we display in Figure 13.11.

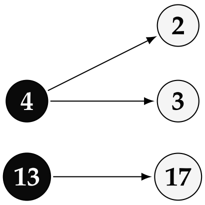

Figure 13.11: G_1 for $D(V_L) = \{4, 13\}$.

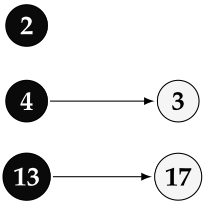

Figure 13.12: G_1 for $D(V_L) = \{2, 4, 13\}$.

Observe that the nodes 2 and 3 cannot be forced. As both nodes can reach the same number of target nodes, we simply choose node 2 as additional leader. The process now repeats itself, we redefine $V_L = \{2, 4, 13\}$ and compute $D(V_L) = \{2, 4, 13\}$. Furthermore, for this leader set, the graph $G_1 = (D(V_L), V_1, E_1)$ is given in Figure 13.12. In this case, the set $V_1 = \{3, 17\}$ of nodes having distance one with respect to $D(V_L)$ is forced. Therefore, we continue with the second bipartite graph $G_2 = (D(V_L), V_2, E_2)$, given in Figure 13.13.

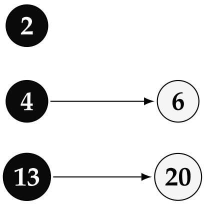

Figure 13.13: G_2 for $D(V_L) = \{2, 4, 13\}$.

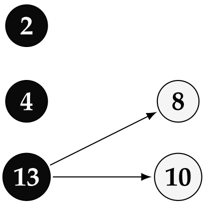

Figure 13.14: G_3 for $D(V_L) = \{2, 4, 13\}$.

The set V_2 is forced by $D(V_L)$ in the graph G_2, hence we continue to investigate the third bipartite graph consisting of nodes having distance three with respect to $D(V_L)$. This graph is displayed in Figure 13.14. As neither node 8 nor 10 can be forced, we have to add another leader. Node 8 can reach 4 target nodes, while node 10 can reach 7 target nodes. Hence, we choose node 10 as additional leader. In other words, we redefine $V_L = \{2, 4, 10, 13\}$. The new bipartite graphs G_1, G_2 and G_3 are given in Figure 13.15.

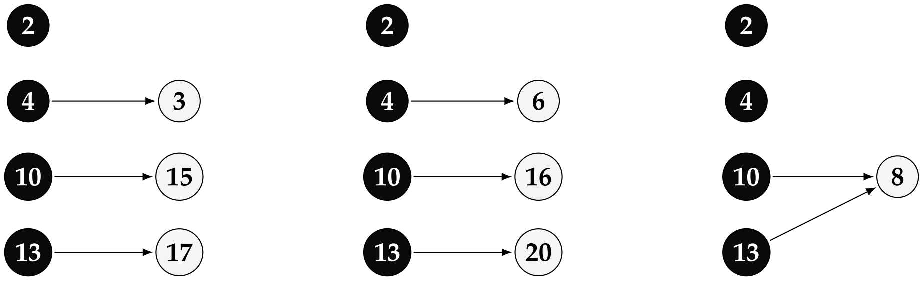

Figure 13.15: Graphs G_1, G_2 and G_3 for $D(V_L) = \{2, 4, 10, 13\}$.

Note that in this case $D(V_L)$ is a zero forcing set in all three bipartite graphs. Furthermore, since $d(D(V_L), v) < 4$ for all $v \in V_T$, Algorithm 1 returns the leader set $V_L = \{2, 4, 10, 13\}$. This choice of leader set guarantees that $(G; V_L; V_T)$ is targeted controllable with respect to $\mathcal{Q}_d(G)$. For the sake of clarity, we display the network graph in Figure 13.16.

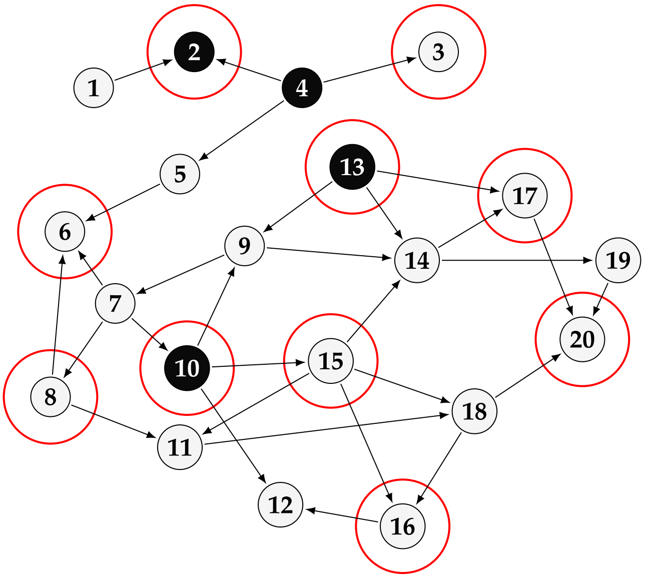

Figure 13.16: Network $G = (V, E)$ with targets $V_T = \{2, 3, 6, 8, 10, 13, 15, 16, 17, 20\}$, and leader set $V_L = \{2, 4, 10, 13\}$.

13.6 conclusions

In this chapter, strong targeted controllability for the class of distance-information preserving matrices has been discussed. We have provided a sufficient graph-theoretic condition for strong targeted controllability, expressed in terms of zero-forcing sets of particular distance-related bipartite graphs. We have shown that this result improves the known sufficient topological condition [141] for strong targeted controllability of the class of distance-information preserving matrices.

Motivated by the observation that the aforementioned sufficient condition is not a one-to-one correspondence, we provided a necessary topological condition for strong targeted controllability. This condition was proved using the fact that the subclass of distance-information preserving matrices is sufficiently rich. Finally, we showed that there is no polynomial-time algorithm to compute minimum leader sets achieving targeted controllability (assuming $P \neq NP$). Therefore, a heuristic leader selection algorithm was given to compute approximate minimum leader sets achieving target control. The algorithm comprises two phases: firstly, it computes a minimum root set of the target set, i.e. a set of vertices from which all target nodes can be reached. Secondly, this minimum root set is greedily extended to a leader set achieving target control.

Both graph-theoretic conditions for strong targeted controllability provided in this chapter are not one-to-one correspondences. Hence, finding a necessary and sufficient topological condition for strong targeted controllability is still an open problem. Furthermore, investigating other system-theoretic concepts like disturbance decoupling and fault detection for the class of distance-information preserving matrices is among the possibilities for future research.

14 CONCLUSIONS

In this thesis we have studied four problems, namely data-driven analysis and control, topology identification, network identifiability and structural controllability. In what follows we highlight the main contributions and provide some ideas for future work.

14.1 CONTRIBUTIONS

In Chapter 2 we have studied Willems' fundamental lemma. This result asserts that under suitable hypotheses all trajectories of a linear system can be expressed in terms of a single (measured) one. We have extended the result to the setting where multiple -possibly short- system trajectories are measured. This result can be applied, for example, to system identification from data sets with missing samples.

In Chapter 3 we have introduced a general framework for data informativity for system analysis and control. We have applied this framework to analyze stability, stabilizability and controllability. This has led to data-driven Hautus tests that can be used to verify controllability and stabilizability of a system directly on the basis of data. We have also studied stabilization by state feedback and linear quadratic regulation (LQR). In the case of stabilization, we saw that the corresponding conditions on the data are weaker than those for system identification. In general, it is thus easier to learn a stabilizing controller than it is to learn a system model from data. In the case of the LQR problem, however, we saw that the data informativity conditions are practically the same as for system identification. For both stabilization and LQR we have established data-driven control design techniques in terms of linear matrix inequalities. In Chapter 4 we have extended the work on data-driven control. In particular, we have treated the suboptimal LQR and $\mathcal{H}_2$ problems. Also for these problems we were able to provide data-driven control design methods. In addition, we have shown by numerical simulations that there is an intuitive trade-off between the number of data samples and the optimal controller performance. In Chapter 5 we have extended the control design methods from Chapters 3 and 4 to a setting involving noisy data. To do so, we have established a generalization of the S-lemma [243] to matrix variables. This result provides verifiable conditions under which one quadratic matrix inequality implies another one. The matrix S-lemma is not only interesting in the setting of Chapter 5, but may also find other applications in the general area of robust control. In Chapter 6 we have aimed at verifying dissipativity properties of a linear system from data. We have studied this problem both for exact and noisy data. In the case of noisy data, we have introduced a new type of dualization lemma. Combined with the matrix S-lemma, this led to new data-driven tests for dissipativity.

