

©Copyright 2023

Shahriar Talebi

Constrained Policy Synthesis:  
Riemannian Flows, Online Regulation, and Distributed Games

Shahriar Talebi

A dissertation  
submitted in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Professor Mehran Mesbahi, Chair

Professor John M. Lee

Professor Maryam Fazel

Professor Lillian J. Ratliff

Professor Amirhossein Taghvaei

Program Authorized to Offer Degree:

Aeronautics & Astronautics

University of Washington

**Abstract**

Constrained Policy Synthesis:  
Riemannian Flows, Online Regulation, and Distributed Games

Shahriar Talebi

Chair of the Supervisory Committee:  
Professor Mehran Mesbahi  
William E. Boeing Department of Aeronautics & Astronautics

This dissertation makes contributions to decision-making processes in both cooperative and non-cooperative environments, spanning several domains from constrained and large-scale dynamical systems to network games and learning for control and estimation problems. First, we examine linearly constrained policy optimization over stabilizing controllers, utilizing a Riemannian metric inherent to optimal control problems. We propose a novel Newton-type algorithm that leverages the manifold’s second-order geometry to ensure local convergence, demonstrating promising results in Structured and Output Linear Quadratic Regulators (LQR) problems. Additionally, we present a distributed model-free policy iteration tailored for large networks of homogeneous systems. This algorithm enables the development of stabilizing distributed feedback controllers through a data-driven approach and the use of a learned stability margin. Addressing online regulation of partially unknown unstable linear systems, we introduce the Data-Guided Regulation (DGR) synthesis procedure, revealing novel geometric and system-theoretic properties while effectively regulating the system’s states. Furthermore, we explore distributed learning in network games using dual averaging, achieving sublinear regret bounds by optimizing global objectives composed of local objective functions and considering network structures. Lastly, we investigate optimal filtering policies for linear systems with unknown noise covariance matrices using noisy

output data, minimizing prediction error through stochastic policy optimization and ensuring theoretical guarantees for biased gradients and stability constraints.

# TABLE OF CONTENTS

	Page
List of Figures . . . . .	iii
List of Tables . . . . .	v
Chapter 1: Summary of Research . . . . .	1
Chapter 2: Policy Optimization over Submanifolds of Stabilizing Controllers . . . . .	3
2.1 Introduction . . . . .	3
2.2 Problem Setup . . . . .	10
2.3 Optimization on Submanifolds of $\mathcal{S}$ and its Geometry . . . . .	11
2.4 Algorithm: Riemannian Newton-type Policy Optimization (RNPO) . . . . .	19
2.5 Analysis of RNPO . . . . .	23
2.6 Applications of RNPO to Distributed and Output LQR Problems . . . . .	32
2.7 Numerical Simulations . . . . .	42
2.8 Remarks and Future Directions . . . . .	51
Chapter 3: Distributed Data-Driven Structured Policy Iteration . . . . .	53
3.1 Introduction . . . . .	53
3.2 Problem Setup . . . . .	56
3.3 Algorithm: Distributed Data-driven Structured Policy Iteration (D3SPI) . . . . .	62
3.4 Analysis of D3SPI . . . . .	66
3.5 Simulation Results . . . . .	85
3.6 Remarks and Future Directions . . . . .	89
Chapter 4: On Regularizability and Online Control of Unstable LTI Systems . . . . .	90
4.1 Introduction . . . . .	91
4.2 Mathematical Preliminaries . . . . .	94

4.3	Problem Setup . . . . .	96
4.4	Regularizable Systems . . . . .	99
4.5	Algorithm: Data-Guided Regulation (DGR) . . . . .	104
4.6	Analysis of DGR . . . . .	107
4.7	The Special Case of $\mathcal{R}(A) \subset \mathcal{R}(B)$ with $\alpha = 0$ . . . . .	118
4.8	Numerical Simulations . . . . .	125
4.9	Remarks and Future Directions . . . . .	131
Chapter 5: Distributed Learning in Network Games: A Dual Averaging Approach		133
5.1	Introduction . . . . .	133
5.2	Mathematical Preliminaries . . . . .	136
5.3	Background and Problem Formulation . . . . .	137
5.4	Algorithm: Team-based Dual Averaging (TDA) . . . . .	139
5.5	Analysis of TDA . . . . .	140
5.6	Numerical Simulations . . . . .	151
5.7	Remarks and Future Directions . . . . .	152
Chapter 6: Optimal Filtering for Linear Systems with Unknown Noise Covariances		157
6.1	Introduction . . . . .	157
6.2	Background and Problem Formulation . . . . .	160
6.3	Estimation-Control Duality Relationship . . . . .	163
6.4	Algorithm: SGD for Learning the Kalman Gain . . . . .	167
6.5	Analysis and Extended Results . . . . .	175
6.6	Numerical Simulations . . . . .	208
6.7	Remarks and Future Directions . . . . .	209
Bibliography . . . . .		212

## LIST OF FIGURES

Figure Number	Page	
2.1	The actual submanifold $\tilde{\mathcal{S}}$ of <i>diagonal</i> state-feedback stabilizing controllers for a system with 2 states and 2 inputs; superimposed with 1) the level-sets of the constrained LQR cost ( $h$ ), and the regions on which its “ <i>Hessian</i> ” is positive definite with respect to 2) the inherent Riemannian geometry ( $\text{Hess}h$ ) and 3) the Euclidean geometry ( $\overline{\text{Hess}h}$ ), respectively; $\tilde{\mathcal{S}}$ is covered by the region on which $\text{Hess}h \succ 0$ . See Example 2.18 in §2.7 for more details. . . . .	6
2.2	The normalized error of iterates and cost values at each iteration for the Structured LQR (SLQR) problem in Example 2.16. . . . .	44
2.3	The normalized error of iterates and cost values at each iteration for the Output-feedback LQR (OLQR) problem in Example 2.16. . . . .	45
2.4	The trajectories of iterates $K = \begin{pmatrix} l_1 & 0 \\ 0 & l_2 \end{pmatrix}$ generated by RNPO (with $\text{Hess}h$ and $\overline{\text{Hess}h}$ ) and Projected Gradient (PG)—from different initial points—for the SLQR problem with sparsity constraint $D^c = \{(1, 2), (2, 1)\}$ . . . . .	47
2.5	The trajectories of iterates $L = \begin{pmatrix} l_1 & l_2 \end{pmatrix}^\top$ generated by RNPO (with $\text{Hess}h$ and $\overline{\text{Hess}h}$ ) and PG—from different initial points—for the OLQR problem with output matrix $C = \begin{pmatrix} 1.0 & 1.0 \end{pmatrix}$ . . . . .	48
2.6	The min, max and median progress of normalized error of iterates and cost values at each iteration, for the SLQR problem with 100 different randomly sampled system parameters and sparsity patterns $D$ . . . . .	49
2.7	The min, max and median of normalized error of iterates and cost values at each iteration of Algorithm 2.1, for the OLQR problem with 100 different randomly sampled system parameters (including output matrix $C$ ). . . . .	50
3.1	Addition of auxiliary links (dashed red) to the subgraph $\mathcal{G}_d$ during the policy learning phase. The size of the subgraph depends on the maximum degree of the original graph $\mathcal{G}$ . . . . .	55
3.2	Convergence of Algorithm 3.1 for 100 randomly sampled system parameters $(A, B)$ . . . . .	87
3.3	Suboptimality of the distributed controller learned by Algorithm 3.1 for a network of $N$ homogeneous plants with path-graphs structures in (a) and (b), and with random 3-regular graphs structures in (c) and (d). . . . .	88

4.1	A unit cube in the domain of $A$ that is mapped to a parallelepiped in its range space. . . . .	110
4.2	Illustration of the upper and lower bounds for instability number of a system with $\sigma_1 = 3$ and $\delta = 0.1$ as in Lemma 4.16. . . . .	112
4.3	A geometric schematic of DGR when $\mathcal{R}(A) \subseteq \mathcal{R}(B)$ . Since $\mathbf{z}_0 := \mathbf{x}_0$ , $\mathbf{z}_t \perp \mathcal{R}(\mathcal{X}_{t-1})$ and $\mathbf{z}_t \in \mathcal{R}(\mathcal{X}_t)$ for $t = 1, 2$ , the set $\{\mathbf{z}_0, \mathbf{z}_1, \mathbf{z}_2\}$ consists of orthogonal vectors. . . . .	119
4.4	Grumman X-29A ( <i>Credits: NASA Photo</i> ), mainly known for its extreme instability while providing high-quality maneuverability; the longitudinal and lateral-directional states are illustrated. . . . .	126
4.5	The state trajectory of X-29 in ND-PA mode with and without DGR for longitudinal control. . . . .	129
4.6	The state trajectory of X-29 in ND-PA mode with and without DGR for lateral-directional control. . . . .	129
4.7	The state trajectory of X-29 in ND-UA mode with and without DGR for longitudinal control. . . . .	130
4.8	The state trajectory of X-29 in ND-UA mode with and without DGR for lateral-directional control. . . . .	130
4.9	The convergence behavior of DGR algorithm for longitudinal and lateral-directional dynamics for both of flight modes. . . . .	131
5.1	Schematic of our game setup including two teams (red A and blue B) with each team having a representative at each node. . . . .	139
5.2	Normalized Average Error (NAE) at each iteration for both TDA and Team-based Mirror-Descent (TMD) algorithms in complete, random 6-regular, and cycle graphs with 50 nodes. . . . .	154
5.3	The random 6-regular networks with 50 nodes simulated in our setup. . . . .	154
5.4	Number of iterations needed for each network so that NAE is less than 0.1. . . . .	155
6.1	Stochastic Gradient Descent (SGD) directly from output data and without prior knowledge of the noise covariances or state information. Mean progress of the normalized estimation error over 20 simulations obtained from data trajectories of a) different batch size $M$ and b) different length $T$ . . . . .	210



## LIST OF TABLES

Table Number	Page
6.1 Differences between SGD algorithms for optimal LQR and optimal estimation problems . . . . .	166

## ACKNOWLEDGMENTS

I would like to express my profound gratitude to the following individuals who have played a significant role in shaping my academic and personal journey:

First and foremost, I am deeply thankful to my Ph.D. advisor, Mehran Mesbahi, for his unwavering encouragement and invaluable guidance throughout this exciting path. I am truly grateful for his enduring support, which has surpassed all my expectations. Prof. Mesbahi's mentorship has not only enabled me to achieve these remarkable milestones but has also imparted numerous technical and professional lessons. From optimization and optimal and robust control theory to learning in control, his expertise has enriched my knowledge and shaped my research direction. Moreover, I would like to acknowledge the immense impact Prof. Mesbahi has had on my personal and professional growth, extending far beyond the confines of academia.

I would also like to extend my sincere appreciation to John M. Lee for his insightful technical advice and inspiring lectures on differential geometry. The knowledge and understanding I gained from these lectures have profoundly influenced my learning experience at UW and will continue to inspire me in the years to come. I am immensely grateful for the deep understanding of mathematical concepts such as topological and Riemannian manifolds, bundles, and complex manifolds that I acquired under the guidance of the figure in the field.

My heartfelt thanks to my M.Sc. advisor, Marwan Simaan, for his unwavering support and encouragement from the early years of my graduate studies until now. Having him by my side has been an invaluable treasure, and it is through his super-

vision that I have been able to lay the foundation for future accomplishments. I am grateful for the opportunity to learn control theory, signal processing, and the captivating field of game theory from such a prominent figure in the field. Prof. Simaan’s impact extends beyond academia, as he has been a pillar of support, providing invaluable insights and encouragement, shaping both my academic achievements and personal growth. I feel truly fortunate to have been mentored and worked with such an exceptional figure, and his influence will resonate with me for years to come.

I am indebted to James V. Burke, Maryam Fazel, and Lillian J. Ratliff for their valuable contributions as committee members. Their insightful discussions on various aspects of my research, including optimization, variational inequality, learning, and games, have greatly improved the outcomes of my dissertation. I would like to extend a special thank you to Lillian for the amazing collaboration that made a significant contribution to Chapter 5. I would also like to express my gratitude to Amirhossein Taghvaei for his role as a committee member and passionate collaborator, whose insights and contributions have helped extend our results to optimal filtering, which forms the backbone of Chapter 6.

My heartfelt thanks go to Kenneth P. Bube, Soumik Pal, and Stefan Steinerberger for their insightful lectures and support in my mathematical education at the UW Department of Mathematics. Ken’s remarkably organized lectures and rigorous notes on real analysis have been instrumental in deepening my understanding of the subject. Soumik’s lectures and follow-up discussions have been an invaluable resource that helped me build a strong foundation in probability and gain a profound understanding of Optimal Transport. I am deeply grateful to Stefan for his joyful lectures and research discussions, which have introduced me to the beauty of “hard analysis” and reciprocated my love for mathematics. I consider myself fortunate to have had the opportunity to interact with each of them.

I would like to acknowledge and express my gratitude to my collaborators for the amazing opportunities and fruitful teamwork we have had: Behçet Açıkmese, Zhihua Qu, Yue Yu, Siavash Alemzadeh, Niyousha Rahimi, Spencer Kraisler, Aditya Deole, Jonathan Becktor, Hesam Talebiyan, Henk van Waarde, and Kuang-Ying (Eddie) Ting. In particular, I extend special thanks to Amir Rahmani and Saptarshi Bandyopadhyay for their exciting research discussions and fruitful collaborations. I would also like to thank the current and former RAIN members: Mengyuan Wang, Josh Holder, Beniamino Pozzan, Jingjing Bu, Mathias Hudoba de Badyn, Dillon Foight, Taylor Reynolds, and Bijan Barzgaran. The time spent with all of you discussing various research questions has created a wonderful research environment, and I am immensely grateful for your contributions. Working with all of you has been an absolute pleasure.

Above all, I want to express my deepest gratitude to my soulmate, Mahlagha (Meli) Sedghi, for her unwavering mental and emotional support, which has been a cornerstone of my personal growth and well-being. Additionally, her engaging scientific discussions have not only ignited my passion for learning but also strengthened our bond as a couple. My beloved Meli! Your compassion and generosity have been a constant source of inspiration. I am eternally grateful for your selfless devotion, including your decision to continue the last two years of your Ph.D. studies remotely so that I could pursue mine in person in Seattle. You are my rock, and I thank you from the bottom of my heart for being there for me every day.

I would also like to extend my thanks to my family and friends for their unwavering support and love. My parents, Ali and Shayesteh, have always believed in me, even when I doubted myself. To my dad, thank you for designing all those Math exams when I felt unchallenged in school. To my mom, thank you for your efforts to improve my understanding of Persian literature despite my playfulness and youthful ignorance.

There are truly not enough words to express my profound gratitude for everything you have done for me. Your unwavering support, love, and understanding have been the bedrock of my life. I am endlessly thankful, and I promise to keep trying to show you just how much you mean to me. Finally, I am grateful to all my friends for the joyful times we spent together, playing, and cooking amazing food. Your love and support mean the world to me, and I am truly thankful for each and every one of you.

## **DEDICATION**

To the women of Iran

## Chapter 1

### SUMMARY OF RESEARCH

Herein, we provide a summary of contributions presented in this dissertation. In Chapter 2, the focus is on linearly constrained policy optimization over the manifold of Schur stabilizing controllers, where a Riemannian metric naturally emerging in optimal control problems is employed. The second-order geometry of the manifold is studied, leading to a Newton-type algorithm with local convergence guarantees. This algorithm leverages the inherent geometry, even without the exponential mapping or a numerically tractable retraction. Two well-known constrained optimal control problems are then addressed, and numerical examples demonstrate the performance of the proposed approach.

Chapter 3 addresses the challenge of controlling large-scale networked systems. A distributed model-free policy iteration algorithm is proposed for designing feedback mechanisms in large networks of homogeneous systems. The algorithm utilizes an underlying information-exchange graph, allowing the distributed controller to synthesize a feedback signal based on data obtained from a small subgraph. The approach includes a learning phase and a stability margin is learned from data. The methodology is evaluated through distributed control scenarios with modeling errors.

Chapter 4 focuses on online regulation of partially unknown linear systems without prior access to an initial stabilizing controller or persistently exciting (PE) input-output data. The notion of “regularizability” is introduced and characterized, along with the Data-Guided Regulation (DGR) synthesis procedure. Leveraging only the knowledge of the input matrix, DGR regulates the system while generating informative data for data-driven stabilization or system identification. The analysis is related to the spectrum and the “instability number”

of the underlying linear system.

In Chapter 5, a distributed no-regret learning algorithm for network games is proposed using the primal-dual method of dual averaging. The algorithm considers a scenario where each player optimizes a global objective formed by local objective functions on a communication graph. Regret analysis is provided for a deterministic network with two teams, and the correlation between the rate of convergence and network structure/connectivity is highlighted. Illustrative examples demonstrate the algorithm's performance.

Chapter 6 addresses learning the optimal filtering policy (Kalman gain) for a linear system with unknown noise covariance matrices using noisy output data. The problem is formulated as stochastic policy optimization, aiming to minimize the output prediction error. Convergence analysis of the stochastic gradient descent algorithm is conducted, considering biased gradients and stability constraints. Error bounds that scale logarithmically with problem dimension are derived, leveraging tools from linear system theory and high-dimensional statistics.



## Chapter 2

# POLICY OPTIMIZATION OVER SUBMANIFOLDS OF STABILIZING CONTROLLERS

In this chapter, we study the linearly constrained policy optimization over the manifold of Schur stabilizing controllers, equipped with a Riemannian metric that emerges naturally in the context of optimal control problems. We provide extrinsic analysis of a generic constrained smooth cost function, which then, facilitates subsuming any such constrained problem into this framework. By studying the second order geometry of this manifold, we provide a Newton-type algorithm with local convergence guarantees that exploits this inherent geometry without using the exponential mapping and even in the absence of a numerically tractable retraction. It hinges upon the developed stability certificate and the linear structure of the constraints. We then apply our methodology to two well-known constrained optimal control problems. Finally, several numerical examples showcase the performance of our approach.

### **2.1 Introduction**

Recently, direct Policy Optimization (PO) for different variants of LQR problems have attracted considerable attention due to its fundamental modeling advantages for feedback dynamical systems; PO for linearly constrained LQR problems ( e.g., the state-feedback SLQR and OLQR), however, is less explored due to the complicated geometry of the feasible set and the non-convexity of the cost function itself. While reparameterization of the LQR problem to a convex setup is possible for unconstrained cases [Mohammadi et al., 2021b], in general, trivial constraints directly on the policy may become completely nontrivial and non-convex

after reparameterization<sup>1</sup>. Also, for constrained optimal LQR problems, the domain is generally non-convex [Ackermann, 1980], it may be disconnected (exponentially in number of states) with potentially multiple local minima on each component, and there is no guarantee that the first-order stationary points be necessarily local minima ( e.g.,see [Feng and Lavaei, 2019] for SLQR problem in continuous-time setup).

Finding the linear output-feedback policy directly for the OLQR problem was first addressed in [Levine and Athans, 1970] but it requires solving nonlinear matrix equations on each iteration. Since then, there has been on-going research efforts to approach this problem from different angles [Anderson and Moore, 2007; Mäkilä and Toivonen, 1987; Moerder and Calise, 1985; Toivonen, 1985; Mårtensson and Rantzer, 2009; Rautert and Sachs, 1997; Iwasaki et al., 1994; Toivonen and Mäkilä, 1987]. In particular, first and second order methods have been adopted for solving SLQR and OLQR problems (see e.g., [Toivonen and Mäkilä, 1987; Mäkilä and Toivonen, 1987] and references therein); However, these methods 1) often utilize backtracking line-search techniques at each iteration, 2) do not provide explicit guarantees for convergence, 2) do not exploit the inherent non-Euclidean geometry, 4) and finally lack a unifying setup that can handle any such linear constraint.

Recently, the state-feedback LQR problem has been studied through the lens of first order methods, in both discrete-time [Bu et al., 2019a] and continuous-time [Bu et al., 2020a] setups. This point of view was initiated when the LQR cost was shown to be *gradient dominant* [Fazel et al., 2018], which facilitates a global convergence guarantee of first order methods for this problem—despite of non-convexity. Since then, PO using first order methods has been investigated for variants of LQR problem such as OLQR [Fatkhullin and Polyak, 2020], model-free setup [Mohammadi et al., 2021a] and risk-constrained LQR [Zhao et al., 2023] just to name a few. The gradient dominant property, however, is only known to be valid with respect to the global optimum of the unconstrained case, and is not necessarily expected for the general constrained LQR problems. By merely using the first order information of

---

<sup>1</sup>There are special exceptions like the quadratic invariance class [Rotkowitz and Lall, 2006].

the cost function, PG techniques—whenever the projection is possible—can be shown to converge to first order stationary points but with a sublinear convergence rate ( e.g., see [Bu et al., 2019a] for SLQR problem). A sublinear rate may be considered to be too slow from a practical point of view, particularly when the second order information of the LQR cost is not out of reach. Despite the complications arising from the non-convexity and even non-connectedness of the feasible set for general constrained optimal control problems, one may consider developing faster convergent algorithms that could facilitate an accelerated exploration of the feasible set for local optima. Note that, the structure on the policy can be enforced through some regularization; however, this approach merely *promotes* the constraints and does not address the problems considered here because the constraints are prescribed as a hard requirement for feasibility of the solution ( e.g., see [Park et al., 2020] for promoting sparsity for SLQR problem). Here, we aim to utilize the second order information of the cost function to improve the convergence rate, and at the same time, subsume *any* linear constraint into a unified setup; e.g., such that the SLQR and OLQR problems can be handled correspondingly.

By ignoring the geometry of the problem, one may aim to optimize for the *linearly* constrained LQR cost by directly utilizing first or second order methods. Since the domain is non-convex, this might be possible by incorporating an Armijo-type backtracking line-search or requiring the initial condition to be close to local optima. Here, we preclude from incorporating any line-search in order to exploit the full potential of the geometry inherent to LQR cost. Adding any back-tracking technique is then considered as an immediate extension to our setup. Also, we aim to pave the road for future consideration of interesting system theoretic criteria that result in *nonlinear* constraints directly on the feedback policy—which is simplified tremendously if seen through the intrinsic geometry of the problem as investigated here (see §2.8 for an example).

Generally, the second order behavior of the cost function can be utilized in order to obtain a descent direction—as in the Newton method—as long as the “Hessian” stays positive definite. But, this second order behavior can be computed using distinct geometries; e.g.,

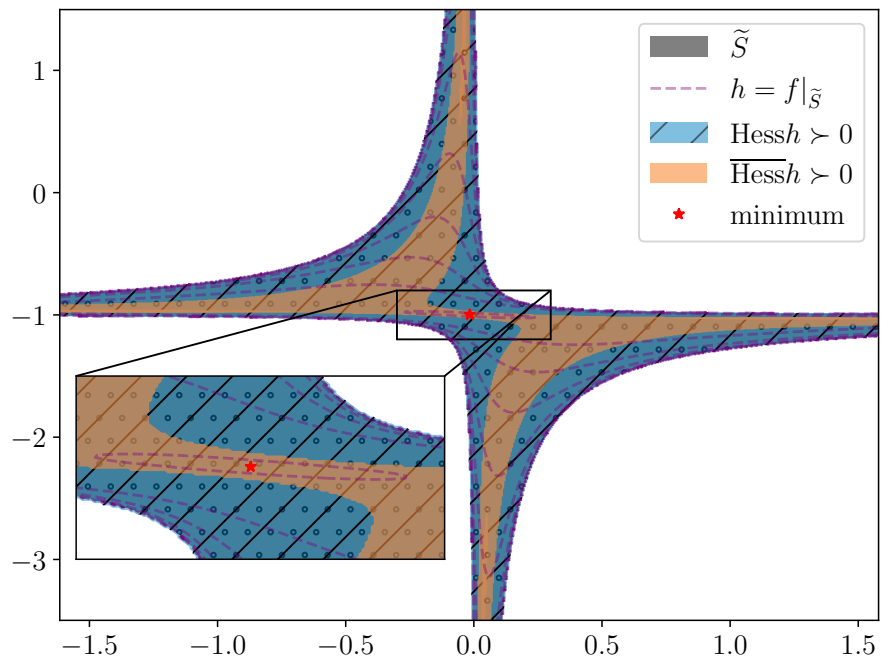


Figure 2.1: The actual submanifold  $\tilde{\mathcal{S}}$  of *diagonal* state-feedback stabilizing controllers for a system with 2 states and 2 inputs; superimposed with 1) the level-sets of the constrained LQR cost ( $h$ ), and the regions on which its “*Hessian*” is positive definite with respect to 2) the inherent Riemannian geometry ( $\text{Hess}h$ ) and 3) the Euclidean geometry ( $\overline{\text{Hess}h}$ ), respectively;  $\tilde{\mathcal{S}}$  is covered by the region on which  $\text{Hess}h \succ 0$ . See Example 2.18 in §2.7 for more details.

with respect to the usual Euclidean geometry (especially for linearly constrained problems) or more interestingly with respect to the non-Euclidean geometry inherent to the cost function itself. A simple—but actual—example of this comparison is depicted in Figure 2.1 over the set of “diagonally” constrained stabilizing controllers (denoted by  $\tilde{\mathcal{S}}$ )—which turns out to be non-convex even for this simple example consisting of two inputs and two states. More specifically, it is the intersection of the 4-dimensional set of stabilizing controllers with a 2-d plane defining the diagonal constraint. Note how the “Euclidean Hessian” of the constrained LQR cost (denoted by  $\overline{\text{Hess}h}$ ) remains positive definite on a much smaller subset of  $\tilde{\mathcal{S}}$ —especially on the vicinity of the minimum (denoted by  $\text{minimum}$ ). Therefore, one would expect that the neighborhood of minimum, on which the Newton updates using Euclidean geometry is guaranteed to converge, should be relatively small. However, if second order behavior of the LQR cost is considered through the lens of the relevant Riemannian geometry, its “Riemannian Hessian” (denoted by  $\text{Hess}h$ ) captures the behavior of the cost function effectively, such that it remains positive definite on a much larger domain—compare to  $\overline{\text{Hess}h}$ . Therefore, one expect a significant difference in performance of second order optimization algorithms utilizing these two distinct geometries.

The machinery developed in the literature for optimization over manifolds heavily depends on access to either the exponential mapping [Gabay, 1982; Smith, 1994] or a *retraction* from its tangent bundle onto the manifold itself [Absil et al., 2009]. However, due to the complicated geometry of the manifold of Schur stabilizing controllers, the exponential mapping is computationally burdensome and a more feasible retraction is not generally available. On the other hand, many interesting constraints for optimal LQR problems (such as OLQR or SLQR) inherits a linear structure. Therefore, it is pertinent to ask if we can still exploit the intrinsic non-Euclidean geometry— e.g., as induced by the quadratic cost and linear dynamics in optimal control problems and illustrated in Figure 2.1—while circumvent the absence of a computationally feasible retraction.

In this chapter, we consider a general optimization problem over the set of linearly constrained stabilizing controllers, which then, can easily be particularized for different con-

strained optimization problems. We introduce a Newton-type algorithm that utilizes both the inherent Riemannian geometry as well as the linear structure of the constraints, and provide convergence analysis to the local minima. Here, in the absence of any computationally feasible retraction, we obtain the so-called *stability certificate* that—together with the linear structure of the constraints—substitute the roll of a retraction by ensuring the feasibility of the next iterate. Finally, as the unit stepsize may not be possible in general, we guarantee a linear convergence rate which will eventually becomes quadratic. Finally, we provide applications of the proposed methodology to the well-known state-feedback SLQR and OLQR problems, followed by numerical examples. Our contributions are as follows:

- We study the second order geometry of the manifold of stabilizing controllers induced by a pertinent Riemannian metric and its associated *connection*<sup>2</sup>.
- We provide extrinsic analysis for first and second-order behavior of a generic smooth cost function constrained to a *Riemannian submanifold*. This, in turn, allows for a generic treatment of constrained optimization problems on the manifold of stabilizing controllers.
- We introduce a Newton-type algorithm with convergence guarantees that exploits the inherent Riemannian geometry in the absence of the exponential mapping or any retraction which hinges upon the developed stability certificate and the linear structure of constraints.
- We apply our methodology to SLQR and OLQR problems by first, computing the second order behavior of the LQR cost with respect to the Riemannian connection, and then, explicating the solution to Newton equation for each case in this geometry.
- While our approach allows for considering any choice of connection, here we focus on

---

<sup>2</sup>A connection in a tangent bundle can be viewed as a generalization of directional derivative in non-Euclidean geometry; see [Lee, 2018].

the associated Riemannian connection and its comparison to the ordinary Euclidean connection.

- Finally, we provide several numerical examples to showcase the performance and advantages of our proposed methodology that exploits this intrinsic geometry by leveraging on the stability certificate.

The rest of the chapter is organized as follows. In §2.2, we introduce the generic problem setup. We provide the analysis of this problem through the lens of differential geometry in §2.3. Next, we present the algorithm and its convergence analysis in §2.4 for optimization on submanifolds of stabilizing controllers with linear structure. Then, its applications to SLQR and OLQR problems are presented in §2.6. Finally, we provide numerical examples in §2.7.

**Notation:** The space of  $m \times n$  matrices over the real field is denoted by  $M(m \times n, \mathbb{R})$  with the trivial smooth structure determined by the atlas consisting of the single chart  $(M(m \times n, \mathbb{R}), \text{vec})$ , where  $\text{vec} : M(m \times n, \mathbb{R}) \rightarrow \mathbb{R}^{mn}$  denotes the vectorization operator which returns a vector obtained by stacking the columns of a matrix—from left to right—on top of one another. We denote the transpose operator by  $(\cdot)^\top$ . The trace and spectral radius of a square matrix is denoted by  $\text{tr}[\cdot]$  and  $\rho(\cdot)$ , respectively. The *Loewner* partial order of symmetric positive (semi-)definite matrices is denoted by  $\succ$  ( $\succcurlyeq$ ). We use the same notation to denote positive (semi-)definiteness of 2-tensor fields. The set of positive integers less than or equal to  $m$  is denoted by  $[m]$ . We denote the subset of stability matrices by  $\mathcal{M} := \{A \in M(n \times n, \mathbb{R}) \mid \rho(A) < 1\}$  which is open in  $M(n \times n, \mathbb{R})$ , and define the “*Lyapunov map*”

$$\mathbb{L} : \mathcal{M} \times M(n \times n, \mathbb{R}) \rightarrow M(n \times n, \mathbb{R})$$

that sends  $(A, \Sigma)$  to the unique solution  $Y$  of the following discrete Lyapunov equation

$$Y = AY A^\top + \Sigma, \tag{2.1}$$

which has the following infinite-sum representation

$$Y = \sum_{i=0}^{\infty} A^i \Sigma (A^\top)^i. \quad (2.2)$$

If  $\Sigma \succeq 0$  ( $\succ 0$ ), then  $Y \succeq 0$  ( $\succ 0$ ). Also, suppose  $\Sigma \succeq 0$  then,  $Y \succ 0$  if and only if  $(A, \Sigma^{1/2})$  is controllable<sup>3</sup>. Finally, for manifolds we follow the notation and results in [Lee, 2013] and [Lee, 2018] unless stated explicitly.

## 2.2 Problem Setup

Given a stabilizable pair  $(A, B)$  with  $A \in M(n \times n, \mathbb{R})$  and  $B \in M(n \times m, \mathbb{R})$ , we define the set of stabilizing controllers as

$$\mathcal{S} := \{K \in M(m \times n, \mathbb{R}) \mid \rho(A + BK) < 1\}.$$

Later, we also introduce a non-Euclidean geometry over  $\mathcal{S}$  using a metric arising naturally in the context of optimal control problems. We are often interested in controller gains  $K$  that lie on a relatively “simple” subset  $\mathcal{K}$  of  $M(m \times n, \mathbb{R})$ , such that  $\tilde{\mathcal{S}} := \mathcal{K} \cap \mathcal{S}$  is an embedded submanifold of  $\mathcal{S}$ . A common example of this is a linear subspace of  $M(m \times n, \mathbb{R})$  which characterizes a prescribed sparsity pattern for the admissible controller gains. Another example happens in optimal output-feedback control problems which will be considered in Section 2.6.5.

In this chapter we are concerned with the following optimization problem:

$$\begin{aligned} \min_K f(K) \in C^\infty(\mathcal{S}) \\ \text{s.t. } K \in \tilde{\mathcal{S}}, \end{aligned} \quad (2.3)$$

where  $\tilde{\mathcal{S}}$  is an embedded submanifold of  $\mathcal{S}$ , and especially when it is endowed with a linear structure. Later, we intend to use this linear structure together with an appropriate

---

<sup>3</sup>See [Gajic and Qureshi, 2008] and references therein for the existence, uniqueness and infinite-sum representation of the solution to the Lyapunov equation.



Riemannian geometry of  $\mathcal{S}$  to circumvent the absence of a (computationally feasible) global retraction from  $T\mathcal{S}$  onto  $\mathcal{S}$  (or  $\tilde{\mathcal{S}}$ )—which is due to the complicated geometry of  $\mathcal{S}$ .

In order to handle a generic embedded submanifold  $\tilde{\mathcal{S}}$ , we study the behavior of the restricted function  $h := f|_{\tilde{\mathcal{S}}}$  from an extrinsic point of view, which then can be applied to any such submanifold. Also, in most practical regimes, the function  $f$  is not convex and the constraint submanifold  $\tilde{\mathcal{S}}$  might be disconnected. Thus, here we focus on local convergence results that aims to exploit the inherent geometry of the problem in order to achieve fast convergence rates with relatively reasonable computational complexity. Finally, the existence of local (or global) optima for the general constrained minimization problem in (2.3), is out of scope of this work and thus, is assumed a priori throughout this chapter.

### **2.3 Optimization on Submanifolds of $\mathcal{S}$ and its Geometry**

In order to address this problem, we analyze the domain manifold using machinery borrowed from differential geometry. Note that embedded submanifolds  $\tilde{\mathcal{S}}$  endowed with a linear structure, can be investigated without using such machinery. However, neither the corresponding results can be generalized to submanifolds with “nonlinear” structure, nor the geometry induced by the cost function can be exploited for the optimization algorithm.

Before we proceed, it is worth noting that if we were to directly apply the results developed for optimization over the manifolds (such as [Absil et al., 2009]) it would have been necessary to access a retraction from the tangent bundle  $T\tilde{\mathcal{S}}$  onto  $\tilde{\mathcal{S}}$ . Unfortunately, in general such mapping is not available here, due to the complicated geometry of  $\mathcal{S}$ . Additionally, we will see that the Riemannian exponential map, with respect to the inherent geometry associated with optimal control problems, involves a system of ODE’s whose coefficients are solutions to different Lyapunov equations. Therefore, even though it is possible to compute the exponential mapping, in general it is hard to justify its computational complexity. Nonetheless, we show how we can circumvent this issue if the Riemannian tangential projection onto  $T\tilde{\mathcal{S}}$  is available—which is a rather computationally feasible operator in general.

### 2.3.1 Analysis of the domain manifold

It is known that  $\mathcal{S}$  is contractible [Bu et al., 2019b], and unbounded when  $m \geq 2$  with the topological boundary  $\partial\mathcal{S} = \{K \in M(m \times n, \mathbb{R}) \mid \rho(A + BK) = 1\}$  as a subset of  $M(m \times n, \mathbb{R})$ . Also,  $\mathcal{S}$  is open in  $M(m \times n, \mathbb{R})$  (which is a consequence of continuity of eigenvalues for matrices with smooth entries<sup>4</sup> and a passing to the quotient argument), and therefore a submanifold without boundary.<sup>5</sup>

In this chapter we focus on  $\mathcal{S}$  as a manifold on its own. Since  $\mathcal{S}$  can be covered by a single smooth chart, the tangent bundle of  $\mathcal{S}$ , denoted by  $T\mathcal{S}$ , is diffeomorphic to  $\mathcal{S} \times \mathbb{R}^{mn}$  which is also diffeomorphic to  $\mathcal{S} \times M(m \times n, \mathbb{R})$ , under the map  $\text{Id}_{\mathcal{S}} \times \text{vec}^{-1}$ . We refer to this composition of diffeomorphisms as “the usual identification of the tangent bundle” (or  $T_K\mathcal{S} \cong M(m \times n, \mathbb{R})$  at any point  $K \in \mathcal{S}$ ) if we need to identify any element of  $T\mathcal{S}$  (or  $T_K\mathcal{S}$ ). In particular, let us denote the coordinates of this global chart by  $(x^{i,j})$  for  $\mathcal{S}$ , its associated global coordinate frame by  $(\frac{\partial}{\partial x^{i,j}})$  or simply  $(\partial_{i,j})$ , and its dual coframe by  $(dx^{i,j})$ , where  $i = 1, \dots, m$  and  $j = 1, \dots, n$ . Also, the  $(k, \ell)$ -th element of any matrix  $A \in M(m \times n, \mathbb{R})$  is denoted by  $[A]_{k,\ell}$  or  $[A]^{k,\ell}$  depending on viewing  $A$  as a point or a tangent vector, respectively. Then, for example, under the usual identification of tangent bundle, for any fixed  $i$  and  $j$  we identify  $\partial_{i,j}$  as a matrix in  $M(m \times n, \mathbb{R})$  whose elements are  $[\partial_{i,j}]^{k,\ell} = 1$  if  $k = i$  and  $\ell = j$ , and otherwise  $[\partial_{i,j}]^{k,\ell} = 0$ . We also use the Einstein summation convention as explained in [Lee, 2018] for double indices; for example, we write  $x^{i,j}\partial_{i,j}$  to denote  $\sum_{i=1}^m \sum_{j=1}^n x^{i,j}\partial_{i,j}$ .

Before we proceed with studying the domain manifold, we present a technical lemma that will be used frequently throughout the chapter.

**Lemma 2.1.** *The subset  $\mathcal{M}$  is an open submanifold of  $M(n \times n, \mathbb{R})$ , the Lyapunov map  $\mathbb{L} : \mathcal{M} \times M(n \times n, \mathbb{R}) \rightarrow M(n \times n, \mathbb{R})$ , is smooth and its differential acts as follows*

$$d\mathbb{L}_{(A,Q)}[E, F] = \mathbb{L}(A, E\mathbb{L}(A, Q)A^\top + A\mathbb{L}(A, Q)E^\top + F),$$

---

<sup>4</sup>See Theorem 5.2 in [Serre, 2010], or Theorem 5.1 in [Kato, 2013].

<sup>5</sup>Cf. [Ohara and Amari, 1992] for another differential geometric study of the set of stabilizing controllers in continuous-time setup.

for any  $(E, F) \in T_{(A,Q)}(\mathcal{M} \times M(n \times n, \mathbb{R}))$  with the identification that follows by  $T_{(A,Q)}(\mathcal{M} \times M(n \times n, \mathbb{R})) \cong T_A \mathcal{M} \oplus T_Q M(n \times n, \mathbb{R}) \cong M(n \times n, \mathbb{R}) \oplus M(n \times n, \mathbb{R})$ . Furthermore, for any  $A \in \mathcal{M}$  and  $Q, \Sigma \in M(n \times n, \mathbb{R})$  (not necessarily symmetric) we have

$$\text{tr} [\mathbb{L}(A^\top, Q)\Sigma] = \text{tr} [Q \mathbb{L}(A, \Sigma)].$$

We refer to the last property of the Lyapunov map involving a trace as “*Lyapunov-trace*” property.