In Chapter 7 we have studied the problem of identifying the topology of a heterogeneous network of linear systems. We have considered both the *identifiability*, as well as the *reconstruction* aspect of this problem. Our identifiability results have recovered and generalized a result for the special case of networks of single integrators [163, 226]. We have also seen that homogeneous networks of single-input single-output systems have quite special identifiability properties that do not extend to the general case of heterogeneous networks. Our topology identification scheme has leveraged Willems' fundamental lemma (Chapter 2) to identify the network's Markov parameters. Then, the network interconnection matrix has been reconstructed by solving a *generalized Sylvester equation* involving the Markov parameters. We have proven that the network topology can be uniquely reconstructed in this way, under the assumptions of topological identifiability and persistently exciting inputs. In Chapter 8 we have investigated a more specific network setup, where the dynamics of each node is a single integrator, and the network is *autonomous*. In this case, excitation has to be secured through the initial conditions of the network. The more specialized setting of Chapter 8 has allowed us to come up with more specific reconstruction methods, in terms of Lyapunov equations.

In Chapter 9 we have focussed on the problem of assessing global identifiability of networks with *known* graph structure. To do so, we have introduced a new graph-theoretic concept called the graph simplification process. We have used this process to assess the identifiability of a subset of the network's transfer functions, and of all the transfer functions in the network. A noteworthy fact is that identifiability can often be secured with a limited number of measured nodes. The number of measured nodes required for full network identifiability is, however, lower bounded by the maximum degree of the network. In a special case, we have established an interesting analogy between global identifiability (Chapter 9) and the concept of *generic* identifiability [81]. Indeed, generic identifiability is related to the existence of certain *vertex-disjoint paths* [81]. On the other hand, global identifiability depends on the existence of a unique set of vertex-disjoint paths. In Chapter 10 we have further explored global identifiability, but for a different class of undirected networks described by state-space systems. In this case we have provided sufficient conditions for identifiability in terms of so-called *zero forcing sets*. The results of Chapter 10 have revealed that in the more specialized (undirected) setup, identifiability can be achieved not only with a limited number of measured nodes, but also with a limited number of *excited* nodes.

In Chapter 11 we have focussed on strong structural controllability. We have studied this problem in the general setting of zero/nonzero/arbitrary structured systems. We have provided both algebraic and graph-theoretic necessary and sufficient conditions for controllability in this setting. We have also shown that our results generalize all related work in the literature. In addition, we have seen that seemingly incomparable results in [207] and [142] follow from our main results; our work has thus revealed an overarching theory. In Chapter 12 we have extended the work on zero/nonzero/arbitrary structured systems. In this chapter, we have focussed on strong structural input-state observability, output controllability and controllability of differential algebraic equations. We have been able to provide conditions for these strong structural properties, using results on the addition and multiplication of zero/nonzero/arbitrary pattern matrices. Finally, in Chapter 13 we have studied strong structural output controllability in a network setting. Here, the output of the network consists of the states of a subset of network nodes, called target nodes. Output controllability is often referred to as *targeted controllability* in this context. We have followed up on the work of [141] by studying targeted controllability for a subclass of state matrices, called distance-information preserving matrices. For this subclass, we have been able to come up with more powerful sufficient conditions for strong structural targeted controllability. We have also provided necessary conditions for targeted controllability, as well as a strategy for the selection of input nodes.

14.2 OUTLOOK

In what follows, we provide some ideas for future work. In Chapters 3, 4, 5 and 6, we have gained insight in the conditions on the data that are necessary for different analysis and control problems. So far, our results are applicable to discrete-time, linear time-invariant systems. However, extensions to continuous-time systems and nonlinear dynamics are also of interest. One could study, for example, model classes of nonlinear systems, where the involved nonlinear functions are unknown linear combinations of known basis functions.

An interesting follow-up question is how to *generate* informative data. In other words, how can we design inputs for a system such that the resulting measured data are informative. In view of our results, a particularly relevant question is how to design experiments such that data-based linear matrix inequalities are feasible. This problem is especially interesting in the noisy data setting in Chapters 5 and 6 where persistency of excitation is generally not sufficient to guarantee that data are informative.

Another topic for future work is related to the noise model employed in Chapter 5. In this chapter, we have assumed a type of energy bound on the noise. It would be interesting to investigate whether control design is also possible for other types of bounds, for example, bounds on the norm of individual noise samples. We believe that also for this kind of noise models, applying a type of S-procedure can be a promising approach for control design.

As also mentioned in the introduction, the problem of topology identification (Chapters 7 and 8) is not only relevant in the systems and control community but also in physics. The types of models that we have considered so far, however, are rather control-oriented. It would thus be of interest to explore topology identification for different types of dynamics, for example, those described by the Schrödinger equation and Liouville-von Neumann equation appearing in quantum mechanics. We believe that there is potential for a system-theoretic approach also in this area.

Related to the identifiability results in Chapters 9 and 10, there is a *synthesis* problem that is largely open. The general problem involves selecting locations for the external excitation signals and measured signals such that the network is globally identifiable. In the context of Chapter 10, input and output sets can be designed by computing a (minimal) zero forcing set for the graph. Minimal zero forcing sets are known for special types of graphs such as path, cycle, complete and tree graphs. For general graphs, however, finding a minimal zero forcing set is known to be NP-hard [2]. In the setting of Chapter 9 it is not yet clear how to choose a (minimal) set of measured nodes such that the network is globally identifiable. We do note that a greedy approach was recently developed to select excited/measured nodes ensuring *generic* identifiability [31].

In the study of structured systems, dependencies among entries are very difficult to deal with. We have seen this problem in Chapters 12 and 13 when studying output controllability. Indeed, the pattern class of the product of two pattern matrices is generally not equal to the product of the pattern classes of the individual matrices (Chapter 12). The test for output controllability in Section 12.3 is thus conservative in general. Dependencies among entries in pattern matrices have been studied in the context of structural controllability [94, 112]. There are, however, still many open questions in this direction. For instance, it is not yet clear how to handle dependencies due to pattern matrix multiplication.

BIBLIOGRAPHY

[1] W. Aangenent, D. Kostic, B. de Jager, R. van de Molengraft, and M. Steinbuch. Data-based optimal control. In *Proceedings of the American Control Conference*, pages 1460–1465, June 2005.

[2] A. Aazami. Hardness results and approximation algorithms for some problems on graphs. *PhD thesis, University of Waterloo*, 2008.

[3] J. Adebayo, T. Southwick, V. Chetty, E. Yeung, Y. Yuan, J. Gonçalves, J. Grose, J. Prince, G. B. Stan, and S. Warnick. Dynamical structure function identifiability conditions enabling signal structure reconstruction. In *Proceedings of the IEEE Conference on Decision and Control*, pages 4635–4641, Dec 2012.

[4] S. Alemzadeh and M. Mesbahi. Distributed Q-learning for dynamically decoupled systems. In *Proceedings of the American Control Conference*, pages 772–777, July 2019.

[5] A. Allibhoy and J. Cortés. Data-based receding horizon control of linear network systems. *https://arxiv.org/abs/2003.09813*, 2020.

[6] D. Alpago, F. Dörfler, and J. Lygeros. An extended Kalman filter for data-enabled predictive control. *https://arxiv.org/abs/2003.08269*, 2020.

[7] K. J. Åström and B. Wittenmark. *Adaptive Control*. Addison-Wesley, 1989.

[8] W Bachmann. Strenge strukturelle steuerbarkeit und beobachtbarkeit von mehrgrößensystemen / strong structural controllability and observability of multi-variable systems. *Regelungstechnik*, 29(1-12):318–323, 1981.

[9] G. Baggio, D. S. Bassett, and F. Pasqualetti. Data-driven control of complex networks. *https://arxiv.org/abs/2003.12189*, 2020.

[10] G. Baggio, V. Katewa, and F. Pasqualetti. Data-driven minimum-energy controls for linear systems. *IEEE Control Systems Letters*, 3(3):589–594, July 2019.

[11] R. H. Bartels and G. W. Stewart. Solution of the matrix equation AX + XB = C. *Communications of the ACM*, 15(9):820–826, 1972.

[12] G. Basile and G. Marro. *Controlled and Conditioned Invariants in Linear System Theory*. Prentice Hall, 1992.

[13] G. Battistelli and P. Tesi. Detecting topology variations in dynamical networks. In *Proceedings of the IEEE Conference on Decision and Control*, pages 3349–3354, 2015.

[14] G. Battistelli and P. Tesi. Detecting topology variations in networks of linear dynamical systems. *IEEE Transactions on Control of Network Systems*, 5(3):1287–1299, Sept 2018.

[15] A. S. Bazanella, M. Gevers, J. M. Hendrickx, and A. Parraga. Identifiability of dynamical networks: Which nodes need be measured? In *Proceedings of the IEEE Conference on Decision and Control*, pages 5870–5875, Dec 2017.

[16] J. Berberich and F. Allgöwer. A trajectory-based framework for data-driven system analysis and control. *arxiv.org/abs/1903.10723*, 2019.

[17] J. Berberich, A. Koch, C. W. Scherer, and F. Allgöwer. Robust data-driven state-feedback design. In *American Control Conference*, pages 1532–1538, 2020.

[18] J. Berberich, J. Köhler, M. A. Müller, and F. Allgöwer. Data-driven model predictive control with stability and robustness guarantees. *arxiv.org/abs/1906.04679*, 2019.

[19] D. S. Bernstein. *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton University Press, 2011.

[20] A. Bisoffi, C. De Persis, and P. Tesi. Data-based guarantees of set invariance properties. *arxiv.org/abs/1911.12293*, 2019.

[21] A. Bisoffi, C. De Persis, and P. Tesi. Data-based stabilization of unknown bilinear systems with guaranteed basin of attraction. *https://arxiv.org/abs/2004.11630*, 2020.

[22] A. Bouhamidi and K. Jbilou. A note on the numerical approximate solutions for generalized Sylvester matrix equations with applications. *Applied Mathematics and Computation*, 206(2):687–694, 2008.

[23] S. J. Bradtke. Reinforcement learning applied to linear quadratic regulation. In *Advances in Neural Information Processing Systems*, pages 295–302, 1993.

[24] D. Burgarth, D. D'Alessandro, L. Hogben, S. Severini, and M. Young. Zero forcing, linear and quantum controllability for systems evolving on networks. *IEEE Transactions on Automatic Control*, 58:2349–2354, 2013.

[25] L. Campestrini, D. Eckhard, A. S. Bazanella, and M. Gevers. Data-driven model reference control design by prediction error identification. *Journal of the Franklin Institute*, 354(6):2628–2647, 2017.

[26] M.C. Campi, A. Lecchini, and S.M. Savaresi. Virtual reference feedback tuning: a direct method for the design of feedback controllers. *Automatica*, 38(8):1337–1346, 2002.

[27] R. Carvalho, L. Buzna, F. Bono, E. Gutiérrez, W. Just, and D. Arrowsmith. Robustness of trans-european gas networks. *Physical Review E*, 80:016106, 2009.

[28] G. Cavraro and V. Kekatos. Graph algorithms for topology identification using power grid probing. *IEEE Control Systems Letters*, 2(4):689–694, Oct 2018.

[29] A. Chapman and M. Mesbahi. On strong structural controllability of networked systems: A constrained matching approach. In *Proceedings of the American Control Conference*, pages 6126–6131, 2013.

[30] A. Chapman, M. Nabi-Abdolyousefi, and M. Mesbahi. Controllability and observability of network-of-networks via cartesian products. *IEEE Transactions on Automatic Control*, 59(10):2668–2679, Oct 2014.

[31] X. Cheng, S. Shi, and P. M. J. Van den Hof. Allocation of excitation signals for generic identifiability of dynamic networks. In *Proceedings of the IEEE Conference on Decision and Control*, pages 5507–5512, Dec 2019.

[32] A. Chiuso and G. Pillonetto. A Bayesian approach to sparse dynamic network identification. *Automatica*, 48(8):1553–1565, 2012.

[33] V. Chvatal. A Greedy Heuristic for the Set-Covering Problem. *Mathematics of Operations Research*, 4(3):233–235, 1979.

[34] A. Clauset, C. Moore, and M. E. J. Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101, 2008.

[35] C. Commault. Structural controllability of networks with dynamical structured nodes. *IEEE Transactions on Automatic Control*, 65(6):2736–2742, 2020.

[36] C. Commault and A. Kibangou. Generic controllability of networks with identical siso dynamical nodes. *IEEE Transactions on Control of Network Systems*, 7(2):855–865, 2020.

[37] C. Commault and J. van der Woude. A classification of nodes for structural controllability. *IEEE Transactions on Automatic Control*, 64(9):3877–3882, 2019.

[38] C. Commault, J. van der Woude, and T. Boukhobza. On the fixed controllable subspace in linear structured systems. *Systems and Control Letters*, 102:42–47, 2017.

[39] C. Commault, J. Van der Woude, and P. Frasca. Functional target controllability of networks: structural properties and efficient algorithms. *to appear in IEEE Transactions on Network Science and Engineering*, 2020.

[40] J. Coulson, J. Lygeros, and F. Dörfler. Data-enabled predictive control: In the shallows of the DeePC. In *Proceedings of the European Control Conference*, pages 307–312, June 2019.

[41] M. Coutino, E. Isufi, T. Maehara, and G. Leus. State-space network topology identification from partial observations. *IEEE Transactions on Signal and Information Processing over Networks*, 6:211–225, 2020.

[42] N. J. Cowan, E. J. Chastain, D. A. Vilhena, J. S. Freudenberg, and C. T. Bergstrom. Nodal dynamics, not degree distributions, determine the structural controllability of complex networks. *PLOS ONE*, 7(6):1–5, 2012.

[43] L. Dai. *Singular Control Systems*, volume 118 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag Berlin Heidelberg, 1989.

[44] T. Dai and M. Sznaier. A moments based approach to designing MIMO data driven controllers for switched systems. In *Proceedings of the IEEE Conference on Decision and Control*, pages 5652–5657, 2018.

[45] J. Dancis. A quantitative formulation of Sylvester's law of inertia. III. *Linear Algebra and its Applications*, 80:141–158, 1986.

[46] A. G. Dankers. System identification in dynamic networks. *PhD thesis, Delft University of Technology*, 2014.

[47] C. De Persis and P. Tesi. Formulas for data-driven control: Stabilization, optimality, and robustness. *IEEE Transactions on Automatic Control*, 65(3):909–924, March 2020.

[48] C. De Persis and P. Tesi. Low-complexity learning of linear quadratic regulators from noisy data. *https://arxiv.org/abs/2005.01082*, 2020.

[49] F. de Roo, A. Tejada, H. van Waarde, and H. L. Trentelman. Towards observer-based fault detection and isolation for branched water distribution networks without cycles. In *Proceedings of the European Control Conference*, pages 3280–3285, 2015.

[50] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu. On the sample complexity of the linear quadratic regulator. *Foundations of Computational Mathematics*, Aug 2019.

[51] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, Dec 1959.

[52] J. Dion, C. Commault, and J. van der Woude. Generic properties and control of linear structured systems: a survey. *Automatica*, 39(7):1125–1144, 2003.

[53] F. Dörfler, J. W. Simpson-Porco, and F. Bullo. Electrical networks and algebraic graph theory: Models, properties, and applications. *Proceedings of the IEEE*, 106(5):977–1005, May 2018.

[54] M. Egerstedt, S. Martini, M. Cao, K. Camlibel, and A. Bicchi. Interacting with networks: How does structure relate to controllability in single-leader, consensus networks? *IEEE Control Systems*, 32(4):66–73, 2012.

[55] W. Favoreel, B. De Moor, and M. Gevers. SPC: Subspace predictive control. *IFAC Proceedings Volumes*, 32(2):4004–4009, 1999.

[56] W. Favoreel, B. De Moor, P. Van Overschee, and M. Gevers. Model-free subspace-based LQG-design. In *Proceedings of the American Control Conference*, pages 3372–3376, June 1999.

[57] M. Fazlyab and V. M. Preciado. Robust topology identification and control of LTI networks. In *Proceedings of the IEEE Global Conference on Signal and Information Processing*, pages 918–922, 2014.

[58] U. Feige. A threshold of ln n for approximating set cover. *Journal of the ACM*, 45(4):634–652, 1998.

[59] M. Fliess and C. Join. Model-free control. *International Journal of Control*, 86(12):2228–2252, 2013.

[60] S. Formentin, A. Karimi, and S. M. Savaresi. Optimal input design for direct data-driven tuning of model-reference controllers. *Automatica*, 49(6):1874–1882, 2013.

[61] P. A. Fuhrmann and U. Helmke. *The Mathematics of Networks of Linear Systems*. Springer, 2015.

[62] K. Furuta and M. Wongsaisuwan. Discrete-time LQG dynamic controller design using plant Markov parameters. *Automatica*, 31(9):1317–1324, 1995.

[63] J. Gao, Y. Y. Liu, R. M. D'Souza, and A. L. Barabási. Target control of complex networks. *Nature Communications*, 5, 2014.

[64] M. Gevers. Identification for control: From the early achievements to the revival of experiment design. *European Journal of Control*, 11(4):335–352, 2005.

[65] M. Gevers, A. S. Bazanella, X. Bombois, and L. Miskovic. Identification and the information matrix: How to get just sufficiently rich? *IEEE Transactions on Automatic Control*, 54(12), Dec 2009.

[66] M. Gevers, A. S. Bazanella, D. F. Coutinho, and S. Dasgupta. Identifiability and excitation of polynomial systems. In *Proceedings of the IEEE Conference on Decision and Control*, pages 4278–4283, Dec 2013.

[67] K. Glover and L. Silverman. Characterization of structural controllability. *IEEE Transactions on Automatic Control*, 21(4):534–537, 1976.

[68] C. D. Godsil. Control by quantum dynamics on graphs. *Physical Review A*, 81(5):052316, 2010.

[69] G. Golub, S. Nash, and C. Van Loan. A Hessenberg-Schur method for the problem AX + XB = C. *IEEE Transactions on Automatic Control*, 24(6):909–913, Dec 1979.

[70] J. Gonçalves and S. Warnick. Necessary and sufficient conditions for dynamical structure reconstruction of LTI networks. *IEEE Transactions on Automatic Control*, 53(7):1670–1674, 2008.