We can show that that many optimal control problems such as SLQR and OLQR (and even Linear Quadratic Gaussian (LQG) control) share a similar cost structure, as

$$f(K) = \frac{1}{2} \text{tr} [P_K \Sigma_K],$$

where  $P_K = \mathbb{L}(A_K^\top, Q + K^\top R K)$  with  $A_K := A + BK$ , and

$$\Sigma_K := \Sigma_1 + K^\top \Sigma_2 K, \tag{2.4}$$

for some  $\Sigma_1, \Sigma_2 \succeq 0$  with appropriate dimensions. Motivated by this, we define a covariant 2-tensor field on  $\mathcal{S}$  which will be proved to be a Riemannian metric. For that, let  $\mathfrak{X}(\mathcal{S})$  denote the set of all vector fields over  $\mathcal{S}$ .

**Lemma 2.2.** *Let  $\langle \cdot, \cdot \rangle_Y : \mathfrak{X}(\mathcal{S}) \times \mathfrak{X}(\mathcal{S}) \rightarrow C^\infty(\mathcal{S})$  denote the mapping that, under the usual identification of the tangent bundle, sets*

$$\langle V, W \rangle_Y \Big|_K := \text{tr} [(V_K)^\top W_K \mathbb{L}(A_K, \Sigma_K)], \quad \forall K \in \mathcal{S},$$

*for any  $V, W \in \mathfrak{X}(\mathcal{S})$ . Then, this mapping is well-defined and induced by a smooth symmetric covariant 2-tensor field.*

Now, let  $T^2(T^*\mathcal{S})$  denote the bundle of covariant 2-tensor fields on  $\mathcal{S}$  and define  $g : \mathcal{S} \rightarrow T^2(T^*\mathcal{S})$  to be its smooth section that sends  $K$  to  $\langle \cdot, \cdot \rangle_Y \Big|_K$ . Then,  $g$  is in fact a Riemannian metric under mild conditions formalized below.

**Proposition 2.3.** *If  $(A_K, \Sigma_K^{1/2})$  is controllable and  $\Sigma_K \succeq 0$  for all  $K \in \mathcal{S}$ , then the tuple  $(\mathcal{S}, g)$  is a Riemannian manifold, and with respect to the dual coframe  $(dx^{i,j})$  we have  $g = g_{(i,j)(k,\ell)} dx^{i,j} \otimes dx^{k,\ell}$  with each  $g_{(i,j)(k,\ell)} \in C^\infty(\mathcal{S})$  satisfying*

$$g_{(i,j)(k,\ell)}(K) = \begin{cases} [Y_K]_{\ell,j}, & \text{if } i = k, \\ 0, & \text{otherwise,} \end{cases}$$

where  $Y_K := \mathbb{L}(A_K, \Sigma_K)$ . Furthermore, the inverse “matrix”  $g^{(i,j)(k,\ell)}$  satisfies

$$g^{(i,j)(k,\ell)}(K) = \begin{cases} [Y_K^{-1}]_{\ell,j}, & \text{if } i = k, \\ 0, & \text{otherwise.} \end{cases}$$

*Remark 2.4.* The premise of Proposition 2.3 is satisfied if  $\Sigma_K \succ 0$  for all  $K \in \mathcal{S}$ ; e.g., when  $\Sigma_1 \succ 0$  and  $\Sigma_2 \succeq 0$ .

### 2.3.2 Riemannian Connection on $T\mathcal{S}$

First, consider a Riemannian submanifold  $(\tilde{\mathcal{S}}, \tilde{g})$  with  $\tilde{g} := \iota_{\tilde{\mathcal{S}}}^* g$  where  $\iota_{\tilde{\mathcal{S}}}^*$  denotes the pull-back by inclusion. In order to understand the second order behavior—i.e. the “Hessian”—of a smooth function on  $\tilde{\mathcal{S}}$ , we need to study the notion of connection in  $T\mathcal{S}$ , and then explain how that would be related to the one in the tangent bundle  $T\tilde{\mathcal{S}}$ . Recall that, by Fundamental Theorem of Riemannian Geometry, there exists a unique connection  $\nabla : \mathfrak{X}(\mathcal{S}) \times \mathfrak{X}(\mathcal{S}) \rightarrow \mathfrak{X}(\mathcal{S})$  in  $T\mathcal{S}$  that is compatible with  $g$  and symmetric; i.e., for all  $U, V, W \in \mathfrak{X}(\mathcal{S})$  we have:

- $\nabla_U \langle V, W \rangle_Y = \langle \nabla_U V, W \rangle_Y + \langle V, \nabla_U W \rangle_Y,$
- $\nabla_U V - \nabla_V U \equiv [U, V],$

where  $[V, W] \in \mathfrak{X}(\mathcal{S})$  denotes the Lie bracket of vector fields  $V$  and  $W$ .

Note that, merely the restriction of  $\nabla$  to  $\mathfrak{X}(\tilde{\mathcal{S}}) \times \mathfrak{X}(\tilde{\mathcal{S}})$  would not be a connection in  $T\tilde{\mathcal{S}}$  because its range does not necessary lie in  $\mathfrak{X}(\tilde{\mathcal{S}})$ . However, we can denote the (Riemannian) *tangential* and *normal projections* by  $\pi^\top : T\mathcal{S}|_{\tilde{\mathcal{S}}} \rightarrow T\tilde{\mathcal{S}}$  and  $\pi^\perp : T\mathcal{S}|_{\tilde{\mathcal{S}}} \rightarrow N\tilde{\mathcal{S}}$ , respectively,

with  $N\tilde{\mathcal{S}}$  indicating the normal bundle of  $\tilde{\mathcal{S}}$ . Subsequently, by The Gauss Formula, if  $\tilde{\nabla} : \mathfrak{X}(\tilde{\mathcal{S}}) \times \mathfrak{X}(\tilde{\mathcal{S}}) \rightarrow \mathfrak{X}(\tilde{\mathcal{S}})$  denotes the Riemannian connection in the tangent bundle  $T\tilde{\mathcal{S}}$ , then we can compute it as follows:

$$\tilde{\nabla}_U V = \pi^\top \nabla_U V, \quad (2.5)$$

for any  $U, V \in \mathfrak{X}(\tilde{\mathcal{S}})$ , where they are extended arbitrarily to smooth vector fields on a neighborhood of  $\tilde{\mathcal{S}}$  in  $\mathcal{S}$ .

Next, for computational purposes, we would like to obtain the Christoffel symbols associated with  $g$  (denoted by  $\Gamma_{(k,\ell)(p,q)}^{(i,j)}$ ) and thus completely determine the connection  $\nabla$ .

**Proposition 2.5.** *Consider a point  $K \in \mathcal{S}$  and, under the usual identification of  $T\mathcal{S}$ , define*

6

$$dY_K(p, q) := \mathbb{L}\left(A_K, B\partial_{(p,q)}Y_K A_K^\top + A_K Y_K \partial_{(p,q)}^\top B^\top + \partial_{(p,q)}^\top \Sigma_2 K + K^\top \Sigma_2 \partial_{(p,q)}\right),$$

for each  $(p, q) \in [m] \times [n]$  where  $Y_K = \mathbb{L}(A_K, \Sigma_K)$ . Then, the Christoffel symbols associated with the metric  $g$  can be obtained in the global coordinate frame  $(\partial_{(i,j)})$  as follows:

- If  $k \neq i \neq p \neq k$ , then  $\Gamma_{(k,\ell)(p,q)}^{(i,j)}(K) = 0$ .

- If  $k = i \neq p$ , then

$$\Gamma_{(k,\ell)(p,q)}^{(i,j)}(K) = \frac{1}{2} [dY_K(p, q) Y_K^{-1}]_{(\ell,j)}.$$

- If  $p = i \neq k$ , then

$$\Gamma_{(k,\ell)(p,q)}^{(i,j)}(K) = \frac{1}{2} [dY_K(k, \ell) Y_K^{-1}]_{(q,j)}.$$

- If  $p = k \neq i$ , then

$$\Gamma_{(k,\ell)(p,q)}^{(i,j)}(K) = -\frac{1}{2} \sum_s [dY_K(i, s)]_{(q,\ell)} [Y_K^{-1}]_{(s,j)}.$$

---

<sup>6</sup>In the proof, we show that  $dY_K(p, q) = \partial_{(p,q)} Y_K$ , denoting the action of the tangent vector  $\partial_{(p,q)}|_K$  on the map  $K \rightarrow (A_K, \Sigma_K) \xrightarrow{\mathbb{L}} Y_K$ .

- Finally, if  $p = k = i$ , then

$$\Gamma_{(k,\ell)(p,q)}^{(i,j)}(K) = \frac{1}{2} \sum_s \left( [dY_K(i, \ell)]_{(q,s)} + [dY_K(i, q)]_{(\ell,s)} - [dY_K(i, s)]_{(q,\ell)} \right) [Y_K^{-1}]_{(s,j)}.$$

*Remark 2.6.* Note that, with respect to the global coordinates of  $(\mathcal{S}, g, \nabla)$ , the *geodesic equation* is a system of  $(mn)$  second-order ordinary differential equations whose varying coefficients involve  $(mn)^3$  Christoffel symbols  $\Gamma_{(i,\ell)(i,q)}^{(i,j)}$  as obtained above. Therefore, computing the Riemannian Exponential mapping is computationally burdensome and this is one of the reasons we avoid using it as a retraction.

### 2.3.3 Extrinsic analysis of a smooth function constrained to a Riemannian submanifold

In this subsection, we study the the gradient and Hessian operator of a constrained smooth function from an extrinsic point of view, which is yet to be defined. In other words, we consider  $(\tilde{\mathcal{S}}, \tilde{g})$  as a Riemannian submanifold of  $(\mathcal{S}, g)$  where  $\tilde{g} = \iota_{\tilde{\mathcal{S}}}^* g$  with  $\iota_{\tilde{\mathcal{S}}}^*$  denoting the pull-back by inclusion of  $\tilde{\mathcal{S}}$  into  $\mathcal{S}$ . Then, by considering any smooth function  $f$  on  $\mathcal{S}$ , we can define its restriction to  $\tilde{\mathcal{S}}$  as

$$h := f|_{\tilde{\mathcal{S}}},$$

and ask how its gradient and Hessian operator is related to that of  $f$ . For proceeding to answer this question, we utilize the Riemannian connection in order to analyze the second order behaviour of  $f$  (or that of  $h$ ).

Recall from [Lee, 2018] that the gradient of  $f$  with respect to the Riemannian metric  $g$ , denoted by  $\text{grad}f \in \mathfrak{X}(\mathcal{S})$ , is defined to be the unique vector field satisfying

$$\langle V, \text{grad}f \rangle_Y = Vf,$$

for any  $V \in \mathfrak{X}(\mathcal{S})$ . Additionally, for any normal vector field  $N$  ( i.e. a smooth section of  $N\tilde{\mathcal{S}}$  ) the “*Weingarten map* in the direction of  $N$ ” is denoted by  $\mathbb{W}_N : \mathfrak{X}(\tilde{\mathcal{S}}) \rightarrow \mathfrak{X}(\tilde{\mathcal{S}})$  which defines a smooth bundle homomorphism from  $T\tilde{\mathcal{S}}$  to itself. Finally, define the “*Hessian operator*” of  $f \in C^\infty(\mathcal{S})$  as the map  $\text{Hess}f : \mathfrak{X}(\mathcal{S}) \rightarrow \mathfrak{X}(\mathcal{S})$  defined by

$$\text{Hess}f[U] := \nabla_U \text{grad}f,$$

for any  $U \in \mathfrak{X}(\mathcal{S})$ . Note that we use the same notation to denote the gradient and Hessian operator defined on the submanifold  $\tilde{\mathcal{S}}$ ; for further discussions regarding the Hessian operator refer to Section 2.3.4. Next, we formalize this abstract extrinsic examination as follows.

**Proposition 2.7.** *Suppose  $\tilde{\mathcal{S}}$  is an embedded Riemannian submanifold of  $\mathcal{S}$  both equipped with their Riemannian connections. Let  $f \in C^\infty(\mathcal{S})$  be any smooth function, then  $h := f|_{\tilde{\mathcal{S}}}$  is smooth on  $\tilde{\mathcal{S}}$  and we have*

$$\text{grad}h = \pi^\top (\text{grad}f|_{\tilde{\mathcal{S}}}).$$

Furthermore, under usual identification of  $T\tilde{\mathcal{S}} \subset T\mathcal{S}$ , for any  $V \in \mathfrak{X}(\tilde{\mathcal{S}})$  we have

$$\text{Hess}h[V] = \pi^\top (\text{Hess}f[V]|_{\tilde{\mathcal{S}}}) + \mathbb{W}_{\pi^\perp(\text{grad}f|_{\tilde{\mathcal{S}}})}(V),$$

where  $V$  is arbitrarily extended to smooth vector fields on a neighborhood of  $\tilde{\mathcal{S}}$  in  $\mathcal{S}$ .

#### 2.3.4 More on the Hessian Operator

The Hessian operator (denoted by  $\text{Hess}f$ ) as introduced in Section 2.3.3 is well-defined and the value of  $\text{Hess}f[U]$  at any  $K \in \mathcal{S}$  depends only on  $U_K$  due to this property for the connection. Note that for any  $U, W \in \mathfrak{X}(\mathcal{S})$ ,

$$\begin{aligned} \langle \text{Hess}f[U], W \rangle_Y &= U \langle \text{grad}f, W \rangle_Y - \langle \text{grad}f, \nabla_U W \rangle_Y \\ &= U(Wf) - (\nabla_U W)f \\ &= W(Uf) - (\nabla_W U)f \end{aligned} \tag{2.6}$$

where the first equality is because the Riemannian connection is compatible with the metric, the second equality is by definition of  $\text{grad}f$ , and the last equality is due to symmetry of the Riemannian connection. Thus, by (2.6) we can conclude that the Hessian operator is self-adjoint, i.e.

$$\langle \text{Hess}f[U], W \rangle_Y = \langle U, \text{Hess}f[W] \rangle_Y.$$

Similarly, we can define  $\text{Hess}h$  for any smooth function  $h \in C^\infty(\tilde{\mathcal{S}})$  where we consider the submanifold  $\tilde{\mathcal{S}} \subset \mathcal{S}$  with the induced Riemannian metric and associated Riemannian connection of  $\mathcal{S}$ .

Next, for the computational purposes, we would like to also introduce the “covariant Hessian of  $f$ ” with respect to  $g$ , denoted by  $\nabla^2 f$ . It is a 2-tensor field obtained by taking total covariant derivative of  $f$  twice. The Riemannian connection is symmetric so is the covariant Hessian. Also, it is easy to see that covariant Hessian and Hessian operator are related as follows

$$\nabla^2 f[W, U] = \nabla_U \nabla_W f - (\nabla_U W)f = \langle \text{Hess}f[U], W \rangle_Y, \quad (2.7)$$

where the last equality follows by (2.6).

Furthermore, recall that  $\text{grad}f = (df)^\sharp$ , where  $\sharp$  denote the index raising operator called the “sharp” operator. On the other hand,  $\text{Hess}f : \mathfrak{X}(\mathcal{S}) \rightarrow \mathfrak{X}(\mathcal{S})$  can be viewed as the total covariant derivative of  $\text{grad}f$ , i.e.  $\text{Hess}f = \nabla \text{grad}f$ . But then

$$\text{Hess}f = \nabla(df)^\sharp = (\nabla(df))^\sharp = (\nabla^2 f)^\sharp, \quad (2.8)$$

where the equality in the middle follows by the fact the index raising commute with the covariant derivative operator, and the last equality is due to the definition of connection for a smooth function  $f \in C^\infty(\mathcal{S})$ . Note that in (2.8), the index raising refer to the second argument of  $\nabla^2 f$ . However, as the covariant Hessian of any smooth function is a symmetric 2-tensor field, the index raising could be respect to any of the entries. Finally, similar definitions and relations as discussed above are available for  $h$  as a smooth function on the embedded Riemannian submanifold  $\tilde{\mathcal{S}}$  with the induced metric and corresponding connection, which are omitted for brevity.

### 2.3.5 The choice of connection

On the manifold  $(\mathcal{S}, g, \nabla)$ , computing the exponential map requires finding solution to a system of ODE’s of dimension  $(mn)$ . Furthermore, it does not necessarily provide an exponential map of a submanifold  $\tilde{\mathcal{S}}$  (unless it happens to be totally geodesic). To avoid the computation complexity of Riemannian exponential map, it seems reasonable to perform updates by using simpler “retractions” from tangent bundle to the manifold (such as the one



defined in [Absil et al., 2009]), however, in general we do not have access to such a retraction in our setup. Another computational complexity of utilizing the Riemannian connection associated with the Riemannian metric  $g$  is the  $(mn)^2$ -number of Lyapunov equations involved in obtaining the Christoffel symbols at each point.

On the other hand, for the applications in which the submanifold appears as  $\tilde{\mathcal{S}} = \mathcal{S} \cap \mathcal{K}$  where  $\mathcal{K}$  is an affine subspace of  $M(m \times n, \mathbb{R})$ , it might seem reasonable to consider the ambient manifold  $(\mathcal{S}, g, \bar{\nabla})$  where  $\bar{\nabla}$  refers to the so-called “*Euclidean connection*”; i.e. the connection whose coefficients (with respect to the global coordinates) all vanish ( $\bar{\Gamma}_{(k,\ell)(p,q)}^{i,j} \equiv 0$  on  $\mathcal{S}$ ). This results in a simpler “Hessian” operator which, however, does not respect the geometry of  $(\mathcal{S}, g)$  simply because  $\bar{\nabla}$  is not compatible with the metric  $g$ —in contrast to its associated Riemannian connection. Nonetheless, for completeness, we also define the “*Euclidean Hessian operator*” of  $f \in C^\infty(\mathcal{S})$  as the map  $\overline{\text{Hess}}f : \mathfrak{X}(\mathcal{S}) \rightarrow \mathfrak{X}(\mathcal{S})$  defined by

$$\overline{\text{Hess}}f[U] := \bar{\nabla}_U \text{grad}f,$$

for any  $U \in \mathfrak{X}(\mathcal{S})$ . It enjoys similar properties as that of  $\text{Hess}f$ , but contains different information about the second order behavior of  $f$  ( e.g., see Figure 2.1 for a comparison).

#### 2.4 Algorithm: Riemannian Newton-type Policy Optimization (RNPO)

Here, we propose an algorithm for learning local optima of smooth cost functions, constrained to submanifolds of  $\mathcal{S}$  that are endowed with a linear structure; that is where  $\tilde{\mathcal{S}} = \mathcal{S} \cap \mathcal{K}$  where  $\mathcal{K}$  entails a linear structure in  $M(m \times n, \mathbb{R})$ . We abstain from using the exponential mapping (due to its computational complexity), in spite of the fact that no other retraction from the tangent space onto the manifold  $\mathcal{S}$  is known. Instead, we exploit this linear structure together with an active stability certificate which guarantees stability of the next iteration by adjusting the stepsize.

In what follows, we first introduce the stability certificate and then propose our algorithm. We show how to use this certificate to determine stepsizes guaranteeing a linear convergence rate, and more importantly, the existence of neighborhoods containing local minima on which

the algorithm reaches a quadratic rate of convergence.

#### 2.4.1 The stability certificate and the algorithm

We know that  $\mathcal{S}$  is open in  $M(m \times n, \mathbb{R})$ , nonetheless, we provide the following lemma which quantifies this fact with respect to the problem parameters, which can be utilized later for analysis of any generic iterative algorithm.

**Lemma 2.8.** *Suppose  $K \in \mathcal{S}$  and choose  $Q_K \in M(n \times n, \mathbb{R})$  such that  $Q_K \succ 0$ . Given any direction  $G \in T_K \mathcal{S} \cong M(m \times n, \mathbb{R})$ , if step-size  $\eta$  is small enough such that*

$$0 \leq \eta \leq s_K := \frac{\lambda_{\min}(Q_K)}{2\lambda_{\max}(\mathbb{L}(A_K^\top, Q_K)) \|BG\|_2},$$

then  $K^+ := K + \eta G \in \mathcal{S}$  and  $s_K$  is referred to as the “stability certificate” at  $K$ .

**Proof.** Since  $K$  is stabilizing, for any such  $Q_K$ , there exists  $P \succ 0$  satisfying

$$P = \mathbb{L}(A_K^\top, Q_K) = A_{K^+}^\top P A_{K^+} + L,$$

where

$$\begin{aligned} L &= Q_K - \eta G^\top B^\top P A_K - \eta A_K^\top P B G - \eta^2 G^\top B^\top P B G \\ &\succeq Q_K - a A_K^\top P A_K - (1 + 1/a) \eta^2 G^\top B^\top P B G \\ &= (1 + a)(Q_K) - aP - (1 + 1/a) \eta^2 G^\top B^\top P B G, \end{aligned}$$

for any  $a > 0$ . Now, by recalling the infinite sum representation of  $P$  and the fact that  $Q_K \succeq 0$  we conclude that  $\lambda_{\max}(P) \geq \lambda_{\max}(Q_K) \geq \lambda_{\min}(Q_K)$ . Thus, we choose  $a$  as follows: if  $\lambda_{\max}(P) > \lambda_{\min}(Q_K)$  then choose it so that

$$1/2a + 1 = \lambda_{\max}(P)/\lambda_{\min}(Q_K);$$

otherwise, choose it large enough such that  $a > 2\eta^2 \|BG\|_2^2$ . Then, either way we must have

$$\lambda_{\min}(L) \geq \lambda_{\min}(Q_K)/2 - \left[ \frac{2\lambda_{\max}^2(P)}{\lambda_{\min}(Q_K)} - \lambda_{\max}(P) \right] \eta^2 \|BG\|_2^2.$$

Therefore, if

$$|\eta| \leq \lambda_{\min}(Q_K)/(2\lambda_{\max}(P)\|BG\|_2)$$

then  $L > 0$ , implying that  $K^+$  is stabilizing by Lyapunov Stability Criterion.  $\square$

*Remark 2.9.* The proceeding lemma also provides a “conditioning” of the optimization problem at hand in terms of system parameters  $A, B$ . In other words, for any choice of  $Q_K \succ 0$  at any  $K \in \mathcal{S}$ , the quotient  $\lambda_{\max}(\mathbb{L}(A_K^\top, Q_K))/\lambda_{\min}(Q_K)$  represents a condition number indicating information about the manifold at  $K$ . In a sense, this is capturing properties of the “Riemannian curvature” of  $(\mathcal{S}, g, \nabla)$  where its precise characterization is deferred to our future work.

Next, we propose an algorithm with convergence guarantees with at least a linear rate (specially when the iterates are far away from the local optima) and eventually a Q-quadratic rate (when it gets close enough to the local optima). The complication here is that we do not have access to a retraction with a reasonable computational complexity (see Remark 2.6). We claim that, starting close enough to a local minimum, a Newton-type method using Riemannian metric and the Euclidean/Riemannian connection must converge quadratically if one could have used stepsize  $\eta = 1$ . This is in fact due to the exponential mapping with respect to the Euclidean connection that serves as a retraction with desirable properties. However, the stability certificate suggest that at least away from the local minimum it might not be possible to use such a large stepsize. Therefore, a stepsize rule is deduced—which hinges on the stability certificate—and the algorithm is summarized in Algorithm 2.1.

Hereafter, we refer to the solution  $G \in T_K\tilde{\mathcal{S}}$  of the following equation as the *Newton direction* on the submanifold  $\tilde{\mathcal{S}}$ :

$$\text{Hess}h_K[G] = -\text{grad}h_K,$$

where  $h = f|_{\tilde{\mathcal{S}}}$ ; and similarly *Euclidean Newton direction* if  $\text{Hess}h$  is replaced by  $\overline{\text{Hess}h}$ .

---

**Algorithm 2.1:** Riemannian Newton-type Policy Optimization (RNPO)
 

---

- 1: **Initialization:** Problem parameters  $(A, B)$ , the linear constraint  $\mathcal{K}$  and an initial feasible stabilizing controller  $K_0 \in \tilde{\mathcal{S}} = \mathcal{S} \cap \mathcal{K}$
- 2: Choose a smooth mapping  $K \rightarrow Q_K \succ 0$
- 3: **For**  $t \geq 0$ , **do**
- 4:     Solve for the Newton direction  $G_t$  on  $\tilde{\mathcal{S}}$ :

$$\text{Hess}h_{K_t}[G_t] = -\text{grad}h_{K_t}$$

- 5:     Use  $Q_{K_t}$  to obtain a stability certificate  $s_{K_t}$
- 6:     Compute step-size  $\eta_t$ :

$$\eta_t = \min \{s_{K_t}, 1\}$$

- 7:     Update:

$$K_{t+1} = K_t + \eta_t G_t$$

- 8:      $t \leftarrow t + 1$
- 

The update in Line 7 is possible due to the linear structure of  $\tilde{\mathcal{S}}$  induced by  $\mathcal{K}$ . Furthermore, the Hessian operator  $\text{Hess}h$  can be replaced by its Euclidean counterpart  $\overline{\text{Hess}h}$ .

## 2.5 Analysis of RNPO

Herein, we establish the local linear-quadratic convergence of RNPO algorithm on the submanifold  $\tilde{\mathcal{S}}$  using differential geometric techniques [Lee, 2018; Absil et al., 2009; Gabay, 1982]. Herein, avoiding the exponential map induced by the Riemannian connection for updating the iterates, and using stability certificate add another layer of complication for convergence analysis. To proceed, we say  $K^*$  is a critical point of  $h$  if  $\text{grad}h_{K^*} = 0$  and it is “nondegenerate” if  $\text{Hess}h_{K^*}$  is nondegenerate, i.e.

$$\langle \text{Hess}h_{K^*}[G], G \rangle_{Y_{K^*}} = 0 \implies G = 0 \in T_{K^*}\tilde{\mathcal{S}}.$$

**Lemma 2.10.** *Suppose  $K^*$  is a nondegenerate local minimum of  $h := f|_{\tilde{\mathcal{S}}}$ . Then  $\text{grad}h_{K^*} = 0$  and there exists a neighborhood of  $K^*$  on which  $\text{Hess}h$  is positive definite. Furthermore,  $\text{Hess}h_{K^*} = \overline{\text{Hess}h_{K^*}}$ .*

The next theorem provides a local convergence guarantee for RNPO algorithm.

**Theorem 2.11.** *Suppose  $K^*$  is a nondegenerate local minimum of  $h := f|_{\tilde{\mathcal{S}}}$  over the submanifold  $\tilde{\mathcal{S}} = \mathcal{S} \cap \mathcal{K}$  for some linear constraint  $\mathcal{K}$ . Then, there exists a neighborhood  $\mathcal{U}^* \subset \tilde{\mathcal{S}}$  of  $K^*$  with the following property: whenever  $K_0 \in \mathcal{U}^*$ , the sequence  $\{K_t\}$  generated by RNPO remains in  $\mathcal{U}^*$  (therefore, is stabilizing), and it converges to  $K^*$  at least at a linear rate and eventually a quadratic rate.*

*Remark 2.12.* The above theorem implies that there exist neighborhoods containing each nondegenerate local minimum of the constrained cost function  $h$  on which the convergence of RNPO is guaranteed. The usefulness of this result is that the initial iterate  $K_0$  is not required to be in a (relatively much smaller) neighborhood of the optimality on which the step-size  $\eta = 1$  is feasible. Instead, by carefully incorporating the stability certificate as obtained in Lemma 2.8, we can obtain a larger basin of attraction for optimality (see Figure 2.5 in §2.7). Finally, even though the convergence rate is at least linear initially, as the algorithm proceeds a quadratic convergence rate is guaranteed.

### 2.5.1 Theoretical Analysis of the Results in §2.3

**Proof of Lemma 2.1.** Since  $\rho : M(n \times n, \mathbb{R}) \rightarrow \mathbb{R}$  is a continuous map,  $\mathcal{M}$  is an open subset of  $M(n \times n, \mathbb{R})$  and thus an open submanifold. For each  $A \in \mathcal{M}$ , by Lyapunov Stability Criterion, there exists a unique solution  $Y$  to (2.1) which satisfies (2.2). But, as for each  $A \in \mathcal{M}$  the series converges, each element of  $Y$  can be written as a convergent power series of elements of  $A$  and  $\Sigma$ . Therefore, each element of  $Y$  is a real analytic function of several variables (as defined in [Krantz and Parks, 2002]) on the open subset  $\mathcal{M} \times M(n \times n, \mathbb{R}) \subset M(n \times n, \mathbb{R})^2$ . So, we conclude that  $\mathbb{L}$  is a well-defined smooth map.<sup>7</sup> Next, under the identification in the premise, we can calculate

$$d\mathbb{L}_{(A,Q)}[E, F] = d\mathbb{L}_{(A,Q)}[E, 0] + d\mathbb{L}_{(A,Q)}[0, F].$$

But  $\mathbb{L}$  is linear in the second variable, so  $d\mathbb{L}_{(A,Q)}[0, F] = \mathbb{L}(A, F)$ . Also since  $\mathcal{M}$  is open, for small enough  $\varepsilon$ ,  $\gamma : [0, \varepsilon] \rightarrow \mathcal{M} \times M(n \times n, \mathbb{R})$  with  $\gamma(t) = (A + tE, Q)$  is a well-defined smooth curve starting at  $(A, Q)$  whose initial velocity is  $(E, 0)$ . Then,

$$d\mathbb{L}_{(A,Q)}[E, 0] = d/dt|_{t=0} \mathbb{L} \circ \gamma(t).$$

Let  $Y_t := \mathbb{L} \circ \gamma(t)$  and  $Y := \mathbb{L} \circ \gamma(0)$ , then we obtain that

$$Y_t - Y = \mathbb{L}(A, t(EYA^\top + AYE^\top) + \mathcal{O}(t^2)) = t\mathbb{L}(A, EYA^\top + AYE^\top) + \mathcal{O}(t^2),$$

where the first equality is by direct algebraic manipulation and the second one follows by linearity of  $\mathbb{L}$  in the second coordinate. Therefore,  $d\mathbb{L}_{(A,Q)}[E, 0] = \mathbb{L}(A, EYA^\top + AYE^\top)$ , and the first claim follows by adding the two computed differentials and using linearity of  $\mathbb{L}$  in the second coordinate again. Finally, note that any square matrix has an spectrum identical to its transpose, therefore if  $A \in \mathcal{M}$  then  $A^\top \in \mathcal{M}$ , and thus the last property follows by the convergent series representations of  $\mathbb{L}(A^\top, Q)$  and  $\mathbb{L}(A, \Sigma)$ , and cyclic permutation property of trace.  $\square$

---

<sup>7</sup>An alternative argument can be provided by the closed form solution of (2.1) and its “vectorization” involving rational functions of several variables with non-vanishing denominators—cf. Lemma 3.6 in [Buet al., 2019a].

**Proof of Lemma 2.2.** First, we show that it is well-defined. Indeed, by Lemma 2.1 for each  $K \in \mathcal{S}$ ,  $\mathbb{L}(A_K, \Sigma_K)$  is uniquely determined, symmetric and smooth in  $K$ , since  $A_K$  is stabilizing. Also,  $\mathbb{L}(A_K, \Sigma_K) \in M(n \times n, \mathbb{R})$  is positive semidefinite definite by observing its series representation as in (2.2) and the fact that  $\Sigma_K \succeq 0$ . Next,  $\text{tr}[(V_K)^\top W_K \mathbb{L}(A_K, \Sigma_K)]$  is a smooth function of elements of  $V_K, W_K$  and  $\mathbb{L}(A_K, \Sigma_K)$ . Therefore, for any  $V, W \in \mathfrak{X}(\mathcal{S})$ , the function  $\langle V, W \rangle_Y$  as defined in the premise is well-defined and smooth on  $\mathcal{S}$ . Additionally, by linearity of trace, we observe that  $\langle \cdot, \cdot \rangle_Y$  is multilinear over  $C^\infty(\mathcal{S})$ , i.e.

$$\langle fU + hV, W \rangle_Y = f \langle U, W \rangle_Y + h \langle V, W \rangle_Y,$$

for any  $f, h \in C^\infty(\mathcal{S})$  and  $U, V \in \mathfrak{X}(\mathcal{S})$ , and similarly for the second entry. Therefore, by Tensor Characterization Lemma, it is induced by a smooth covariant 2-tensor field. Finally, it is symmetric because for any  $K \in \mathcal{S}$

$$\langle V, W \rangle_Y |_{K} = \text{tr}[(\mathbb{L}(A_K, \Sigma_K))^\top (W_K)^\top V_K] = \text{tr}[(W_K)^\top V_K (\mathbb{L}(A_K, \Sigma_K))^\top] = \langle W, V \rangle_Y |_{K},$$

where we used the symmetry of  $\mathbb{L}(A_K, \Sigma_K)$ .  $\square$

**Proof of Proposition 2.3.** We know that  $\mathcal{S}$  is a smooth manifold, and by Lemma 2.2 and Smoothness Criteria for Tensor Fields,  $g$  is a smooth symmetric covariant 2-tensor field. So it is left to show that it is positive definite at each point  $K \in \mathcal{S}$ . But  $\Sigma_K \succeq 0$  and  $(A_K, \Sigma_K)$  is controllable, therefore  $Y_K = \mathbb{L}(A_K, \Sigma_K)$  is a positive definite matrix, implying that  $g_K(E, E) = \text{tr}[(EY_K^{\frac{1}{2}})^\top EY_K^{\frac{1}{2}}] \geq 0$ , for any  $E \in T_K\mathcal{S}$  with equality if and only if  $E$  is the zero element. Next, to calculate the coordinate representation of  $g$ , for each coordinate pairs  $(i, j)$  and  $(k, \ell)$  we have

$$g_{(i,j)(k,\ell)}(K) = g_K(\partial_{i,j}|_K, \partial_{k,\ell}|_K) = \text{tr}[(\partial_{i,j}|_K)^\top \partial_{k,\ell}|_K Y_K],$$

under the usual identification of  $T_K\mathcal{S}$ . Since under this identification each  $\partial_{i,j}$  corresponds to an element of  $M(m \times n, \mathbb{R})$  with entry 1 in  $(i, j)$ -th coordinate and zero elsewhere, the expression for  $g_{(i,j)(k,\ell)}$  follows by direct computation of the last equality. Finally, by definition of “inverse matrix”, for each  $(i, j)$  and  $(k, \ell)$  we must have  $\sum_{r,s} g^{(i,j)(r,s)} g_{(r,s)(k,\ell)} = 1$  if  $(i, j) =$

$(k, \ell)$  and 0 otherwise. Next, for each  $i = k$ , let  $[g^{(k, \cdot)(k, \cdot)}]$  denote the matrix with  $g^{(k, j)(k, s)}$  as its  $(j, s)$ -th entry. Then, by the expression for  $g_{(i, j)(k, \ell)}$ , it must satisfy  $[g^{(k, \cdot)(k, \cdot)}] Y_K = I_n$ , and therefore  $[g^{(k, \cdot)(k, \cdot)}] = Y_K^{-1}$  as  $Y_K \succ 0$ . The rest follows by doing a similar calculation for each  $i \neq k$  and noting the zero pattern in the expression for  $g_{(i, j)(k, \ell)}$ .  $\square$

**Proof of Proposition 2.5.** For each  $(i, j), (k, \ell), (p, q) \in [m] \times [n]$ , we know that

$$\Gamma_{(k, \ell)(p, q)}^{(i, j)} = \sum_{r, s} \frac{g^{(i, j)(r, s)}}{2} \left( \partial_{(k, \ell)} g_{(r, s)(p, q)} + \partial_{(p, q)} g_{(r, s)(k, \ell)} - \partial_{(r, s)} g_{(k, \ell)(p, q)} \right), \quad (2.9)$$

where  $g^{(i, j)(r, s)}$  denotes the inverse matrix of  $g_{(i, j)(r, s)}$ . If  $k \neq i \neq p \neq k$ , then by sparsity pattern in the expression for  $g_{(i, j)(k, \ell)}$  in Proposition 2.3 we obtain  $\Gamma_{(k, \ell)(p, q)}^{(i, j)} = 0$ . Next, if  $k = i \neq p$ , then we have

$$\sum_s \frac{g^{(i, j)(i, s)}}{2} \left( \partial_{(p, q)} g_{(i, s)(i, \ell)} \right) = \frac{1}{2} \left( \partial_{(p, q)} [Y_K]_{(\ell, \cdot)} \right) [Y_K^{-1}]_{(\cdot, j)}.$$

Next, for any fix  $i, p$  and  $q$ , let  $\Gamma_{(i, \cdot)(p, q)}^{(i, \cdot)}$  denote the  $n \times n$  matrix with  $\Gamma_{(i, \ell)(p, q)}^{(i, j)}$  as its  $(\ell, j)$  entry. Then it must satisfy

$$\Gamma_{(i, \cdot)(p, q)}^{(i, \cdot)} = \frac{1}{2} \left( \partial_{(p, q)} Y_K \right) Y_K^{-1},$$

where  $\partial_{(p, q)} Y_K$  indicates the action of tangent vector  $\partial_{(p, q)}$  on the composite map  $K \rightarrow (A_K, \Sigma_K) \xrightarrow{\mathbb{L}} Y_K$ . By Lemma 2.1, we can calculate

$$\partial_{(p, q)} Y_K = d\mathbb{L}_{(A_K, \Sigma_K)} \left[ B \partial_{(p, q)}, \partial_{(p, q)}^\top \Sigma_2 K + K^\top \Sigma_2 \partial_{(p, q)} \right] = dY_K(p, q),$$

under the usual identification of  $T_K \mathcal{S}$ . This proves the second case. The third case follows by the symmetry of the Riemannian connection, i.e.  $\Gamma_{(k, \ell)(p, q)}^{(i, j)} = \Gamma_{(p, q)(k, \ell)}^{(i, j)}$ . Next, if  $p = k \neq i$ , then by Proposition 2.3, we get

$$\sum_s \frac{g^{(i, j)(i, s)}}{2} \left( \partial_{(i, s)} g_{(k, \ell)(k, q)} \right) = \frac{1}{2} \sum_s \left[ \partial_{(i, s)} Y_K \right]_{(q, \ell)} [Y_K^{-1}]_{(s, j)},$$

with  $\partial_{(i, s)} Y_K = dY_K(i, s)$  computed similarly. Therefore, (2.9) simplifies for  $\Gamma_{(k, \ell)(k, q)}^{(i, j)}$  as claimed. Finally, if  $k = i = p$  then similarly we get

$$\sum_s \frac{g^{(i, j)(i, s)}}{2} \left( \partial_{(i, \ell)} g_{(i, s)(i, q)} + \partial_{(i, q)} g_{(i, s)(i, \ell)} - \partial_{(i, s)} g_{(i, \ell)(i, q)} \right),$$



which simplifies (2.9) for  $\Gamma_{(i,\ell)(i,q)}^{(i,j)}$ .  $\square$

**Proof of Proposition 2.7.** Note that  $f$  is smooth and  $\tilde{\mathcal{S}}$  is an embedded submanifold of  $\mathcal{S}$ . Therefore,  $h : \tilde{\mathcal{S}} \rightarrow \mathbb{R}$  is smooth by restriction and we can define  $\text{grad}h$  and  $\text{Hess}h$  on  $\tilde{\mathcal{S}}$ . But,  $\text{grad}h \in \mathfrak{X}(\tilde{\mathcal{S}})$  is the unique vector field on  $\tilde{\mathcal{S}}$  such that  $\tilde{g}(W, \text{grad}h) = Wh$  for any  $W \in \mathfrak{X}(\tilde{\mathcal{S}})$ . Unravelling the definition implies that for any  $K \in \tilde{\mathcal{S}} \subset \mathcal{S}$ ,

$$\begin{aligned} dh_K(W_K) &= df_K(d\iota_{\tilde{\mathcal{S}}}(W_K)) \\ &= g_K(d\iota_{\tilde{\mathcal{S}}}(W_K), \text{grad}f_K) \\ &= g_K(d\iota_{\tilde{\mathcal{S}}}(W_K), d\iota_{\tilde{\mathcal{S}}}(\pi^\top \text{grad}f_K)), \end{aligned}$$

because  $h = f \circ \iota_{\tilde{\mathcal{S}}}$  and thus  $dh = df \circ d\iota_{\tilde{\mathcal{S}}}$ , where the last equality follows by the fact that  $\iota_{\tilde{\mathcal{S}}}(W_K) \in T_K\mathcal{S}$  is tangent to  $\tilde{\mathcal{S}}$ . By definition of tangential projection,  $\pi^\top(\text{grad}f|_{\tilde{\mathcal{S}}})$  is then a smooth vector field on  $\tilde{\mathcal{S}}$  that satisfies

$$Wh = \tilde{g}(W, \pi^\top \text{grad}f|_{\tilde{\mathcal{S}}}),$$

for any  $W \in \mathfrak{X}(\tilde{\mathcal{S}})$ . Therefore, the first claim follows by uniqueness of the gradient. Next, note that the Hessian operator of  $h \in C^\infty(\tilde{\mathcal{S}})$  is defined as  $\text{Hess}h[V] := \tilde{\nabla}_V \text{grad}h$ , for any  $V \in \mathfrak{X}(\tilde{\mathcal{S}})$ . But then, the first claim together with (2.5) and the linearity of connection imply that

$$\begin{aligned} \text{Hess}h[V] &= \pi^\top \nabla_V(\pi^\top(\text{grad}f|_{\tilde{\mathcal{S}}})) \\ &= \pi^\top(\text{Hess}f[V]|_{\tilde{\mathcal{S}}}) - \pi^\top \nabla_V(\pi^\perp(\text{grad}f|_{\tilde{\mathcal{S}}})), \end{aligned}$$

where all  $V$ ,  $\pi^\top(\text{grad}f|_{\tilde{\mathcal{S}}})$  and  $\pi^\perp(\text{grad}f|_{\tilde{\mathcal{S}}})$  are extended arbitrarily to smooth vector fields on a neighborhood of  $\tilde{\mathcal{S}}$  in  $\mathcal{S}$ . Finally, the extrinsic expression of  $\text{Hess}h$  follows by The Weingarten Equation.  $\square$

**Proof of Lemma 2.10.** Let  $\tilde{\gamma} : (-\varepsilon, \varepsilon) \rightarrow \tilde{\mathcal{S}}$  denote the smooth geodesic curve on the submanifold  $\tilde{\mathcal{S}}$  with  $\tilde{\gamma}(0) = K^*$  and  $\tilde{\gamma}'(0) = F$  for an arbitrary  $F \in T_{K^*}\tilde{\mathcal{S}}$ . Define  $\ell(t) :=$

$h \circ \tilde{\gamma}(t) : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}$  which it is smooth by composition. Then,  $K^*$  is a strict local minimum for  $h$ , so is  $t = 0$  for  $\ell(t)$  following by continuity of  $\tilde{\gamma}$ . Therefore,

$$0 = \ell'(0) = dh_{\tilde{\gamma}(0)} \circ \tilde{\gamma}'(0) = \langle \text{grad}h_{K^*}, F \rangle_{Y_{K^*}},$$

and as  $F$  was an arbitrary tangent vector, we conclude that  $\text{grad}h_{K^*} = 0$ . Next, Taylor's formula for  $\ell$  at  $t = 0$  yields

$$\ell(t) = \ell(0) + t \langle \text{grad}h_{K^*}, F \rangle_{Y_{K^*}} + \frac{1}{2} \ell''(s) t^2,$$

for some  $s \in (0, t)$ . As  $\text{grad}h_{K^*} = 0$  and  $t = 0$  is a local minimum of  $\ell(t)$ , we must have  $\ell''(s) \geq 0$ ; by tending  $t \rightarrow 0$ , smoothness of  $\ell$  implies that  $\ell''(0) \geq 0$ . But,

$$\ell''(t) = \tilde{D}_t \langle \text{grad}h_{\tilde{\gamma}(t)}, \tilde{\gamma}'(t) \rangle_{Y_{\tilde{\gamma}(t)}} = \left\langle \tilde{D}_t \text{grad}h_{\tilde{\gamma}(t)}, \tilde{\gamma}'(t) \right\rangle_{Y_{\tilde{\gamma}(t)}}$$

where  $\tilde{D}_t$  denotes the covariant derivative along  $\tilde{\gamma}$  on  $\tilde{\mathcal{S}}$ , and the last equality follows by its compatibility with the metric and the fact that  $\tilde{\gamma}$  is a geodesic (so that  $\tilde{D}_t \tilde{\gamma}'(t) \equiv 0$ ). As  $\text{grad}h|_{\tilde{\gamma}(t)} \in \mathfrak{X}(\tilde{\gamma})$  is clearly extendable, we conclude that

$$\ell''(t) = \left\langle \nabla_{\tilde{\gamma}'(t)} \text{grad}h|_{\tilde{\gamma}(t)}, \tilde{\gamma}'(t) \right\rangle_{Y_{\tilde{\gamma}(t)}} = \langle \text{Hess}h_{\tilde{\gamma}(t)}[\tilde{\gamma}'(t)], \tilde{\gamma}'(t) \rangle_{Y_{\tilde{\gamma}(t)}},$$

and thus particularly  $\ell''(0) = \langle \text{Hess}h_{K^*}[F], F \rangle_{Y_{K^*}}$ . Since  $F$  was arbitrary and  $K^*$  is non-degenerate,  $\ell''(0) \geq 0$  implies that  $\text{Hess}h_{K^*}$  is positive definite. Next, existence of a neighborhood at  $K^*$  on which  $\text{Hess}h$  is positive definite follows by smoothness—in particular continuity—of the operator  $\text{Hess}h_K$  in  $K$ . Finally, let  $\tilde{\nabla}$  and  $\overline{\tilde{\nabla}}$  denote the connections on  $T\tilde{\mathcal{S}}$  induced, respectively, by the connections  $\nabla$  and  $\overline{\nabla}$  on  $T\mathcal{S}$ . Then by The Difference Tensor Lemma, the difference tensor between  $\tilde{\nabla}$  and  $\overline{\tilde{\nabla}}$ —defined as  $D(U, V) := \tilde{\nabla}_U V - \overline{\tilde{\nabla}}_U V$  for any  $U, V \in \mathfrak{X}(\tilde{\mathcal{S}})$ —is indeed a (1,2)-tensor field. That means, as  $\text{grad}h_{K^*} = 0$ ,

$$\text{Hess}h_{K^*}[U_{K^*}] - \overline{\text{Hess}h_{K^*}}[U_{K^*}] = D(U, \text{grad}h)|_{K^*} = 0.$$

The last claim then follows as  $U \in \mathfrak{X}(\tilde{\mathcal{S}})$  was arbitrary.  $\square$

2.5.2 *Theoretical Analysis of the Results in §2.4*

**Proof of Theorem 2.11.** By Lemma 2.10,  $\text{grad}h_{K^*} = 0$  and there exists a neighborhood  $\mathcal{U}$  of  $K^*$  on which  $\text{Hess}h_K$  is positive. Furthermore, by continuity of  $\text{Hess}h$  (and, if necessary, shrinking  $\mathcal{U}$ ) we can obtain constant positive scalars  $m$  and  $M$  such that for all  $K \in \mathcal{U}$  and  $G \in T_K\tilde{\mathcal{S}}$ ,

$$m\|G\|_{Y_K}^2 \leq \langle \text{Hess}h_K[G], G \rangle_{Y_K} \leq M\|G\|_{Y_K}^2. \quad (2.10)$$

In particular, if  $G_t \in T_{K_t}\tilde{\mathcal{S}}$  is the Newton direction at some point  $K_t \in \mathcal{U}$ , then (by Cauchy-Schwartz inequality at  $K_t$ ) we must have

$$\|G_t\|_{Y_{K_t}} \leq \frac{1}{m} \|\text{grad}h_{K_t}\|_{Y_{K_t}}. \quad (2.11)$$

Next, define the curve  $\gamma : [0, s_{K_t}] \rightarrow \tilde{\mathcal{S}}$  with  $\gamma(\eta) = K_t + \eta G_t$ , and consider a smooth parallel vector field (with respect to the Riemannian connection)  $E(\eta)$  along  $\gamma$ <sup>8</sup>. Also, define  $\phi : [0, s_{K_t}] \rightarrow \mathbb{R}$  with

$$\phi(\eta) := \langle \text{grad}h_{\gamma(\eta)}, E(\eta) \rangle_{Y_{\gamma(\eta)}}.$$

Notice that  $\text{grad}h$  is smooth, so is  $\phi$  and by compatibility with the metric and the fact that  $\text{grad}h_{\gamma(\eta)}$  is clearly extendable, we have

$$\phi'(\eta) = \langle D_\eta \text{grad}h_{\gamma(\eta)}, E(\eta) \rangle_{Y_{\gamma(\eta)}} = \langle \text{Hess}h_{\gamma(\eta)}[G_t], E(\eta) \rangle_{Y_{\gamma(\eta)}},$$

where  $D_\eta$  is the covariant derivative along  $\gamma$  and  $G_t$  is extended to the vector field along  $\gamma$  with constant coordinates in the global coordinate frame. Thus, as

$$\phi(\eta) = \phi(0) + \eta\phi'(0) + \int_0^\eta [\phi'(\tau) - \phi'(0)]d\tau,$$

by direct substitution and the fact that  $G_t$  is the Newton direction at iteration  $t$ , we obtain that

$$\phi(\eta) = (\eta - 1) \langle \text{Hess}h_{K_t}[G_t], E(0) \rangle_{Y_{K_t}} + \int_0^\eta \langle [\text{Hess}h_{\gamma(\tau)} - \mathcal{P}_{0,\tau}^\gamma \text{Hess}h_{\gamma(0)}]G_t, E(\tau) \rangle_{Y_{\gamma(\tau)}} d\tau,$$

---

<sup>8</sup>We particularly refer to [Lee, 2018] for the idea of “parallel vector fields along curves” and “parallel transport”.

where  $\mathcal{P}_{0,\tau}^\gamma$  denotes the parallel transport from 0 to  $\tau$  along  $\gamma$ . Again, as every parallel transport map along  $\gamma$  is a linear isometry we claim that

$$\begin{aligned} \langle \text{grad}h_{K_{t+1}}, E(\eta_t) \rangle_{Y_{K_{t+1}}} &= (\eta_t - 1) \langle \mathcal{P}_{0,\eta_t}^\gamma \text{Hess}h_{K_t}[G_t], E(\eta_t) \rangle_{Y_{K_{t+1}}} \\ &\quad + \int_0^{\eta_t} \langle \mathcal{P}_{\tau,\eta_t}^\gamma [\text{Hess}h_{\gamma(\tau)} - \mathcal{P}_{0,\tau}^\gamma \text{Hess}h_{\gamma(0)}]G_t, E(\eta_t) \rangle_{Y_{K_{t+1}}} d\tau. \end{aligned}$$

Note that, for each  $\tau \in [0, s_{K_t}]$ ,  $\text{Hess}h_{\gamma(\tau)}$  is a self-adjoint operator that is smooth in  $\tau$  as  $\gamma$  is. So, by (2.10), we obtain

$$\|\mathcal{P}_{0,\eta_t}^\gamma \text{Hess}h_{K_t}[G_t]\|_{Y_{K_{t+1}}} \leq M \|G_t\|_{Y_{K_t}},$$

and by smoothness there exist a constant  $L > 0$  such that

$$\|\mathcal{P}_{\tau,\eta_t}^\gamma [\text{Hess}h_{\gamma(\tau)} - \mathcal{P}_{0,\tau}^\gamma \text{Hess}h_{\gamma(0)}]G_t\|_{Y_{K_{t+1}}} \leq \tau L \|G_t\|_{Y_{K_t}}^2$$

where we used the isometry of parallel transport again in obtaining the bounds. Therefore, by choosing the parallel vector field  $E(\eta)$  along  $\gamma$  such that  $E(\eta_t) = \text{grad}h_{K_{t+1}}$  we obtain that

$$\|\text{grad}h_{K_{t+1}}\|_{Y_{K_{t+1}}} \leq M |1 - \eta_t| \|G_t\|_{Y_{K_t}} + \frac{\eta_t}{2} L \|G_t\|_{Y_{K_t}}^2.$$

Combining this with (2.11) yields

$$\|\text{grad}h_{K_{t+1}}\|_{Y_{K_{t+1}}} \leq \frac{M}{m} |1 - \eta_t| \|\text{grad}h_{K_t}\|_{Y_{K_t}} + \frac{L}{2m^2} \eta_t \|\text{grad}h_{K_t}\|_{Y_{K_t}}^2. \quad (2.12)$$

Next, let  $F_{t+1} \in T_{K_{t+1}} \widetilde{\mathcal{S}}$  be tangent vector that  $\xi(\eta) = \widetilde{\text{exp}}_{K_{t+1}}[\eta F_{t+1}]$  is the minimum-length geodesic in  $\widetilde{\mathcal{S}}$  joining  $\xi(0) = K_{t+1}$  to  $\xi(1) = K^*$ , where  $\widetilde{\text{exp}}$  denotes the exponential map on  $\widetilde{\mathcal{S}}$ . This is certainly possible (by shrinking  $\mathcal{U}$  if necessary) because geodesics are locally-minimizing [Lee, 2018]. Similar to the function  $\phi$ , define  $\psi : [0, 1] \rightarrow \mathbb{R}$  with

$$\psi(\eta) := \langle \text{grad}h_{\xi(\eta)}, E(\eta) \rangle_{Y_{\xi(\eta)}},$$

for some parallel vector  $E(\eta)$  along  $\xi$ . Then, similarly

$$\psi'(\eta) = \langle D_\eta \text{grad}h_{\xi(\eta)}, E(\eta) \rangle_{Y_{\xi(\eta)}} = \langle \text{Hess}h_{\xi(\eta)}[\xi'(\eta)], E(\eta) \rangle_{Y_{\xi(\eta)}}.$$

The velocity of any geodesic is a parallel vector field along itself, so by choosing  $E(\eta) = \xi'(\eta)$  and using the fundamental lemma of calculus for  $\psi$  we obtain that

$$\psi(1) = \langle \text{grad}h_{K_{t+1}}, F_{t+1} \rangle_{Y_{K_{t+1}}} + \int_0^1 \langle \text{Hess}h_{\xi(\tau)}[\xi'(\tau)], \xi'(\tau) \rangle_{Y_{\xi(\tau)}} d\tau.$$

Note that  $\psi(1) = 0$  and  $\|\xi'(\tau)\|_{Y_{\xi(\tau)}} = \|F_{t+1}\|_{Y_{K_{t+1}}}$  for all  $\tau$  as  $\xi$  is a geodesic. Thus, by using (2.10), we conclude that

$$m\|F_{t+1}\|_{Y_{K_{t+1}}} \leq \|\text{grad}h_{K_{t+1}}\|_{Y_{K_{t+1}}} \leq M\|F_{t+1}\|_{Y_{K_{t+1}}}. \quad (2.13)$$

Finally, combining (2.12) and (2.13) at two iterations  $t+1$  and  $t$ , and noticing  $\text{dist}(K_{t+1}, K^*) = \|F_{t+1}\|_{Y_{K_{t+1}}}$  imply that

$$\text{dist}(K_{t+1}, K^*) \leq |1 - \eta_t| \frac{M^2}{m^2} \text{dist}(K_t, K^*) + \eta_t \frac{LM^2}{2m^3} \text{dist}(K_t, K^*)^2, \quad (2.14)$$

where  $\text{dist}(\cdot, \cdot)$  denotes the Riemannian distance function between two points. Next, note that the mapping  $K \rightarrow Q_K$  is chosen to be smooth such that  $Q_K \succ 0$ , therefore as a result of Lemma 2.1, the mapping  $K \rightarrow \mathbb{L}(A_K^\top, Q_K)$  is smooth by composition. By smoothness (in particular continuity) of this mapping and the continuity of the maximum eigenvalue (utilized in the definition of stability certificate  $s_K$  in Lemma 2.8), we can shrink  $\mathcal{U}$ —if necessary—to obtain a positive constant  $c > 0$  such that

$$s_{K_t} \geq \frac{c}{\|G_t\|_{Y_{K_t}}} \geq \frac{cm/M}{\text{dist}(K_t, K^*)}, \quad (2.15)$$

where the last inequality follows by combining (2.11), (2.13) and the fact that  $\text{dist}(K_t, K^*) = \|F_t\|_{Y_{K_t}}$ . Now, pick  $r \in (0, 1)$ ; if we set  $\mathcal{U}^* \subset \mathcal{U} \subset \tilde{\mathcal{S}}$  such that for any  $K_0 \in \mathcal{U}^*$  we have

$$\text{dist}(K_0, K^*) < \min\left\{\frac{cm}{M(1-r/2)}, \frac{m^3r}{LM^2}\right\}$$

then by the choice of stepsize  $\eta_t = \min\{s_{K_t}, 1\}$  and the lower-bound in (2.15), we can claim that

$$|1 - \eta_0| \frac{M^2}{m^2} + \eta_0 \frac{LM^2}{2m^3} \text{dist}(K_0, K^*) < r.$$

But then, (2.14) implies that

$$\text{dist}(K_1, K^*) \leq r \text{dist}(K_0, K^*).$$

Therefore,  $K_1 \in \mathcal{U}^*$  as  $r < 1$ , and thus by induction we conclude a linear convergence rate to  $K^*$ . Consequently, (2.15) implies that  $s_{K_t} \geq 1$  for large enough  $t$ , and thus by the choice of step-size (2.14) simplifies to

$$\text{dist}(K_{t+1}, K^*) \leq \frac{LM^2}{2m^3} \text{dist}(K_t, K^*)^2,$$

guaranteeing a quadratic convergence rate. Finally, Lemma 2.10 implies that a critical point is nondegenerate with respect to the induced Riemannian connection on  $T\tilde{\mathcal{S}}$  if and only if it is so with respect to the Euclidean one. The proof for RNPO with  $\overline{\text{Hess}}$  then follows similarly by redefining  $\phi$  and  $\psi$  using the Euclidean metric under the usual identification of the tangent bundle, which is omitted for brevity.  $\square$

## 2.6 Applications of RNPO to Distributed and Output LQR Problems

In this section we provide a couple of applications of our methodology for optimizing the LQR cost over two different submanifold induces by Structured LQR (SLQR) and Output-feedback LQR (OLQR) problems.

### 2.6.1 Background

Consider a discrete-time linear time-invariant system dynamics

$$\begin{aligned} \mathbf{x}_{k+1} &= A\mathbf{x}_k + B\mathbf{u}_k, \\ \mathbf{y}_k &= C\mathbf{x}_k \end{aligned} \tag{2.16}$$

where  $A \in M(n \times n, \mathbb{R})$ ,  $B \in M(n \times m, \mathbb{R})$  and  $C \in M(d \times n, \mathbb{R})$  are the system parameters for some integers  $n$ ,  $m$  and  $d$ ;  $\mathbf{x}_k$  and  $\mathbf{u}_k$  denote the states vector and inputs vector, respectively, and  $\mathbf{x}_0$  is given. Conventionally, the Linear Quadratic Regulators (LQR) problem is

to design a sequence of input signals  $\mathbf{u} = (\mathbf{u}_k)_0^\infty \in \ell_2$  that minimizes the following quadratic cost

$$J_{\mathbf{x}_0}(\mathbf{u}) = \frac{1}{2} \sum_{k=0}^{\infty} \mathbf{x}_k^\top Q \mathbf{x}_k + \mathbf{u}_k^\top R \mathbf{u}_k, \quad (2.17)$$

subject to the dynamics in (2.16), where  $Q = Q^\top \succcurlyeq 0$  and  $R = R^\top \succ 0$  are prescribed cost parameters. Using Dynamic Programming or Calculus of Variations, it is well known<sup>9</sup> that the optimal state-feedback solution  $\mathbf{u}^*$  to this problem reduces to solving the Discrete-time Algebraic Riccati Equation (DARE) for the optimal cost matrix  $P_{\text{LQR}}$ . This results in a linear state-feedback optimal control  $\mathbf{u}_k^* = K_{\text{LQR}} \mathbf{x}_k$  where  $K_{\text{LQR}} \in \mathbb{M}(m \times n, \mathbb{R})$  is the optimal LQR gain (policy) obtaining from  $P_{\text{LQR}}$ . Also, the associated optimal cost can be obtained as  $J_{\mathbf{x}_0}(\mathbf{u}^*) = \frac{1}{2} \mathbf{x}_0^\top P_{\text{LQR}} \mathbf{x}_0$ .

Naturally, one could think of the LQR cost as a map  $K \mapsto J_{\mathbf{x}_0}(\mathbf{u} = K\mathbf{x})$ , however, this would depend on  $\mathbf{x}_0$  and generally, its value can be still finite while  $K$  is not necessarily stabilizing (i.e. when  $K \notin \mathcal{S}$  is not feasible). Instead, in order to avoid the dependency on the initial state while considering the constraints on the linear feedback control policy directly, we pose the following constrained optimization problem

$$\begin{aligned} \min_K f(K) &:= \mathbb{E}_{\mathbf{x}_0 \sim \mathcal{D}} J_{\mathbf{x}_0}(\mathbf{u}) & (2.18) \\ \text{s.t. } \mathbf{x}_{k+1} &= A\mathbf{x}_k + B\mathbf{u}_k, \quad \forall k \geq 0, \\ \mathbf{u} &= K\mathbf{x}, \\ K &\in \tilde{\mathcal{S}}, \end{aligned}$$

where  $\tilde{\mathcal{S}}$  is an embedded submanifold of  $\mathcal{S}$ , and  $\mathcal{D}$  denotes a distribution of zero-mean multivariate random variables of dimension  $n$  with covariance matrix  $\Sigma_1$  so that  $0 \prec \Sigma_1 = \Sigma_1^\top \in \mathbb{M}(n \times n, \mathbb{R})$ .

Next, we can reformulate the problem in (2.18) as follows. For each stabilizing controller

---

<sup>9</sup>See, e.g., Section 22.7 in [Goodwin et al., 2001].

$K \in \mathcal{S}$ , from (2.16) and (2.17) we have that

$$J_{\mathbf{x}_0}(\mathbf{u} = K\mathbf{x}) = \frac{1}{2} \sum_{k=0}^{\infty} \mathbf{x}_0^\top (A_K^k)^\top [Q + K^\top RK] A_K^k \mathbf{x}_0,$$

where  $A_K := A + BK$ . Since,  $A_K$  is a stability matrix, the sum  $\sum_{k=0}^{\infty} (A_K^k)^\top [Q + K^\top RK] A_K^k$  converges, which is equal to the unique solution of the following Lyapunov equation:

$$P_K := \mathbb{L}(A_K^\top, K^\top RK + Q).$$

Therefore,  $f(K)$  can be calculated as

$$f(K) = \frac{1}{2} \mathbb{E}_{\mathbf{x}_0 \sim \mathcal{D}} \text{tr} [P_K \mathbf{x}_0 \mathbf{x}_0^\top] = \frac{1}{2} \text{tr} [P_K \Sigma_1].$$

Thus, the problem in (2.18) reduces to

$$\begin{aligned} \min_K f(K) &= \frac{1}{2} \text{tr} [P_K \Sigma_1] \\ \text{s.t.} \quad &K \in \tilde{\mathcal{S}}. \end{aligned} \tag{2.19}$$

This reformulation of the LQR cost function has been adopted before (see e.g., [Fazel et al., 2018; Mårtensson, 2012; Bu et al., 2019a]) but the inherent geometry of the submanifold  $\tilde{\mathcal{S}}$  has been overlooked. If there is no constraint on the controller, i.e.  $\tilde{\mathcal{S}} = \mathcal{S}$ , then a well-known quasi-Newton algorithm—due to Hewer—is known to converge to the optimal state-feedback control at a quadratic rate [Hewer, 1971]. Otherwise,  $\tilde{\mathcal{S}}$  may have disconnected components, and in general, the constrained cost function may introduce stationary points that are not local minima. Nonetheless, in this section, we apply the techniques developed in §2.3 and Algorithm 2.1 to the constraint arising in the well-known SLQR and OLQR problems. Note that both of these problems can be cast as an optimization in (2.19) with  $\tilde{\mathcal{S}}$  denoting a specific submanifold of  $\mathcal{S}$  that will be detailed in Section 2.6.4 and Section 2.6.5, respectively.

### 2.6.2 Solving for the Newton direction

In order to solve for the Newton direction at any  $K \in \tilde{\mathcal{S}}$ , suppose the tuple  $(\tilde{\partial}_{(p,q)}|_{(p,q) \in D})$  denotes a smooth local frame for  $\tilde{\mathcal{S}}$  on a neighborhood of  $K$  where  $D$  is a subset of  $[m] \times [n]$



depending on the dimension of  $\tilde{\mathcal{S}}$ .<sup>10</sup> In fact, by Proposition 2.7, for any  $G = [G]^{k,\ell} \tilde{\partial}_{(k,\ell)}|_K \in T_K \tilde{\mathcal{S}}$  (interpreted as a subspace of  $T_K \mathcal{S}$ ), finding Newton direction on  $\tilde{\mathcal{S}}$  reduces to solving the following system of  $|D|$ -linear equations (for each index  $(p, q) \in D$ )

$$\sum_{(k,\ell) \in D} [G]^{k,\ell} h_{;(k,\ell)(p,q)}(K) = - \left\langle \pi^\top (\text{grad} f|_K), \tilde{\partial}_{(p,q)}|_K \right\rangle_{Y_K},$$

where  $h_{;(k,\ell)(p,q)}$  denote the coordinate functions of  $\tilde{\nabla}^2 h$  with respect to the local coframe dual to  $(\tilde{\partial}_{(p,q)}|_{(p,q) \in D})$ . Thus, by (2.7),  $h_{;(k,\ell)(p,q)}(K)$  can be computed as follows

$$h_{;(k,\ell)(p,q)}(K) = \left\langle \text{Hess} h_K [\tilde{\partial}_{(k,\ell)}|_K], \tilde{\partial}_{(p,q)}|_K \right\rangle_{Y_K};$$

or with  $\text{Hess} h$  replaced by  $\overline{\text{Hess} h}$ , depending on the choice of connection.

### 2.6.3 Analysis of the special cost function

Now we turn our attention towards the analysis of the following cost function over the Riemannian manifolds  $(\mathcal{S}, g)$  and  $(\tilde{\mathcal{S}}, \tilde{g})$ . To particularize the results obtained so far to this special case, we set  $\Sigma_2 = 0$  in the definition of  $\Sigma_K$  in (2.4).

**Proposition 2.13.** *Consider the Riemannian manifold  $(\mathcal{S}, g, \nabla)$  and the cost function  $f : \mathcal{S} \rightarrow \mathbb{R}$  defined as  $f(K) = \frac{1}{2} \text{tr} [P_K \Sigma_1]$  where  $P_K = \mathbb{L}(A_K^\top, K^\top R K + Q)$ . Then,  $f$  is smooth and under the usual identification of the tangent bundle*

$$\text{grad} f_K = R K + B^\top P_K A_K.$$

Furthermore,  $\text{Hess} f$  and  $\overline{\text{Hess} f}$  are both self-adjoint operators which can be characterized as follows: for any  $E, F \in T_K \mathcal{S}$ ,

$$\begin{aligned} \langle \text{Hess} f_K [E], F \rangle_{Y_K} &= \langle B^\top (S_K|_F) A_K, E \rangle_{Y_K} + \langle (R + B^\top P_K B) E + B^\top (S_K|_E) A_K, F \rangle_{Y_K} \\ &\quad - \left\langle \text{grad} f_K, [E]^{k,\ell} [F]^{p,q} \Gamma_{(k,\ell)(p,q)}^{i,j}(K) \partial_{i,j} \right\rangle_{Y_K}, \end{aligned}$$

---

<sup>10</sup>As  $\tilde{\mathcal{S}}$  is an embedded submanifold of  $\mathcal{S}$ , each  $\tilde{\partial}_{(p,q)}|_K$  can be also interpreted as an element of  $T_K \mathcal{S}$ .

and

$$\langle \overline{\text{Hess}f}[E], F \rangle_{Y_K} = \langle B^\top(S_K|_F)A_K, E \rangle_{Y_K} + \langle (R + B^\top P_K B)E + B^\top(S_K|_E)A_K, F \rangle_{Y_K},$$

with  $\Gamma_{(k,\ell)(p,q)}^{i,j}$  denoting the Christoffel symbols of  $g$  and

$$S_K|_E := \mathbb{L}(A_K^\top, E^\top \text{grad}f_K + (\text{grad}f_K)^\top E).$$

The next corollary provides a similar result for the LQR cost restricted to an embedded submanifold of  $\mathcal{S}$ , which is a consequence of Proposition 2.7 and Proposition 2.13.

**Corollary 2.14.** *Under the premise of Proposition 2.13, define  $h = f|_{\tilde{\mathcal{S}}}$  where  $\tilde{\mathcal{S}} \subset \mathcal{S}$  is an embedded Riemannian submanifold with the induced metric and connection. Then,  $h$  is smooth and under the usual identification of tangent bundle*

$$\text{grad}h_K = \pi^\top (RK + B^\top P_K A_K).$$

Furthermore,  $\text{Hess}h$  is a self-adjoint operator and can be characterized as follows: for any  $E, F \in T_K \tilde{\mathcal{S}} \subset T_K \mathcal{S}$ ,

$$\begin{aligned} \langle \text{Hess}h_K[E], F \rangle_{Y_K} &= \langle B^\top(S_K|_F)A_K, E \rangle_{Y_K} + \langle (R + B^\top P_K B)E + B^\top(S_K|_E)A_K, F \rangle_{Y_K} \\ &\quad - \left\langle \text{grad}h_K, [E]^{k,\ell} [F]^{p,q} \Gamma_{(k,\ell)(p,q)}^{i,j}(K) \partial_{i,j} \right\rangle_{Y_K}, \end{aligned}$$

where  $\Gamma_{(k,\ell)(p,q)}^{i,j}$  denotes the Christoffel symbols associated with  $g$  and

$$S_K|_E := \mathbb{L}(A_K^\top, E^\top \text{grad}f_K + (\text{grad}f_K)^\top E).$$

*Remark 2.15.* If  $Q \succ 0$ , then we can choose the mapping  $K \rightarrow Q_K$  to be  $Q_K = Q + K^\top R K$ , and thus the stability certificate  $s_K$  as defined in Lemma 2.8 must satisfy

$$s_K \geq \frac{\lambda_{\min}(Q) \lambda_{\min}(\Sigma_1)}{4f(K) \|BG\|_2},$$

where  $f$  denotes the LQR cost. This is because  $R, Q, \Sigma_1 \succ 0$  and so is  $P_K \succ 0$ , and thus by the trace inequality

$$f(K) \geq \frac{1}{2} \lambda_{\min}(\Sigma_1) \lambda_{\max}(P_K).$$

The claimed lower-bound on stability certificate then follows by combining the last inequality with the definition of stability certificate. Otherwise if  $Q \succeq 0$ , one can leverage on the observability of the pair  $(A, Q^{1/2})$  to get similar results.

#### 2.6.4 State-feedback SLQR problem

Any required sparsity pattern on the controller gain  $K$  imposes a linear constraint set, denoted by  $\mathcal{K}_D$ , which indicates a linear subspace of  $M(m \times n, \mathbb{R})$  with nonzero entries only for a prescribed subset  $D$  of entries, i.e., for any  $K \in \mathcal{K}_D$  and  $(i, j) \notin D$  we must have  $[K]_{i,j} = 0$ . Let the tuple  $(x^{(i,j)})_{(i,j) \in [m] \times [n]}$  denote the component functions of the global smooth chart  $(M(m \times n, \mathbb{R}), \text{vec})$ , and define  $\Phi : M(m \times n, \mathbb{R}) \rightarrow \mathbb{R}^{mn-|D|}$  with

$$\Phi(K) = \sum_{(i,j) \notin D} [K]_{i,j} x^{(i,j)}.$$

Then, it is easy to see that  $\Phi$  is a smooth submersion, and so is  $\Phi|_{\mathcal{S}}$  because  $\mathcal{S}$  is an open submanifold of  $M(m \times n, \mathbb{R})$ . Therefore, as  $\tilde{\mathcal{S}} = \mathcal{S} \cap \mathcal{K}_D = (\Phi|_{\mathcal{S}})^{-1}(0)$ , by Submersion Level Set Theorem we conclude that  $\tilde{\mathcal{S}}$  is a properly embedded submanifold of dimension  $|D|$ .

Furthermore, at any point  $K \in \tilde{\mathcal{S}}$  and for any tangent vector  $E \in T_K \mathcal{S}$ , we can compute the tangential projection  $\pi^\top : T_K \mathcal{S} \rightarrow T_K \tilde{\mathcal{S}}$  as follows:

$$\pi^\top E = \arg \min_{\tilde{E} \in T_K \tilde{\mathcal{S}}} \left\langle E - \tilde{E}, E - \tilde{E} \right\rangle_{Y_K}. \quad (2.20)$$

As  $\mathcal{K}_D$  is a linear subspace of  $M(m \times n, \mathbb{R})$ , we can identify  $T_K \tilde{\mathcal{S}}$  with  $\mathcal{K}_D$  itself (due to dimensional reasons). Then, it is not hard to show that the unique solution  $\tilde{E}^*$  to the minimization above (with linear constraint and strongly convex cost function, as  $Y_K \succ 0$ ), must satisfy  $E - \tilde{E}^* \perp \mathcal{K}_D$  with respect to the Riemannian metric at  $K$ ; or equivalently,

$$\text{Proj}_{\mathcal{K}_D} \left[ (E - \tilde{E}^*) Y_K \right] = 0,$$

where  $\text{Proj}_{\mathcal{K}_D}$  denotes the Euclidean projection onto the sparsity pattern  $\mathcal{K}_D$ . Note that at each  $K \in \tilde{\mathcal{S}}$ , the last equality consists of  $|D|$  nontrivial linear equations involving  $|D|$  unknowns (as the nonzero entries of  $\tilde{E}^*$ ), which can be solved efficiently.

Finally, if  $\tilde{\partial}_{(i,j)}$  (as described in section 2.6.2) is taken to be  $\tilde{\partial}_{(i,j)} = \partial_{(i,j)}$  for  $(i,j) \in D$ , then  $(\tilde{\partial}_{(i,j)}|_{(i,j) \in D})$  forms a global smooth frame for  $T\tilde{\mathcal{S}}$ . And thus, for each  $(k,\ell), (p,q) \in D$ , the coordinates  $h_{;(k,\ell)(p,q)}(K)$  simplifies to

$$h_{;(k,\ell)(p,q)}(K) = \left\langle (R + B^\top P_K B) \partial_{(k,\ell)} + B^\top (S_K|_{\partial_{(k,\ell)}}) A_K, \partial_{(p,q)} \right\rangle_{Y_K} \\ + \left\langle B^\top (S_K|_{\partial_{(p,q)}}) A_K, \partial_{(k,\ell)} \right\rangle_{Y_K} - \left\langle \pi^\top \text{grad} f_K, \Gamma_{(k,\ell)(p,q)}^{i,j}(K) \partial_{i,j} \right\rangle_{Y_K}.$$

### 2.6.5 Output-feedback LQR (OLQR) problem

The OLQR problem can be formulated as the optimization problem in (2.18) with the submanifold  $\tilde{\mathcal{S}} = \mathcal{S} \cap \mathcal{K}_C$  with the constraint set  $\mathcal{K}_C$  defined as

$$\mathcal{K}_C := \{K \in M(m \times n, \mathbb{R}) \mid K = LC, L \in M(m \times d, \mathbb{R})\},$$

where  $C \in M(d \times n, \mathbb{R})$  is the prescribed output matrix. Note that  $\mathcal{K}_C$  is a linear subspace of  $M(m \times n, \mathbb{R})$  whose dimension depends on the rank of  $C$ . For simplicity of presentation, we suppose  $C$  has full rank equal to  $d \leq n$ . Now, define  $\Psi : M(m \times n, \mathbb{R}) \rightarrow M(m \times n, \mathbb{R})$  as follows:

$$\Psi(K) = K(I_n - C^\dagger C),$$

where  $\dagger$  denotes the Moore–Penrose inverse. Note that  $\Psi$  is a linear map that is surjective onto its range, denoted by  $\mathcal{R}$ , which is a  $m(n-d)$  dimensional linear subspace of  $M(m \times n, \mathbb{R})$ . Therefore,  $\Phi : \mathcal{S} \rightarrow \mathcal{R}$  defined as the restriction of  $\Psi$  both in domain and codomain is a smooth submersion. Finally, as  $CC^\dagger C = C$ , we can observe that  $\text{Ker}(\Psi) = \mathcal{K}_C$ . Therefore,  $\tilde{\mathcal{S}} = \mathcal{K}_C \cap \mathcal{S} = \Phi^{-1}(0)$  and thus, by Submersion Level Set Theorem,  $\tilde{\mathcal{S}}$  is a properly embedded submanifold of  $\mathcal{S}$  with dimension  $md$ . We also conclude that at each  $K \in \tilde{\mathcal{S}}$ , we can canonically identify the tangent space at  $K$  as follows:

$$T_K \tilde{\mathcal{S}} = \text{Ker}(d\Phi_K) \cong \mathcal{K}_C.$$

Next, at any  $K \in \tilde{\mathcal{S}}$  and for any  $E \in T_K \mathcal{S}$ , the tangential projection of  $E$ , denoted by  $\tilde{E}^* = \pi^\top E$ , is the unique solution of a minimization similar to (2.20). But, under the above

identification, it must satisfy  $E - \tilde{E}^* \perp \mathcal{K}_C$  (with respect to the Riemannian metric), or equivalently,

$$\text{tr} \left[ C^\top L^\top (E - \tilde{E}^*) Y_K \right] = 0, \quad \forall L \in \text{M}(m \times d, \mathbb{R}).$$

Here,  $Y_K = \mathbb{L}(A_K, \Sigma_1)$  is positive definite and since  $C$  is assumed to be full-rank,  $CY_K C^\top$  is positive definite. So, we conclude that  $\pi^\top E = L^* C$  with  $L^* \in \text{M}(m \times d, \mathbb{R})$  being the unique solution of the following linear equation

$$L^* C Y_K C^\top = E Y_K C^\top.$$

Finally, at each point  $K \in \tilde{\mathcal{S}}$ , we denote the global coordinate functions of  $\text{M}(m \times d, \mathbb{R})$  by the tuple  $(\bar{x}^{i,j})$  for  $(i, j) \in D := [m] \times [d]$  and its corresponding global coordinate frame by  $(\bar{\partial}_{(i,j)})$ . Under the identification of  $T_K \tilde{\mathcal{S}} \cong \mathcal{K}_C$  explained above, we claim that the choice of  $(\tilde{\partial}_{(i,j)} = \bar{\partial}_{(i,j)} C)$  for  $(i, j) \in D$  forms a global smooth frame for  $T\tilde{\mathcal{S}}$  because they are linearly independent global vector fields on  $\tilde{\mathcal{S}}$  as  $C$  has full-rank. But then, the coordinates of the covariant Hessian  $h_{;(k,\ell)(p,q)}(K)$  with respect to this frame can be computed by substituting  $E = \bar{\partial}_{(k,\ell)} C$  and  $F = \bar{\partial}_{(p,q)} C$  in Corollary 2.14 for each  $(k, \ell), (p, q) \in D$ —similar to the SLQR case. It is worth noting that the sparsity pattern in  $E, F$  and Christoffel symbols can simplify the computations which is omitted here for brevity.