[71] G. R. Gonçalves da Silva, A. S. Bazanella, C. Lorenzi, and L. Campestrini. Data-driven LQR control design. *IEEE Control Systems Letters*, 3(1):180–185, 2019.

[72] S. Gracy, F. Garin, and A. Y. Kibangou. Structural and strongly structural input and state observability of linear network systems. *IEEE Transactions on Control of Network Systems*, 5(4):2062–2072, 2018.

[73] M. Grewal and K. Glover. Identifiability of linear and nonlinear dynamical systems. *IEEE Transactions on Automatic Control*, 21(6):833–837, Dec 1976.

[74] R. Guimerà and M. Sales-Pardo. Missing and spurious interactions and the reconstruction of complex networks. *Proceedings of the National Academy of Sciences*, 106(52):22073–22078, 2009.

[75] M. Guo, C. De Persis, and P. Tesi. Learning control for polynomial systems using sum of squares. *https://arxiv.org/abs/2004.00850*, 2020.

[76] A. Haber and M. Verhaegen. Subspace identification of large-scale interconnected systems. *IEEE Transactions on Automatic Control*, 59(10):2754–2759, Oct 2014.

[77] A. Haber and M. Verhaegen. Sparse solution of the Lyapunov equation for large-scale interconnected systems. *Automatica*, 73(11):256–268, 2016.

[78] C. Hartung, G. Reissig, and F. Svaricek. Characterization of sign controllability for linear systems with real eigenvalues. In *Proceedings of the Australian Control Conference*, pages 450–455, 2013.

[79] S. Hassan-Moghaddam, N. K. Dhingra, and M. R. Jovanović. Topology identification of undirected consensus networks via sparse inverse covariance estimation. In *Proceedings of the IEEE Conference on Decision and Control*, pages 4624–4629, 2016.

[80] D. Hayden, Y. Yuan, and J. Gonçalves. Network identifiability from intrinsic noise. *IEEE Transactions on Automatic Control*, 62(8):3717–3728, Aug 2017.

[81] J. M. Hendrickx, M. Gevers, and A. S. Bazanella. Identifiability of dynamical networks with partial node measurements. *IEEE Transactions on Automatic Control*, 64(6):2240–2253, June 2019.

[82] D. Hershkowitz and H. Schneider. Ranks of zero patterns and sign patterns. *Linear and Multilinear Algebra*, 34(1):3–19, 1993.

[83] L. Hewing, K. P. Wabersich, M. Menner, and M. N. Zeilinger. Learning-based model predictive control: Toward safe learning in control. *Annual Review of Control, Robotics, and Autonomous Systems*, 3(1):269–296, 2020.

[84] H. Hjalmarsson. From experiment design to closed-loop control. *Automatica*, 41(3):393–438, 2005.

[85] H. Hjalmarsson, M. Gevers, S. Gunnarsson, and O. Lequin. Iterative feedback tuning: theory and applications. *IEEE Control Systems Magazine*, 18(4):26–41, Aug 1998.

[86] H. Hjalmarsson, S. Gunnarsson, and M. Gevers. A convergent iterative restricted complexity control design scheme. In *Proceedings of the IEEE Conference on Decision and Control*, pages 1735–1740, Dec 1994.

[87] L. Hogben. Minimum rank problems. *Linear Algebra and its Applications*, 432(8):1961–1974, 2010.

[88] C. M. Holcomb and R. R. Bitmead. Subspace identification with multiple data records: unlocking the archive. *arxiv.org/abs/1704.02635*, 2017.

[89] Z. Hou and Z. Wang. From model-based control to data-driven control: Survey, classification and perspective. *Information Sciences*, 235:3–35, 2013.

[90] L. Huang, J. Coulson, J. Lygeros, and F. Dörfler. Data-enabled predictive control for grid-connected power converters. In *Proceedings of the IEEE Conference on Decision and Control*, pages 8130–8135, Dec 2019.

[91] A. Iannelli and R. S. Smith. A multiobjective LQR synthesis approach to dual control for uncertain plants. *IEEE Control Systems Letters*, 4(4):952–957, 2020.

[92] V. N. Ioannidis, Y. Shen, and G. B. Giannakis. Semi-blind inference of topologies and dynamical processes over dynamic graphs. *IEEE Transactions on Signal Processing*, 67(9):2263–2274, 2019.

[93] J. C. Jarczyk, F. Svaricek, and B. Alt. Strong structural controllability of linear systems revisited. In *Proceedings of the IEEE Conference on Decision and Control and European Control Conference*, pages 1213–1218, 2011.

[94] J. Jia, H. L. Trentelman, W. Baar, and M. K. Camlibel. Strong structural controllability of systems on colored graphs. *To appear in IEEE Transactions on Automatic Control*, 2019.

[95] J. Jia, H. L. Trentelman, and M. K. Camlibel. Fault detection and isolation for linear structured systems. *IEEE Control Systems Letters*, 4(4):874–879, 2020.

[96] J. Jia, H. J. van Waarde, H. L. Trentelman, and M. K. Camlibel. A unifying framework for strong structural controllability. *to appear in IEEE Transactions on Automatic Control*, 2020.

[97] A. Julius, M. Zavlanos, S. Boyd, and G. J. Pappas. Genetic network identification using convex programming. *IET Systems Biology*, 3(3):155–166, 2009.

[98] S. Kar and J. M. F. Moura. Distributed average consensus in sensor networks with random link failures. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 1013–1016, 2007.

[99] L. H. Keel and S. P. Bhattacharyya. Controller synthesis free of analytical models: Three term controllers. *IEEE Transactions on Automatic Control*, 53(6):1353–1369, July 2008.

[100] A. Koch, J. Berberich, and F. Allgöwer. Provably robust verification of dissipativity properties from data. *https://arxiv.org/abs/2006.05974*, 2020.

[101] A. Koch, J. Berberich, and F. Allgöwer. Verifying dissipativity properties from noise-corrupted input-state data. *https://arxiv.org/pdf/2004.07270.pdf*, 2020.

[102] F. Koerts, M. Bürger, A. J. van der Schaft, and C. De Persis. Topological and graph-coloring conditions on the parameter-independent stability of second-order networked systems. *SIAM Journal on Control and Optimization*, 55(6):3750–3778, 2017.

[103] G. Kossinets. Effects of missing data in social networks. *Social Networks*, 28(3):247–268, 2006.

[104] S. G. Krantz and H. R. Parks. *A Primer of Real Analytic Functions*. Birkhäuser Boston, 2002.

[105] T. Kuwae, E. Miyoshi, S. Hosokawa, K. Ichimi, J. Hosoya, T. Amano, T. Moriya, M. Kondoh, R. C. Ydenberg, and R. W. Elner. Variable and complex food web structures revealed by exploring missing trophic links between birds and biofilm. *Ecology Letters*, 15(4):347–356, 2012.

[106] P. Lancaster and H. K. Farahat. Norms on direct sums and tensor products. *Mathematics of Computation*, 26(118):401–414, 1972.

[107] V. Latora and M. Marchiori. Vulnerability and protection of infrastructure networks. *Physical Review E*, 71(1):015103, 2005.

[108] H. J. LeBlanc, H. Zhang, X. Koutsoukos, and S. Sundaram. Resilient asymptotic consensus in robust networks. *IEEE Journal on Selected Areas in Communications*, 31(4):766–781, 2013.

[109] J. Li, X. Chen, S. Pequito, G. J. Pappas, and V. M. Preciado. Structural target controllability of undirected networks. In *Proceedings of the IEEE Conference on Decision and Control*, pages 6656–6661, Dec 2018.

[110] C.T. Lin. Structural controllability. *IEEE Transactions on Automatic Control*, 19(3):201–208, 1974.

[111] D. Liu, P. Yan, and Q. Wei. Data-based analysis of discrete-time linear systems in noisy environment: Controllability and observability. *Information Sciences*, 288:314–329, 2014.

[112] F. Liu and A. S. Morse. Structural controllability of linear systems. In *Proceedings of the IEEE Conference on Decision and Control*, pages 3588–3593, 2017.

[113] Y. Y. Liu, J. J. Slotine, and A. L. Barabasi. Controllability of complex networks. *Nature*, 473(7346):167–173, 2011.

[114] L. Ljung. *System Identification: Theory for the User*. Prentice Hall information and system sciences series. Prentice Hall PTR, 1999.

[115] J. Lofberg. YALMIP: a toolbox for modeling and optimization in MATLAB. In *IEEE International Conference on Robotics and Automation*, pages 284–289, 2004.

[116] Z.-Q. Luo, J. F. Sturm, and S. Zhang. Multivariate nonnegative quadratic mappings. *SIAM Journal on Optimization*, 14(4):1140–1162, 2004.

[117] J.H. Maddocks. Restricted quadratic forms, inertia theorems, and the Schur complement. *Linear Algebra and its Applications*, 108:1–36, 1988.

[118] G. Mao, B. Fidan, and B. D. O. Anderson. Wireless sensor network localization techniques. *Computer Networks*, 51(10):2529–2553, 2007.