### 2.6.6 Theoretical Analysis of the Results in §2.6

**Proof of Proposition 2.13.** By definition,  $f : \mathcal{S} \rightarrow \mathbb{R}$  can be viewed as composition of the following maps:

$$f : K \xrightarrow{\Phi} (A_K^\top, K^\top R K + Q) \xrightarrow{\mathbb{L}} P_K \xrightarrow{\Psi} \frac{1}{2} \text{tr} [P_K \Sigma_1]. \quad (2.21)$$

Since the first and last map are smooth ( i.e. linear or quadratic in  $K$ ), we conclude that  $f \in C^\infty(\mathcal{S})$  by composition and Lemma 2.1. For any  $K \in \mathcal{S}$ , we can compute its differential at  $K$ , denoted by  $df_K$ , using the chain rule as

$$df_K(E) = d\Psi_{P_K} \circ d\mathbb{L}_{(A_K^\top, K^\top R K + Q)} \circ d\Phi_K(E),$$

for any  $E \in T_K \mathcal{S}$ . But  $\Psi$  is a linear map, and under the usual identification of the tangent bundle we obtain

$$d\Phi_K(E) = (E^\top B^\top, E^\top RK + K^\top RE).$$

Therefore, by Lemma 2.1 we claim that

$$d(\mathbb{L} \circ \Phi)_K(E) = \mathbb{L}(A_K^\top, E^\top(B^\top P_K A_K + RK) + (K^\top R + A_K^\top P_K B)E), \quad (2.22)$$

and thus

$$df_K(E) = \Psi \circ \mathbb{L}(A_K^\top, E^\top(B^\top P_K A_K + RK) + (K^\top R + A_K^\top P_K B)E). \quad (2.23)$$

But, by Lyapunov-trace property we get that

$$df_K(E) = \langle E, RK + B^\top P_K A_K \rangle_{Y_K},$$

with  $Y_K = \mathbb{L}(A_K, \Sigma_1)$  which is well-defined and unique as  $A_K$  is a stability matrix. As  $df_K(E) = Ef$ , the expression for  $\text{grad} f \in \mathfrak{X}(\mathcal{S})$  then follows by its definition. Next, as the Hessian operator is self-adjoint (see section 2.3.4), in order to obtain  $\text{Hess} f$  we can compute (2.6) for any  $U, W \in \mathfrak{X}(\mathcal{S})$ . As  $\text{Hess} f[U]|_K$  only depends on the value of  $U$  at  $K$ , it suffices to obtain  $\text{Hess} f_K[U_K]$  at each  $K \in \mathcal{S}$  with  $U_K = E$  for arbitrary  $E \in T_K \mathcal{S}$ . To do so, we compute  $\langle \text{Hess} f_K[E], F \rangle_Y$  for an arbitrary vector  $F \in T_K \mathcal{S}$  by extending  $F$  to the vector field  $W$  along the curve  $\gamma : t \rightarrow K + tE$  with constant coordinates with respect to the global coordinate frame  $(\partial_{(i,j)})$ . As  $\langle \text{Hess} f[U], W \rangle_Y|_K$  only depends on the value of  $W_K = F$  and  $U_K = E$ , it doesn't matter how we have extended these vector fields. By properties of the Riemannian connection and the fact that  $W$  can be extended with constant coordinates, here we can compute  $\nabla_U W|_K$  in the global coordinate frame  $(\partial_{i,j})$  and obtain that

$$\nabla_U W|_K = [E]^{k,\ell} [F]^{p,q} \Gamma_{(k,\ell)(p,q)}^{i,j}(K) \partial_{i,j}, \quad (2.24)$$

where  $\Gamma_{(k,\ell)(p,q)}^{i,j}(K)$  denotes the Christoffel symbols associated with the Riemannian metric  $g$  at the point  $K \in \mathcal{S}$ . Therefore, from (2.6) we have that

$$\langle \text{Hess} f[U], W \rangle_Y|_K = Er - \langle \text{grad} f_K, \nabla_U W|_K \rangle_{Y_K}, \quad (2.25)$$

where  $r := \langle \text{grad} f, W \rangle_Y \in C^\infty(\mathcal{S})$ . By the expression obtained for  $\text{grad} f$  and the fact that  $W$  has constant coordinates, the mapping  $K \rightarrow r(K)$  can be decomposed as:

$$K \xrightarrow{\text{Id} \times \text{Id}} (K, K) \xrightarrow{\text{Id} \times (\mathbb{L} \circ \Phi)} (K, P_K) \xrightarrow{\Xi} (A_K^\top, (\text{grad} f_K)^\top F + F^\top \text{grad} f_K) \xrightarrow{\Psi \circ \mathbb{L}} r(K),$$

where we used the Lyapunov-trace property and invariance of trace under transpose to justify the last mapping. Also note that  $\Phi$  and  $\Psi$  are defined in (2.21) and  $\Xi : M(m \times n, \mathbb{R}) \times M(n \times n, \mathbb{R}) \rightarrow M(n \times n, \mathbb{R}) \times M(n \times n, \mathbb{R})$  is defined as above. Therefore, under the usual identification of tangent bundle, for any  $(E, G) \in T_{(K, P_K)}(M(m \times n, \mathbb{R}) \times M(n \times n, \mathbb{R}))$  we can compute that

$$d\Xi_{(K, P_K)}[E, G] = \left( E^\top B^\top, E^\top (R + B^\top P_K B) F + A_K^\top G B F + F^\top (R + B^\top P_K B) E + F^\top B^\top G^\top A_K \right). \quad (2.26)$$

Therefore, by the chain rule, for any  $E \in T_K \mathcal{S}$  we have

$$dr_K(E) = \Psi \circ d\mathbb{L} \circ d\Xi \circ (E, d(\mathbb{L} \circ \Phi)_K(E)),$$

where the base points of differentials are understood and dropped for brevity. But then, by (2.22) and (2.26) we get that

$$\begin{aligned} d\Xi[E, d(\mathbb{L} \circ \Phi)_K(E)] &= \left( E^\top B^\top, E^\top (R + B^\top P_K B) F \right. \\ &\quad \left. + A_K^\top (S_K|_E) B F + F^\top (R + B^\top P_K B) E + F^\top B^\top (S_K|_E) A_K \right), \end{aligned}$$

where  $S_K|_E$  is defined in the premise. Therefore,

$$\begin{aligned} dr_K(E) &= \Psi \circ \mathbb{L} \left( A_K^\top, E^\top (R + B^\top P_K B) F \right. \\ &\quad \left. + A_K^\top (S_K|_E) B F + A_K^\top (S_K|_F) B E + E^\top B^\top (S_K|_F) A_K \right), \end{aligned}$$

and by Lyapunov-trace property we can simplify it as follows

$$\begin{aligned} dr_K(E) &= \frac{1}{2} \text{tr} [E^\top B^\top (S_K|_F) A_K Y_K + A_K^\top (S_K|_F) B E Y_K] \\ &\quad + \frac{1}{2} \text{tr} [(E^\top (R + B^\top P_K B) + A_K^\top (S_K|_E) B) F Y_K] \\ &\quad + \frac{1}{2} \text{tr} [F^\top ((R + B^\top P_K B) E + B^\top (S_K|_E) A_K) Y_K], \end{aligned}$$

where  $Y_K = \mathbb{L}(A_K, \Sigma_1)$ . Noting that  $Y_K$ ,  $P_K$ ,  $S_K|_E$  and  $S_K|_F$  are all symmetric, using the cyclic permutation property of trace, we get that

$$dr_K(E) = \langle (R + B^\top P_K B)E + B^\top (S_K|_E)A_K, F \rangle_{Y_K} + \langle B^\top (S_K|_F)A_K, E \rangle_{Y_K}. \quad (2.27)$$

Then, the expression for  $\text{Hess}f$  follows by substituting (2.27) and (2.24) in (2.25). Finally, the expression of  $\overline{\text{Hess}f}$  can be obtained similarly by threading through the definitions.  $\square$

**Proof of Corollary 2.14.** Smoothness of  $h$  and the expression of its gradient follows immediately by Proposition 2.7 and Proposition 2.13. In order to compute  $\text{Hess}h_K$ , we can combine its extrinsic representation as obtained in Proposition 2.7 with (2.25), and use the definition of Weingarten map to obtain that

$$\begin{aligned} \langle \text{Hess}h_K[E], F \rangle_{Y_K} &= \langle \pi^\top (\text{Hess}f[U]|_{\tilde{\mathcal{S}}}), W \rangle_Y |_{Y_K} + \left\langle \mathbb{W}_{\pi^\perp(\text{grad}f|_{\tilde{\mathcal{S}})}(U)}, W \right\rangle_Y |_{Y_K} \\ &= Er - \langle \text{grad}f_K, \nabla_U W|_K \rangle_{Y_K} + \langle \pi^\perp \text{grad}f_K, \pi^\perp \nabla_U W|_K \rangle_{Y_K} \\ &= Er - \langle \pi^\top \text{grad}f_K, \nabla_U W|_K \rangle_{Y_K}, \end{aligned}$$

for any  $E, F \in T_K \tilde{\mathcal{S}} \subset \mathcal{S}$ , which are extended to vector fields on a neighborhood in  $\mathcal{S}$  with constant coordinates with respect to the global coordinate frame. The claimed expression of  $\text{Hess}h_K$  then follows by substituting (2.24) and (2.27) into the last expression.  $\square$

## 2.7 Numerical Simulations

In this section, we provide numerical examples for optimizing the LQR cost over submanifolds induced by SLQR and OLQR problems. Recall that, for each of these problems, we can compute the coordinate functions of the covariant Hessian  $h_{;(k,\ell)(p,q)}(K)$  with respect to the corresponding coordinate frame described in the previous subsections. Therefore, finding the Newton direction  $G$  at any point  $K \in \tilde{\mathcal{S}}$  reduces to solving the system of linear equations for the unknowns  $[G]^{k,\ell}$ , as described in section 2.6.2, and forming the Newton direction as  $G = [G]^{k,\ell} \tilde{\partial}_{(k,\ell)}|_K \in T_K \tilde{\mathcal{S}}$ .

For each of SLQR and OLQR problems, we have simulated three different algorithms, the first two are the variants of Algorithm 2.1 where we use Riemannian connection or



Euclidean connection to compute  $\text{Hess}h$  or  $\overline{\text{Hess}}h$ , respectively. Note, despite the fact that  $\text{Hess}h_{K^*} = \overline{\text{Hess}}h_{K^*}$  whenever  $\text{grad}h_{K^*} = 0$  (as shown in Lemma 2.10), it is not necessarily the case where  $\text{grad}h$  does not vanish; therefore, we expect  $\text{Hess}h$  and  $\overline{\text{Hess}}h$  to contain very different information on neighborhoods of isolated local minima which directly influence the performance of RNPO as will be discussed below. The third one is the Projected Gradient (PG) algorithm as studied in [Bu et al., 2019a]. Here, the step size for the PG algorithm is a constant value that guarantees the iterates stay stabilizing as suggested therein. However, as the convergence rate of the PG is guaranteed to be sublinear, the improvement of error from optimality is not relatively significant.

**Example 2.16** (Convergence of RNPO versus PG). Assume the following parameters

$$(A|B|Q|\Sigma) = \left( \begin{array}{ccc|cc|ccc|ccc} 0.8 & 1 & 0 & 0 & 1 & 10 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0.6 & 1 & 0 & 1 & 0 & 5 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0.1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 10 \end{array} \right),$$

$R = I_2$  and  $C = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ . Also, consider a prescribed sparsity pattern as a hard constraint for the SLQR problem that is described by  $D := \{(1,1), (1,3), (2,2)\}$  as specified in Section 2.6.4. The initial controller for SLQR and OLQR problems are chosen as  $K_0 = \begin{pmatrix} -0.2 & 0.0 & 0.2 \\ 0.0 & -0.5 & 0.0 \end{pmatrix}$  and  $K_0 = \begin{pmatrix} 0 & 0 \\ 0 & -0.6 \end{pmatrix} C$ , respectively, which are points in the corresponding  $\tilde{\mathcal{S}}$ . Then, the result of applying Algorithm 2.1 to each of these problems are provided in Figure 2.2 and Figure 2.3, respectively.

As guaranteed in the Theorem 2.11, the linear-quadratic convergence behavior of RNPO is observed in the both problems. Especially, the quadratic behavior starts after 32 steps and 23 iterations for SLQR and OLQR examples, respectively. In both cases, using Riemannian connection in Algorithm 2.1 (blue curves) has a superior convergence rate compared to the case of using the Euclidean connection (orange curves); this was expected as the Riemannian connection is compatible with the metric induced by the geometry inherent to the cost

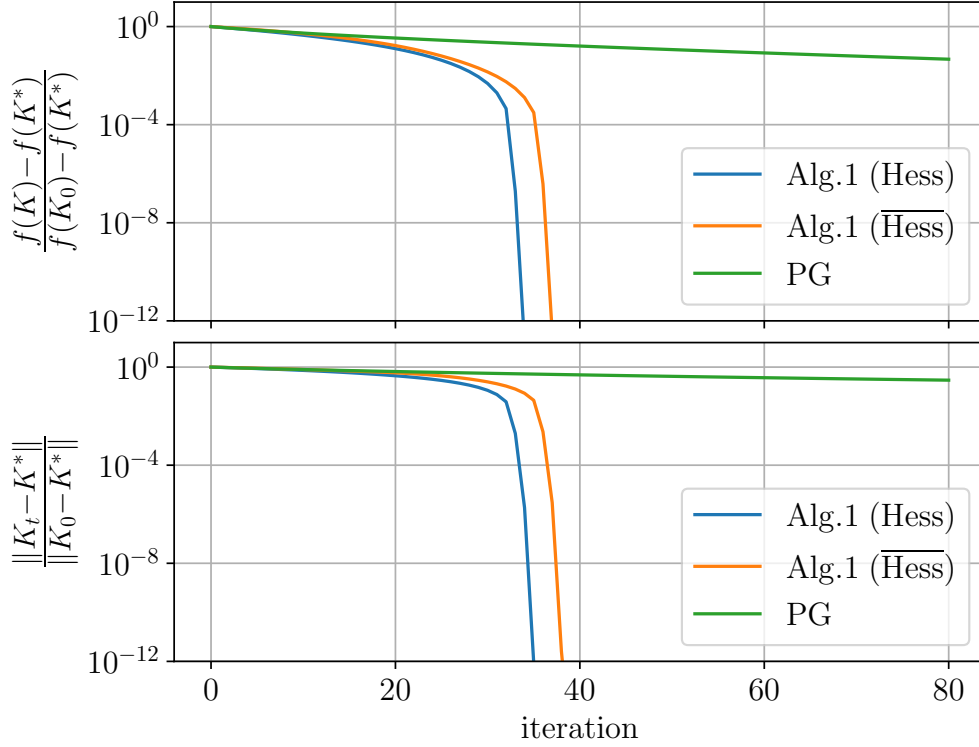


Figure 2.2: The normalized error of iterates and cost values at each iteration for the SLQR problem in Example 2.16.

function itself. However, this relative advantage in sample complexity requires computation of Christoffel symbols.

**Example 2.17** (Trajectories of RNPO using Hess versus  $\overline{\text{Hess}}$ ). In order to illustrate how the performance of RNPO is different in terms of using the Riemannian connection ( $\text{Hess}h$ ) versus Euclidean connection ( $\overline{\text{Hess}}h$ ), we consider another example with system parameters  $(A|B|Q|R|\Sigma) =$

$$\left( \begin{array}{cc|cc|cc|cc|cc} 0.8 & 1.0 & 0.0 & 1.0 & 10.0 & 0.0 & 0.1 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.9 & 1.0 & 0.0 & 0.0 & 0.5 & 0.0 & 0.1 & 0.0 & 5.0 \end{array} \right).$$

We run RNPO and PG algorithms for both OLQR and SLQR problems involving two decision variables, so that we can plot the trajectories of iterates over the level curves of the

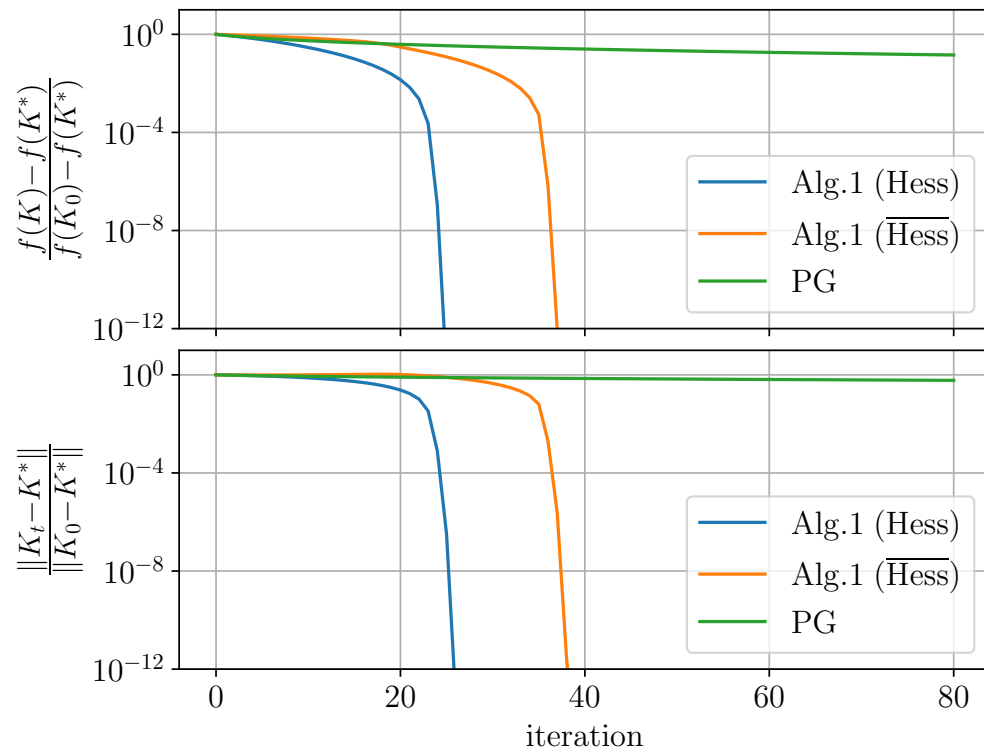


Figure 2.3: The normalized error of iterates and cost values at each iteration for the OLQR problem in Example 2.16.

associated cost function from different initial points (as illustrated in Figure 2.5 and Figure 2.4, respectively).

In Figure 2.4, first, note that RNPO with  $\overline{\text{Hess}}$  does not converge if initialized away from the local minimum (and away from the line  $l_2 = -1$ ) simply because the Euclidean Hessian fails to be positive definite therein (see Figure 2.1). On the other hand, RNPO with  $\text{Hess}$  successfully captures the inherent geometry of the problem here and converges from all initializations. These exemplify how RNPO can exploit the connection compatible with the metric (inherent to the cost function) in order to provide more effective iterate updates. Second, the square marker on each trajectory of RNPO indicates the first time stepsize  $\eta_t = 1$  is guaranteed to be stabilizing ( i.e.  $s_{K_t} \geq 1$ ). It can be seen that the neighborhood of the local minimum (zoomed in)—on which the identity stepsize is possible—is relatively very small. Whereas, by using the stability certificate, the specific choice of stepsize adopted here enables RNPO to handle initialization further away from the local minimum. Finally, PG algorithm has a sublinear rate, therefore, despite its progress in the beginning, it becomes so slow that it does not converge practically.

Next, notice that in both Figure 2.4 and Figure 2.5, the trajectories of RNPO with  $\text{Hess}h$  is much more favorable in comparison to RNPO with  $\overline{\text{Hess}h}$ ; especially, if it is initialized from points further away from the local minimum and closer to the boundary. Also, the iterates of RNPO with  $\text{Hess}$  converges much faster than that of  $\overline{\text{Hess}}$ ; especially, if initialized from points closer to the boundary. Additionally, similar to Figure 2.4, the region on which the unit stepsize is guaranteed to be stabilizing is relatively very small in Figure 2.5. Similarly, due to sublinear convergence rate of PG, it does not converge practically.

**Example 2.18** (Randomly selected system parameters). Next, we consider an example with  $n = 6$  number of states and  $m = 3$  number of inputs, and simulate the behavior of RNPO and PG for 100 randomly sampled system parameters. Particularly, the parameters  $(A, B)$  are sampled from a zero-mean unit-variance normal distribution where  $A$  is scaled so that the open-loop system is stable, i.e.,  $K_0 = 0$  is stabilizing, and the pair is checked to be

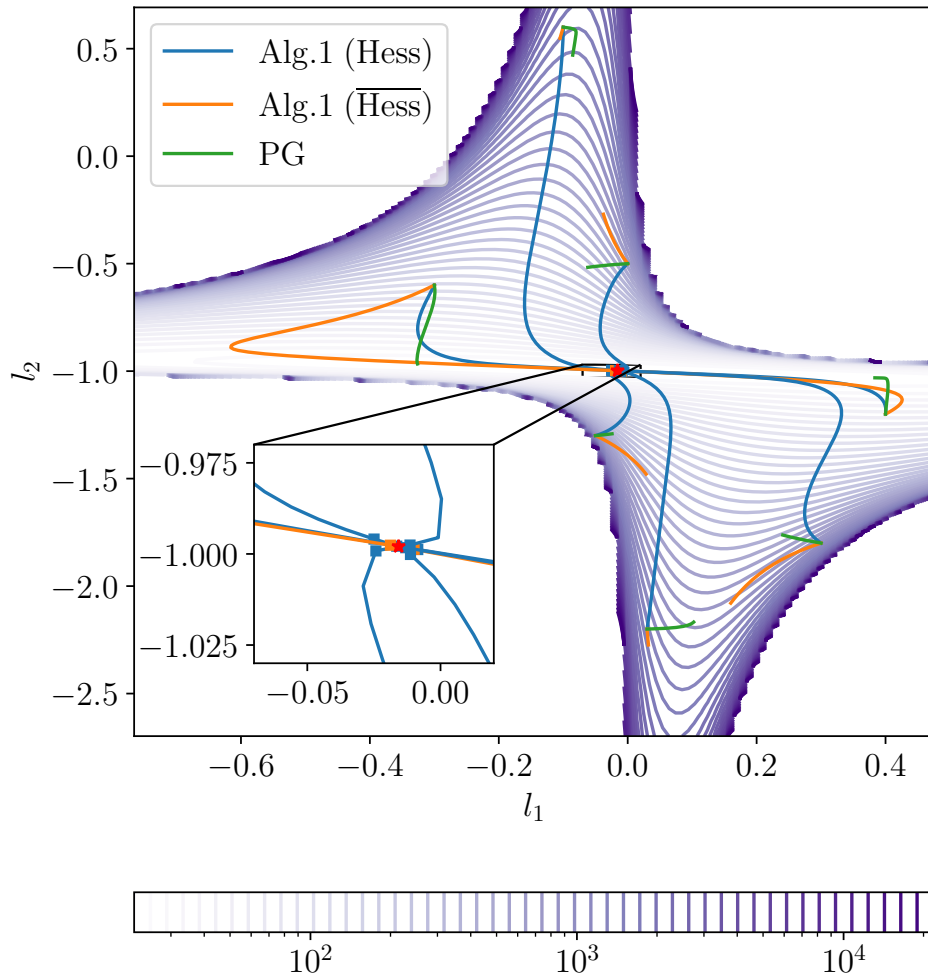


Figure 2.4: The trajectories of iterates  $K = \begin{pmatrix} l_1 & 0 \\ 0 & l_2 \end{pmatrix}$  generated by RNPO (with  $\text{Hess}h$  and  $\overline{\text{Hess}h}$ ) and PG—from different initial points—for the SLQR problem with sparsity constraint  $D^c = \{(1, 2), (2, 1)\}$ .

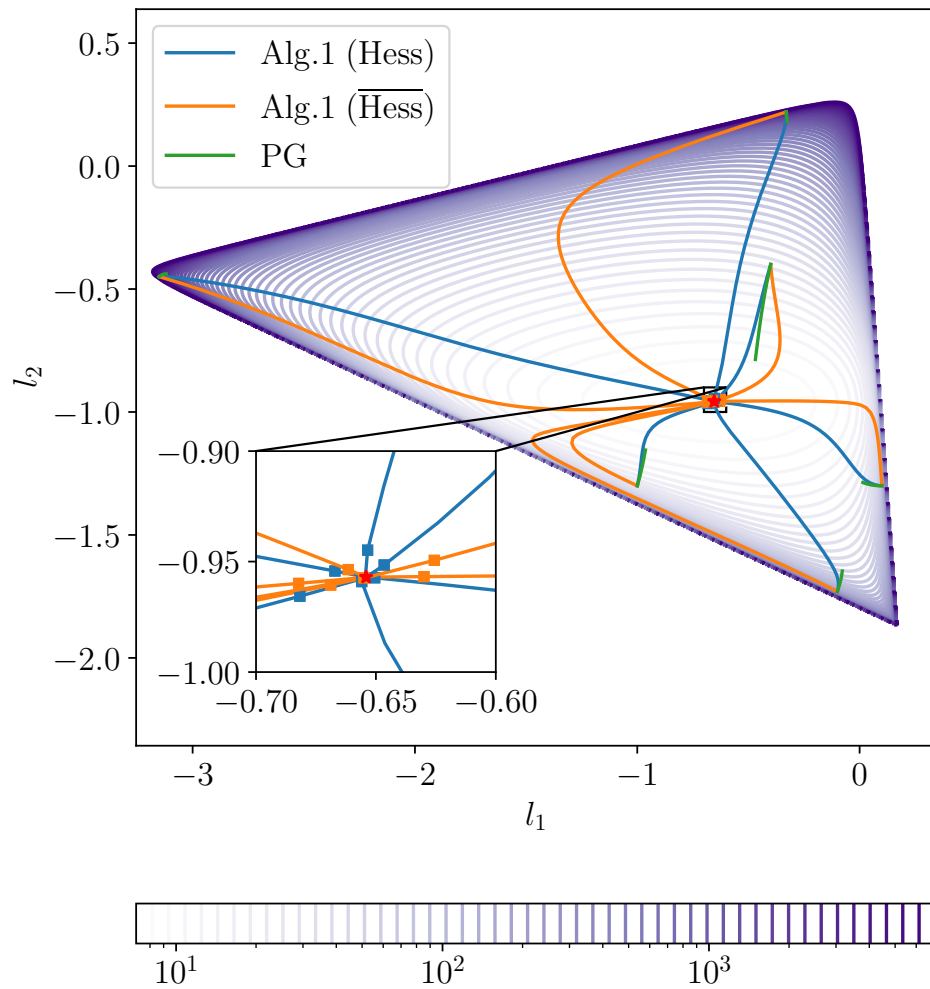


Figure 2.5: The trajectories of iterates  $L = \begin{pmatrix} l_1 & l_2 \end{pmatrix}^T$  generated by RNPO (with  $\text{Hess}h$  and  $\overline{\text{Hess}h}$ ) and PG—from different initial points—for the OLQR problem with output matrix  $C = \begin{pmatrix} 1.0 & 1.0 \end{pmatrix}$ .

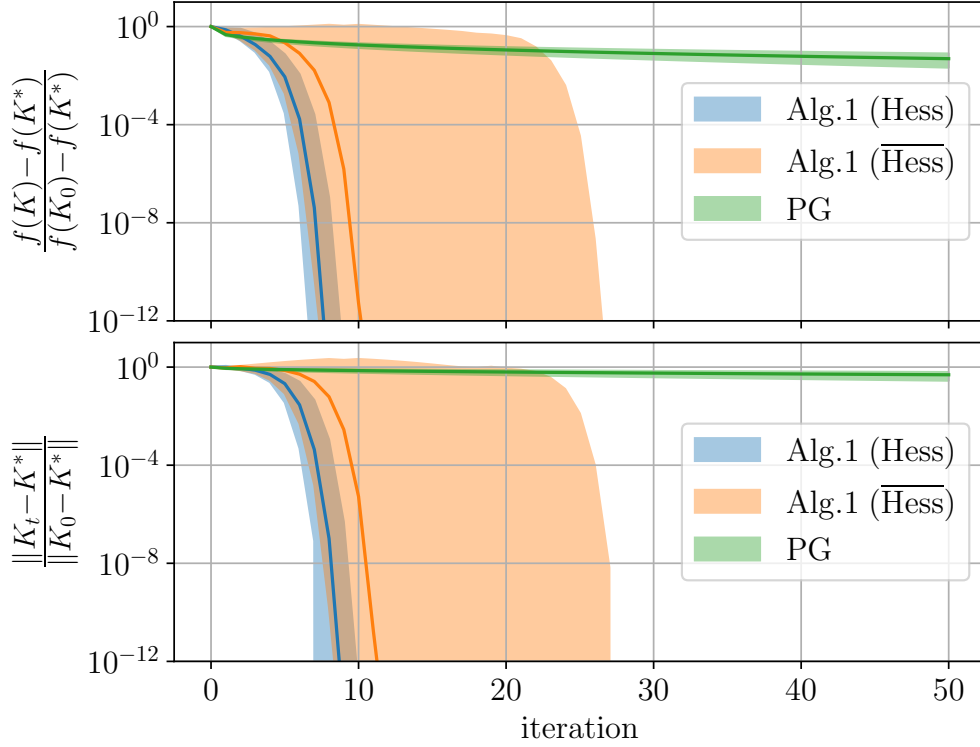


Figure 2.6: The min, max and median progress of normalized error of iterates and cost values at each iteration, for the SLQR problem with 100 different randomly sampled system parameters and sparsity patterns  $D$ .

controllable. Also, we choose  $Q = \Sigma = I_n$  and  $R = I_m$  in order to compare the convergence behaviors across different samples consistently.

For the SLQR problem, we randomly sample for the sparsity pattern  $D$  so that at least half of the entries are zero and all of the them have converged from  $K_0 = 0$  in less than 30 iterations. For the OLQR problem, we also randomly sample the output matrix  $C$  with  $d = 2$ , where %98 and %92 of them have converged from  $K_0 = 0$  in less than 50 iterations using Hess and  $\overline{\text{Hess}}$ , respectively. Finally, the minimum, maximum and median progress of the three algorithms for both SLQR and OLQR problems are illustrated in Figure 2.6 and Figure 2.7, respectively.

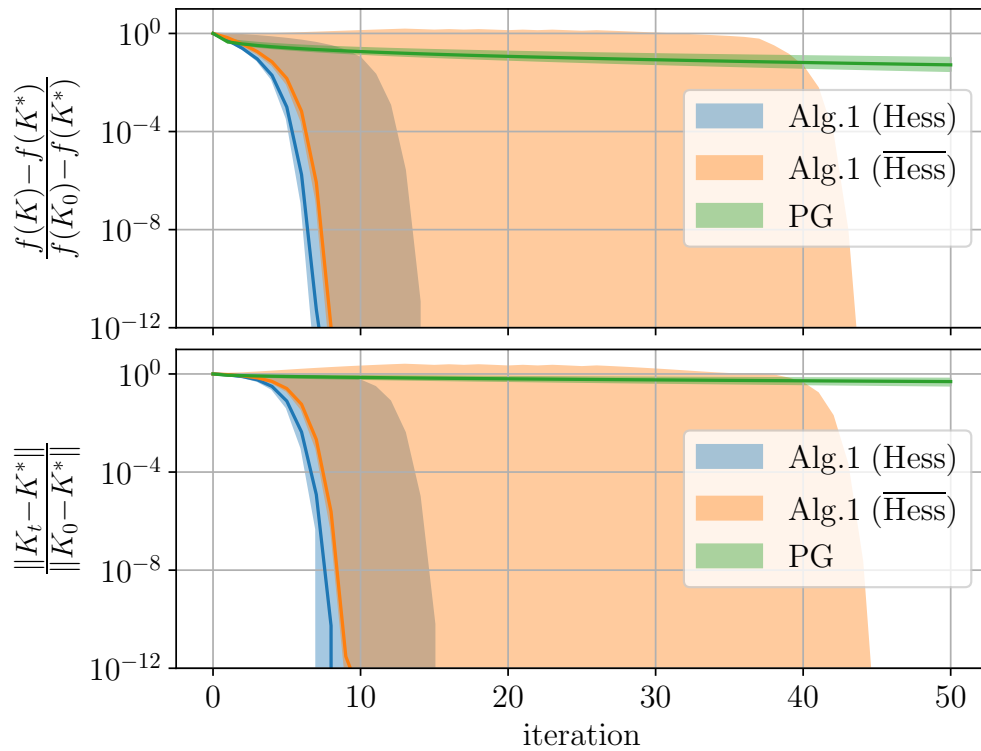


Figure 2.7: The min, max and median of normalized error of iterates and cost values at each iteration of Algorithm 2.1, for the OLQR problem with 100 different randomly sampled system parameters (including output matrix  $C$ ).



## 2.8 Remarks and Future Directions

In this chapter, we considered the problem of optimizing a smooth function over submanifolds of Schur stabilizing controllers  $\mathcal{S}$ . In order to treat this problem in a general format, we have studied the first and second order behavior of a smooth function when constraint to an embedded submanifold from an extrinsic point of view. Then, using the second order information of the restricted function, we have developed an algorithm that guarantees convergence to local minima at least with a linear rate and eventually with a quadratic rate. Combining this approach with backtracking line-search techniques or positive definite modifications of the Hessian operator [Dennis Jr and Schnabel, 1996; Nocedal and Wright, 2006] can be considered as immediate future directions that guarantee a global convergence behavior.

Even though the proposed algorithm depend on the linear structure of  $\tilde{\mathcal{S}}$ , the machinery developed here can be utilized for considering more complicated submanifolds which is the topic of our future work. As an example and in contrast to the other two (SLQR and OLQR) problems considered, we can explain how a constraint on the “average input energy” would translate to a *nonlinear* constraint that is related quiet deeply to the inherent geometry of the LQR problem. For that, we define the *average input energy*, denote by  $E_{\mathbf{u}}$ , as a measure for the average consumption of energy to be

$$E_{\mathbf{u}} := \mathbb{E}_{\mathbf{x}_0 \sim \mathcal{D}} \|\mathbf{u}\|_{\ell_2}^2,$$

where  $\|\cdot\|_{\ell_2}$  refers to  $\ell_2$ -norm. By the choice of static linear policy, i.e.  $\mathbf{u} = K\mathbf{x}$  for any  $K \in \mathcal{S}$ , and thus considering the closed-loop system dynamics  $\mathbf{x}_k = (A_K)^k \mathbf{x}_0$ , we can show, in the next paragraph, that under the usual identification of  $T_K \mathcal{S}$  with  $M(m \times n, \mathbb{R})$  we have  $E_{\mathbf{u}}(K) = \|K\|_{Y_K}^2$ . But then,  $E_{\mathbf{u}} : \mathcal{S} \rightarrow \mathbb{R}$  is smooth by composition, and then every regular level set of  $E_{\mathbf{u}}$  translates to an upperbound on the average input energy. Therefore, by Regular Level Set Theorem, it introduces a constraint as an embedded submanifold of  $\mathcal{S}$  which has a nonlinear but simple structure whenever considered in the associated Riemannian geometry. However, solving this problem still requires an efficient retraction that would

substitute the linear updates possible in SLQR and OLQR problems. Our approach allows integration of such retractions which introduces a new approach for solving constrained optimization problems over the manifold of Schur stabilizing controllers—such as the constraint LQR problems.

In particular, under the choice of linear policy, i.e.  $\mathbf{u} = K\mathbf{x}$  for some  $K \in \mathcal{S}$ , and by considering the closed-loop system dynamics  $\mathbf{x}_k = (A_K)^k \mathbf{x}_0$ , we can compute the expected total input energy as a function of  $K$  as follows

$$\begin{aligned} E_u(K) &= \mathbb{E}_{\mathbf{x}_0 \sim \mathcal{D}} \sum_{k=0}^{\infty} \|K\mathbf{x}_k\|_F^2 \\ &= \mathbb{E}_{\mathbf{x}_0 \sim \mathcal{D}} \sum_{k=0}^{\infty} \text{tr} [K^\top K (A_K)^k \mathbf{x}_0 \mathbf{x}_0^\top (A_K^\top)^k] \\ &= \sum_{k=0}^{\infty} \text{tr} [K^\top K (A_K)^k \Sigma_1 (A_K^\top)^k] \\ &= \text{tr} [K^\top K Y_K], \end{aligned}$$

where  $Y_K = \mathbb{L}(A_K, \Sigma_1)$  and  $\|\cdot\|_{\ell_2}$  for a matrix valued signal refers to the  $\|\text{vec}(\cdot)\|_{\ell_2}$ . Also, the result of this chapter has been mainly adapted from [Talebi and Mesbahi, 2022].

As another future directions of this work, we point out the problem of learning the constrained policy without the knowledge of system parameters. Also, uncertainty in the problem parameters often occurs in large networked systems, requiring a scalable solution strategy. This is the topic of the next chapter, where we consider learning the problem of structured controller for a network of homogeneous systems.

## Chapter 3

# DISTRIBUTED DATA-DRIVEN STRUCTURED POLICY ITERATION

Control of networked systems, comprised of interacting agents, is often achieved through modeling their underlying interactions. Constructing accurate models of such interactions—in the meantime—can become prohibitive in applications. Data-driven control methods avoid such complexities by directly synthesizing a controller from the observed data. In this chapter, we propose an algorithm referred to as Distributed Data-driven Structured Policy Iteration (D3SPI), for synthesizing an efficient feedback mechanism that respects the sparsity pattern induced by the underlying interaction network. In particular, our algorithm uses temporary “auxiliary” links to boost information exchange for a (smaller) sub-network during the “learning phase.” We then proceed to show that the updated policy results in a stabilizing structured policy for the entire network. This is followed by stability and convergence analysis for the proposed distributed policies throughout the learning phase, exploiting a construct referred to as the “patterned linear semigroup.” The performance of D3SPI is then demonstrated using representative simulation scenarios.

### 3.1 Introduction

In recent years, there has been a renewed interest in distributed control of large-scale systems. The unprecedented interdependence and size of the data generated by such systems have necessitated a distributed approach to policy computation in order to influence or direct their behavior and performance. In these scenarios, collective actions are often synthesized via local decisions, informed by a *structured* information exchange mechanism. An important roadblock for centralized control design methods, is thereby, their scalability and

shortcomings in utilizing the underlying structure of large-scale interconnected systems.<sup>1</sup>

Structured control synthesis in the meantime is an NP-hard constrained optimization problem [Papadimitriou and Tsitsiklis, 1986]. Hence, distributed control design for large-scale systems has often been pursued not necessarily to characterize optimal policies per se, but to devise efficient (possibly suboptimal) control mechanisms that exploit the inherent system structure. In parallel, recent advances in measurement technologies have made available of an unprecedented amount of data, motivating how offline and online data-processing can be leveraged for data-driven decision-making on high-dimensional complex systems.

In this work, we propose the linear-quadratic regulator (LQR)-based algorithm, coined Distributed Data-driven Structured Policy Iteration (D3SPI), to iteratively learn stabilizing controllers for unknown but identical linear dynamical systems that are connected via a network induced by the coupling in their performance. The setup is a particular realization of cooperative game-theoretic decision-making (see remarks under Footnote 2). This class of synthesis problems is motivated by applications such as formation flight [Stipanovic et al., 2004] and distributed camera systems [Borrelli et al., 2005], where the dynamics of the network nodes (agents) cannot be precisely parameterized. D3SPI is built upon a data-driven learning phase on a subgraph in a large system. This subgraph includes the agent with maximum degree in the network and requires enabling *auxiliary* links within this subgraph in order to iteratively learn a stabilizing structured controller (optimal for the subgraph) for the entire network. This “extension” synthesis procedure utilizes a symmetry property of the networked systems, that we refer to as *patterned linear semigroup* (see Section 3.2.1).

The remainder of the chapter is organized as follows. In §3.2 we introduce the problem setup and motivation behind our work, and provide an overview of the related literature (§3.2.2). In §3.3, we present and analyze the D3SPI algorithm, followed by the theoretical analysis in §3.4. Illustrative examples are provided in §3.5, followed by concluding remarks in §3.6.

---

<sup>1</sup> $\mathcal{O}(n^3)$  complexity of solving the *Algebraic Riccati Equation* [Bini et al., 2011] and scalability issues of *Model Predictive Control* [Camponogara et al., 2002] are among such examples.

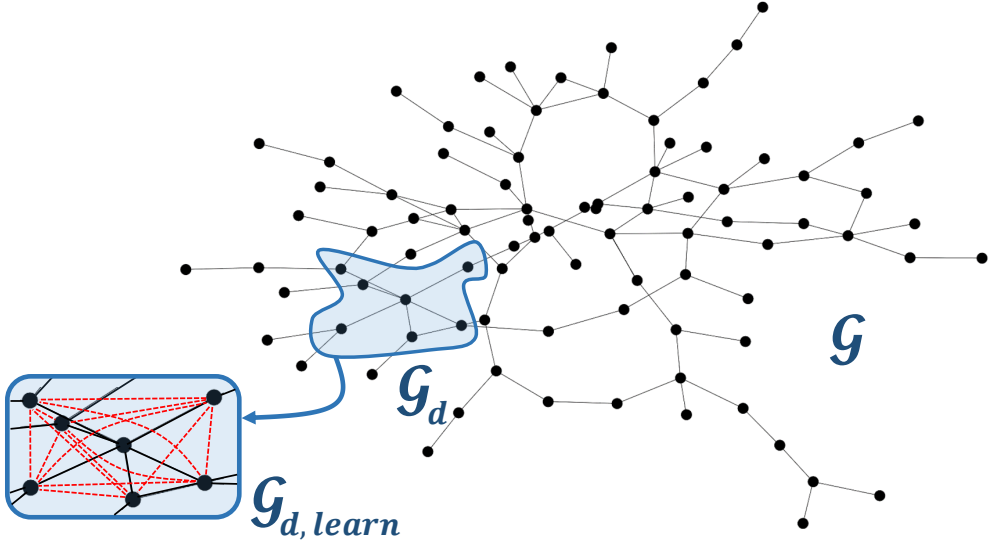


Figure 3.1: Addition of auxiliary links (dashed red) to the subgraph  $\mathcal{G}_d$  during the policy learning phase. The size of the subgraph depends on the maximum degree of the original graph  $\mathcal{G}$ .

**Notation.** The operator  $\text{diag}(\cdot)$  makes a square diagonal matrix out of the elements of its argument;  $\text{vech}(\cdot)$  on the other hand, takes a square matrix and stacks the lower left triangular half (including the diagonal) into a single vector. We use  $N \succ 0$  ( $\succeq 0$ ) to declare  $N$  as a positive-(semi)definite matrix. The  $i$ th eigenvalue and spectral radius of  $M$  are denoted by  $\lambda_i(M)$  and  $\rho(M)$ ;  $M$  is called Schur stable when  $\rho(M) < 1$ . We say that an  $n$ -dimensional linear system parameterized by the pair  $(A, B)$  is *controllable* if the controllability matrix  $\mathcal{C} = [B \ AB \ \dots \ A^{n-1}B]$  has a full-rank. We denote the *Kronecker product* of two matrices by  $\otimes$ . For a block matrix  $\mathbf{F}$ , by  $[\mathbf{F}]_{rk}$  we imply the  $r$ th row and  $k$ th column “block” component with appropriate dimensions. An (undirected) graph is characterized by  $\mathcal{G} = (\mathcal{V}_{\mathcal{G}}, \mathcal{E}_{\mathcal{G}})$  where  $\mathcal{V}_{\mathcal{G}}$  is the set of nodes and  $\mathcal{E}_{\mathcal{G}} \subseteq \mathcal{V}_{\mathcal{G}} \times \mathcal{V}_{\mathcal{G}}$  denotes the set of edges. An edge exists from node  $i$  to  $j$  if (the unordered pair)  $(i, j) \in \mathcal{E}_{\mathcal{G}}$ ; this is also specified by writing  $j \in \mathcal{N}_i$ , where  $\mathcal{N}_i$  is the set of neighbors of node  $i$ . We designate the maximum degree of  $\mathcal{G}$  by  $d_{\max}(\mathcal{G})$ . Finally, the graph  $\mathcal{G}$  can be represented using matrices such as the Laplacian  $\mathcal{L}_{\mathcal{G}}$  or the adjacency  $\mathcal{A}_{\mathcal{G}}$ .

### 3.2 Problem Setup

Consider a network of identical agents with interdependencies induced by a network-level objective. In particular, we assume that the system contains  $N$  agents forming a graph  $\mathcal{G} = (\mathcal{V}_{\mathcal{G}}, \mathcal{E}_{\mathcal{G}})$ , where each node of the graph in  $\mathcal{V}_{\mathcal{G}}$  represents a linear discrete-time system,

$$\mathbf{x}_{i,t+1} = A\mathbf{x}_{i,t} + B\mathbf{u}_{i,t}, \quad i = 1, 2, \dots, N, \quad (3.1)$$

with  $\mathbf{x}_{i,t} \in \mathbb{R}^n$  and  $\mathbf{u}_{i,t} \in \mathbb{R}^m$  denoting the state and input of agent  $i$  at time-step  $t$ , respectively. The unknown system matrices  $A \in \mathbb{R}^{n \times n}$  and  $B \in \mathbb{R}^{n \times m}$  are assumed to form a controllable pair. The network dynamics can compactly be represented as,

$$\hat{\mathbf{x}}_{t+1} = \hat{\mathbf{A}}\hat{\mathbf{x}}_t + \hat{\mathbf{B}}\hat{\mathbf{u}}_t, \quad (3.2)$$

where  $\hat{\mathbf{x}}_t \in \mathbb{R}^{Nn}$  and  $\hat{\mathbf{u}}_t \in \mathbb{R}^{Nm}$  are comprised of the states and inputs of entire network,  $\hat{\mathbf{x}}_t = [\mathbf{x}_{1,t}^\top \dots \mathbf{x}_{N,t}^\top]^\top$ ,  $\hat{\mathbf{u}}_t = [\mathbf{u}_{1,t}^\top \dots \mathbf{u}_{N,t}^\top]^\top$ , with  $\hat{\mathbf{A}} \in \mathbb{R}^{Nn \times Nn}$  and  $\hat{\mathbf{B}} \in \mathbb{R}^{Nn \times Nm}$  are in block diagonal forms  $\hat{\mathbf{A}} = \mathbf{I}_N \otimes A$  and  $\hat{\mathbf{B}} = \mathbf{I}_N \otimes B$ . The agents' interconnections are represented by edges  $\mathcal{E}_{\mathcal{G}} \subseteq \mathcal{V}_{\mathcal{G}} \times \mathcal{V}_{\mathcal{G}}$  that can facilitate a distributed feedback design. We do not assume that  $\mathcal{G}$  is necessarily connected; the motivation for this becomes apparent subsequently. Let  $\mathcal{N}_i$  denote the set of neighbors of node  $i$  in  $\mathcal{G}$  (excluding itself). Then, based on the underlying communication graph and for any choice of positive integers  $a$  and  $b$ , we define a linear subspace of  $\mathbb{R}^{a \times bN}$  as,

$$\mathcal{U}_{a,b}^N(\mathcal{G}) := \{\mathbf{M} \in \mathbb{R}^{a \times bN} \mid [\mathbf{M}]_{ij} = \mathbf{0} \text{ if } j \notin \mathcal{N}_i \cup \{i\}, \quad [\mathbf{M}]_{ij} \in \mathbb{R}^{a \times b}, \quad i, j = 1, \dots, N\}.$$

Without having access to the system parameters  $A$  and  $B$ , we are interested in designing linear feedback gains, consistent with the desired sparsity pattern induced by the network, using data generated by (3.2). More precisely, given an initial condition  $\hat{\mathbf{x}}_1$ , the distributed (structured) optimal control problem assumes the form,

$$\begin{aligned} \min_{\hat{\mathbf{K}}} \quad & \sum_{t=1}^{\infty} \hat{\mathbf{x}}_t^\top \hat{\mathbf{Q}} \hat{\mathbf{x}}_t + \hat{\mathbf{u}}_t^\top \hat{\mathbf{R}} \hat{\mathbf{u}}_t \\ \text{s.t.} \quad & (3.2), \quad \hat{\mathbf{u}}_t = \hat{\mathbf{K}} \hat{\mathbf{x}}_t, \quad \hat{\mathbf{K}} \in \mathcal{U}_{m,n}^N(\mathcal{G}), \end{aligned} \quad (3.3)$$

where  $\hat{\mathbf{K}}$  stabilizes the pair  $(\hat{\mathbf{A}}, \hat{\mathbf{B}})$  (i.e.,  $\rho(\hat{\mathbf{A}} + \hat{\mathbf{B}}\hat{\mathbf{K}}) < 1$ ),  $\hat{\mathbf{R}} = \mathbf{I}_N \otimes R$  and  $\hat{\mathbf{Q}} = \mathbf{I}_N \otimes Q_1 + \mathcal{L}_{\mathcal{G}} \otimes Q_2$  for some given cost matrices  $Q_1 \succ 0$ ,  $Q_2 \succeq 0$ ,  $R \succ 0$ . Note that  $\hat{\mathbf{Q}} \in \mathcal{U}_{n,n}^N(\mathcal{G})$  is positive definite. Such interdependence induced through the cost has been considered in a graph-based distributed control framework; see for instance [Borrelli and Keviczky, 2008; Deshpande et al., 2011; Wang et al., 2017; Massioni and Verhaegen, 2009]. In a nutshell, the first term in  $\hat{\mathbf{Q}}$  encodes the cost pertinent to state regulation for each agent, while the second term, captures the “disagreement” cost between the neighbors.<sup>2</sup>

In this work, we propose a data-guided suboptimal solution to (3.3) not relying on knowledge of the system parameters  $\hat{\mathbf{A}}$  and  $\hat{\mathbf{B}}$ . Indeed, unknown system parameters may further complicate the considered setup. Instead, our approach relies on the system’s input-state time series for synthesizing distributed feedback control on  $\mathcal{G}$ . A summary of the challenges for analyzing this problem is listed as follows: I) The constrained optimization problem in (3.3) is in general NP-hard [Papadimitriou and Tsitsiklis, 1986]. Based on the complete knowledge of the system parameters, this problem has been investigated under variety of assumptions [Gupta et al., 2005; Rotkowitz and Lall, 2005; Bamieh et al., 2002; Borrelli and Keviczky, 2008], or approached directly with the aid of projected gradient-based policy updates [Mårtensson and Rantzer, 2009; Bu et al., 2019a]. II) In the meantime, policies obtained via data-driven approaches, do not necessarily respect the hard constraints on  $\hat{\mathbf{K}} \in \mathcal{U}_{m,n}^N(\mathcal{G})$

---

<sup>2</sup>One instance of such an interactive cost among agents appears in the cooperative game setup where agent  $i$  aims to solve the minimization problem,

$$\min_{(\mathbf{u}_{i,t})_{t=0}^{\infty} \in \ell_2} \mathbf{J}_i(\hat{\mathbf{x}}_1, \hat{\mathbf{u}}_t) = N \sum_{t=0}^{\infty} \left( \mathbf{x}_{i,t}^{\top} Q_1 \mathbf{x}_{i,t} + \mathbf{u}_{i,t}^{\top} R \mathbf{u}_{i,t} + \sum_{j \in \mathcal{N}_i} (\mathbf{x}_{j,t} - \mathbf{x}_{i,t})^{\top} Q_2 (\mathbf{x}_{j,t} - \mathbf{x}_{i,t}) \right).$$

Then, it is well-known that the set of Pareto front solution of this game can be obtained by minimizing the parametric cost function,

$$\min_{(\mathbf{u}_{1,t})_{t=0}^{\infty}, \dots, (\mathbf{u}_{N,t})_{t=0}^{\infty} \in \ell_2} \sum_{i=1}^N \alpha_i \mathbf{J}_i(\hat{\mathbf{x}}_1, \hat{\mathbf{u}}_t),$$

parameterized by  $\alpha_1, \dots, \alpha_N$  where  $\alpha_i \in [0, 1]$  and  $\sum_i \alpha_i = 1$  (see e.g. [Engwerda, 2005b]). Therefore, a cost such as in (3.3) can be viewed as a special case of the fair Pareto optimal solution with the choice of  $\alpha_i = 1/N$  for all  $i$ .

posed in (3.3). In particular, we point out that “projection” onto the intersection of the constraint imposed by the network and stabilizing controllers is not straightforward due to the intricate geometry of the set of stabilizing controllers [Bu et al., 2020b]. III) Another key challenge in adopting data-driven methods for the entire network is rooted in the “curse of dimensionality” inherent in the design and analysis of large-scale systems. In fact, even collecting data from the entire network can be prohibitive. IV) Finally, it is often impossible in applications to pause the operation of the network for data collection or decision-making purposes. Therefore, an attractive feature of a synthesis process for large networks would be its online realizability.

### 3.2.1 Structures in the Problem and our Approach

As in this work the sparsity requirement  $\hat{\mathbf{K}} \in \mathcal{U}_{m,n}^N(\mathcal{G})$  is considered as a *hard constraint* for control synthesis, and as such, the corresponding optimization is NP-hard in general, we shift our attention from the optimal solution of (3.3) towards a “reasonable” suboptimal stabilizing distributed controller with a reasonable computational cost. We begin by introducing a set that plays an important role in our subsequent analysis. For any integer  $r \geq 2$ , first consider the following linear subspace of  $\mathbb{R}^{rn \times rn}$ :

$$\mathbf{L}(r \times n, \mathbb{R}) := \{ \mathbf{N}_r \mid \mathbf{N}_r = \mathbf{I}_r \otimes (A - B) + \mathbb{1}_r \mathbb{1}_r^T \otimes B, \quad A, B \in \mathbb{R}^{n \times n} \}.$$

Now, we define the *patterned linear semigroup* as<sup>3</sup>

$$\text{PL}(r \times n, \mathbb{R}) := \{ \mathbf{N}_r \in \mathbf{L}(r \times n, \mathbb{R}) \cap \text{GL}(rn, \mathbb{R}) \mid A \in \text{GL}(n, \mathbb{R}) \cap \mathbb{S}^n, B \in \mathbb{S}^n \},$$

where  $\mathbb{S}^n$  denotes the set of symmetric  $n \times n$  matrices. We note that the patterned linear semigroup  $\text{PL}(r \times n, \mathbb{R})$  is a sub-semigroup of  $\text{GL}(rn, \mathbb{R})$ , and  $\mathbf{L}(r \times n, \mathbb{R})$  is a linear subspace

---

<sup>3</sup>The proposed patterned linear semigroup and characterizing its interplay with the Lyapunov equation is considered as a key contribution in our approach. Note that, despite the claim in [Talebi et al., 2021a],  $\text{PL}(r \times n, \mathbb{R})$  requires further refinement to become a linear “subgroup” of  $\text{GL}(rn, \mathbb{R})$ . Nonetheless, the current characterization is sufficient for our purposes and further extensions are deferred to our future work.



of  $\mathbb{R}^{rn \times rn}$ , closed under matrix multiplication; see the proof of Lemma 3.13. The following observation underscores the relevance of patterned linear semigroup in system analysis.

**Proposition 3.1.** *For a Schur stable matrix  $A \in L(r \times n, \mathbb{R})$  and  $0 \prec Q \in \mathbb{S}^{rn}$ , let  $P$  denote the unique solution to the corresponding discrete-time Lyapunov equation, i.e.,  $P = A^\top P A + Q$ . Then,  $P \in \text{PL}(r \times n, \mathbb{R})$  if and only if  $Q \in \text{PL}(r \times n, \mathbb{R})$ .*

The invariance of the Lyapunov equation under the action of patterned linear semigroup has important implications for our data-driven synthesis for large-scale networks; see Proposition 3.5 and its implications in Algorithm 3.1.

In what follows, we summarize the key ingredients of our approach to distributed data-driven policy iteration: I) Inspired by a Q-learning-based policy iteration algorithm, we propose a model-free structured policy iteration scheme for obtaining “reasonable” performance for the synthesis problem (3.3). The main challenge with this approach is to ensure obtaining policies that conform to the sparsity constraint  $\hat{\mathbf{K}} \in \mathcal{U}_{m,n}^N(\mathcal{G})$  at each iteration. This issue is particularly relevant when an arbitrary “projection” of the iterated policy on the set  $\mathcal{U}_{m,n}^N(\mathcal{G})$  fails to be stabilizing, reflecting the intricate geometry of the set of stabilizing controllers [Bu et al., 2020b]. In order to overcome this challenge, we propose learning two distinct facets of control, resembling the “individual” and “cooperative” components. In this direction, we simultaneously learn a stability margin for each of these components, allowing us to extend the synthesis procedure to the entire network. II) Inherent to our algorithm is a synthesis subproblem whose dimension is related only to the maximum node in underlying graph rather than the dimension of the original network. In particular, we will reason that for the learning phase, our method only requires data collection from a (specific) smaller sub-network  $\mathcal{G}_d \subseteq \mathcal{G}$  with size  $d = d_{\max}(\mathcal{G}) + 1$ . This subgraph is substantially smaller than the original graph whenever  $d_{\max}(\mathcal{G})$  is significantly smaller than  $N$ , reflecting the empirical feature of many real-world networks. III) Finally, we note that terminating distributed processes on real-world networks for the purpose of data collection is often infeasible. As such, we allow temporary links on the subgraph  $\mathcal{G}_d$  *only during the learning phase* of the proposed

algorithm—links that are subsequently removed. We note that for the learning phase, our method requires observations only from this portion of the graph where temporary links are augmented in order to make  $\mathcal{G}_d$  a clique. The distributed control underpinnings of method proposed in this work follows its model-based analogue previously studied in [Borrelli and Keviczky, 2008; Wang et al., 2017]. In this work, our aim is to extend the setup proposed in these aforementioned works to propose a model-free structured policy iteration algorithm which is not only computationally efficient, but also practical for large-scaled networked systems.

### 3.2.2 *Related Literature*

Distributed control is a well-established area of research in systems theory. The roots of the field trace back to the socioeconomics literature in 1970’s [McFadden, 1969] and early works in the control literature followed suite later during that decade [Wang and Davison, 1973]. The main motivation for these works was lack of scalability in centralized planning and control, due to information or computational limitations [Sandell et al., 1978; Ioannou, 1986]. Fast forward a few decades, sufficient graph-theoretic conditions were provided for stability of formations comprised of identical vehicles [Fax and Murray, 2004] and, along the same lines, graph-based distributed controller synthesis was further examined independently in works such as [Massioni and Verhaegen, 2009; Deshpande et al., 2012; Borrelli and Keviczky, 2008; Mårtensson and Rantzer, 2009]. The topic was also studied from the perspective of spatial invariance [Bamieh et al., 2002; Motee and Jadbabaie, 2008] and a compositional layered design [Chapman et al., 2017; Alemzadeh and Mesbahi, 2018]. Moreover, from an agent-level perspective, the problem has been tackled for both homogeneous systems [Massioni and Verhaegen, 2009; Borrelli and Keviczky, 2008; Wang et al., 2017] and more recently heterogeneous ones [Stürz et al., 2021].

Having access to the underlying system model is a common assumption in the literature on distributed control, where the goal is to find a distributed feedback mechanism that conforms to an underlying network topology. However, deriving dynamic models from first

principles could be restrictive for large-scale systems. Such restrictions also hold for parametric perturbations that occur due to inefficient modeling or other unknown design factors. For instance, even the LQR solution with its strong input robustness properties, may have small stability margins for general parameter perturbations [Dorato et al., 1994]. Robust synthesis approaches could alleviate this issue when the perturbations follow specific models, in both centralized [Khargonekar et al., 1990] and distributed [Li et al., 2012] cases. However, if the original estimates of system parameters are inaccurate or the perturbations violate the presumed model, then both stability and optimality of the proposed feedback mechanisms can be compromised. Data-driven control, on the other hand, circumvents such drawbacks and utilize the available data generated by the system when its model is unavailable. This point of view has historically been examined in the context of adaptive control and system identification [Ljung, 1999], particularly, when asymptotic properties of the synthesized system are of interest. For more recent works that have adopted a non-asymptotic outlook on data-driven control, we mention [Van Waarde et al., 2020; Alaeddini et al., 2018; Dean et al., 2019a] that used batched data for synthesis, as well as online iterative procedures [Talebi et al., 2020; Oymak and Ozay, 2019]. Furthermore, in regards to the adaptive nature of such algorithms, there is a close connection between online data-driven control and reinforcement learning [Lewis et al., 2012; Bradtke et al., 1994]. In these latter works, policy iteration has been extended to approximate LQR by avoiding the direct solution of Algebraic Riccati Equation (ARE); yet majority of these works do not have favorable scaling properties.

Control and estimation for large-scale systems offers its unique set of challenges due higher levels of uncertainty, scalability issues, and modeling errors. Nevertheless, model-free synthesis for large-scale systems, as a discipline, is still in its infancy. From a control theoretic perspective, the work [Luo et al., 2019] addresses some of the aforementioned issues using ideas from mean-field multiagent systems and with the key assumption of partial exchangeability. The work [Alemzadeh and Mesbahi, 2019] on the other hand, provides a decentralized LQR algorithm based on network consensus that has low complexity, but potentially a high cost of implementation. Lastly, SDP projection-based analysis has been

examined in [Chang and Shahrampour, 2021], where each agent will have a sublinear regret as compared with the best fixed controller in hindsight. The problem has also been considered from a game-theoretic standpoint [Li et al., 2017; Talebi et al., 2019; Nowé et al., 2012], where agents can have conflicting objectives.

### 3.3 Algorithm: Distributed Data-driven Structured Policy Iteration (D3SPI)

In this section, we present and discuss the main algorithm of the chapter, namely, D3SPI. Given the underlying communication graph  $\mathcal{G}$ , the networked system is considered as a black-box, whereas the designer is capable of inserting input signals to the system and observe states. The goal of D3SPI is then to find a data-guided suboptimal solution for (3.3) without knowledge of system parameters  $\hat{\mathbf{A}}$  and  $\hat{\mathbf{B}}$ . To this end, our approach involves considering the synthesis problem on a subgraph  $\mathcal{G}_d \subseteq \mathcal{G}$ , with the associated time-series data. Before presenting the main algorithm, we formalize two useful notions in order to facilitate the presentation.

**Definition 3.2.** Given a subgraph  $\mathcal{G}' \subseteq \mathcal{G}$  and a node labeling, let  $\text{Policy}(\mathcal{V}_{\mathcal{G}'})$  denote the concatenation of policies of the agents in  $\mathcal{V}_{\mathcal{G}'}$ , i.e.,

$$\text{Policy}(\mathcal{V}_{\mathcal{G}'}) := [\mathbf{u}_1^\top \quad \mathbf{u}_2^\top \quad \cdots \quad \mathbf{u}_{|\mathcal{V}_{\mathcal{G}'}|}^\top]^\top,$$

where  $\mathbf{u}_i$  is the feedback control policy of agent  $i$  in the subgraph  $\mathcal{G}'$  as a mapping from  $\{\mathbf{x}_j | j \in \mathcal{N}_i \cup \{i\}\}$  to  $\mathbb{R}^m$ . Furthermore, we use  $\text{Policy}(\mathcal{V}_{\mathcal{G}'})|_t$  to denote the realization of these policies at time  $t$ . Similarly, we define,

$$\text{State}(\mathcal{V}_{\mathcal{G}'}) := [\mathbf{x}_1^\top \quad \mathbf{x}_2^\top \quad \cdots \quad \mathbf{x}_{|\mathcal{V}_{\mathcal{G}'}|}^\top]^\top.$$

The D3SPI algorithm is introduced in Algorithm 3.1 with the following standard assumption.

**Assumption 3.3.** The initial controller  $K_1$  is stabilizing for the controllable pair  $(A, B)$ , and  $\mathbf{e}_t$  in Algorithm 3.2 is such that  $\text{Policy}(\mathcal{V}_{\mathcal{G}'})|_t$  remains persistently exciting.

We refer to the main loop of the algorithm in Line 5 as the *learning phase*. During the learning phase, we include temporary “auxiliary” links to  $\mathcal{G}_d$  and make the communication graph a clique. We show such distinction by  $\mathcal{G}_{d,\text{learn}}$ , where  $|\mathcal{V}_{\mathcal{G}_d}| = |\mathcal{V}_{\mathcal{G}_{d,\text{learn}}}|$  but  $\mathcal{G}_{d,\text{learn}}$  is a clique. Inherent to D3SPI is a policy iteration on  $\mathcal{G}_{d,\text{learn}}$  that characterizes components  $K_k$  and  $L_k$ , intuitively representing “self” and “cooperative” controls at iteration  $k$ , respectively. In particular, during the learning phase, we utilize these control components in order to design and update an effective stabilizing controller for the rest of the network  $\mathcal{G} \setminus \mathcal{G}_{d,\text{learn}}$ .

We do so by ensuring that during the learning phase, information is exchanged unidirectionally from  $\mathcal{G}_{d,\text{learn}}$  to the rest of the network; hence, the policy of the agents in  $\mathcal{G} \setminus \mathcal{G}_{d,\text{learn}}$  is dependent on those in  $\mathcal{G}_{d,\text{learn}}$ , and not vice versa. After the learning phase terminates, we remove the temporary links added during the learning phase (re-initialize to the original network), and synthesize a suboptimal stabilizing control for the entire network. In the learning phase of D3SPI, we use a Recursive Least Squares (RLS)-based recursion to estimate the unknown parameters in the cost matrix at iteration  $k$ , referred to as  $\tilde{\mathbf{H}}_k$ . This process is performed in Subgraph Policy Evaluation (SPE) (Algorithm 3.2) subroutine by inputting (sub-)graph  $\mathcal{G}$ ,  $\mathcal{G}_{d,\text{learn}}$ , the mapping policy( $\mathcal{V}_{\mathcal{G}_d}$ ), and the previous estimate of  $\tilde{\mathbf{H}}_{k-1}$ . As will be discussed in Section 3.4.1,  $\tilde{\mathbf{H}}_k$  contains the required information to determine the two control components  $K_k$  and  $L_k$  from the data. We extract this square matrix through a recursive update on the vector  $\boldsymbol{\theta}_{k-1}$ , derived from half-vectorization of  $\tilde{\mathbf{H}}_{k-1}$ , solving RLS for the linear equation  $\mathcal{R}(\tilde{\mathbf{x}}_t, \tilde{\mathbf{u}}_t) = \boldsymbol{\zeta}_t^\top \boldsymbol{\theta}_{k-1}$ , where  $\mathcal{R}(\tilde{\mathbf{x}}_t, \tilde{\mathbf{u}}_t)$  denotes the local cost and  $\boldsymbol{\zeta}_t \in \mathbb{R}^p$  contains the data measurements. We use subscript  $k$  for policy update and  $t$  for data collection. The adaptive nature of the algorithm involves the exploration signal  $\mathbf{e}_t$  to be augmented to the policy vector in order to provide persistence of excitation.

In our setup,  $\mathbf{e}_t$  is sampled from a normal distribution  $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \Sigma)$  where the choice of the variance  $\Sigma$  is problem-specific. In practice, excitation of the input is a subtle task and has been realized in a variety of forms such as random noise [Bradtke et al., 1994], sinusoidal signals [Jiang and Jiang, 2012], and exponentially decaying noise [Lewis and Vamvoudakis, 2010]. We denote by  $\mathcal{P}$  the projection factor that is reset to  $\mathcal{P}_0 \succ 0$  for each iteration. Con-

---

**Algorithm 3.1:** Distributed Data-driven Structured Policy Iteration (D3SPI)

---

- 1: **Initialization** ( $t \leftarrow 1, k \leftarrow 1, \tau_1 \leftarrow 0, \Sigma \leftarrow \mathbf{I}$ )
- 2: Choose  $\mathcal{G}_d \subseteq \mathcal{G}$  with  $d = d_{\max}(\mathcal{G}) + 1$ . Get  $Q_1, Q_2, R$ , and set  $Q_d \leftarrow Q_1 + dQ_2$
- 3: Set  $\mathcal{P}_0 \leftarrow \beta \mathbf{I}_{p(p+1)/2}$  for large  $\beta > 0$ ,  $K_1$  stabilizing (3.1),  $L_1 = \mathbf{0}$ , and  $\Delta K_1 = K_1$
- 4: Get  $\tilde{\mathbf{x}}_1 \in \mathbb{R}^{dn}$  and  $\tilde{\mathbf{H}}_0 \in \mathbb{R}^{p \times p}$ ,  $p = d(n+m)$  and turn on temporary links in  $\mathcal{G}_d$
- 5: **While** ( $K_k, L_k$ ) **has not converged, do** (“learning phase”)
- 6: Set  $\text{Policy}_k(\mathcal{V}_{\mathcal{G}})$  such that for each  $i \in \mathcal{V}_{\mathcal{G}}$ ,

$$\begin{aligned} \mathbf{u}_i &\leftarrow \Delta K_k \mathbf{x}_i + L_k \sum_{j \in \mathcal{N}_i} \frac{\tau_k}{d-1} \mathbf{x}_j, & \text{if } i \in \mathcal{V}_{\mathcal{G} \setminus \mathcal{G}_d} \\ \mathbf{u}_i &\leftarrow \Delta K_k \mathbf{x}_i + L_k \sum_{j \in \mathcal{V}_{\mathcal{G}_d}} \mathbf{x}_j, & \text{if } i \in \mathcal{V}_{\mathcal{G}_d} \end{aligned}$$

- 7: Evaluate  $\tilde{\mathbf{H}}_k$  from Algorithm 3.2:  $\tilde{\mathbf{H}}_k \leftarrow \text{SPE}(\mathcal{G}, \mathcal{G}_{d,\text{learn}}, \text{Policy}_k(\mathcal{V}_{\mathcal{G}}), \tilde{\mathbf{H}}_{k-1}, \mathcal{P}_0)$
- 8: Recover  $X_1, X_2, Y_1, Y_2, Z_1$  and  $Z_2$  from  $\tilde{\mathbf{H}}_k$

$$\begin{aligned} X_1 &\leftarrow \tilde{\mathbf{H}}_k[1:n, 1:n], & Y_1 &\leftarrow \tilde{\mathbf{H}}_k[dn+1:dn+m, dn+1:dn+m] \\ X_2 &\leftarrow \tilde{\mathbf{H}}_k[1:n, n+1:2n], & Y_2 &\leftarrow \tilde{\mathbf{H}}_k[dn+1:dn+m, dn+m+1:dn+2m+1] \\ Z_1 &\leftarrow \tilde{\mathbf{H}}_k[dn+1:dn+m, 1:n], & Z_2 &\leftarrow \tilde{\mathbf{H}}_k[dn+1:dn+m, n+1:2n] \\ \Delta X &\leftarrow X_1 - X_2, & \Delta Y &\leftarrow Y_1 - Y_2, & \Delta Z &\leftarrow Z_1 - Z_2. \end{aligned}$$

- 9: Update the control components

$$\begin{aligned} F^{-1} &\leftarrow Y_1 - (d-1)Y_2(Y_1 + (d-2)Y_2)^{-1}Y_2, & G &\leftarrow (Y_1 + (d-1)Y_2)^{-1}Y_2(Y_1 - Y_2)^{-1} \\ K_{k+1} &\leftarrow -FZ_1 + (d-1)GZ_2, & L_{k+1} &\leftarrow -FZ_2 + GZ_1 + (d-2)GZ_2, & \Delta K_{k+1} &\leftarrow K_{k+1} - L_{k+1} \end{aligned}$$

- 10: Obtain the stability margin  $\tau_{k+1} \leftarrow \sqrt{\gamma_{k+1}^2 / (1 + \gamma_{k+1})}$  by updating

$$\begin{aligned} \Xi_{k+1} &\leftarrow \Delta X - Q_d + \Delta K_{k+1}^\top \Delta Z + \Delta Z^\top \Delta K_{k+1} + \Delta K_{k+1}^\top (\Delta Y - R) \Delta K_{k+1} \\ \gamma_{k+1} &\leftarrow \lambda_{\min} \left( \Delta K_{k+1}^\top R \Delta K_{k+1} + Q_d \right) / \lambda_{\max} \left( \Xi_{k+1} + L_{k+1}^\top (\Delta Y - R) L_{k+1} \right) \end{aligned}$$

- 11: Go to Line 5 and set  $k \leftarrow k + 1$

- 12: Switch OFF the temporary links and retrieve  $\mathcal{G}_d$

- 13: Set  $\text{Policy}_k(\mathcal{V}_{\mathcal{G}})$  such that for each  $i \in \mathcal{V}_{\mathcal{G}}$ ,  $\mathbf{u}_i \leftarrow \Delta K_k \mathbf{x}_i + \frac{\tau_k}{d-1} L_k \sum_{j \in \mathcal{N}_i} \mathbf{x}_j$
-

vergence of SPE—guaranteed based on the persistence of excitation condition—is followed by the update of  $\tilde{\mathbf{H}}_k$  that encodes the necessary information to obtain  $K_k$  and  $L_k$ . This is achieved by recovering the block matrices  $X_1, X_2, Y_1, Y_2, Z_1$ , and  $Z_2$  from  $\tilde{\mathbf{H}}_k$  that are further utilized to form intermediate variables  $F$  and  $G$ . Matrix inversions on line 9 of Algorithm 3.1 will be justified in Section 3.4 Lemma 3.13. Such recovery of meaningful blocks from  $\tilde{\mathbf{H}}_k$  is due to the specific matrix structure resulting from adding extra links to  $\mathcal{G}_d$ ; this point will be discussed further subsequently. Each iteration loop is completed by updating the parameters  $\gamma_k$  and  $\tau_k$ , that prove instrumental in the stability analysis of the proposed controller. With the convergence of D3SPI,  $\mathcal{G}_d$  is retrieved by removing the temporary links and the structured policy is extended to the entire graph  $\mathcal{G}$ .

Let us point out a few remarks on the computational complexity of the proposed algorithm. First, note that the inverse operations on line 9 occur on matrices of size  $m \times m$ , and hence computationally inexpensive. Furthermore, the complexity of finding extreme singular values—as on line 10 in Algorithm 3.1—is known to be  $\mathcal{O}(n^2)$  [Comon and Golub, 1990]. Hence, the computational complexity of D3SPI is mainly due to the SPE recursion that is equivalent to the complexity of RLS for the number of unknown system parameters in  $\mathcal{G}_d$ , i.e., the computational cost is  $\mathcal{O}(d^2(n+m)^2)$  [Haykin, 2002]. This implies that the computational complexity of the algorithm is *fixed* for any number of agents  $N$ , as long as the maximum degree of the graph retains its order.

*Remark 3.4.* Adding temporary links within the subgraph  $\mathcal{G}_d$  is an effective means of learning optimal  $K_k$  and  $L_k$  for the subgraph  $\mathcal{G}_{d,\text{learn}}$  by utilizing dynamical interdependencies among the agents. Although initializing  $K_k$  such that (3.1) is Schur stable is a standard assumption in data-driven control, obtaining this initial gain for an unknown system is nontrivial. While we invoke this assumption in this work, the interested reader is referred to [Talebi et al., 2020; Chen and Hazan, 2020] for more recent works pertaining to this assumption and related system theoretic issues [Yu et al., 2021].

### 3.4 Analysis of D3SPI

In this section, we provide convergence and stability analyses for the D3SPI algorithm. In this direction, we first study the structure and stability margins of each local controller and proceed to establish stability properties of the proposed controller for the entire network throughout the learning process. Lastly, we show the convergence of D3SPI to a stabilizing suboptimal distributed controller followed by the derivation of a suboptimality bound characterized by the problem parameters. For clarity, we defer some of the subtleties of the analysis and detailed proofs to the end of this section.

First, let us demonstrate how a specific structure and stability of the controller for the subgraph  $\mathcal{G}_{d,\text{learn}}$ , when properly initialized, can be preserved throughout the D3SPI algorithm.

**Proposition 3.5.** *Let  $\tilde{\mathbf{K}}_k := \mathbf{I}_d \otimes (K_k - L_k) + \mathbb{1}\mathbb{1}^\top \otimes L_k$  for all  $k \geq 1$ , with  $K_k$  and  $L_k$  as in Algorithm 3.1. Under Assumption 3.3 and throughout the learning phase (for all  $k \geq 1$ ),  $\tilde{\mathbf{K}}_k$  is stabilizing for the system in  $\mathcal{G}_{d,\text{learn}}$  and  $\text{Policy}_k(\mathcal{V}_{\mathcal{G}_{d,\text{learn}}})|_t = \tilde{\mathbf{K}}_k \text{State}(\mathcal{V}_{\mathcal{G}_{d,\text{learn}}})|_t$ , for all  $t$ . Furthermore,  $\Delta K_k := K_k - L_k$  stabilizes the dynamics of a single agent, i.e.,  $A + B\Delta K_k$  is Schur stable.*

Note that Proposition 3.5 proves the existence of a stabilizing controller  $\Delta K_k$  and its corresponding cost-to-go matrix  $\Delta P_k$ . In the sequel, our goal is to design a distributed suboptimal controller for the entire networked system on  $\mathcal{G}$  based on the components that shape  $\Delta K_k$ . This extension is built upon the stability margin derived next.