[119] I. Mareels and J. W. Polderman. *Adaptive Systems: An Introduction*. Systems & Control: Foundations & Applications. Birkhäuser Boston, 1996.

[120] I. Markovsky. A software package for system identification in the behavioral setting. *Control Engineering Practice*, 21(10):1422–1436, 2013.

[121] I. Markovsky. The most powerful unfalsified model for data with missing values. *Systems & Control Letters*, 95:53–61, 2016.

[122] I. Markovsky. A missing data approach to data-driven filtering and control. *IEEE Transactions on Automatic Control*, 62(4):1972–1978, April 2017.

[123] I. Markovsky and R. Pintelon. Identification of linear time-invariant systems from multiple experiments. *IEEE Transactions on Signal Processing*, 63(13):3549–3554, July 2015.

[124] I. Markovsky and P. Rapisarda. On the linear quadratic data-driven control. In *European Control Conference*, pages 5313–5318, July 2007.

[125] I. Markovsky and P. Rapisarda. Data-driven simulation and control. *International Journal of Control*, 81(12):1946–1959, 2008.

[126] I. Markovsky and K. Usevich. Structured low-rank approximation with missing data. *SIAM Journal on Matrix Analysis and Applications*, 34(2):814–830, 2013.

[127] I. Markovsky, J. C. Willems, P. Rapisarda, and B. L. M. De Moor. Algorithms for deterministic balanced subspace identification. *Automatica*, 41(5):755–766, 2005.

[128] I. Markovsky, J. C. Willems, P. Rapisarda, and B. L. M. De Moor. Data driven simulation with applications to system identification. *IFAC Proceedings Volumes*, 38(1):970–975, 2005.

[129] D. Materassi and G. Innocenti. Topological identification in networks of dynamical systems. *IEEE Transactions on Automatic Control*, 55(8):1860–1871, 2010.

[130] D. Materassi and M. V. Salapaka. On the problem of reconstructing an unknown topology via locality properties of the Wiener filter. *IEEE Transactions on Automatic Control*, 57(7):1765–1777, 2012.

[131] T. M. Maupong, J. C. Mayo-Maldonado, and P. Rapisarda. On Lyapunov functions and data-driven dissipativity. *IFAC-PapersOnLine*, 50(1):7783–7788, 2017.

[132] T. M. Maupong and P. Rapisarda. Data-driven control: A behavioral approach. *Systems & Control Letters*, 101:37–43, 2017.

[133] H. Mayeda and T. Yamada. Strong structural controllability. *SIAM Journal on Control and Optimization*, 17(1):123–138, 1979.

[134] T. Menara, D. S. Bassett, and F. Pasqualetti. Structural controllability of symmetric networks. *IEEE Transactions on Automatic Control*, 64(9):3740–3747, 2019.

[135] T. Menara, G. Bianchin, M. Innocenti, and F. Pasqualetti. On the number of strongly structurally controllable networks. In *Proceedings of the American Control Conference*, pages 340–345, 2017.

[136] T. Menara, V. Katewa, D. S. Bassett, and F. Pasqualetti. The structured controllability radius of symmetric (brain) networks. In *Proceedings of the American Control Conference*, pages 2802–2807, 2018.

[137] Mehran Mesbahi and Magnus Egerstedt. *Graph Theoretic Methods in Multiagent Networks*. Princeton University Press, 2010.

[138] M. Milanese and A. Vicino. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica*, 27(6):997–1009, 1991.

[139] S. Mishra. On the maximum uniquely restricted matching for bipartite graphs. *Electronic Notes in Discrete Mathematics*, 37:345–350, 2011.

[140] N. Monshizadeh. Amidst data-driven model reduction and control. *IEEE Control Systems Letters*, 4(4):833–838, 2020.

[141] N. Monshizadeh, M. K. Camlibel, and H. L. Trentelman. Strong targeted controllability of dynamical networks. In *Proceedings of the IEEE Conference on Decision and Control*, pages 4782–4787, 2015.

[142] N. Monshizadeh, S. Zhang, and M. K. Camlibel. Zero forcing sets and controllability of dynamical systems defined on graphs. *IEEE Transactions on Automatic Control*, 59(9):2562–2567, 2014.

[143] M. Moonen, B. De Moor, L. Vandenberghe, and J. Vandewalle. On- and off-line identification of linear state-space models. *International Journal of Control*, 49(1):219–232, 1989.

[144] S. Moothedath, K. Yashashwi, P. Chaporkar, and M. N. Belur. Target controllability of structured systems. In *Proceedings of the European Control Conference*, pages 3484–3489, 2019.

[145] F. Morbidi and A. Y. Kibangou. A distributed solution to the network reconstruction problem. *Systems & Control Letters*, 70:85–91, 2014.

[146] S. S. Mousavi, M. Haeri, and M. Mesbahi. Robust strong structural controllability of networks with respect to edge additions and deletions. In *Proceedings of the American Control Conference*, pages 5007–5012, 2017.

[147] S. S. Mousavi, M. Haeri, and M. Mesbahi. On the structural and strong structural controllability of undirected networks. *IEEE Transactions on Automatic Control*, 63(7):2234–2241, 2018.

[148] S. S. Mousavi, M. Haeri, and M. Mesbahi. Strong structural controllability of signed networks. In *Proceedings of the IEEE Conference on Decision and Control*, pages 4557–4562, 2019.

[149] S. S. Mousavi, M. Haeri, and M. Mesbahi. Strong structural controllability under network perturbations. *https://arxiv.org/abs/1904.09960*, 2019.

[150] S. Mukherjee, H. Bai, and A. Chakrabortty. On model-free reinforcement learning of reduced-order optimal control for singularly perturbed systems. In *IEEE Conference on Decision and Control*, pages 5288–5293, Dec 2018.

[151] K. Murota. *Systems analysis by graphs and matroids: structural solvability and controllability*. Algorithms and combinatorics. Springer-Verlag, 1987.

[152] S. Nabavi and A. Chakrabortty. A graph-theoretic condition for global identifiability of weighted consensus networks. *IEEE Transactions on Automatic Control*, 61(2):497–502, Feb 2016.

[153] M. Nabi-Abdolyousefi and M. Mesbahi. Network identification via node knock-out. In *Proceedings of the IEEE Conference on Decision and Control*, pages 2239–2244, 2010.

[154] M. Nabi-Abdolyousefi and M. Mesbahi. Network identification via node knockout. *IEEE Transactions on Automatic Control*, 57(12):3214–3219, Dec 2012.

[155] H. Niu, H. Gao, and Z. Wang. A data-driven controllability measure for linear discrete-time systems. In *Proceedings of the IEEE Data Driven Control and Learning Systems Conference*, pages 455–466, May 2017.

[156] B. A. Ogunnaike. A contemporary industrial perspective on process control theory and practice. *Annual Reviews in Control*, 20:1–8, 1996.

[157] D. D. Olesky, M. Tsatsomeros, and P. van den Driessche. Qualitative controllability and uncontrollability by a single entry. *Linear Algebra and its Applications*, 187:183–194, 1993.

[158] R. Olfati-Saber, J. A. Fax, and R. M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233, 2007.

[159] A. Olshevsky. Minimal controllability problems. *IEEE Transactions on Control of Network Systems*, 1(3):249–258, 2014.

[160] J. B. Orlin, K. Madduri, K. Subramani, and M. Williamson. A faster algorithm for the single source shortest path problem with few distinct positive lengths. *Journal of Discrete Algorithms*, 8(2):189–198, June 2010.

[161] S. Oymak and N. Ozay. Non-asymptotic identification of LTI systems from a single trajectory. *https://arxiv.org/abs/1806.05722*, 2018.

[162] B. Pang, T. Bian, and Z. Jiang. Data-driven finite-horizon optimal control for linear time-varying discrete-time systems. In *Proceedings of the IEEE Conference on Decision and Control*, pages 861–866, Dec 2018.

[163] P. E. Paré, V. Chetty, and S. Warnick. On the necessity of full-state measurement for state-space network reconstruction. In *IEEE Global Conference on Signal and Information Processing*, pages 803–806, 2013.

[164] U. S. Park and M. Ikeda. Stability analysis and control design of LTI discrete-time systems by the direct use of time series data. *Automatica*, 45(5):1265–1271, 2009.

[165] F. Pasqualetti, S. Zampieri, and F. Bullo. Controllability metrics, limitations and algorithms for complex networks. In *2014 American Control Conference*, pages 3287–3292, 2014.

[166] M. Penrose. *Random Geometric Graphs*. Oxford studies in probability. Oxford University Press, 2003.

[167] S. Pequito, S. Kar, and A. P. Aguiar. A framework for structural input/output and control configuration selection in large-scale systems. *IEEE Transactions on Automatic Control*, 61(2):303–318, Feb 2016.

[168] S. Pequito, N. Popli, S. Kar, M. D. Ilić, and A. P. Aguiar. A framework for actuator placement in large scale power systems: Minimal strong structural controllability. In *Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, pages 416–419, 2013.

[169] I. Pólik and T. Terlaky. A survey of the S-lemma. *SIAM Review*, 49(3):371–418, 2007.

[170] N. Popli, S. Pequito, S. Kar, A. P. Aguiar, and M. Ilić. Selective strong structural minimum-cost resilient co-design for regular descriptor linear systems. *Automatica*, 102:80–85, 2019.

[171] A. Rahmani, M. Ji, M. Mesbahi, and M. Egerstedt. Controllability of multi-agent systems from a graph-theoretic perspective. *SIAM Journal on Control and Optimization*, 48(1):162–186, 2009.

[172] K. R. Ramaswamy, G. Bottegal, and P. M. J. Van den Hof. Local module identification in dynamic networks using regularized kernel-based methods. In *Proceedings of the IEEE Conference on Decision and Control*, pages 4713–4718, Dec 2018.

[173] A. C. M. Ran and R. Vreugdenhil. Existence and comparison theorems for algebraic Riccati equations for continuous- and discrete-time systems. *Linear Algebra and its Applications*, 99:63–83, 1988.

[174] P. Rapisarda, A. R. F. Everts, and M. K. Camlibel. Fault detection and isolation for systems defined over graphs. In *Proceedings of the IEEE Conference on Decision and Control*, pages 3816–3821, Dec 2015.

[175] K. J. Reinschke, F. Svaricek, and H-D Wend. On strong structural controllability of linear systems. In *Proceedings of the IEEE Conference on Decision and Control*, pages 203–208, 1992.

[176] K. J. Reinschke and G. Wiedemann. Digraph characterization of structural controllability for linear descriptor systems. *Linear Algebra and its Applications*, 266:199–217, 1997.

[177] G. Reissig, C. Hartung, and F. Svaricek. Strong structural controllability and observability of linear time-varying systems. *IEEE Transactions on Automatic Control*, 59(11):3087–3092, 2014.

[178] A. Romer, J. Berberich, J. Köhler, and F. Allgöwer. One-shot verification of dissipativity properties from input-output data. *IEEE Control Systems Letters*, 3(3):709–714, July 2019.

[179] A. Romer, J. M. Montenbruck, and F. Allgöwer. Determining dissipation inequalities from input-output samples. *IFAC-PapersOnLine*, 50(1):7789–7794, 2017.

[180] A. Romer, J. M. Montenbruck, and F. Allgöwer. Sampling strategies for data-driven inference of passivity properties. In *Proceedings of the IEEE Conference on Decision and Control*, pages 6389–6394, 2017.

[181] M. Rotulo, C. De Persis, and P. Tesi. Data-driven linear quadratic regulation via semidefinite programming. *arxiv.org/abs/1911.07767*, 2019.

[182] M. G. Safonov and T. Tsao. The unfalsified control concept and learning. *IEEE Transactions on Automatic Control*, 42(6):843–847, June 1997.

[183] J. R. Salvador, D. Muñoz de la Peña, T. Alamo, and A. Bemporad. Data-based predictive control via direct weight optimization. *IFAC-PapersOnLine*, 51(20):356–361, 2018.

[184] B. M. Sanandaji, T. L. Vincent, and M. B. Wakin. Exact topology identification of large-scale interconnected dynamical systems from compressive observations. In *Proceedings of the American Control Conference*, pages 649–656, 2011.

[185] L. Scardovi and R. Sepulchre. Synchronization in networks of identical linear systems. *Automatica*, 45(11):2557–2562, 2009.

[186] C. W. Scherer. A full block S-procedure with applications. In *Proceedings of the IEEE Conference on Decision and Control*, pages 2602–2607, 1997.

[187] C. W. Scherer. LPV control and full block multipliers. *Automatica*, 37(3):361–375, 2001.

[188] C. W. Scherer and S. Weiland. *Lecture Notes DISC Course on Linear Matrix Inequalities in Control.* 1999.

[189] S. Segarra, M. T. Schaub, and A. Jadbabaie. Network inference from consensus dynamics. In *Proceedings of the IEEE Conference on Decision and Control*, pages 3212–3217, Dec 2017.

[190] S. Shahrampour and V. M. Preciado. Topology identification of directed dynamical networks via power spectral analysis. *IEEE Transactions on Automatic Control*, 60(8):2260–2265, 2015.

[191] B. Shali. Strong structural properties of structured linear systems. *Master's thesis, University of Groningen*, 2019.

[192] Y. Shen, B. Baingana, and G. B. Giannakis. Kernel-based structural equation models for topology identification of directed networks. *IEEE Transactions on Signal Processing*, 65(10):2503–2516, 2017.

[193] G. Shi and R.E. Skelton. Markov data-based LQG control. *ASME Journal of Dynamical Systems, Measurement, and Control*, 122(3):551–559, 1998.

[194] R. Shields and J. Pearson. Structural controllability of multiinput linear systems. *IEEE Transactions on Automatic Control*, 21(2):203–212, 1976.

[195] V. Simoncini. A new iterative method for solving large-scale Lyapunov matrix equations. *SIAM Journal on Scientific Computing*, 29(3):1268–1288, 2007.

[196] V. Simoncini. Computational methods for linear matrix equations. *SIAM Review*, 58(3):377–441, 2016.

[197] R. E. Skelton and G. Shi. The data-based LQG control problem. In *Proceedings of the IEEE Conference on Decision and Control*, pages 1447–1452, Dec 1994.

[198] R.E. Skelton, T. Iwasaki, and D.E. Grigoriadis. *A Unified Algebraic Approach To Control Design*. Taylor & Francis, 1997.

[199] D. C. Soreide, R. K. Bogue, L. J. Ehernberger, and H. R. Bagley. Coherent lidar turbulence for gust load alleviation. In *Optical Instruments for Weather Forecasting*, volume 2832, pages 61–75. International Society for Optics and Photonics, 1996.

[200] T. H. Summers, F. L. Cortesi, and J. Lygeros. On submodularity and controllability in complex dynamical networks. *IEEE Transactions on Control of Network Systems*, 3(1):91–101, 2016.

[201] S. Sundaram and C. N. Hadjicostis. Structural controllability and observability of linear systems over finite fields with applications to multi-agent systems. *IEEE Transactions on Automatic Control*, 58(1):60–73, 2013.

[202] M. Suzuki, N. Takatsuki, J. I. Imura, and K. Aihara. Node knock-out based structure identification in networks of identical multi-dimensional subsystems. In *Proceedings of the European Control Conference*, pages 2280–2285, 2013.

[203] P. Tabuada and L. Fraile. Data-driven stabilization of SISO feedback linearizable systems. *https://arxiv.org/abs/2003.14240*, 2020.

[204] P. Tabuada, W. Ma, J. Grizzle, and A. D. Ames. Data-driven control for feedback linearizable single-input systems. In *Proceedings of the IEEE Conference on Decision and Control*, pages 6265–6270, 2017.

[205] H. G. Tanner. On the controllability of nearest neighbor interconnections. In *Proceedings of the IEEE Conference on Decision and Control*, pages 2467–2472, 2004.

[206] M. Timme and J. Casadiego. Revealing networks from dynamics: an introduction. *Journal of Physics A: Mathematical and Theoretical*, 47(34):343001, Aug 2014.

[207] M. Trefois and J. C. Delvenne. Zero forcing number, constrained matchings and strong structural controllability. *Linear Algebra and its Applications*, 484:199–218, 2015.

[208] H. L. Trentelman, A. A. Stoorvogel, and M. Hautus. *Control Theory for Linear Systems*. Springer Verlag, London, UK, 2001.

[209] S. Tu and B. Recht. The gap between model-based and model-free methods on the linear quadratic regulator: An asymptotic viewpoint. *https://arxiv.org/abs/1812.03565*, 2018.

[210] V. Tzoumas, A. Jadbabaie, and G. J. Pappas. Minimal reachability problems. In *Proceedings of the IEEE Conference on Decision and Control*, pages 4220–4225, 2015.

[211] V. Tzoumas, M. A. Rahimian, G. J. Pappas, and A. Jadbabaie. Minimal actuator placement with bounds on control effort. *IEEE Transactions on Control of Network Systems*, 3(1):67–78, 2016.

[212] J. Umenberger, M. Ferizbegovic, T. B. Schön, and H. Hjalmarsson. Robust exploration in linear quadratic reinforcement learning. In *Advances in Neural Information Processing Systems 32*, pages 15336–15346. Curran Associates, Inc., 2019.

[213] P. A. Valdés-Sosa, J. M. Sánchez-Bornot, A. Lage-Castellanos, M. Vega-Hernández, J. Bosch-Bayard, L. Melie-García, and E. Canales-Rodríguez. Estimating brain functional connectivity with sparse multivariate autoregression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360:969–981, 2005.

[214] P. M. J. Van den Hof, A. Dankers, P. S. C. Heuberger, and X. Bombois. Identification of dynamic models in complex networks with prediction error methods-Basic methods for consistent module estimates. *Automatica*, 49(10):2994–3006, 2013.

[215] A. van der Schaft. *L2-Gain and passivity techniques in nonlinear control*. Springer, 2017.