**Proposition 3.6.** *At each iteration  $k \geq 1$ , let  $K_k$ ,  $L_k$  and  $\tau_k$  be obtained via Algorithm 3.1. Then,  $A + B(K_k - \alpha L_k)$  is Schur stable for all  $\alpha$  satisfying  $|\alpha - 1| \leq \tau_k$ .*

The stability margin  $\tau_k$  in Proposition 3.6 is upper-bounded by the stability margin of the pair  $(A + B(K_k - L_k), B)$ . This implies that if the original closed-loop system for an agent does not have a favorable stability margin, then  $\tau_k$  can be small—reducing the influence of the



agent's neighbors on its policy (Line 13 of Algorithm 3.1). Nonetheless, Proposition 3.6 provides model-free stability gain margins  $\tau_k$  at each iteration of the algorithm for the dynamics of a single agent in  $\mathcal{G}$ . In our analysis, we take advantage of these margins to characterize stability guarantees for the controller proposed during the learning phase of D3SPI. This is captured in the following result.

**Theorem 3.7.** *Suppose  $K_k$ ,  $L_k$  and  $\tau_k$  are defined as in Algorithm 3.1. Then, under Assumption 3.3, the control policy  $\text{Policy}_k(\mathcal{V}_{\mathcal{G}})$  designed during the learning phase (line 6) stabilizes the network  $\mathcal{G}$  at each iteration of the learning phase and for any choice of  $\mathcal{V}_{\mathcal{G}_d}$ .*

Theorem 3.7 establishes that the proposed feedback mechanism stabilizes the entire network, facilitating control of agents *outside* of  $\mathcal{G}_d$ , during the learning phase. In the meantime, the practicality and suboptimality of the algorithm depend on its convergence addressed next.

**Theorem 3.8.** *Under Assumption 3.3 and (long enough) finite termination of Algorithm 3.2, Algorithm 3.1 converges, i.e.,  $K_k \rightarrow K^*$ ,  $L_k \rightarrow L^*$ , and  $\tau_k \rightarrow \tau^*$  as  $k \rightarrow \infty$ , where  $\tilde{\mathbf{K}}^* = \mathbf{I}_d \otimes (K^* - L^*) + \mathbb{1}\mathbb{1}^\top \otimes L^*$  is the optimal solution to the infinite-horizon state-feedback LQR problem with system parameters  $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{Q}}, \tilde{\mathbf{R}})$  defined as  $\tilde{\mathbf{A}} = \mathbf{I}_d \otimes A$ ,  $\tilde{\mathbf{B}} = \mathbf{I}_d \otimes B$ ,  $\tilde{\mathbf{Q}} = \mathbf{I}_d \otimes (Q_1 + dQ_2) - \mathbb{1}\mathbb{1}^\top \otimes Q_2$ , and  $\tilde{\mathbf{R}} = \mathbf{I}_d \otimes R$ .*

Finally, we note that as the temporary links introduced during the learning phase are removed, the structure of the agents' interaction is once again the original network  $\mathcal{G}$ . As such, it is vital to provide stability guarantees after Algorithm 3.1 terminates and components of the control design have converged. This issue is addressed in the following corollary whose proof is similar to that of Theorem 3.7 and thus omitted for brevity.

**Corollary 3.9.** *Suppose that  $K^*$ ,  $L^*$ ,  $\gamma^*$ , and  $\tau^*$  are given as in Theorem 3.8 under a convergent Algorithm 3.1. Then  $\text{Policy}(\mathcal{V}_{\mathcal{G}})$  (defined on Line 13), stabilizes the entire networked system in (3.2).*

We conclude this section by exploring the suboptimality of the proposed policy. Given the problem parameters, let  $\hat{\mathbf{K}}_{\text{struc}}^*$  denote the globally optimal distributed solution for the

---

**Algorithm 3.2:** Subgraph Policy Evaluation (SPE)
 

---

- 1: **Input:** Graph  $\mathcal{G}$ , subgraph  $\mathcal{G}' \subseteq \mathcal{G}$ ,  $\text{Policy}(\mathcal{V}_{\mathcal{G}})$ ,  $\mathbf{H}$ ,  $\mathcal{P}$
- 2: **Output:** Updated cost matrix  $\mathbf{H}^+$  associated with  $\mathcal{G}'$
- 3: **While**  $\mathbf{H}$  **has not converged**, **do**
- 4: Set  $\tilde{\mathbf{x}}_t \leftarrow \text{State}(\mathcal{V}_{\mathcal{G}'}|_t)$  and  $\tilde{\mathbf{u}}_t \leftarrow \text{Policy}(\mathcal{V}_{\mathcal{G}'}|_t)$
- 5: Choose  $\mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \Sigma)$  and update  $\text{Policy}(\mathcal{V}_{\mathcal{G}'}|_t)$  as  $\tilde{\mathbf{u}}_t \leftarrow \tilde{\mathbf{u}}_t + \mathbf{e}_t$  for all  $i \in \mathcal{V}_{\mathcal{G}'}$
- 6: Run  $\mathcal{G}$  under policy  $\text{Policy}(\mathcal{V}_{\mathcal{G}})$
- 7: Collect  $\text{State}(\mathcal{V}_{\mathcal{G}'}|_{t+1})$  only from  $\mathcal{G}'$  and set

$$\tilde{\mathbf{x}}_{t+1} \leftarrow \text{State}(\mathcal{V}_{\mathcal{G}'}|_{t+1}), \tilde{\mathbf{u}}_{t+1} \leftarrow \text{Policy}(\mathcal{V}_{\mathcal{G}'}|_{t+1})$$

- 8: Set

$$\phi_t \leftarrow [\tilde{\mathbf{x}}_t^\top \tilde{\mathbf{u}}_t^\top]^\top - [\tilde{\mathbf{x}}_{t+1}^\top \tilde{\mathbf{u}}_{t+1}^\top]^\top$$

- 9: Compute  $\zeta_t = \text{vech}(\phi_t \phi_t^\top)$  and

$$\mathcal{R}(\tilde{\mathbf{x}}_t, \tilde{\mathbf{u}}_t) = \tilde{\mathbf{x}}_t^\top (\mathbf{I} \otimes Q_d - \mathbf{1}\mathbf{1}^\top \otimes Q_2) \tilde{\mathbf{x}}_t + \tilde{\mathbf{u}}_t^\top (\mathbf{I} \otimes R) \tilde{\mathbf{u}}_t$$

- 10: Set  $\boldsymbol{\theta} \leftarrow \text{vech}(\mathbf{H})$  and update

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \mathcal{P} \zeta_t (\mathcal{R}(\tilde{\mathbf{x}}_t, \tilde{\mathbf{u}}_t) - \zeta_t^\top \boldsymbol{\theta}) / (1 + \zeta_t^\top \mathcal{P} \zeta_t),$$

$$\mathcal{P} \leftarrow \mathcal{P} - \mathcal{P} \zeta_t \zeta_t^\top \mathcal{P} / (1 + \zeta_t^\top \mathcal{P} \zeta_t)$$

- 11: Find  $\mathbf{H}^+ = \text{vech}^{-1}(\boldsymbol{\theta})$ , update  $\mathbf{H} \leftarrow \mathbf{H}^+$  and  $t \leftarrow t + 1$
-

structured LQR problem in (3.3) with the associated cost matrix  $\hat{\mathbf{P}}_{\text{struct}}^*$ . Given any other stabilizing structured policy  $\hat{\mathbf{K}}$  associated with cost matrix  $\hat{\mathbf{P}}$ , we define the *optimality gap* as  $\text{gap}(\hat{\mathbf{K}}) := \text{Tr}[\hat{\mathbf{P}} - \hat{\mathbf{P}}_{\text{struct}}^*]$ . The following theorem provides an upperbound on the optimality gap of structured policy learned by D3SPI based on the problem parameters. In particular, when the system is “contractible,” the derived upperbound depends on the difference of the distributed controller with that of *unstructured* optimal LQR controller.

**Theorem 3.10.** *Let  $\hat{\mathbf{K}}^*$  be the structured policy learned by Algorithm 3.1 at convergence, corresponding to the cost matrix  $\hat{\mathbf{P}}^*$ . Moreover, let  $\hat{\mathbf{K}}_{\text{lqr}}$  denote the optimal (unstructured) solution to the infinite-horizon state-feedback LQR problem with parameters  $(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{Q}}, \hat{\mathbf{R}})$  with the cost matrix  $\hat{\mathbf{P}}_{\text{lqr}}$ . If  $\hat{\mathbf{A}}_{\hat{\mathbf{K}}_{\text{lqr}}} = \hat{\mathbf{A}} + \hat{\mathbf{B}}\hat{\mathbf{K}}_{\text{lqr}}$  is contractible then  $0 \leq \text{gap}(\hat{\mathbf{K}}^*) \leq \text{Tr}(\mathbf{M})/[1 - \sigma_{\max}^2(\hat{\mathbf{A}}_{\hat{\mathbf{K}}_{\text{lqr}}})]$ , where  $\mathbf{M} := (\hat{\mathbf{R}} + \hat{\mathbf{B}}^\top \hat{\mathbf{P}}^* \hat{\mathbf{B}})(\hat{\mathbf{K}}^* \hat{\mathbf{K}}^{*\top} - \hat{\mathbf{K}}_{\text{lqr}} \hat{\mathbf{K}}_{\text{lqr}}^\top) + 2\hat{\mathbf{A}}^\top \hat{\mathbf{P}}^* \hat{\mathbf{B}}(\hat{\mathbf{K}}^* - \hat{\mathbf{K}}_{\text{lqr}})$ .*

*Remark 3.11.* First, note how the converged policy by Theorem 3.8 is related to the optimal LQR policy on the fully connected subgraph  $\mathcal{G}_{d,\text{learn}}$ . Second, contractibility of the pair  $(A, B)$  is more restrictive than regularizability of the system [Talebi et al., 2020], a notion that has recently been employed in iterative data-guided control methods [Lale et al., 2020b; Agarwal et al., 2019]. Contractibility also facilitates the validity of assuming access to the initial stabilizing controller.

Next, we provide the building blocks needed for the proofs and analysis of our algorithm. We first provide some insights on how the setup is connected to the classic model-based LQR machinery and some previously established results that we leverage from the literature. The main proofs then follows.

#### 3.4.1 Underlying Model of the Subsystem $\mathcal{G}_d$

The configuration of the synthesis problem in D3SPI intertwines an online recursion on the subsystem corresponding to  $\mathcal{G}_d$  and the original system  $\mathcal{G}$ . In particular, during the learning phase, considering the same cost structure and problem parameters as in (3.3)—but

for the completed subgraph  $\mathcal{G}_{d,\text{learn}}$ — results in the  $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{Q}}, \tilde{\mathbf{R}})$  parameters as defined in Theorem 3.8. Then, similarly from (3.1) the dynamics of the subgraph  $\mathcal{G}_{d,\text{learn}}$  assumes the form,

$$\tilde{\mathbf{x}}_{t+1} = \tilde{\mathbf{A}}\tilde{\mathbf{x}}_t + \tilde{\mathbf{B}}\tilde{\mathbf{u}}_t, \quad (3.4)$$

where  $\tilde{\mathbf{x}}_t$  and  $\tilde{\mathbf{u}}_t$  are formed from concatenation of state and control signals in  $\mathcal{G}_{d,\text{learn}}$ — recall that  $\tilde{\mathbf{u}}_t$  is also denoted by  $\text{Policy}(\mathcal{V}_{\mathcal{G}_{d,\text{learn}}})|_t$  in the algorithm to emphasize the temporal implementation of a specific policy in Algorithm 3.2. From the Bellman equation [Bellman, 1966] for the LQR problem with these parameters, the cost matrix  $\tilde{\mathbf{P}}_k$  of  $\mathcal{G}_{d,\text{learn}}$  is correlated with the one-step LQR cost as,

$$\tilde{\mathbf{x}}_t^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{x}}_t = \mathcal{R}(\tilde{\mathbf{x}}_t, \tilde{\mathbf{u}}_t) + \tilde{\mathbf{x}}_{t+1}^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{x}}_{t+1}, \quad (3.5)$$

where  $\mathcal{R}(\tilde{\mathbf{x}}_t, \tilde{\mathbf{u}}_t) = \tilde{\mathbf{x}}_t^\top \tilde{\mathbf{Q}} \tilde{\mathbf{x}}_t + \tilde{\mathbf{u}}_t^\top \tilde{\mathbf{R}} \tilde{\mathbf{u}}_t$  and  $\tilde{\mathbf{P}}_k$  satisfies the Lyapunov equation,

$$\tilde{\mathbf{P}}_k = (\tilde{\mathbf{A}} + \tilde{\mathbf{B}}\tilde{\mathbf{K}}_k)^\top \tilde{\mathbf{P}}_k (\tilde{\mathbf{A}} + \tilde{\mathbf{B}}\tilde{\mathbf{K}}_k) + \tilde{\mathbf{Q}} + \tilde{\mathbf{K}}_k^\top \tilde{\mathbf{R}} \tilde{\mathbf{K}}_k, \quad (3.6)$$

and  $\tilde{\mathbf{K}}_k$  is the controller policy at iteration  $k$ . The dynamic programming solution to the LQR problem suggests a linear feedback form  $\tilde{\mathbf{u}}_t = \tilde{\mathbf{K}}_k \tilde{\mathbf{x}}_t$  for the subsystem  $\mathcal{G}_{d,\text{learn}}$  at each iteration. Combining (3.4) and (3.5) with some rearrangements result in

$$\begin{aligned} \tilde{\mathbf{x}}_t^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{x}}_t &= \tilde{\mathbf{z}}_t^\top \begin{bmatrix} \tilde{\mathbf{Q}} + \tilde{\mathbf{A}}^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{A}} & \tilde{\mathbf{A}}^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{B}} \\ \tilde{\mathbf{B}}^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{A}} & \tilde{\mathbf{R}} + \tilde{\mathbf{B}}^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{B}} \end{bmatrix} \tilde{\mathbf{z}}_t \\ &=: \tilde{\mathbf{z}}_t^\top \begin{bmatrix} [\tilde{\mathbf{H}}_k]_{11} & [\tilde{\mathbf{H}}_k]_{12} \\ [\tilde{\mathbf{H}}_k]_{21} & [\tilde{\mathbf{H}}_k]_{22} \end{bmatrix} \tilde{\mathbf{z}}_t = \tilde{\mathbf{z}}_t^\top \tilde{\mathbf{H}}_k \tilde{\mathbf{z}}_t, \end{aligned} \quad (3.7)$$

where  $\tilde{\mathbf{z}}_t = [\tilde{\mathbf{x}}_t^\top \tilde{\mathbf{u}}_t^\top]^\top$ . Then, the following policy update (due to Hewer) is guaranteed to converge to the optimal LQR policy under controllability assumption [Hewer, 1971]:

$$\tilde{\mathbf{K}}_{k+1} = - \left( \tilde{\mathbf{R}} + \tilde{\mathbf{B}}^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{B}} \right)^{-1} \tilde{\mathbf{B}}^\top \tilde{\mathbf{P}}_k \tilde{\mathbf{A}} = - [\tilde{\mathbf{H}}_k]_{22}^{-1} [\tilde{\mathbf{H}}_k]_{21}, \quad (3.8)$$

which is also reconstructed by information in  $\tilde{\mathbf{H}}_k$ . Furthermore, the cost matrix in (3.6) can also be reconstructed by the same information:

$$\tilde{\mathbf{P}}_k = [\tilde{\mathbf{H}}_k]_{11} + [\tilde{\mathbf{H}}_k]_{12}\tilde{\mathbf{K}}_k + \tilde{\mathbf{K}}_k^\top[\tilde{\mathbf{H}}_k]_{21} + \tilde{\mathbf{K}}_k^\top[\tilde{\mathbf{H}}_k]_{22}\tilde{\mathbf{K}}_k. \quad (3.9)$$

Hence,  $\tilde{\mathbf{H}}_k$  provides the required information to perform both policy update and policy evaluation steps in a policy iteration algorithm. We will see that because of the particular structure of our setup,  $\tilde{\mathbf{H}}_k$  enjoys a special block pattern captured by the proposed patterned linear semigroup, justifying the recovery of the block matrices  $X_1, X_2, Y_1, Y_2, Z_1$ , and  $Z_2$  from  $\tilde{\mathbf{H}}_k$  in D3SPI. D3SPI leverages this idea to implicitly learn  $\tilde{\mathbf{H}}_k$  from data (by adapting the idea of [Bradtke et al., 1994]) and exploit these matrix blocks in order to find a suboptimal solution to the main distributed problem in (3.3).

Here, in addition to the policy update from data as in (3.8), we show that the same information can be used to also learn a gain margin directly from data (see Proposition 3.6). This gain margin is then used to guarantee the stability of the entire network (see Theorem 3.7).

Finally, for technical reasons, recall that the infinite-horizon state-feedback LQR problem with parameters  $(A, B, Q, R)$  can be cast as the minimization of

$$f_\Sigma(K) := \text{Tr}[P_K \Sigma] \quad (3.10)$$

over the static stabilizing policy  $K$ , for some initial state distribution with covariance  $\Sigma \succ 0$ , where  $P_K$  is cost matrix associated with  $K$  satisfying the following Lyapunov equation [Levine et al., 1971; Mårtensson and Rantzer, 2009; Bu et al., 2019a]:

$$P_K = (A + BK)^\top P_K (A + BK) + Q + K^\top R K.$$

Herein, we set  $\Sigma = \mathbf{I}$  and consider  $f_{\mathbf{I}}(K)$ .

### 3.4.2 Main Theoretical Results

In the remainder of this section, we first restate some well-known facts to make the chapter self-contained, and then propose a few additional algebraic facts for our analysis whose proofs are deferred to Appendix 3.4.3. We then continue with the proof of the main results.

**Lemma 3.12.** *The following relations hold:*

1. ([Horn and Johnson, 2012]) *When  $X \succ 0$ ,*

$$\begin{aligned} M^\top X N + N^\top X M &\succeq -(aM^\top X M + \frac{1}{a}N^\top X N), \\ M^\top X N + N^\top X M &\preceq aM^\top X M + \frac{1}{a}N^\top X N, \end{aligned}$$

where  $M, N \in \mathbb{R}^{n \times m}$  with  $n \geq m$  and  $a > 0$ .

2. ([Lancaster, 1970, Lyapunov Equation]) *Suppose that  $A \in \mathbb{R}^{n \times n}$  has spectral radius less than 1, i.e.,  $\rho(A) < 1$ . Then  $A^\top X A + Q - X = 0$  has a unique solution,  $X = \sum_{j=0}^{\infty} (A^\top)^j Q A^j$ . In this case, if  $Q \succ 0$ , then  $X \succ 0$ .*

3. ([Horn and Johnson, 2012, Block matrix inverse formula (0.8.5.6)]) *The following equation holds for matrices  $A, B, C$ , and  $D$  with compatible dimensions,*

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} H^{-1} & -H^{-1}BD^{-1} \\ -D^{-1}CH^{-1} & D^{-1} + D^{-1}CH^{-1}BD^{-1} \end{bmatrix},$$

where  $D$  and  $H = A - BD^{-1}C$  are invertible.

4. (Matrix Inversion Lemma [Woodbury, 1950]) *The following holds,*

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1},$$

for matrices  $A, U, C$ , and  $V$  with compatible dimensions where  $A, C$ , and  $A + UCV$  are invertible.

Finally, we provide the main technical lemma that streamlines the properties of the patterned linear semigroup under algebraic manipulation which will be frequently used in the proof of Proposition 3.5 and Theorem 3.8.

**Lemma 3.13.** *Suppose  $\mathbf{N}_r \in \text{PL}(r \times n, \mathbb{R})$  for some  $n$  and  $r \geq 2$  such that  $\mathbf{N}_r = \mathbf{I}_r \otimes (A - B) + \mathbf{1}_r \mathbf{1}_r^\top \otimes B$ , for some  $A \in \text{GL}(n, \mathbb{R}) \cap \mathbb{S}^{n \times n}$  and  $B \in \mathbb{S}^{n \times n}$ . Then the following hold:*

1.  $\det(\mathbf{N}_r) = \det(A - B)^{r-1} \det(A + (r - 1)B)$ .
2. If  $\mathbf{N}_r \succ 0$ , then we have  $A + (\ell - 1)B \succ 0$  for all  $\ell = 0, 1, \dots, r$ . Furthermore,  $A - \ell B (A + (\ell - 1)B)^{-1} B$  is invertible for  $\ell = 1, 2, \dots, r - 1$ .
3. If  $\mathbf{N}_r \succ 0$ , then  $\mathbf{N}_r^{-1} \in \text{PL}(r \times n, \mathbb{R})$ , i.e.,

$$\mathbf{N}_r^{-1} = \mathbf{I}_r \otimes (F_r + G_r) - \mathbb{1}_r \mathbb{1}_r^\top \otimes G_r,$$

with  $F_r$  and  $G_r$  defined as,

$$\begin{aligned} F_r &= (A - (r - 1)B(A + (r - 2)B)^{-1}B)^{-1}, \\ G_r &= (A + (r - 1)B)^{-1}B(A - B)^{-1}. \end{aligned}$$

4. If  $\mathbf{M}_r = \mathbf{I}_r \otimes (C - D) + \mathbb{1}_r \mathbb{1}_r^\top \otimes D$  then,

$$\mathbf{N}_r \mathbf{M}_r = \mathbf{I}_r \otimes (A - B)(C - D) + \mathbb{1}_r \mathbb{1}_r^\top \otimes (B(C - D) + (A - B)D + rBD).$$

#### *Proof of Proposition 3.1*

For any Schur stable matrix  $A$ , and any symmetric positive definite matrix  $Q$  there exists a unique positive definite solution  $P$  to the discrete Lyapunov equation described by  $P = \sum_{j=0}^{\infty} (A^\top)^j Q A^j$  (Lemma 3.12.2). Note, that  $\text{PL}(r \times n, \mathbb{R}) \subset \mathbb{S}^{rn \times rn}$  by construction. Therefore, since  $\text{PL}(r \times n, \mathbb{R})$  is a linear group, each summand also falls in  $\text{PL}(r \times n, \mathbb{R})$  whenever  $Q \in \text{PL}(r \times n, \mathbb{R})$ . Thus, as the infinite sum preserves the structure,  $P \in \text{PL}(r \times n, \mathbb{R})$  whenever  $Q \in \text{PL}(r \times n, \mathbb{R})$ . Conversely, if  $P \in \text{PL}(r \times n, \mathbb{R})$ , then  $Q = P - A^\top P A^\top$  also must lie in  $\text{PL}(r \times n, \mathbb{R})$  as  $Q \succ 0$ . This completes the proof.  $\square$

#### *Proof of Proposition 3.5*

At iteration  $k$  of the learning phase in Algorithm 3.1, the first claim is a direct consequence of the structure of  $\mathcal{G}_{d,\text{learn}}$  during the learning phase, where  $\mathcal{G}_{d,\text{learn}} = \mathcal{K}(\mathcal{G}_d)$  and hence

$\mathbf{u}_i = (K_k - L_k)\mathbf{x}_i + L_k \sum_{j \in \mathcal{V}_{\mathcal{G}_d, \text{learn}}} \mathbf{x}_j$  for all  $i \in \mathcal{V}_{\mathcal{G}_d, \text{learn}}$ , which, in turn, results in  $\tilde{\mathbf{u}}_t = \tilde{\mathbf{K}}_k \tilde{\mathbf{x}}_t$  with  $\tilde{\mathbf{K}}_k$  as claimed. The stability of the policy  $\tilde{\mathbf{K}}_k$  for the pair  $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$  throughout the learning phase is then argued in Section 3.4.2 under Assumption 3.3.

Next, the cost matrix  $\tilde{\mathbf{P}}_k$  associated with  $\tilde{\mathbf{K}}_k$  satisfies the Lyapunov equation (3.6). We can verify that  $\tilde{\mathbf{A}} + \tilde{\mathbf{B}}\tilde{\mathbf{K}}_k \in \text{L}(d \times n, \mathbb{R})$  which is also Schur stable, and  $\tilde{\mathbf{Q}} + \tilde{\mathbf{K}}_k^\top \tilde{\mathbf{R}} \tilde{\mathbf{K}}_k \in \text{PL}(d \times n, \mathbb{R})$ . Thus, by Proposition 3.1, we conclude that  $\tilde{\mathbf{P}}_k \in \text{PL}(d \times n, \mathbb{R})$ . So, let

$$\tilde{\mathbf{P}}_k = \text{I}_d \otimes (P_1 - P_2) + \mathbb{1}\mathbb{1}^\top \otimes P_2, \quad (3.11)$$

for some  $P_1, P_2$ . Note that  $\tilde{\mathbf{K}}_k$  is stabilizing and  $\tilde{\mathbf{Q}} + \tilde{\mathbf{K}}_k^\top \tilde{\mathbf{R}} \tilde{\mathbf{K}}_k \succ 0$ , therefore  $\tilde{\mathbf{P}}_k \succ 0$  from (3.6). But then, by Lemma 3.13.Item 2 and the structure of  $\tilde{\mathbf{P}}_k$  from (3.11), we claim that  $\Delta P_k := P_1 - P_2 \succ 0$ . Next, one can also verify that

$$\begin{aligned} (\tilde{\mathbf{A}} + \tilde{\mathbf{B}}\tilde{\mathbf{K}}_k)^\top \tilde{\mathbf{P}}_k (\tilde{\mathbf{A}} + \tilde{\mathbf{B}}\tilde{\mathbf{K}}_k) &= \text{I}_d \otimes (A_{\Delta K_k}^\top (\Delta P_k) A_{\Delta K_k}) + \mathbb{1}\mathbb{1}^\top \otimes (\star), \\ \tilde{\mathbf{Q}} + \tilde{\mathbf{K}}_k^\top \tilde{\mathbf{R}} \tilde{\mathbf{K}}_k &= \text{I}_d \otimes (Q_d + (\Delta K_k)^\top R (\Delta K_k)) + \mathbb{1}\mathbb{1}^\top \otimes (\star), \end{aligned}$$

where  $\Delta K_k = K_k - L_k$ ,  $A_{\Delta K_k} = A + B(\Delta K_k)$ ,  $Q_d = Q_1 + dQ_2$ , and  $(\star)$  is hiding extra terms. But then, by (3.6) and (3.11) we obtain that  $\Delta P_k$  must satisfy:

$$\Delta P_k = A_{\Delta K_k}^\top (\Delta P_k) A_{\Delta K_k} + Q_d + \Delta K_k^\top R \Delta K_k, \quad (3.12)$$

which itself is a Lyapunov equation. Finally, since  $Q_d + \Delta K_k^\top R \Delta K_k \succ 0$  and  $\Delta P_k \succ 0$ , by Lyapunov Stability Criterion we conclude that  $A_{\Delta K_k}$  is Schur stable. This completes the proof.  $\square$

### *Proof of Proposition 3.6*

Define the Lyapunov candidate function  $V_k(\mathbf{x}_t) = \mathbf{x}_t^\top \Delta P_k \mathbf{x}_t$  with  $\Delta P_k \succ 0$  as in (3.12) and where  $\mathbf{x}_t$  contains the states of the closed-loop system  $\mathbf{x}_{t+1} = (A + B(K_k - \alpha L_k))\mathbf{x}_t$  for some scalar  $\alpha$ . We show that for the given choice of  $\alpha$ ,  $V_k$  is decreasing. We define

$$\Delta V_k(\mathbf{x}_t) := V_k(\mathbf{x}_{t+1}) - V_k(\mathbf{x}_t) = \mathbf{x}_t^\top \Gamma_k \mathbf{x}_t,$$



where

$$\Gamma_k = (A_{\Delta K_k} + (1 - \alpha)L_k)^\top \Delta P_k (A_{\Delta K_k} + (1 - \alpha)L_k) - \Delta P_k.$$

Suppose  $\alpha < 1$ , then from (3.12),

$$\begin{aligned} \Gamma_k &= - (Q_d + \Delta K_k^\top R \Delta K_k) + (1 - \alpha)^2 L_k^\top B^\top \Delta P_k B L_k \\ &\quad + (1 - \alpha) (A_{\Delta K_k}^\top \Delta P_k B L_k + L_k^\top B^\top \Delta P_k A_{\Delta K_k}) \\ &\preceq - (Q_d + \Delta K_k^\top R \Delta K_k) + (1 - \alpha)^2 L_k^\top B^\top \Delta P_k B L_k \\ &\quad + (1 - \alpha) (a A_{\Delta K_k}^\top \Delta P_k A_{\Delta K_k} + (1/a) L_k^\top B^\top \Delta P_k B L_k) \\ &= - (Q_d + \Delta K_k^\top R \Delta K_k) + (1 - \alpha) a A_{\Delta K_k}^\top \Delta P_k A_{\Delta K_k} \\ &\quad + (1 - \alpha) (1/a + 1 - \alpha) L_k^\top B^\top \Delta P_k B L_k \end{aligned}$$

where the inequality holds for any  $a > 0$  due to Lemma 3.12. Let  $\beta = (1 - \alpha)/2$  and choose  $a = \beta + \sqrt{\beta^2 + 1}$ . Then,  $(1/a + 1 - \alpha) = a$  and thus

$$\lambda_{\max}(\Gamma_k) \leq -\lambda_{\min}[Q_d + \Delta K_k^\top R \Delta K_k] + (1 - \alpha) a \lambda_{\max}[A_{\Delta K_k}^\top \Delta P_k A_{\Delta K_k} + L_k^\top B^\top \Delta P_k B L_k]$$

Now, using the parameters (3.13) constructing the blocks of (3.7), we obtain

$$L_k^\top B^\top \Delta P_k B L_k = L_k^\top (\Delta Y - R) L_k,$$

$$A_{\Delta K_k}^\top \Delta P_k A_{\Delta K_k} = \Delta X - Q_d + \Delta K_k^\top \Delta Z + \Delta Z^\top \Delta K_k + \Delta K_k^\top (\Delta Y - R) \Delta K_k =: \Xi_k.$$

Thus, the latter bound can be obtained completely from data as

$$\lambda_{\max}(\Gamma_k) \leq ((1 - \alpha)a - \gamma_k) \lambda_{\max}[\Xi_k + L_k^\top (\Delta Y - R) L_k]$$

with

$$\gamma_k := \lambda_{\min}[Q_d + \Delta K_k^\top R \Delta K_k] / \lambda_{\max}[\Xi_k + L_k^\top (\Delta Y - R) L_k],$$

which coincides with updates in Line 10 of Algorithm 3.1.

Finally, from the hypothesis  $1 - \tau_k < \alpha < 1$  with  $\tau_k = \sqrt{\gamma_k^2 / (1 + \gamma_k)}$ , we obtain that  $\gamma_k^2 - 4\beta^2 \gamma_k - 4\beta^2 > 0$ . But, since  $\gamma_k > 0$ , this second-order term in  $\gamma_k$  is positive only if

$$\gamma_k > 2\beta^2 + 2\beta \sqrt{\beta^2 + 1} = 2\beta(\beta + \sqrt{\beta^2 + 1}) = (1 - \alpha)a.$$

Therefore,  $\Delta V_k(\mathbf{x}_t) < 0$  for  $1 - \tau_k < \alpha < 1$ . Similar reasoning for  $1 < \alpha < 1 + \tau_k$  also shows that  $\Delta V_k(\mathbf{x}_t) < 0$  which completes the proof.  $\square$

*Proof of Theorem 3.7*

From Definition 3.2 (according to a consistent choice of labeling of the nodes so that the last  $d$  nodes are chosen as  $\mathcal{G}_d$ ) and Proposition 3.5, the feedback policy in line 6 of Algorithm 3.1 can be cast in the compact form,

$$\begin{aligned} \text{Policy}_k(\mathcal{V}_{\mathcal{G}}) &= \begin{bmatrix} \text{Policy}_k(\mathcal{V}_{\mathcal{G} \setminus \mathcal{G}_d, \text{learn}}) \\ \text{Policy}_k(\mathcal{V}_{\mathcal{G}_d, \text{learn}}) \end{bmatrix} \\ &= \begin{bmatrix} \hat{\mathbf{K}}_{\mathcal{G} \setminus \mathcal{G}_d} & \hat{\mathbf{L}}_{\mathcal{G} \setminus \mathcal{G}_d} \\ \mathbf{0} & \tilde{\mathbf{K}}_k \end{bmatrix} \begin{bmatrix} \text{State}_k(\mathcal{V}_{\mathcal{G} \setminus \mathcal{G}_d, \text{learn}}) \\ \text{State}_k(\mathcal{V}_{\mathcal{G}_d, \text{learn}}) \end{bmatrix} \\ &=: \hat{\mathbf{K}}_k \text{State}_k(\mathcal{G}), \end{aligned}$$

where  $\hat{\mathbf{L}}_{\mathcal{G} \setminus \mathcal{G}_d} := \frac{\tau_k}{d-1} [\mathcal{A}_{\mathcal{G}}]_{12} \otimes L_k$ ,  $\tilde{\mathbf{K}}_k := \mathbf{I}_d \otimes (K_k - L_k) + \mathbb{1}_d \mathbb{1}_d^T \otimes L_k$ ,  $\hat{\mathbf{K}}_{\mathcal{G} \setminus \mathcal{G}_d} := \mathbf{I}_{N-d} \otimes K_k - (\mathbf{I}_{N-d} - \frac{\tau_k}{d-1} \mathcal{A}_{\mathcal{G} \setminus \mathcal{G}_d}) \otimes L_k$ ,  $\mathcal{A}_{\mathcal{G}}$  denotes the adjacency matrix of  $\mathcal{G}$ , and  $[\mathcal{A}_{\mathcal{G}}]_{12}$  is its submatrix capturing the interconnection of  $\mathcal{G} \setminus \mathcal{G}_d$  and  $\mathcal{G}_d$ :

$$\mathcal{A}_{\mathcal{G}} = \begin{bmatrix} \mathcal{A}_{\mathcal{G} \setminus \mathcal{G}_d} & [\mathcal{A}_{\mathcal{G}}]_{12} \\ * & \mathcal{A}_{\mathcal{G}_d} \end{bmatrix}.$$

Note that the structure of  $\hat{\mathbf{K}}_k$  emanates from the fact that the information exchange is unidirectional during the learning phase. Now consider the closed-loop system of  $\mathcal{G}$ ,

$$\hat{\mathbf{A}}_{\mathcal{G}|cl} := \hat{\mathbf{A}} + \hat{\mathbf{B}}\hat{\mathbf{K}}_k = \begin{bmatrix} \mathbf{I}_{N-d} \otimes A + (\mathbf{I}_{N-d} \otimes B)\hat{\mathbf{K}}_{\mathcal{G} \setminus \mathcal{G}_d} & (\mathbf{I}_{N-d} \otimes B)\hat{\mathbf{L}}_{\mathcal{G} \setminus \mathcal{G}_d} \\ \mathbf{0} & \tilde{\mathbf{A}}_{\tilde{\mathbf{K}}_k} \end{bmatrix},$$

where  $\tilde{\mathbf{A}}_{\tilde{\mathbf{K}}_k} := \tilde{\mathbf{A}} + \tilde{\mathbf{B}}\tilde{\mathbf{K}}_k$  is the closed-loop system of  $\mathcal{G}_d$ . Define  $S = \mathbf{I}_{N-d} - \frac{\tau_k}{d-1} \mathcal{A}_{\mathcal{G} \setminus \mathcal{G}_d}$  and let  $J$  be the Jordan form of  $S$  according to the similarity transformation  $\mathcal{T} \in \mathbb{R}^{(N-d) \times (N-d)}$

such that  $\mathcal{T}S\mathcal{T}^{-1} = J$ . Now consider the following similarity transformation of  $\hat{\mathbf{A}}_{\mathcal{G}}|_{cl}$ ,

$$\begin{aligned} & \begin{bmatrix} \mathcal{T} \otimes \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_d \otimes \mathbf{I}_n \end{bmatrix} \left( \hat{\mathbf{A}}_{\mathcal{G}}|_{cl} \right) \begin{bmatrix} \mathcal{T} \otimes \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_d \otimes \mathbf{I}_n \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbf{I}_{N-d} \otimes (A + BK_k) - J \otimes BL_k & * \\ \mathbf{0} & \tilde{\mathbf{A}}_{\tilde{\mathbf{K}}_k} \end{bmatrix}. \end{aligned}$$

Note that  $\mathbf{I}_{N-d} \otimes (A + BK_k) - J \otimes BL_k$  is a block upper triangular matrix whose diagonal blocks are equal to  $A + B(K_k - \lambda_i(S)L_k)$  for  $i = 1, \dots, N - d$ . We already know from Proposition 3.5 that  $\tilde{\mathbf{A}}_{\tilde{\mathbf{K}}_k}$  is Schur stable. Hence, in order to show that  $\hat{\mathbf{A}}_{\mathcal{G}}|_{cl}$  is Schur stable, it suffices to show that  $\rho(A + B(K - \lambda_i(S)L)) < 1$  for  $i = 1, \dots, N - d$ . Recall that  $|\lambda_i(\mathcal{A}_{\mathcal{G}})| \leq d_{\max}$  [Bollobás, 2013], thus by definition of  $S$  and the fact that  $d_{\max} = d - 1$ , we conclude that  $|\lambda_i(S) - 1| \leq \tau_k$ . The rest of the proof now follows directly from Proposition 3.6.  $\square$

### *Proof of Theorem 3.8*

At iteration  $k$  of the learning phase in Algorithm 3.1, consider  $\tilde{\mathbf{H}}_k$  and its corresponding blocks as defined in (3.7). First, we consider the structure of the stabilizing feedback policy  $\tilde{\mathbf{K}}_k$  as shown in Proposition 3.5, together with that of system parameters  $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{Q}}, \tilde{\mathbf{R}})$ , and apply Lemma 3.13 and Proposition 3.1 to conclude that  $[\tilde{\mathbf{H}}_k]_{11}, [\tilde{\mathbf{H}}_k]_{22} \in \text{PL}(d \times n, \mathbb{R})$  and  $[\tilde{\mathbf{H}}_k]_{21} \in \text{L}(d \times n, \mathbb{R})$ . Thus, we get

$$\begin{aligned} [\tilde{\mathbf{H}}_k]_{11} &= \mathbf{I}_d \otimes (X_1 - X_2) + \mathbb{1}\mathbb{1}^\top \otimes X_2, \\ [\tilde{\mathbf{H}}_k]_{22} &= \mathbf{I}_d \otimes (Y_1 - Y_2) + \mathbb{1}\mathbb{1}^\top \otimes Y_2, \\ [\tilde{\mathbf{H}}_k]_{21} &= \mathbf{I}_d \otimes (Z_1 - Z_2) + \mathbb{1}\mathbb{1}^\top \otimes Z_2, \end{aligned}$$

which coincides with the recovery of  $X_i, Y_i$  and  $Z_i$  for  $i = 1, 2$  in Line 8 of Algorithm 3.1. We can also unravel the structure of block matrices constructing  $\tilde{\mathbf{H}}_k$  and obtain that

$$\begin{aligned} X_1 &= Q_1 + (d-1)Q_2 + A^\top P_1 A, & Y_2 &= B^\top P_2 B, \\ X_2 &= -Q_2 + A^\top P_2 A, & Z_1 &= B^\top P_1 A, \\ Y_1 &= R + B^\top P_1 B, & Z_2 &= B^\top P_2 A, \end{aligned} \tag{3.13}$$

with  $P_1$  and  $P_2$  as in (3.11). Then, by Lemma 3.13, we similarly get:

$$[\tilde{\mathbf{H}}_k]_{22}^{-1} = \mathbf{I}_d \otimes (F - G) + \mathbb{1}\mathbb{1}^\top \otimes G,$$

where (by Lemma 3.13 Item 3)  $F$  and  $G$  must satisfy

$$\begin{aligned} F^{-1} &= Y_1 - (d-1)Y_2(Y_1 + (d-2)Y_2)^{-1}Y_2, \\ G &= \left(Y_1 + (d-1)Y_2\right)^{-1}Y_2\left(Y_1 - Y_2\right)^{-1}, \end{aligned}$$

which coincides with the definitions in Line 9 of Algorithm 3.1. Finally, by definition of  $K_{k+1}$  and  $L_{k+1}$  in Line 9 and Lemma 3.13 Item 4, one can verify that  $\tilde{\mathbf{K}}_{k+1} = -[\tilde{\mathbf{H}}_k]_{22}^{-1}[\tilde{\mathbf{H}}_k]_{21}$ , which coincides with the policy iteration in the Hwer's algorithm [Hewer, 1971] for the system in  $\mathcal{G}_{d,\text{learn}}$  (see also Section 3.4.1). Note that by assumption the pair  $(A, B)$  is controllable, so is the system  $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$  in  $\mathcal{G}_{d,\text{learn}}$ . Therefore, these updates are guaranteed to remain stabilizing and converge to the claimed optimal LQR policy  $\tilde{\mathbf{K}}^*$  provided that we have access to the true parameters  $\tilde{\mathbf{H}}_k$ .

Next, consider LQR cost  $f_I(\tilde{\mathbf{K}})$  as in (3.10) but for the problem parameters  $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{Q}}, \tilde{\mathbf{R}})$ . For completing the proof of convergence, it is left to argue that there exists a large enough integer  $C$  such that, at each iteration  $k$  of the learning phase, the recursive least square in Algorithm 3.2 provides a more accurate estimation, denoted by  $\boldsymbol{\theta}_k^\circ$ , of the true parameters  $\boldsymbol{\theta}_k = \text{vech}(\tilde{\mathbf{H}}_k)$ , and the LQR cost  $f_I(\tilde{\mathbf{K}}) = \text{Tr}(\tilde{\mathbf{P}}_{\tilde{\mathbf{K}}})$  decreases. This claim essentially follows by [Bradtke et al., 1994, Theorem 1] which we try to summarize for completeness. For that, at iteration  $k$ , let us denote the policy obtained using the estimated parameters by  $\tilde{\mathbf{K}}_k^\circ$  which in turn estimates the true policy  $\tilde{\mathbf{K}}_k$ . We then define a ‘‘Lyapunov’’ function candidate

$$s_k := f_I(\tilde{\mathbf{K}}_{k-1}^\circ) + \|\boldsymbol{\theta}_{k-2} - \boldsymbol{\theta}_{k-2}^\circ\|.$$

Following the same induction reasoning as that of [Bradtke et al., 1994, Theorem 1] and under persistently exciting input, there exists an integer  $C$  such that

$$s_{k+1} = s_k - \varepsilon_1(C) \|\boldsymbol{\theta}_{k-1} - \boldsymbol{\theta}_{k-1}^\circ\| - \varepsilon_2(C) \|\tilde{\mathbf{K}}_k - \tilde{\mathbf{K}}_{k-1}^\circ\|^2,$$

for some positive constants  $\varepsilon_1(C)$  and  $\varepsilon_2(C)$  that are independent of  $k$ . But then,  $s_{k+1} \leq s_k$  and

$$\varepsilon_1(C) \sum_{k=2}^{\infty} \|\boldsymbol{\theta}_{k-1} - \boldsymbol{\theta}_{k-1}^\circ\| \leq s_1; \quad \varepsilon_2(C) \sum_{k=2}^{\infty} \|\tilde{\mathbf{K}}_k - \tilde{\mathbf{K}}_{k-1}^\circ\|^2 \leq s_1.$$

Also,  $s_1$  is bounded as  $\tilde{\mathbf{K}}_1$  stabilizes  $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ . This guarantees that, first,  $\tilde{\mathbf{K}}_k^\circ$  remains stabilizing as  $f_1(\tilde{\mathbf{K}}_{k-1}^\circ) \leq s_0$  and  $\tilde{\mathbf{Q}} \succ 0$ ; second, the estimates  $\boldsymbol{\theta}_{k-1}^\circ$  become more accurate; and third,  $\tilde{\mathbf{K}}_k^\circ \rightarrow \tilde{\mathbf{K}}^*$  as  $\tilde{\mathbf{K}}_{k+1} \rightarrow \tilde{\mathbf{K}}^*$  at  $k \rightarrow \infty$  [Hewer, 1971]. This completes the proof.  $\square$

### *Proof of Theorem 3.10*

Consider the LQR cost  $f_I(\hat{\mathbf{K}})$  as in (3.10) but for the problem parameters  $(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{Q}}, \hat{\mathbf{R}})$ . In the “unstructured” case (i.e. ignoring the constraint  $K \in \mathcal{U}_{m,n}^N(\mathcal{G})$ ), we know that the optimal LQR cost matrix for the entire networked system satisfies [Goodwin et al., 2001],

$$\hat{\mathbf{P}}_{\text{lqr}}^* = \hat{\mathbf{A}}_{\hat{\mathbf{K}}_{\text{lqr}}}^\top \hat{\mathbf{P}}_{\text{lqr}} \hat{\mathbf{A}}_{\hat{\mathbf{K}}_{\text{lqr}}} + \hat{\mathbf{K}}_{\text{lqr}}^\top \hat{\mathbf{R}} \hat{\mathbf{K}}_{\text{lqr}} + \hat{\mathbf{Q}}, \quad (3.14)$$

where  $\hat{\mathbf{K}}_{\text{lqr}} = \arg \min_K f_I(K)$ . Moreover, the cost matrix  $\hat{\mathbf{P}}^*$ —associated with the structured policy  $\hat{\mathbf{K}}^*$  learned by Algorithm 3.1—satisfies

$$\hat{\mathbf{P}}^* = \mathbf{A}_{\hat{\mathbf{K}}^*}^\top \hat{\mathbf{P}}^* \mathbf{A}_{\hat{\mathbf{K}}^*} + \hat{\mathbf{K}}^{*\top} \hat{\mathbf{R}} \hat{\mathbf{K}}^* + \hat{\mathbf{Q}}, \quad (3.15)$$

where  $\mathbf{A}_{\hat{\mathbf{K}}^*} = \hat{\mathbf{A}} + \hat{\mathbf{B}} \hat{\mathbf{K}}^*$  and  $\hat{\mathbf{K}}^* \in \mathcal{U}_{m,n}^N(\mathcal{G})$ . Finally, let  $\hat{\mathbf{K}}_{\text{struc}}^* \in \arg \min_{K \in \mathcal{U}_{m,n}^N(\mathcal{G})} f_I(K)$  denote a “structured” stabilizing optimal LQR policy which is associated with the cost matrix  $\hat{\mathbf{P}}_{\text{struc}}^*$ . We know such a policy exists since the smooth cost is lower-bounded and  $\hat{\mathbf{K}}_1 = I_N \otimes K_1 \in \mathcal{U}_{m,n}^N(\mathcal{G})$  is a feasible point of this optimization—as  $K_1$  is assumed to be

stabilizing for the single pair  $(A, B)$ . Therefore,

$$\text{Tr} \left[ \hat{\mathbf{P}}_{\text{lqr}}^* \right] = f_1 \left( \hat{\mathbf{K}}_{\text{lqr}}^* \right) \leq f_1 \left( \hat{\mathbf{K}}_{\text{struc}}^* \right) \leq f_1 \left( \hat{\mathbf{K}}^* \right) = \text{Tr} \left[ \hat{\mathbf{P}}^* \right],$$

where the last inequality above follows by the fact that  $\hat{\mathbf{K}}^*$  is a feasible solution to the structured problem by construction, *i.e.*,  $\hat{\mathbf{K}}^* \in \mathcal{U}_{m,n}^N(\mathcal{G})$ . Therefore,  $0 \leq \text{gap}(\hat{\mathbf{K}}^*) \leq \text{Tr} \left[ \hat{\mathbf{P}}^* - \hat{\mathbf{P}}_{\text{lqr}}^* \right]$ . But then, one can obtain from (3.14), (3.15), and some algebraic manipulation,

$$\hat{\mathbf{P}}^* - \hat{\mathbf{P}}_{\text{lqr}}^* = \mathbf{A}_{\hat{\mathbf{K}}_{\text{lqr}}^*}^\top \left( \hat{\mathbf{P}}^* - \hat{\mathbf{P}}_{\text{lqr}}^* \right) \mathbf{A}_{\hat{\mathbf{K}}_{\text{lqr}}^*} + \mathbf{M}',$$

where

$$\mathbf{M}' = \mathbf{A}_{\hat{\mathbf{K}}^*}^\top \hat{\mathbf{P}}^* \mathbf{A}_{\hat{\mathbf{K}}^*} - \mathbf{A}_{\hat{\mathbf{K}}_{\text{lqr}}^*}^\top \hat{\mathbf{P}}^* \mathbf{A}_{\hat{\mathbf{K}}_{\text{lqr}}^*} + \hat{\mathbf{K}}^{*\top} \hat{\mathbf{R}} \hat{\mathbf{K}}^* - \hat{\mathbf{K}}_{\text{lqr}}^{*\top} \hat{\mathbf{R}} \hat{\mathbf{K}}_{\text{lqr}}^*.$$

Since  $\hat{\mathbf{P}}^* - \hat{\mathbf{P}}_{\text{lqr}}^* \succ 0$  and  $\hat{\mathbf{A}}_{\hat{\mathbf{K}}_{\text{lqr}}^*}$  is contractible by the hypothesis, from the first part and Theorem 1 in [Mori et al., 1982] we obtain,

$$\text{gap}(\hat{\mathbf{K}}^*) \leq \frac{\text{Tr}(\mathbf{M}')}{1 - \sigma_{\max}^2(\hat{\mathbf{A}}_{\hat{\mathbf{K}}_{\text{lqr}}^*})} = \frac{\text{Tr}(\mathbf{M})}{1 - \sigma_{\max}^2(\hat{\mathbf{A}}_{\hat{\mathbf{K}}_{\text{lqr}}^*})},$$

where the last equality follows by the cyclic permutation property of trace and definition of  $\mathbf{A}_{\hat{\mathbf{K}}^*}$ .  $\square$

### 3.4.3 Proof of Lemma 3.13

First, we show that the following algebraic identities hold which will be used in the proof of Lemma 3.13.

**Lemma 3.14.** *Suppose  $A$  and  $B$  are symmetric matrices such that  $A$ ,  $A-B$ , and  $A+(n-1)B$  are all invertible for some integer  $n$ . Then the following relations hold:*

1.  $(A + nB)(A + (n-1)B)^{-1}(A - B) = A - nB(A + (n-1)B)^{-1}B.$
2.  $(A + nB)(A + (n-1)B)^{-1}(A - B) = (A - B)(A + (n-1)B)^{-1}(A + nB).$
3.  $(A + nB)(A - B)^{-1}B = B(A - B)^{-1}(A + nB).$

*Proof:* These claims follow by the algebraic manipulations below: First,

$$\begin{aligned}
& (A + nB)(A + (n - 1)B)^{-1}(A - B) \\
&= \left( (A + (n - 1)B) + B \right) (A + (n - 1)B)^{-1}(A - B) \\
&= \left( I + B(A + (n - 1)B)^{-1} \right) (A - B) \\
&= A - B + B(A + (n - 1)B)^{-1}A - B(A + (n - 1)B)^{-1}B \\
&= A + B \left( (A + (n - 2)B)^{-1}A - I \right) - B(A + (n - 1)B)^{-1}B \\
&= A + B(A + (n - 2)B)^{-1} \left( A - (A + (n - 2)B) \right) \\
&\quad - B(A + (n - 1)B)^{-1}B \\
&= A - (n - 1)B(A + (n - 1)B)^{-1}B - B(A + (n - 1)B)^{-1}B \\
&= A - nB(A + (n - 1)B)^{-1}B.
\end{aligned}$$

Second,

$$\begin{aligned}
& (A + nB)(A + (n - 1)B)^{-1}(A - B) \\
&= -n(A - B)(A + (n - 1)B)^{-1}(A - B) \\
&\quad + (n + 1)A(A + (n - 1)B)^{-1}(A - B) \\
&= -n(A - B)(A + (n - 1)B)^{-1}(A - B) \\
&\quad + (n + 1)(I + (n - 1)BA^{-1})^{-1}(A - B) \\
&= -n(A - B)(A + (n - 1)B)^{-1}(A - B) \\
&\quad + (n + 1)(A - B) \left( (I + (n - 1)BA^{-1})(A - B) \right)^{-1}(A - B) \\
&= -n(A - B)(A + (n - 1)B)^{-1}(A - B) \\
&\quad + (n + 1)(A - B) \left( (A - B)(I + (n - 1)A^{-1}B) \right)^{-1}(A - B) \\
&= -n(A - B)(A + (n - 1)B)^{-1}(A - B) \\
&\quad + (n + 1)(A - B)(A + (n - 1)B)^{-1}A \\
&= (A - B)(A + (n - 1)B)^{-1}(A + nB).
\end{aligned}$$

Finally,

$$\begin{aligned}
& (A + nB)(A - B)^{-1}B \\
&= (A - B + (n + 1)B)(A - B)^{-1}B \\
&= (\mathbf{I} + (n + 1)B(A - B)^{-1})B = B(\mathbf{I} + (n + 1)(A - B)^{-1}B) \\
&= B(A - B)^{-1}(A + nB). \quad \square
\end{aligned}$$

*Proof of Lemma 3.13*

Part 1) We prove the claim by induction. First, note that both  $\mathbf{N}_r$  and its principle submatrix  $A$  are invertible. For  $r = 2$ , by Schur complement of  $\mathbf{N}_2$ , we get

$$\begin{aligned}
\det(\mathbf{N}_2) &= \det(A) \det(A - BA^{-1}B) \\
&= \det(A) \det(\mathbf{I} - BA^{-1}) \det(\mathbf{I} + BA^{-1}) \det(A) \\
&= \det(A - B) \det(A + B).
\end{aligned}$$

Now, suppose the claim holds for  $r = p$ . Then, for  $r = p + 1$ , similarly by Schur complement we get

$$\begin{aligned}
\det(\mathbf{N}_{p+1}) &= \det(A) \det(\mathbf{N}_p - \mathbb{1}\mathbb{1}^\top \otimes BA^{-1}B) \\
&= \det(A) \det(A - B)^{p-1} \det(A - BA^{-1}B + (p - 1)(B - BA^{-1}B)), \\
&= \det(A) \det(A - B)^{p-1} \det(A - B + pBA^{-1}(A - B)), \\
&= \det(A) \det(A - B)^p \det(\mathbf{I} + pBA^{-1}), \\
&= \det(A - B)^p \det(A + pB),
\end{aligned}$$

where the second equality follows by applying the induction hypothesis to  $\mathbf{N}_p - \mathbb{1}\mathbb{1}^\top \otimes BA^{-1}B$  and some algebraic manipulation. This completes the proof.

Part 2) From item 1 of the this lemma,

$$\det(\mathbf{N}_r - \lambda \mathbf{I}_r \otimes \mathbf{I}) = \det(A - \lambda \mathbf{I} - B)^{r-1} \det(A - \lambda \mathbf{I} + (r - 1)B),$$



implying that the spectrum of  $\mathbf{N}_r$  coincides with that of  $A - B$  and  $A + (r - 1)B$ —modulo algebraic multiplicities. Hence,  $\mathbf{N}_r \succ 0$  results in  $A - B \succ 0$  and  $A + (r - 1)B \succ 0$ . Furthermore,  $\mathbf{N}_r \succ 0$  if and only if its principal submatrices are positive definite. So, by applying the latter result to principal submatrices, we claim that  $A + \ell B \succ 0$  for  $\ell = 0, \dots, r - 2$ . Lastly, from item 1 of Lemma 3.14, for  $\ell = 1, \dots, r - 1$  we have

$$A - \ell B(A + (\ell - 1)B)^{-1}B = (A + \ell B)(A + (\ell - 1)B)^{-1}(A - B),$$

which, by the first part of this claim, is invertible as a multiplication of invertible matrices.

3) Since  $\mathbf{N}_r$  and  $A$  are invertible, the Schur complement  $\mathbf{N}_{r-1} - L_{r-1}A^{-1}L_{r-1}^\top$  is also invertible where  $L_{r-1} = \mathbb{1}_{r-1} \otimes B$ . We prove the claim by induction on  $r$ . For  $r = 2$ , by [Horn and Johnson, 2012, Block matrix inverse formula (0.8.5.6)],

$$\mathbf{N}_2^{-1} = \begin{bmatrix} H^{-1} & -H^{-1}BA^{-1} \\ -A^{-1}BH^{-1} & A^{-1} + A^{-1}BH^{-1}BA^{-1} \end{bmatrix},$$

where  $H = A - BA^{-1}B$  is the Schur complement of  $A$ . By [Horn and Johnson, 2012, Woodbury inversion formula (0.7.4.1)],  $H^{-1} = A^{-1} + A^{-1}BH^{-1}BA^{-1}$ , establishing the recurrence of diagonal blocks. Also,  $\mathbf{N}_2$  is symmetric, so is  $\mathbf{N}_2^{-1}$  and thus establishing that  $\mathbf{N}_2 \in \text{PL}(2 \times n, \mathbb{R})$ . Now, from item 1 in Lemma 3.14 with  $n = 1$ , we get that

$$\begin{aligned} A^{-1}BH^{-1} &= A^{-1}B(A - B)^{-1}A(A + B)^{-1} \\ &= A^{-1}B(A + B)^{-1}A(A - B)^{-1} \\ &= A^{-1}B(I + A^{-1}B)^{-1}(A - B)^{-1} \\ &= (I - (I + A^{-1}B)^{-1})(A - B)^{-1} = G_2, \end{aligned}$$

where we also used  $(A - B)^{-1}A(A + B)^{-1} = (A + B)^{-1}A(A - B)^{-1}$  derived from Lemma 3.14 item 2. Hence,

$$\mathbf{N}_2^{-1} = I_2 \otimes (H^{-1} + H^{-1}BA^{-1}) - \mathbb{1}_2 \mathbb{1}_2^\top \otimes (H^{-1}BA^{-1}).$$

Assume that the claim holds for  $r = p$ . To extend the result to  $r = p + 1$ , again by [Horn and Johnson, 2012, Block matrix inverse formula (0.8.5.6)] and [Horn and Johnson, 2012,

Woodbury inversion formula (0.7.4.1)],

$$\mathbf{N}_{p+1}^{-1} = \begin{bmatrix} A & L_p^\top \\ L_p & \mathbf{N}_p \end{bmatrix}^{-1} = \begin{bmatrix} P^{-1} & -P^{-1}L_p^\top\mathbf{N}_p^{-1} \\ -\mathbf{N}_p^{-1}L_pP^{-1} & (\mathbf{N}_p - L_pA^{-1}L_p)^{-1} \end{bmatrix},$$

where  $P = A - L_p^\top\mathbf{N}_p^{-1}L_p$  and  $L_p = \mathbb{1}_p \otimes B$ . Let  $\mathbf{N}_p^{-1} = \mathbf{I}_p \otimes (F_p + G_p) - \mathbb{1}_p\mathbb{1}_p^\top \otimes G_p$  where,

$$F_p = \left( A - (p-1)B(A + (p-2)B)^{-1}B \right)^{-1},$$

$$G_p = (A + (p-1)B)^{-1}B(A - B)^{-1},$$

where the inversions are valid from item 2 of the current Lemma. By simplification we get  $P = A - pB(F_p - (p-1)G_p)B$  and from Lemma 3.14 items 1 and 2,

$$\begin{aligned} F_p - (p-1)G_p &= \left( A - (p-1)B(A + (p-2)B)^{-1}B \right)^{-1} \\ &\quad - (p-1)(A + (p-1)B)^{-1}B(A - B)^{-1} \\ &= (A + (p-1)B)^{-1}(A + (p-2)B)(A - B)^{-1} \\ &\quad - (p-1)(A + (p-1)B)^{-1}B(A - B)^{-1} \\ &= (A + (p-1)B)^{-1}, \end{aligned}$$

where the first term in the second equation undergoes the first two items in Lemma 3.14 consecutively. The latter equality results in  $P = A - pB(A + (p-1)B)^{-1}B$ . Next, considering the off-diagonal blocks of  $\mathbf{N}_{p+1}^{-1}$ , with some simplification, each block of  $P^{-1}L_p^\top\mathbf{N}_p^{-1}$  is equivalent to  $P^{-1}B(F_p - (p-1)G_p)$  and using the previous reasoning and Lemma 3.14 it can be simplified to,

$$\begin{aligned} P^{-1}B(F_p - (p-1)G_p) &= P^{-1}B(A + (p-1)B)^{-1} \\ &= (A + pB)^{-1}(A + (p-1)B)(A - B)^{-1}B(A + (p-1)B)^{-1} \\ &= (A + pB)^{-1}B(A - B)^{-1}. \end{aligned}$$

Similarly, each block of  $\mathbf{N}_p^{-1}L_pP^{-1}$  is also equal to  $(A + (p-1)B)^{-1}B(A - B)^{-1}$ . Therefore, it only remains to show that the blocks of  $(\mathbf{N}_p - L_pA^{-1}L_p)^{-1}$  are consistent with the desired pattern in  $\mathbf{N}_{p+1}^{-1}$ . Note that  $\mathbf{N}_p - L_pA^{-1}L_p = \mathbf{I} \otimes (A - B) + \mathbb{1}\mathbb{1}^\top \otimes (B - BA^{-1}B)$ . Hence,

with some algebraic rigor we can show that each diagonal term of  $(\mathbf{N}_p - L_p A^{-1} L_p)^{-1}$  is  $(A + pB)^{-1}(A + (p-1)B)(A - B)^{-1}$  and each off-diagonal becomes  $-(A + pB)B(A - B)^{-1}$ . Hence,

$$\mathbf{N}_{p+1}^{-1} = \mathbf{I} \otimes (F_{p+1} + G_{p+1}) - \mathbb{1}\mathbb{1}^\top G_{p+1},$$

with  $F_{p+1}$  and  $G_{p+1}$  defined as,

$$\begin{aligned} F_{p+1} &= (A + pB)^{-1}(A + (p-1)B)(A - B)^{-1} \\ G_{p+1} &= (A + pB)B(A - B)^{-1}. \end{aligned}$$

4) With direct multiplication and using the mixed-product property of Kronecker products,

$$\begin{aligned} \mathbf{N}_r \mathbf{M}_r &= \mathbf{I}_r \otimes (A - B)(C - D) + \mathbb{1}_r \mathbb{1}_r^\top \otimes B(C - D) \\ &\quad + \mathbb{1}_r \mathbb{1}_r^\top \otimes (A - B)D + r \mathbb{1}_r \mathbb{1}_r^\top \otimes BD \\ &= \mathbf{I}_r \otimes (A - B)(C - D) \\ &\quad + \mathbb{1}_r \mathbb{1}_r^\top \otimes (B(C - D) + (A - B)D + rBD). \quad \square \end{aligned}$$

### 3.5 Simulation Results

In this section, we examine the performance and convergence of D3SPI. In order to assess the suboptimality of the synthesized controllers, we report the trace of cost matrices,  $\text{Tr}(\hat{\mathbf{P}}_k)$ , associated with the proposed distributed controller learned by D3SPI at iteration  $k$ . As the optimal distributed design is unknown, we compare these results against the optimal cost for the unconstrained LQR problem,  $\text{Tr}(\hat{\mathbf{P}}_{\text{LQR}})$ , obtained via the solution of the Algebraic Riccati Equation with parameters  $(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{Q}}, \hat{\mathbf{R}})$ . Note that this is an infeasible solution to the problem in (3.3); nevertheless, it provides a theoretical lowerbound to evaluate the performance of any feasible solution—including the optimal one.<sup>4</sup>

---

<sup>4</sup>All the simulations were run on a 3.2 GHz Quad-Core Intel Core i5 CPU and in MATLAB. The scripts are publicly available at <https://github.com/shahriarta/D3SPI>.

### 3.5.1 Convergence–Randomly Selected Parameters

In the first example, we sample continuous-time system parameters of a single agent  $(A, B)$  from a zero-mean normal distribution with unit covariance, such that  $A \in \mathbb{R}^{5 \times 5}$  and  $B \in \mathbb{R}^{5 \times 3}$ . We then consider a path-graph of 10 agents and demonstrate how Algorithm 3.1 converges for this network using different instances of the system parameters. The continuous-time system dynamics of a single agent is discretized with a sampling rate of  $\Delta T = 0.1s$ . We set  $Q_1 = 0.2$ ,  $Q_2 = I_n$  and  $R = I_m$  in (3.3). We assume a random exploration signal sampled from a normal distribution  $e_t \sim \mathcal{N}(0, \sigma^2)$ , where the variance  $\sigma^2$  is chosen accordingly for different input channels.

Figure 3.2 shows the performance of the synthesized controller by illustrating the normalized suboptimality error for the entire network with respect to the (infeasible) LQR controller. This figure depicts the simulation results for 100 random system parameters and shows the progress of the proposed method at each iteration in Algorithm 3.1. The actual simulations are plotted in faded color, whereas their statistical characteristics are plotted in solid black. Note that almost all realizations of the network have converged after 5 iterations of the learning phase, due to its quadratic convergence rate. Assessing the suboptimality of the proposed controller compared against the (infeasible) centralized LQR controller, reveals an average improvement by a factor of 200.

### 3.5.2 Network of homogeneous plants

In this example, we apply D3SPI to two other simulation scenarios involving homogeneous networks of agents with unknown and unstructured model uncertainties. In particular, we use the dynamics of plants with continuous-time system parameters  $(A, B)$  (as reported in Appendix F of [Hung and MacFarlane, 1982]), in conjunction with random  $d$ -regular graph topologies of different sizes. We then examine the efficacy of D3SPI by illustrating the cost associated with the proposed distributed controller as a function of nodes in the graph. The rest of the problem parameters are chosen identical to the setup in Section 3.5.1. Uncertainty

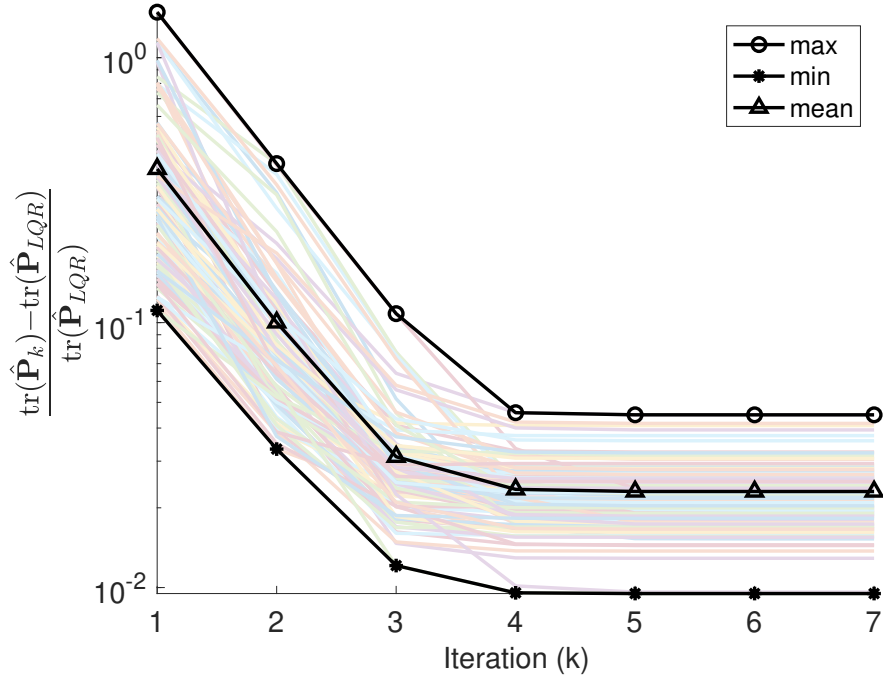
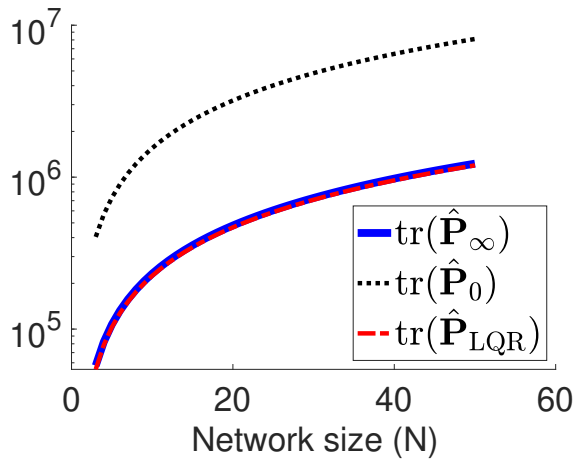


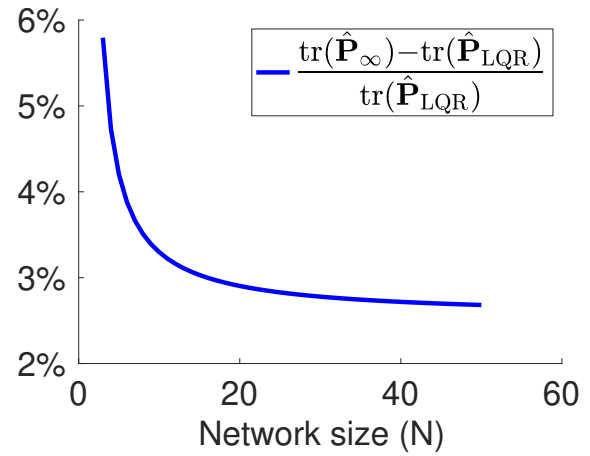
Figure 3.2: Convergence of Algorithm 3.1 for 100 randomly sampled system parameters  $(A, B)$ .

in the model are introduced in this example as follows. Each agent follows an unknown LTI dynamics similar to (3.1) with  $A$  replaced by  $A + \Delta A$ , where entries of  $\Delta A$  are sampled from a normal distribution  $\mathcal{N}(0, 0.05)$ . Assuming that one has access to a stabilizing controller  $K_1$  for the system with nominal parameters  $A$  and  $B$ , we set the initial stabilizing controller to be the LQR optimal controller with parameters  $(A, B, Q_1, R)$ .

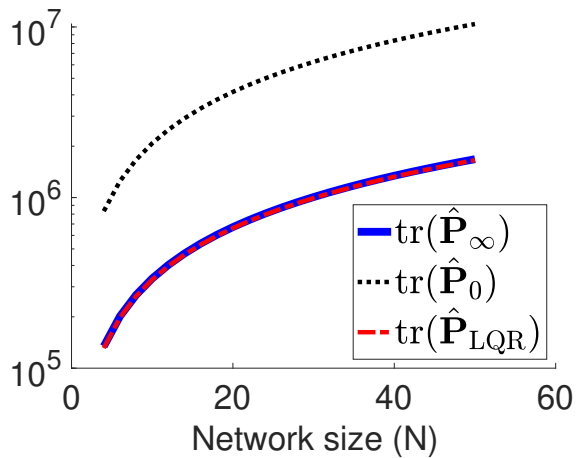
Figure 3.3 shows the results of the second simulation example, illustrating how the cost of the proposed controller changes with respect to the number of nodes in a path graph with different number of nodes. Figure 3.3a compares the cost associated with our design  $\hat{\mathbf{P}}_\infty$  against the cost of the initial controller  $\hat{\mathbf{P}}_1$ , and the (infeasible) LQR controller  $\hat{\mathbf{P}}_{\text{LQR}}$ . Figure 3.3b illustrates the evolution of the normalized suboptimality of our proposed algorithm (with respect to the infeasible LQR controller) as a function of number of nodes in the corresponding graphs. Figure 3.3c and Figure 3.3d show similar results for the random 3-regular graph topologies with even number of nodes.



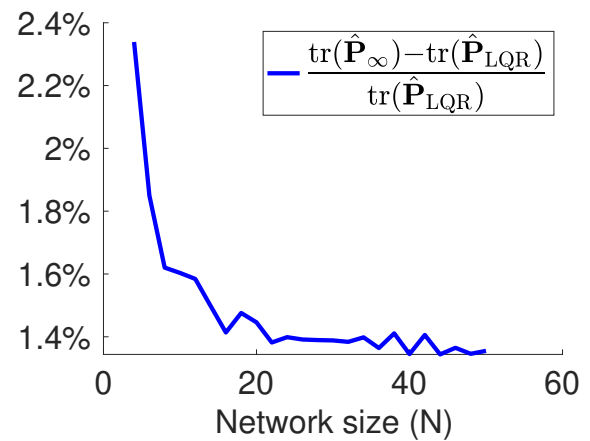
(a)



(b)



(c)



(d)

Figure 3.3: Suboptimality of the distributed controller learned by Algorithm 3.1 for a network of  $N$  homogeneous plants with path-graphs structures in (a) and (b), and with random 3-regular graphs structures in (c) and (d).

As it can be validated from Figure 3.3, the cost associated with our final proposed distributed controller has significantly improved the optimality of the initial controller. In particular, the normalized suboptimality errors of our final design are less than 6% and 2.4% for path-graph and random 3-regular graph topologies, respectively. Furthermore, this normalized error generally decrease as the number of nodes in the corresponding graphs increase.

### **3.6 Remarks and Future Directions**

In this chapter, we proposed a model-free distributed policy iteration algorithm to find structured controllers for large-scale networked systems from data. In this direction, we incorporated ideas from distributed control and data-driven optimal control literature to address the model-free distributed control problem. We then analyzed the structure of the distributed policy iteration algorithm and show how the data-driven nature of our algorithm can address unknown model uncertainty.

One way to extend the results of this work is to consider a heterogeneous setup where agents have different dynamics and design parameters. Such conditions are not trivial to handle and add another layer of complexity to the analysis which is the topic of our follow-up work. Besides, we acknowledge that the performance of our current method depends on the maximum degree of the underlying graph. Therefore, this methodology is computationally efficient when the node with highest degree has significantly less interactions than the size of the graph. Also, the result of this chapter has been mainly adapted from [Alemzadeh et al., 2021].

Finally, we note that both the solution strategies in the previous two chapters are building on the assumption of having access to an initially stabilizing controller. When the underlying system is open-loop unstable, with unknown parameters, assuming direct access to such controller may seem unreasonable. This is, in fact, the subject of study in the following chapter where we provide an online stabilization algorithm with theoretical guarantees that can help to bootstrap any learning algorithm out of instability issues.

## Chapter 4

**ON REGULARIZABILITY AND ONLINE CONTROL OF  
UNSTABLE LTI SYSTEMS**

Recently, data-driven methods for control of dynamic systems have received considerable attention in system theory and machine learning as they provide a mechanism for feedback synthesis from the observed time-series data. Learning, say through direct policy updates, often requires assumptions such as knowing *a priori* that the initial policy (gain) is stabilizing (e.g., when the open-loop system is stable) or a persistently exciting (PE) input-output data is available. In this chapter, we examine online regulation of (possibly unstable) partially unknown linear systems with no prior access to an initial stabilizing controller nor a PE input-output data; we instead leverage the knowledge of the input matrix for online regulation. First, we introduce and characterize the notion of “regularizability” for linear systems that gauges the extent by which a system can be regulated in finite-time in contrast to its asymptotic behavior (commonly characterized by stabilizability/controllability). Next, having access only to the input matrix, we propose the DGR synthesis procedure that—as its name suggests—regulates the underlying state while also generating informative data that can subsequently be used for data-driven stabilization or system identification. The analysis is also related in spirit, to the spectrum and the “instability number” of the underlying linear system, a novel geometric property studied in this chapter. We further elucidate our results by considering special cases for system parameters as well as improving the computational performance of DGR via a rank-one update. Finally, we demonstrate the utility of the proposed approach via an example involving online regulation of the X-29 aircraft.



## 4.1 Introduction

Feedback control is ubiquitous in modern technology including applications where it provides means of stabilization in addition to performance. Control of open-loop unstable plants arising for instance, in industrial and flight control applications, underscores the importance of stabilization with robustness guarantees. As such, control of unstable systems is an ongoing research topic, particularly in the context of safety-critical systems. It is well-known that unstable systems are fundamentally more difficult to control [Stein, 2003]; in fact, practical closed-loop systems with unstable subsystems are only locally stable [Sree and Chidambaram, 2006]. Yet, most of the existing synthesis literature has focused on model-based control where the designer has to discern fundamental limitations stemming from process instabilities [Skogestad et al., 2002].

Recent interest in model-free stabilization in the meantime, has been motivated by novel sensing technologies, robust machine learning, and efficient computational methods to reason about control and estimation of uncertain systems- all from measured (online) data [Hou et al., 2017; Sedghi et al., 2020]. Safety-critical systems have in fact necessitated non-asymptotic analysis on data-driven methods [Faradonbeh et al., 2019; Dean et al., 2019b]. In particular, there has been a growing interest in examining finite-time control of unknown linear dynamical systems from time-series or a single trajectory [Alaeddini et al., 2018; Sarkar et al., 2019; Berberich et al., 2020a; Oymak and Ozay, 2019; Fattahi et al., 2019; Wagenmaker and Jamieson, 2020; Tsiamis and Pappas, 2019]. Parallel to asymptotic analysis in traditional adaptive control and system identification (sysID) [Ljung, 2001; Narendra and Annaswamy, 2012; Åström and Wittenmark, 2013], model-based finite-time control has benefited from a least-squares approach to identification followed by robust synthesis—see for example [Dean et al., 2017]. In this direction, probabilistic bounds on the estimation error related to the required run-time have been examined. While it has been shown that model-based methods require fewer measurements for certain control problems in general [Tu and Recht, 2019],<sup>1</sup>

---

<sup>1</sup>That is, first finding a model estimate from data and then use that estimate for control design.

data collection required for sysID can be expensive or impractical due to resource limitations and safety constraints. Furthermore, some of the aforementioned studies rely on *a priori* information about the system, such as estimates of system parameters [Dean et al., 2019b], an initial stabilizing controller [Kim and Ng, 2005; Bradtke et al., 1994; Fazel et al., 2018], or assuming an open-loop stable system [Oymak and Ozay, 2019; Sarkar et al., 2019].

It is known that an input-output trajectory of a controllable linear time invariant (LTI) system can be parameterized by (offline) data trajectories generated from a persistently exciting (PE) input [Willems et al., 2005]. Building on this fact, there has been recent works on stabilization of LTI systems directly from the available data (e.g., see [De Persis and Tesi, 2019; Coulson et al., 2019; Baros et al., 2020; Van Waarde et al., 2020; Yu et al., 2021; Berberich et al., 2020b]). However, ensuring a PE input-output data may not be practical for data-driven control or identification of unstable systems even in low dimensions without recourse to resets [De Persis and Tesi, 2019].<sup>2</sup> Hence, existing data-guided methods might not be directly applicable for safety critical control such as online flight control [Lozano et al., 2004] or infrastructure recovery [González et al., 2017]. Our work is motivated by such applications, requiring no reliance on an initial stabilizing controller nor a PE input-output data for data-guided control. In this direction, we focus on instances where the input matrix of the LTI system is known. This point of view has been adopted by the desire to ensure satisfactory performance for online data-guided control based on a single trajectory—even when the underlying system is unstable—from the onset.