[216] C. F. Van Loan. The ubiquitous Kronecker product. *Journal of Computational and Applied Mathematics*, 123(1):85–100, 2000.

[217] P. van Overschee and B. de Moor. *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. Kluwer Academic Publishers, 1996.

[218] H. J. Van Waarde, M. K. Camlibel, and M. Mesbahi. From noisy data to feedback controllers: non-conservative design via a matrix S-lemma. *https://arxiv.org/abs/2006.00870*, 2020.

[219] H. J. van Waarde, M. K. Camlibel, and H. L. Trentelman. A distance-based approach to strong target control of dynamical networks. *IEEE Transactions on Automatic Control*, 62(12):6266–6277, 2017.

[220] H. J. van Waarde, C. De Persis, M. K. Camlibel, and P. Tesi. Willems' fundamental lemma for state-space systems and its extension to multiple datasets. *IEEE Control Systems Letters*, 4(3):602–607, 2020.

[221] H. J. Van Waarde, J. Eising, H. L. Trentelman, and M. K. Camlibel. Data informativity: a new perspective on data-driven analysis and control. *to appear in IEEE Transactions on Automatic Control*, 2020.

[222] H. J. van Waarde and M. Mesbahi. Data-driven parameterizations of suboptimal LQR and H2 controllers. In *to appear in Proceedings of the IFAC World Congress*, 2020.

[223] H. J. van Waarde, P. Tesi, and M. K. Camlibel. Identifiability of undirected dynamical networks: A graph-theoretic approach. *IEEE Control Systems Letters*, 2(4):683–688, Oct 2018.

[224] H. J. van Waarde, P. Tesi, and M. K. Camlibel. Topological conditions for identifiability of dynamical networks with partial node measurements. *IFAC-PapersOnLine*, 51(23):319–324, 2018.

[225] H. J. van Waarde, P. Tesi, and M. K. Camlibel. Topology identification of heterogeneous networks of linear systems. In *Proceedings of the IEEE Conference on Decision and Control*, pages 5513–5518, Dec 2019.

[226] H. J. van Waarde, P. Tesi, and M. K. Camlibel. Topology reconstruction of dynamical networks via constrained Lyapunov equations. *IEEE Transactions on Automatic Control*, 64(10):4300–4306, 2019.

[227] M. Verhaegen and P. Dewilde. Subspace model identification part 1. the output-error state-space model identification class of algorithms. *International Journal of Control*, 56(5):1187–1210, 1992.

[228] M. Verhaegen and V. Verdult. *Filtering and system identification: a least squares approach*. Cambridge university press, 2007.

[229] H. Wai, A. Scaglione, B. Barzel, and A. Leshem. Joint network topology and dynamics recovery from perturbed stationary points. *IEEE Transactions on Signal Processing*, 67(17):4582–4596, 2019.

[230] G. C. Walsh and Hong Ye. Scheduling of networked control systems. *IEEE Control Systems Magazine*, 21(1):57–65, Feb 2001.

[231] W.-X Wang, Y.-C Lai, C. Grebogi, and J. Ye. Network reconstruction based on evolutionary-game data via compressive sensing. *Physical Review X*, 1:021021, Dec 2011.

[232] Z. Wang and D. Liu. Data-based controllability and observability analysis of linear discrete-time systems. *IEEE Transactions on Neural Networks*, 22(12):2388–2392, Dec 2011.

[233] H. Weerts, P. M. J. Van den Hof, and A. Dankers. Single module identifiability in linear dynamic networks. In *Proceedings of the IEEE Conference on Decision and Control*, pages 4725–4730, 2018.

[234] H. H. M. Weerts, P. M. J. Van den Hof, and A. G. Dankers. Identifiability of dynamic networks with part of the nodes noise-free. *IFAC-PapersOnLine*, 49(13):19–24, 2016.

[235] H. H. M. Weerts, P. M. J. van den Hof, and A. G. Dankers. Identifiability of linear dynamic networks. *Automatica*, 89:247–258, 2018.

[236] H. H. M. Weerts, P. M. J. Van den Hof, and A. G. Dankers. Prediction error identification of linear dynamic networks with rank-reduced noise. *Automatica*, 98:256–268, 2018.

[237] R. West, A. Paranjape, and J. Leskovec. Mining missing hyperlinks from human navigation traces: A case study of wikipedia. In *Proceedings of the International Conference on World Wide Web*, pages 1242–1252, 2015.

[238] P. Wieland, R. Sepulchre, and F. Allgöwer. An internal model principle is necessary and sufficient for linear output synchronization. *Automatica*, 47(5):1068–1074, 2011.

[239] E. P. Wigner. The unreasonable effectiveness of mathematics in the natural sciences. *Communications on Pure and Applied Mathematics*, 13(1):1–14, 1960.

[240] J. C. Willems. Dissipative dynamical systems part I: General theory. *Archive for Rational Mechanics and Analysis*, 45:321–351, 1972.

[241] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. M. De Moor. A note on persistency of excitation. *Systems & Control Letters*, 54(4):325–329, 2005.

[242] X. Wu, X. Zhao, J. Lü, L. Tang, and J. A. Lu. Identifying topologies of complex dynamical networks with stochastic perturbations. *IEEE Transactions on Control of Network Systems*, 3(4):379–389, 2016.

[243] V. A. Yakubovich. S-procedure in nonlinear control theory. *Vestnik Leningrad University Mathematics*, 4:73–93, 1977.

[244] T. Yamada and D. Luenberger. Generic controllability theorems for descriptor systems. *IEEE Transactions on Automatic Control*, 30(2):144–152, 1985.

[245] T. Yang, A. Saberi, A. A. Stoorvogel, and H. F. Grip. Output synchronization for heterogeneous networks of introspective right-invertible agents. *International Journal of Robust and Nonlinear Control*, 24(13):1821–1844, 2014.

[246] Y. Yuan, G. Stan, S. Warnick, and J. Gonçalves. Robust dynamical network structure reconstruction. *Automatica*, 47(6):1230–1235, 2011.

[247] S. Zhang, M. Cao, and M. K. Camlibel. Upper and lower bounds for controllable subspaces of networks of diffusively coupled agents. *IEEE Transactions on Automatic Control*, 59(3):745–750, 2014.

[248] B. Zhou, Z. Wang, Y. Zhai, and H. Yuan. Data-driven analysis methods for controllability and observability of a class of discrete LTI systems with delays. In *Proceedings of the IEEE Data Driven Control and Learning Systems Conference*, pages 380–384, May 2018.

[249] J. Zhou and J. Lu. Topology identification of weighted complex dynamical networks. *Physica A: Statistical Mechanics and its Applications*, 386(1):481–491, 2007.

[250] Y. Zi-zong and G. Jin-hai. Some equivalent results with Yakubovich's S-Lemma. *SIAM Journal on Control and Optimization*, 48(7):4474–4480, 2010.

[251] J. G. Ziegler and N. B. Nichols. Optimum settings for automatic controllers. *Transactions of the ASME*, 64:759–768, 1942.

SUMMARY

This thesis focuses on modeling, analysis and control of dynamical systems using data and structure. In particular, we consider four different problems, namely data-driven control, topology identification, network identifiability, and structural controllability.

First, we revisit a result by Willems and coauthors called the fundamental lemma. This result enables a parameterization of all trajectories of a linear system in terms of a single (measured) trajectory. The result has important consequences for system identification and data-driven control. Our contribution is to extend the fundamental lemma to the situation where multiple -possibly short- trajectories are measured instead of a single one. This result is shown to be beneficial, for example, in the identification of linear systems from data sets with missing samples.

Next, we turn our attention to data-driven analysis and control. Here, the goal is to verify system-theoretic properties and to construct controllers of an unknown dynamical system directly using data. We study these problems from the perspective of "data informativity". This means that we are interested in characterizing the conditions on the data that enable data-driven analysis and control. We provide a general definition for data informativity. Thereafter, we study several data-driven analysis problems such as stability, stabilizability, controllability and dissipativity. In addition, we study control problems such as stabilization, $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control. We characterize the informativity of data for each of these analysis and control problems. For the control problems, we also provide direct control design methods. Both the situations in which data are exact and noisy are considered. In the setting of noisy data, our investigation leads to a novel dualization lemma and S-lemma for matrix variables. These results are of independent interest, and may also find other applications in the general area of robust control.

Subsequently, we investigate the problem of topology identification of networked systems. A networked system (or network) is a collection of interconnected dynamical systems. Networks are often represented by a graph, where nodes correspond to dynamical systems and edges represent coupling between systems. In many real-world networks, the graph structure underlying the network is unknown. This is, for example, the case in neural networks and networks of interconnected stock prices. Thus, a relevant problem is that of *topology identification*, which entails the identification of the network graph on the basis of measurements taken from (a subset of) the node systems. We study topology identification for general heterogeneous networks of linear systems, where the node systems are general (possibly distinct) multi-input multi-output systems. We treat both the identifiability and the reconstruction aspect of topology identification.

Identifiability is concerned with the question whether there *exists* a data set with which the network topology can be uniquely determined. Reconstruction, in turn, involves the development of algorithms that identify the network topology on the basis of measured data. We provide conditions for topological identifiability, recovering existing conditions as special cases. Our reconstruction methodology combines Willems' fundamental lemma with the solution to certain generalized Sylvester equations in order to identify the network topology. We also consider topology identification in the more specialized setup of autonomous networks of single integrators. In this case we show that the topology can be reconstructed by solving a certain Lyapunov equation.

In addition to studying topology identification, we study identifiability of networks with *known* graph structure. The assumption of known graph structure is justifiable in engineering systems such as water distribution networks. We first focus on a class of networked systems where the interactions between nodes are described by transfer functions, and each node is influenced by an independent excitation signal but only a subset of node signals is measured. Identifiability then involves the question whether the transfer functions in the network can be uniquely identified using a priori knowledge of the graph structure and measurements of the excitation and node signals. We treat this problem from a purely graph-theoretic perspective. To this end, we define a notion of so-called *global* identifiability. We then introduce a new concept called the graph simplification process. Using this concept, it is possible to provide necessary and sufficient graph-theoretic conditions for global identifiability. We also study a similar notion of global identifiability for a different class of undirected networks described by state-space systems. In this case it turns out to be possible to secure identifiability by a limited number of measured *and* excited nodes.

Finally, we study strong structural controllability of linear systems. We study this problem in a general setup where each entry of the system matrices is not known exactly, but is known to be either zero, nonzero or arbitrary. The problem of strong structural controllability then involves verifying whether all systems with this zero/nonzero/arbitrary structure are controllable in the classical sense. We provide algebraic necessary and sufficient conditions for strong structural controllability in terms of full rank properties of so-called pattern matrices. In addition, we provide graph-theoretic conditions in terms of a color change rule. Besides controllability, we also study strong structural input-state observability and output controllability. For both of these problems, we come up with algebraic conditions in terms of full rank properties of sums and products of pattern matrices. We also provide more specialized results for strong structural output controllability in a network setting. In this setting, we provide both a sufficient and a necessary condition for strong structural output controllability, as well as a method for selecting inputs that guarantee ouput controllability.

SAMENVATTING

Dit proefschrift gaat over het modelleren, analyseren en regelen van dynamische systemen op basis van data en systeemstructuur. In het bijzonder beschouwen we vier verschillende problemen, namelijk datagestuurd regelaar-ontwerp, graafidentificatie, identificeerbaarheid van netwerken, en structurele regelbaarheid.

Ten eerste herzien we een resultaat van Willems en coauteurs dat het fundamentele lemma heet. Dit resultaat maakt het mogelijk om alle trajektorieën van een lineair systeem te parametriseren in termen van een enkele (gemeten) trajektorie. Dit resultaat heeft belangrijke consequenties voor systeemidentificatie en datagestuurd ontwerp van regelaars. Onze bijdrage is om het fundamentele lemma uit te breiden naar de situatie waarin meerdere (mogelijk korte) trajektorieën gemeten worden in plaats van één enkele. Dit resultaat is nuttig, bijvoorbeeld in de identificatie van lineaire systemen op basis van datasets met missende datapunten.

Vervolgens richten we onze aandacht op datagestuurde analyse en regelaar-ontwerp. Het doel is om van een onbekend dynamisch systeem, systeemtheoretische eigenschappen af te leiden en regelaars te construeren op basis van data. We bestuderen deze problemen vanuit het perspectief van “data informativiteit". Dit betekent dat we geïnteresseerd zijn in het karakteriseren van de data-eigenschappen die datagestuurde analyse en regelaar-ontwerp mogelijk maken. We geven een algemene definitie van de informativiteit van data. Vervolgens bestuderen we verschillende datagestuurde analyseproblemen zoals stabiliteit, stabiliseerbaarheid, regelbaarheid en dissipativiteit. Daarnaast bestuderen we regelproblemen zoals stabilisatie, $\mathcal{H}_2$ en $\mathcal{H}_\infty$ regeling. We karakteriseren de informativiteit van data voor elk van deze analyse- en regelproblemen. In het geval van de regelproblemen geven we ook ontwerpmethodes om regelaars direct uit data te vinden. We beschouwen zowel de situatie waarin de data exact zijn, als de situatie waarin de data beïnvloed zijn door ruis. In de tweede situatie leidt ons onderzoek tot een nieuw dualisatielemma en een S-lemma voor matrixvariabelen. Deze resultaten vinden wellicht ook nog andere toepassingen in de algemene richting van robuuste regeltechniek.

Daaropvolgend onderzoeken we het probleem van graafidentificatie van netwerksystemen. Een netwerksysteem (of netwerk) is een collectie van gekoppelde dynamische systemen. Netwerken worden vaak gerepresenteerd door een graaf, waarin knopen corresponderen met dynamische systemen en zijden koppeling tussen systemen weergeven. In veel echte netwerken is de onderliggende graafstructuur van het netwerk onbekend. Dit is bijvoorbeeld het geval in neurale netwerken en in netwerken van gekoppelde aandeelprijzen. Zodoende is *graafidentificatie* een relevant probleem. Het doel hiervan is om de netwerkgraaf te identificeren op basis van metingen van (een deel) van de gekoppelde syste-

men. We bestuderen graafidentificatie voor algemene heterogene netwerken van lineaire systemen, waarbij de individuele systemen mogelijk verschillend zijn en meerdere ingangen en uitgangen hebben. We behandelen zowel het identificeerbaarheids- als het reconstructie-aspect van graafidentificatie. Identificeerbaarheid is begaan met de vraag of er een dataset *bestaat* waarmee we de netwerkgraaf uniek kunnen bepalen. Reconstructie, op zijn beurt, beslaat de ontwikkeling van algoritmes die de netwerkgraaf kunnen identificeren op basis van data. We geven voorwaarden voor graafidentificeerbaarheid, waarbij we reeds bestaande voorwaarden als speciale gevallen terugvinden. Onze reconstructiemethode combineert Willems' fundamentele lemma met de oplossingen van zekere veralgemeniseerde Sylvestervergelijkingen en identificeert op deze wijze de netwerkgraaf. We beschouwen ook graafidentificatie in het speciale geval van autonome netwerken die beschreven worden door enkele integratordynamica. In dit geval laten we zien dat de graaf kan worden gereconstrueerd door een zekere Lyapunov-vergelijking op te lossen.

Naast het bestuderen van graafidentificatie, bestuderen we ook de identificeerbaarheid van netwerken met *bekende* graafstructuur. De aanname dat de graaf bekend is, is bijvoorbeeld geoorloofd in waterdistributienetwerken. We focussen eerst op een klasse van netwerksystemen waarin de interacties tussen knopen worden beschreven door overdrachtsfuncties en elke knoop wordt beïnvloed door een onafhankelijk ingangssignaal, maar slechts een deel van de uitgangssignalen van de knopen wordt gemeten. Identificeerbaarheid houdt dan in dat de overdrachtsfuncties in het netwerk uniek geïdentificeerd kunnen worden, gebruikmakend van de kennis van de graafstructuur en metingen van de ingangs- en uitgangssignalen van het netwerk. We behandelen dit probleem vanuit een puur graaftheoretisch oogpunt. Daarom definiëren we een zogenaamde notie van *globale* identificeerbaarheid. Vervolgens introduceren we een nieuw concept dat het graafvereenvoudigingsproces heet. Door middel van dit concept is het mogelijk om nodige en voldoende graaftheoretische voorwaarden te geven voor globale identificeerbaarheid. We bestuderen ook een vergelijkbare notie van globale identificeerbaarheid voor een klasse van ongerichte netwerken beschreven door toestandsruimtemodellen. In dit geval blijkt het mogelijk te zijn om identificeerbaarheid te garanderen met een beperkt aantal ingangs- en uitgangsknopen.

Tenslotte bestuderen we sterke structurele regelbaarheid van lineaire systemen. We bestuderen een algemene versie van dit probleem, waarin de exacte waarde van elk element van de systeemmatrices onbekend is, maar waar van ieder element wel bekend is of het nul, niet-nul of arbitrair is. Sterke structurele regelbaarheid houdt dan in dat alle systemen met deze nul/niet-nul/arbitrair structuur regelbaar zijn in de klassieke zin. We geven algebraïsche nodige en voldoende voorwaarden voor sterke structurele regelbaarheid in termen van volle rangeigenschappen van symbolische matrices. Daarnaast geven we graaftheoretische voorwaarden in termen van een kleurregel. Naast regelbaarheid bestuderen we ook sterke structurele ingang-toestand waarneembaarheid en uitgangsregelbaarheid. Voor beide problemen bedenken we algebraïsche voorwaarden in termen van volle rangeigenschappen van sommen en producten van symbolische matrices. We geven ook meer gespecialiseerde resultaten voor sterke structurele uitgangsregelbaarheid van netwerksystemen. In dit geval geven we zowel een voldoende als een nodige voorwaarde voor sterke structurele uitgangsregelbaarheid, en een methode om ingangen te selecteren die uitgangsregelbaarheid garanderen.