In order to realize the above program in a systematic way, in the first part of the chapter, we introduce a class of linear systems exhibiting a property called “regularizability;”<sup>3</sup> this notion captures the input “effectiveness” as it relates to finite-time regulation. We then proceed to characterize regularizability using linear matrix inequalities (LMIs), as well as clarify how it relates to spectral properties of the underlying LTI system. Additionally, we

---

<sup>2</sup>For instance, injecting white noise into an unstable system can result in ill-conditioned data matrices, that in turn, leads to numerical issues.

<sup>3</sup>Not to be confused with the notion of “regularity” for singular systems [Ozcaldiran and Lewis, 1990].

show how this system-theoretic notion can be verified in a more transparent manner for a subclass of partially known systems.

In the second part of this chapter, by employing the notion of regularizability, we introduce the Data-Guided Regulation (DGR) algorithm, an online iterative synthesis procedure that utilizes a single trajectory for an otherwise partially unknown (discrete time) LTI system. DGR does not use prior assumptions on the linear state dynamics, nor access to an initial stabilizing controller or an input-output dataset; instead, the algorithm only relies on the knowledge of the input matrix. The knowledge of the input matrix is motivated by scenarios where it is known how the control input affects the state dynamics, yet how the internal states of the system interact is uncertain (for example, consider the problem of controlling an unknown networked system from a given set of nodes). This assumption also proves useful for our setup in order to, 1) ensure a satisfactory performance for the system trajectory from the onset of the regulation process, 2) avoid requiring an initial stabilizing controller, and 3) avoid requiring a PE input-output trajectory from an unstable systems (that is often impractical and leads to ill-conditioned data matrices for post-processing). We postulate that in the case when the input matrix is also unknown, deriving nontrivial guarantees for closed loop performance of unstable systems from the onset might prove to be illusive. Finally, as pointed out above having access to the input structure of a system is pertinent to a number of applications that involve learning [Vrabie et al., 2009; Wagenmaker and Jamieson, 2020; Nozari et al., 2017; Sharf and Zelazo, 2018]; a similar assumption has been adopted for learning and control of nonlinear systems [Jagtap et al., 2020], where the system dynamics is affine in control with known input mapping and unknown state dynamics.

The contribution of the proposed work is as follows: (1) in addition to introducing the notion of regularizability for LTI systems, we show how it is distinct from related properties such as stabilizability. We believe that regularizability is of independent interest particularly as it pertains to online regulation; (2) we derive conditions under which DGR can eliminate unstable modes of the (unknown) system and regulate its state trajectory.<sup>4</sup> DGR essentially

---

<sup>4</sup>Here, regulation is ensured by bounding the norm of the system states during the learning process; see

aims for simultaneous identification and regulation of the hidden unstable modes from a single trajectory in a feedback form. As such, DGR can avoid some of the conditioning issues that arise in processing data generated by an unstable system. Using the notion of regularizability, we then proceed to derive upper bounds on the state trajectories based on a geometric quantity for LTI systems that we refer to as the “instability number;” (3) we show that while DGR performs well for a large class of unstable systems, special structures (e.g., symmetry) further facilitate deriving intuitive bounds on the system trajectory during the learning process; (4) finally, we show that the discrete nature of time-series data enables a recursive approach to DGR synthesis. In this direction a recursive DGR is proposed that circumvents storing the entire data history and avoids demanding operations such as pseudoinverse computation or multiplying large matrices.<sup>5</sup>

The rest of the chapter is organized as follows. In §4.2, we provide an overview of mathematical notions used in the chapter. In §4.3, we introduce the problem setup as well as a motivating example, followed by introducing the notion of regularizability for an LTI model. We further study the properties of regularizable systems in §4.4. Additionally, the DGR algorithm is proposed in §4.5 as the means of online regulation of (possibly) unstable systems. The subsequent part of §4.5 is devoted to the analysis of the DGR-induced closed loop system, deriving upper bounds on the state trajectories, and efficient implementation of DGR. We provide an illustrative example in §4.8.

## 4.2 *Mathematical Preliminaries*

We denote the fields of real and complex numbers by  $\mathbb{R}$  and  $\mathbb{C}$ , respectively, and real  $n \times m$  matrices by  $\mathbb{R}^{n \times m}$ . The  $n \times 1$  vector of all ones is denoted by  $\mathbf{1}$ . The unit vector  $\mathbf{e}_i$  is a

---

Section 4.3 for more details.

<sup>5</sup>A preliminary version of this work is the manuscript [Talebi et al., 2020]. The contributions of the present work as compared with [Talebi et al., 2020] include various LMI characterizations of regularizability, its extension to polytopic uncertain systems, its use in the context of online regulation with a more general setup involving the control cost, recursive and efficient updates of the online regulation algorithm, as well as a more detailed discussion on the examples and relevant literature. Furthermore, the proofs and analysis that are not presented in the conference version have been included in this manuscript.

column vector with identity at its  $i$ th entry and zero elsewhere. The  $n \times n$  identity matrix is denoted by  $\mathbf{I}_n$  (or simply  $\mathbf{I}$ ). The  $\text{diag}(\cdot)$  indicates a diagonal matrix constructed by elements of its argument in the same order starting from upper-left corner.

For a real symmetric matrix  $L$ , we say that  $L \succ 0$  when  $L$  is positive-definite (PD) and  $L \succeq 0$  for the positive-semidefinite (PSD) case. The algebraic multiplicity of an eigenvalue  $\lambda$  is denoted by  $m(\lambda)$ ;  $\lambda$  is called simple if  $m(\lambda) = 1$ . The range and nullspace of a real matrix  $M \in \mathbb{R}^{n \times m}$  are denoted by  $\mathcal{R}(M) \subseteq \mathbb{R}^n$  and  $\mathcal{N}(M) \subseteq \mathbb{R}^m$ , respectively, the dimension of  $\mathcal{R}(M)$  is designated by  $\text{rank}(M)$ , and its transpose by  $M^\top$ . The dimension of a vector space is denoted by **dim**. The span of a set of vectors over the complex field is denoted by  $\text{span}\{\cdot\}$ .

The singular value decomposition of a matrix  $M \in \mathbb{R}^{n \times m}$  is the factorization  $M = U\Sigma V^\top$ , where the unitary matrices  $U \in \mathbb{R}^{n \times n}$  and  $V \in \mathbb{R}^{m \times m}$  consist of the left and right “singular” vectors of  $M$ , and  $\Sigma \in \mathbb{R}^{n \times m}$  is the diagonal matrix of singular values in a descending order. The reduced order matrices  $U_r, V_r$  can be obtained by truncating the factored matrices  $U$  and  $V$  in the SVD to the first  $r$  columns, where  $r = \text{rank}(M)$ . The thin SVD of  $M$  is then the factorization  $M = U_r \Sigma_r V_r^\top$ , where  $\Sigma_r \in \mathbb{R}^{r \times r}$  is now nonsingular. From SVD, one can also construct the Moore-Penrose generalized inverse —*pseudoinverse* for short— of  $M$  as  $M^\dagger = V \Sigma^\dagger U^\top$ , in which  $\Sigma^\dagger$  is obtained from  $\Sigma$  by first replacing each nonzero singular value with its inverse (zero singular values remain intact) followed by a transpose.

A square matrix  $A \in \mathbb{R}^{n \times n}$  is Schur stable if  $\rho(A) < 1$ , where  $\rho(\cdot)$  denotes the spectral radius, i.e., maximum modulus of eigenvalues of its matrix argument. The matrix  $A$  is (complex) diagonalizable if there exist a diagonal matrix  $\Lambda \in \mathbb{C}^{n \times n}$  and a nonsingular matrix  $U \in \mathbb{C}^{n \times n}$  such that  $A = U\Lambda U^{-1}$ . In this case,  $\Lambda$  consists of the eigenvalues of  $A$  with columns of  $U$  as the corresponding eigenvectors. The orthogonal projection of a vector  $\mathbf{v}$  on a linear subspace  $S$  is denoted by  $\Pi_S(\mathbf{v})$ .<sup>6</sup> When the columns of a matrix  $U \in \mathbb{R}^{n \times k}$  form an orthonormal basis for the subspace  $S$ , then  $\Pi_S = UU^\top$ .

The Euclidean norm of a vector  $\mathbf{x} \in \mathbb{R}^n$  is denoted by  $\|\mathbf{x}\| = (\mathbf{x}^\top \mathbf{x})^{1/2}$ . For a matrix

---

<sup>6</sup>We will be working with finite dimensional vector spaces and as such all subspaces are closed.

$M$ , its operator norm is denoted by  $\|M\| = \sup\{\|M\mathbf{u}\| : \|\mathbf{u}\| = 1\}$ . By  $\mathcal{B}_2^r$ , we refer to the  $r$ -dimensional Euclidean ball of unit radius. An  $r$ -dimensional multi-index  $\alpha$  is an  $r$ -tuple of the form  $(\alpha_1, \alpha_2, \dots, \alpha_r)$  with all non-negative integers  $\alpha_i$ , where the sum of its elements is denoted by  $|\alpha| = \sum_{i=1}^r \alpha_i$ ;  $\alpha \in \{0, 1\}^r$  signifies that each  $\alpha_i \in \{0, 1\}$  for  $i = 1, \dots, r$ .

Finally, we say that  $\mathbf{x}_0$  *excites  $k$  modes of a matrix  $A$*  if  $\mathbf{x}_0$  is contained in the (complex-)span of  $k$  eigenvectors of  $A$ , but not in the span of any  $k - 1$  eigenvectors; we refer to those  $k$  eigenvectors (for which  $\mathbf{x}_0$  is in the span of) as the corresponding *excited modes*.

### 4.3 Problem Setup

In this section, we introduce the problem setup and highlight its unique features through an example. Consider a discrete-time LTI model of the form,

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t, \quad \mathbf{x}_0 \text{ given}, \quad (4.1)$$

where  $A \in \mathbb{R}^{n \times n}$  and  $B \in \mathbb{R}^{n \times m}$  are the system parameters and  $\mathbf{x}_t \in \mathbb{R}^n$  and  $\mathbf{u}_t \in \mathbb{R}^m$  denote the state and control inputs at time index  $t$ , respectively. We assume that the system matrix  $A$  is unknown and (possibly) unstable, and that the input matrix  $B$  is known. The problem of interest is to design  $\mathbf{u}_t$  from online state measurements (and not the system matrix  $A$  nor the offline data) such that: I) the system is regulated, with a norm uniformly bounded during the learning process, e.g.,  $\mathbf{x}_t$  evolves in a (safe) region with a quantifiable bounded norm, and the corresponding data matrix does not become ill-conditioned, and II) the system generates informative data for post-processing, for example in the context of data-driven stabilization or system identification.<sup>7</sup>

Considering regulation by having access to the input matrix is of interest in applications where it is known a priori how various control inputs effect the dynamic states, e.g., how the elevator deflection effects the aircraft pitch dynamics, or influencing a diffusive network from certain boundary nodes. Intuitively, this assumption allows an online regulation mechanism

---

<sup>7</sup>We interchangeably use the terms linear independence and *informativity* of data to emphasize that the collected data has useful information content for decision-making; the orthogonal “hidden” signal  $\mathbf{z}_t$  in Lemma 4.14 further exemplifies this perspective.

to have a chance of stabilizing an unknown (and possibly) unstable system in real-time from the onset of the learning process.

The following example motivates our setup and underscores why the data-guided perspective requires introducing new system theoretic notions.

**Example 4.1.** For any positive integer  $n$ , define the system matrix  $A \in \mathbb{R}^{n \times n}$  and the input matrix  $B \in \mathbb{R}^n$  as,

$$A = \begin{pmatrix} \lambda_1 & 1 & 0 & \dots & 0 \\ 0 & \lambda_2 & 1 & & \vdots \\ 0 & 0 & \lambda_3 & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \dots & & 0 & \lambda_n \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Note that for any choice of  $\lambda_i \in \mathbb{R}$ , the pair  $(A, B)$  is controllable (and therefore stabilizable). Furthermore, since the set  $\{\lambda_i\}$  coincides with the spectrum of  $A$ , if any subset of  $\{\lambda_i\}$  are equal, then  $A$  contains the corresponding Jordan block. Moreover, when  $\lambda_i \neq \lambda_j$  ( $i \neq j$ ), then  $A$  is diagonalizable. Let  $\mathbf{x}_0 = \mathbf{e}_1$  and observe that under (4.1), we have  $\mathbf{e}_1^\top \mathbf{x}_t = \lambda_1^t$  for all  $0 \leq t < n$  regardless of the input  $\mathbf{u}_t$ . This implies that, for “any” choice of input, for the first  $n$  iterations, the first state of the system grows exponentially fast with the rate  $\lambda_1$  whenever  $|\lambda_1| > 1$ .

*Remark 4.2.* Example 4.1 constructs a family of controllable systems where no controller can regulate their respective first states—at least for the first  $n$  iterations. That is, a system state will grow exponentially fast regardless of the choice of  $\mathbf{u}_t$ , even when all eigenvalues of  $A$  except  $\lambda_1$  are stable (e.g.,  $|\lambda_i| < 1$  for  $i = 2, \dots, n$ ). Note that in this example, the (right) eigenvector associated with the unstable mode of  $A$  (i.e., the eigen-pair  $(\lambda_1, \mathbf{e}_1)$ ) is orthogonal to  $\mathcal{R}(B) = \mathcal{R}(\mathbf{e}_n)$ . This is despite the fact that the Popov-Belevitch-Hautus (PBH) controllability test holds (i.e., for any left eigenvector  $\mathbf{v}$  of  $A$  we have  $\mathbf{v}^\top B \neq 0$ ). This example highlights that controllability of a pair  $(A, B)$  does not capture “regularizability”

of an unstable linear system, specially when closed loop regulation has to be achieved in a data-guided manner and from the onset of the learning process. Finally, we point out that in the particular case when  $\lambda_i = 0$  for  $i = 2, \dots, n$ , the controllability matrix corresponding to  $(A, B)$  is anti-diagonal with all anti-diagonal elements equal to identity. Therefore, it has singular values/eigenvalues all equal to  $\pm 1$ . This implies that the controllability matrix has condition number equal to identity; as such modes that are difficult to regularize are not distinguished by the controllability matrix.

In order to formalize the behavior of the class of systems mentioned above, we introduce a system theoretic notion that captures the effectiveness of the input as pertinent to online regulation. In order to motivate this notion, note that the dynamics in (4.1) can be represented as,

$$\begin{aligned} \mathbf{x}_{t+1} &= \Pi_{\mathcal{R}(B)^\perp} A \mathbf{x}_t + \Pi_{\mathcal{R}(B)} A \mathbf{x}_t + B \mathbf{u}_t \\ &= \Pi_{\mathcal{R}(B)^\perp} A \mathbf{x}_t + B(B^\dagger A \mathbf{x}_t + \mathbf{u}_t). \end{aligned}$$

Setting  $\mathbf{u}_t = -B^\dagger A \mathbf{x}_t + \bar{\mathbf{u}}_t$ , (4.1) can be rewritten as  $\mathbf{x}_{t+1} = \tilde{A} \mathbf{x}_t + B \bar{\mathbf{u}}_t$  where,

$$\tilde{A} := \Pi_{\mathcal{R}(B)^\perp} A, \quad (4.2)$$

and  $\bar{\mathbf{u}}_t$  is yet to be designed. Note that the signals  $\tilde{A} \mathbf{x}_t$  and  $B \bar{\mathbf{u}}_t$  are now orthogonal. This implies that the control signal would not directly affect the part of dynamics that is generated by  $\Pi_{\mathcal{R}(B)^\perp} A$ . As such, in order to have even the possibility of achieving “some” online performance for this system in finite-time, we require that this part of the dynamics be stable. This observation thereby motivates the following definition.

**Definition 4.3.** The pair  $(A, B)$  is called *regularizable* if  $\tilde{A} := \Pi_{\mathcal{R}(B)^\perp} A$  is Schur stable.

As we will show subsequently, regularizability of a pair  $(A, B)$  is related to the stabilizability of  $(A, B)$  as well as detectability of  $(A, B^\top)$ ; a combination that is not typically encountered in LTI analysis. This connection is intuitive, as regulation of a system in finite-time requires the states to be accessible (for control and observation) through the input



matrix  $B$ . Regularizability also facilitates a new perspective on LTI systems, providing a basis for the analysis of online algorithms such as the one proposed in Section 4.5.

#### 4.4 Regularizable Systems

In order to get a better sense of the notion of regularizability, we study the spectral properties of  $\tilde{A}$  in (4.2) and its relation with system matrices  $A$  and  $B$ . First, the following example highlights why regularizability of a system is distinct from its controllability.

**Example 4.4.** Consider the linear system with  $A$  defined as in Example 4.1 such that  $|\lambda_1| > 1$  and  $|\lambda_i| < 1$  for  $i = 2, \dots, n$ . Note that the pair  $(A, \mathbf{e}_n)$  is controllable (and thus stabilizable); however this pair is not regularizable. On the other hand, the pair  $(A, \mathbf{e}_1)$  is regularizable but not controllable.

Recall that a pair  $(A, B)$  is stabilizable if and only if  $(A^\top, B^\top)$  is detectable. The detectability of  $(A, B^\top)$  is seldom of interest in linear system theory [Hespanha, 2018]; however, we show that it is indeed, a necessary condition for  $(A, B)$  to be regularizable. To this end, we first connect regularizability to the spectral properties of the pair  $(A, B)$ .

**Lemma 4.5.** *Let  $\tilde{A} = \Pi_{\mathcal{R}(B)^\perp} A$ . Then for each right eigenpair  $(\lambda, \mathbf{v})$  of  $A$  the following holds:*

- $(\lambda, \mathbf{v})$  is a right eigenpair of  $\tilde{A}$  whenever  $\mathbf{v} \in \mathcal{R}(B)^\perp$  or  $\lambda = 0$ .
- $(0, \mathbf{v})$  is a right eigenpair of  $\tilde{A}$  whenever  $\mathbf{v} \in \mathcal{R}(B)$ .

The proof of Lemma 4.5 directly follows from the definitions and therefore is omitted. Note that the above lemma does not address the scenario where  $(\lambda, \mathbf{v})$  is an eigenpair of  $A$ , with  $\lambda \neq 0$ , and  $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ , with nontrivial  $\mathbf{v}_1 \in \mathcal{R}(B)$  and  $\mathbf{v}_2 \in \mathcal{R}(B)^\perp$ . The following example illustrates that  $\tilde{A}$ , as a product of matrix  $A$  with an orthogonal projection operator, has a spectral radius distinct from  $A$ .

**Example 4.6.** Consider the system in Example 4.1, where the identity off-diagonal elements of  $A$  are replaced with 10,  $\lambda_1 = 0.9$  and  $\lambda_i = 0$  for all  $i = 2, \dots, n$ , and  $B = \mathbb{1}$ . It is straightforward to show that for all  $n \geq 2$ ,  $A$  is Schur stable with spectral radius of 0.9 while  $\tilde{A}$  is not, i.e.,  $(A, B)$  is not regularizable. In this case, in spite of  $A$  being Schur stable, its operator norm is about 10. Furthermore, the spectral radius of  $\tilde{A}$  would be 4.55 for  $n = 2$  and increases to about 10 as  $n$  increases. This results in a pathological behavior despite the fact that the system is originally stable, e.g., any infinite horizon closed-loop LQR controller for this system would demonstrate undesirable behavior—similar to Example 4.1—when initialized from  $\mathbf{x}_0 = \mathbb{1}$ .<sup>8</sup> Finally, it is worth noting that the controllability matrix of this pair is ill-conditioned in contrast to Example 4.1.

The preceding discussion exemplifies that even for a stable system, it is nontrivial to assert that state trajectories over a finite time horizon are “well-regulated.” It is no surprise then that, in spite of its severe limitations from a system theoretic perspective, most of the recent works on data-guided control focus on *contractible* systems as they streamline composition rules and analysis for consecutive iterations in a learning algorithm [Lale et al., 2020b; Agarwal et al., 2019]. However, the succeeding remark shows why regularizability, as introduced in this chapter, is less restrictive, and thus—by replacing contractility—can mitigate those system theoretic limitations.

*Remark 4.7.* A pair  $(A, B)$  is said to be contractible if there exists a controller  $K$  such that  $\|A - BK\| < 1$ . Noting that

$$A - BK = \tilde{A} + \Pi_{\mathcal{R}(B)}(A - BK),$$

---

<sup>8</sup>One practical remedy to this problem is to split the dynamics into multiple time-scales using, say, a sampling heuristics [Manohar et al., 2019] However, time-scale separation often requires physical insights, non-trivial to identify in general [Naidu and Calise, 2001], let alone for a system with unknown dynamics. We address time-scale identification from finite-time collected data as a potential future direction of our work.

for any vector  $\mathbf{x} \in \mathbb{R}^n$ , (by orthogonality) it follows that,

$$\begin{aligned}\|\tilde{A}\mathbf{x}\|^2 &= \|(A - BK)\mathbf{x}\|^2 - \|\Pi_{\mathcal{R}(B)}(A - BK)\mathbf{x}\|^2 \\ &\leq \|(A - BK)\|^2 \|\mathbf{x}\|^2.\end{aligned}$$

This, in turn, implies that a contractible system is regularizable (as in that case  $\|\tilde{A}\| < 1$ ). In particular, if the original system matrix  $A$  is non-expansive (at least on the subspace  $A^{-1}\{\mathcal{R}(B)^\perp\}$ ), then  $(A, B)$  is regularizable.

The following results further clarifies the relation between regularizable systems and their system theoretic twins.

**Proposition 4.8.** *If  $(A, B)$  is regularizable, then*

- $(A, B)$  is stabilizable, and
- $(A, B^\top)$  is detectable.

**Proof.** For the first claim, note that  $\tilde{A} = A - \Pi_{\mathcal{R}(B)}A = A + BK$ , where  $K := -B^\dagger A$ . Thus if  $(A, B)$  is regularizable then  $K$  is a stabilizing closed loop controller. For the second claim, we establish a contrapositive. Suppose that  $(A, B^\top)$  is not detectable. Hence there exists a right eigenpair  $(\lambda, \mathbf{v})$  of  $A$ , where  $|\lambda| \geq 1$  and  $\mathbf{v} \in \mathcal{N}(B^\top) = \mathcal{R}(B)^\perp$ . Then, Lemma 4.5 implies that  $(\lambda, \mathbf{v})$  must be a right eigenpair of  $\tilde{A}$ . Since  $|\lambda| \geq 1$ ,  $\tilde{A}$  is not Schur stable and therefore  $(A, B)$  is not regularizable.  $\square$

Note that the consequents of Proposition 4.8 are equivalent whenever  $A$  is symmetric, as detectability of  $(A, B^\top)$  is equivalent to stabilizability of  $(A^\top, B)$ . Also, note that Proposition 4.8 provides a necessary condition for regularizability, whereas the following counterexample underscores why the stabilizability of  $(A, B)$  even when combined with detectability of  $(A, B^\top)$ , is not sufficient.

**Example 4.9.** Let the system matrices  $A, B$  be defined as in Example 4.1 and consider the pair  $(A_1, B_1) := (A + A^\top, B)$ . By the structure of  $A_1$ , note that  $(A_1, B_1)$  is controllable. Since

$A_1$  is symmetric,  $(A_1, B_1^\top)$  is also observable. By direct computation we observe that,

$$\tilde{A} = \Pi_{\mathcal{R}(B)^\perp} A = \begin{pmatrix} 2\lambda_1 & 1 & 0 & \dots & 0 \\ 1 & 2\lambda_2 & 1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 1 & 2\lambda_{n-1} & 1 \\ 0 & \dots & 0 & 0 & 0 \end{pmatrix}.$$

Now if any of  $\lambda_i$  for  $i = 1, \dots, n-1$  is say, larger than  $1/2$ , then  $\tilde{A}$  would be unstable, implying that  $(A_1, B_1)$  is not regularizable.

In order to complete our understanding of regularizability, we provide several characterizations using Linear Matrix Inequalities (LMI).

**Proposition 4.10.** *Consider a pair  $(A, B)$ , and denote  $\Pi_\perp := \Pi_{\mathcal{R}(B)^\perp}$ . Then the following are equivalent:*

1. *The pair  $(A, B)$  is regularizable.*
2.  *$\exists P \succ 0$  such that  $\rho(A^\top \Pi_\perp P \Pi_\perp A P^{-1}) < 1$ .*
3.  *$\exists P \succ 0$  such that  $\|P^{1/2} \Pi_\perp A P^{-1/2}\| < 1$ .*
4.  *$\exists P \succ 0$  such that  $A^\top \Pi_\perp P \Pi_\perp A - P \prec 0$ .*

$$5. \exists W \succ 0 \text{ such that } \begin{pmatrix} W & \Pi_\perp A W \\ W A^\top \Pi_\perp & W \end{pmatrix} \succ 0.$$

6.  *$\exists P \succ 0$  and  $G \in \mathbb{R}^{n \times n}$  such that,*

$$\begin{pmatrix} P & A^\top \Pi_\perp G^\top \\ G \Pi_\perp A & G + G^\top - P \end{pmatrix} \succ 0.$$

7.  $\exists P \succ 0$ , and  $G, H \in \mathbb{R}^{n \times n}$  such that,

$$\begin{pmatrix} GA + A^\top G^\top - P & A^\top H^\top - G \\ HA - G^\top & \Pi_\perp P \Pi_\perp - H - H^\top \end{pmatrix} \prec 0.$$

**Proof.** Noting that regularizability of  $(A, B)$  is equivalent to Schur stability of  $\Pi_\perp A$ , the first four equivalences are direct consequences of Theorem 7.7.7 in [Horn and Johnson, 2012]. By Schur complement and constructing a congruence induced by  $\text{diag}(I, P^{-1})$ , (iv) becomes equivalent to (v). The last two equivalences are due to Theorem 1 in [De Oliveira et al., 1999a] and Theorem 1 in [De Oliveira et al., 1999b], respectively.  $\square$

We conclude this section by providing a sufficient condition for guaranteeing when a polytopic uncertain LTI system is regularizable.

**Proposition 4.11.** Consider  $A_i \in \mathbb{R}^{n \times n}$  for  $i = 1, \dots, N$  and suppose there exist matrices  $P_i \succ 0$  and  $G, H \in \mathbb{R}^{n \times n}$  satisfying,

$$\begin{pmatrix} GA_i + A_i^\top G^\top - P_i & A_i^\top H^\top - G \\ HA_i - G^\top & \Pi_S P_i \Pi_S - H - H^\top \end{pmatrix} \prec 0,$$

for some linear subspace  $S \subseteq \mathbb{R}^n$ . Then a pair  $(A, B)$  is regularizable whenever  $A \in \text{convhull}\{A_i\}_1^N$  and

$$\begin{pmatrix} P_i & P_i \Pi_{\mathcal{R}(B)^\perp} \\ \Pi_{\mathcal{R}(B)^\perp} P_i & \Pi_S P_i \Pi_S \end{pmatrix} \succeq 0, \quad \forall i = 1, \dots, N.$$

**Proof.** Since  $A \in \text{convhull}\{A_i\}_1^N$ , there exists scalars  $\alpha_i \in [0, 1]$  with  $\sum_1^N \alpha_i = 1$  such that  $A = \sum_1^N \alpha_i A_i$ . By defining  $P = \sum_1^N \alpha_i P_i$  and taking the convex combinations of the negative definite matrices in the hypothesis with weights  $\alpha_i$  we obtain,

$$\begin{pmatrix} GA + A^\top G^\top - P & A^\top H^\top - G \\ HA - G^\top & \Pi_S P \Pi_S - H - H^\top \end{pmatrix} \prec 0. \quad (4.3)$$

Now by taking the Schur complement of the LMI in the hypothesis involving the input matrix  $B$  it follows that,

$$\Pi_S P_i \Pi_S \succeq \Pi_{\mathcal{R}(B)^\perp} P_i \Pi_{\mathcal{R}(B)^\perp}, \quad \forall i = 1, \dots, N.$$

Convex combinations of these LMIs with the same coefficients lead to,  $\Pi_S P \Pi_S \succeq \Pi_{\mathcal{R}(B)^\perp} P \Pi_{\mathcal{R}(B)^\perp}$ . This, together with the LMI in (4.3) imply the LMI in Proposition 4.10.(vii). As  $P \succ 0$ , we conclude that the pair  $(A, B)$  is regularizable.  $\square$

*Remark 4.12.* Note that the proof above also shows that the last LMI in the statement of Proposition 4.11 is equivalent to

$$\Pi_S P_i \Pi_S \succeq \Pi_{\mathcal{R}(B)^\perp} P_i \Pi_{\mathcal{R}(B)^\perp}, \quad \forall i = 1, \dots, N; \quad (4.4)$$

which is certainly satisfied when  $S = \mathcal{R}(B)^\perp$ . Thus, a direct consequence of Proposition 4.11–together with the characterization in Proposition 4.10.(vii)–is as follows: if there exists an input matrix  $B$  such that  $(A_i, B)$  is regularizable for each  $i = 1, \dots, N$ , then we can conclude that  $(A, B)$  is regularizable for any (unknown) matrix  $A \in \text{convhull}\{A_i\}_1^N$ . This observation does not follow directly from the definition as spectral radius is not subadditive. Moreover, Proposition 4.11 provides the flexibility of working with the linear subspace  $S$  independently of  $\mathcal{R}(B)$ , which proves to be useful for design purposes, e.g., devising an input matrix in order to make a polytopic uncertain system regularizable.

Finally, Proposition 4.11–in view of (4.4)–implies that regularizability is a monotonic system theoretic property with respect to the input, in the sense that enlarging  $\mathcal{R}(B)$  would not destroy its regularizability. In fact, enlarging  $\mathcal{R}(B)$  for a system would make it “more” regularizable (as  $\tilde{A}$  will have smaller spectral radius).

#### 4.5 Algorithm: Data-Guided Regulation (DGR)

The primary focus of this section is devising an online, data-driven feedback controller to regulate the system’s state trajectories, quantified in terms of a signal norm. In this direction, we propose an iterative procedure for updating the feedback gain (policy); the form of the

controller can be motivated by considering, at each iteration  $t$ , the following optimization problem with a “one-step quadratic cost”,<sup>9</sup>

$$\begin{aligned} \min_{\mathbf{u}_t} \quad & \|\mathbf{x}_{t+1}\|^2 + \alpha\|\mathbf{u}_t\|^2 \\ \text{s.t.} \quad & \mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t, \end{aligned} \tag{4.5}$$

where  $\mathbf{x}_t$  is measured over time but the system matrix  $A$  is unknown, and  $\alpha \geq 0$  is a regularization factor for the controller design.<sup>10</sup> In the case of known  $A$ , it is straightforward to characterize the set of minimizers of the above optimization problem through the first order optimality condition,

$$(\alpha I_m + B^\top B)\mathbf{u}_t + B^\top A\mathbf{x}_t = 0;$$

as such, the corresponding input belongs to a linear subspace in  $\mathbb{R}^m$  parameterized by the system matrices and data. The following proposition illustrates why regularizability as presented in §4.4 is pertinent to online regulation of LTI systems.

**Proposition 4.13.** *For every  $\alpha \in [0, \varepsilon)$ , with some small enough  $\varepsilon > 0$ , the minimum norm solution of the iterative optimization (4.5) stabilizes the system (4.1) if and only if the pair  $(A, B)$  is regularizable.*

**Proof.** Given a fixed  $\alpha \geq 0$ , the minimum norm solution to (4.5) at iteration  $t$  is  $\mathbf{u}_t^* = -G_\alpha A\mathbf{x}_t$ , where  $G_\alpha := (\alpha I + B^\top B)^\dagger B^\top$ . Therefore, this iterative solution stabilizes the system in (4.1) if and only if  $A - BG_\alpha A$  is Schur stable. Using properties of the pseudoinverse,  $A - BG_0 A = (I - BB^\dagger(BB^\dagger)^\top)A = (I - BB^\dagger)A = \tilde{A}$ , where  $\tilde{A}$  is as defined in Definition 4.3. The proof now follows by continuity of the spectral radius with respect to  $\alpha$ .  $\square$

---

<sup>9</sup>The setup resembles dead-beat control design, with the caveat that the synthesis is data-guided.

<sup>10</sup>We note that considering a more elaborate form of cost (e.g., finite/infinite horizon LQR cost) for this optimization problem is certainly relevant. However, in this specific problem setup, i.e., no prior knowledge on the matrix  $A$  and absence of any prior input-state data, we have observed no significant numerical advantage in considering a more elaborate cost—particularly for upper bounding the state trajectories from the onset of the learning process.

Note that  $G_\alpha \rightarrow 0$  as  $\alpha \rightarrow \infty$ , implying that  $\mathbf{u}_t \rightarrow 0$  for all  $t$ . As such, in general, the solution to (4.5) is stabilizing when  $\alpha$  is small enough.

The formulation of the optimization problem (4.5) requires the knowledge of system parameters; nonetheless, it forms the basis for the proposed algorithm when  $A$  is unknown and potentially unstable. The corresponding synthesis procedure is detailed in Algorithm 4.1. Specifically, for any  $\alpha \geq 0$ , at iteration  $t$ , DGR sets

$$\mathbf{u}_t^* = -K_t^* \mathbf{x}_t, \quad K_t^* := G_\alpha \mathcal{Y}_t \mathcal{X}_{t-1}^\dagger, \quad (4.6)$$

where  $G_\alpha := (\alpha I + B^\top B)^\dagger B^\top$  and  $\mathcal{X}_{t-1}, \mathcal{Y}_t \in \mathbb{R}^{n \times t}$  are the measured data matrices,

$$\begin{aligned} \mathcal{X}_{t-1} &:= \begin{pmatrix} \mathbf{x}_0 & \dots & \mathbf{x}_{t-1} \end{pmatrix}, \\ \mathcal{Y}_t &:= \begin{pmatrix} \mathbf{x}_1 - B\mathbf{u}_0 & \dots & \mathbf{x}_t - B\mathbf{u}_{t-1} \end{pmatrix}. \end{aligned}$$

Intuitively, collecting more data results in capturing the essential (e.g., unstable) modes in

---

**Algorithm 4.1:** Data-Guided Regulation (DGR)

---

- 1: **Initialization** (at  $t = 0$ )
  - 2: Measure  $\mathbf{x}_0$ ; set  $K_0 = \mathbf{0}$ ,  $G_\alpha = (\alpha I + B^\top B)^\dagger B^\top$
  - 3: Set  $\mathcal{X}_0 = \begin{pmatrix} \mathbf{x}_0 \end{pmatrix}$  and  $\mathcal{Y}_0 = \begin{pmatrix} \end{pmatrix}$
  - 4: **While stopping criterion not met**<sup>11</sup>
  - 5: Compute  $\mathbf{u}_t = -K_t \mathbf{x}_t$
  - 6: Run system (4.1) and measure  $\mathbf{x}_{t+1}$
  - 7: Update  $\mathcal{Y}_{t+1} = \begin{pmatrix} \mathcal{Y}_t & \mathbf{x}_{t+1} - B\mathbf{u}_t \end{pmatrix}$
  - 8:  $K_{t+1} = G_\alpha \mathcal{Y}_{t+1} \mathcal{X}_t^\dagger$
  - 9:  $\mathcal{X}_{t+1} = \begin{pmatrix} \mathcal{X}_t & \mathbf{x}_{t+1} \end{pmatrix}$
  - 10:  $t = t + 1$
- 

the dynamics. As such, it is important to note that DGR is particularly relevant for online

---

<sup>11</sup>The stopping criterion can be application specific. For instance, for sysID generating  $n$  linearly independent data is sufficient, while mere stabilization may require less; see [Van Waarde et al., 2020].



regulation of unstable systems, when the controller does not have access to enough state data for the purpose of identification or stabilization. The proposed technique is close in spirit to modal analysis where regression-based methods are leveraged to extract and control the dominant modes of the system [Simon and Mitter, 1968; Proctor et al., 2016]. The emphasis of DGR, however, is on the significance of each temporal action for safety-critical applications; in these scenarios, it might be rather unrealistic to generate sufficient data from the inherent unstable modes.

From an implementation perspective, the DGR algorithm can become computationally expensive for large-scale systems. This is primary due to steps 7-9 of Algorithm 4.1, where the entire temporal data is stored in  $\mathcal{X}_{t+1}$  and  $\mathcal{Y}_{t+1}$ ; the pseudoinverse operation in the meantime has complexity  $\mathcal{O}(n^2t)$  required at iteration  $t$ . While for the purpose of analysis, we present the basic form of DGR (as in Algorithm 4.1), in Section 4.7.1 we will propose Fast Data-Guided Regulation (F-DGR) to circumvent the complexity of storing and computing on large datasets using a rank-one update on the data matrices, resulting in a recursive evaluation of  $\mathcal{Y}_{t+1}\mathcal{X}_t^\dagger$  (see Algorithm 4.2).

## 4.6 Analysis of DGR

In this section, we provide the performance analysis for DGR in the general setting; as pointed out previously, DGR is particularly relevant when  $t \leq n$ , where  $n$  denotes the dimension of the underlying system. We examine the effects of DGR on the system's state trajectory and deduce effective guarantees in terms of norm upper-bound and informativity of generated data. In addition, we will see how a particular structure of the system matrix  $A$ , such as  $\mathcal{R}(A) \subset \mathcal{R}(B)$  or its diagonalizability, facilitates further insights into the operation of DGR as presented in the next subsection.

First, we show why regularizability is essential for the analysis of the trajectory generated under Algorithm 4.1; in hindsight, justifying its introduction in the first place.

**Lemma 4.14.** *For all  $t > 0$ , the trajectory generated by Algorithm 4.1 satisfies,*

$$\mathbf{x}_{t+1} = \Pi_{\mathcal{R}(B)^\perp} A \mathbf{x}_t + \Pi_{\mathcal{R}(B)} A \mathbf{z}_t + \Delta_\alpha \mathbf{w}_t,$$

where  $\Delta_\alpha := B(B^\dagger - G_\alpha)A$ ,  $\mathbf{z}_0 := \mathbf{x}_0$ ,  $\mathbf{w}_0 = 0$ , and  $\mathbf{z}_t := \Pi_{\mathcal{R}(\mathcal{X}_{t-1})^\perp} \mathbf{x}_t$  and  $\mathbf{w}_t := \Pi_{\mathcal{R}(\mathcal{X}_{t-1})} \mathbf{x}_t$  for  $t > 0$ . Furthermore,  $\{\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_t\}$  is a set of “orthogonal” vectors (possibly including the zero vector), and  $\Delta_0 = 0$ .

**Proof.** Let  $B = U_r \Sigma_r V_r^\top$  be the “thin” SVD of  $B$ , where  $r = \text{rank}(B)$ . Since  $BB^\dagger = U_r U_r^\top = \Pi_{\mathcal{R}(U_r)}$ ,

$$\begin{aligned} \mathbf{x}_{t+1} &= A \mathbf{x}_t + B \mathbf{u}_t \\ &= [A - B G_\alpha \mathcal{Y}_t \mathcal{X}_{t-1}^\dagger] \mathbf{x}_t \\ &= [A - B G_\alpha A \Pi_{\mathcal{R}(\mathcal{X}_{t-1})}] \mathbf{x}_t \\ &= \Pi_{\mathcal{R}(U_r)^\perp} A \mathbf{x}_t + B B^\dagger A (\mathbf{z}_t + \mathbf{w}_t) - B G_\alpha A \mathbf{w}_t \\ &= \Pi_{\mathcal{R}(U_r)^\perp} A \mathbf{x}_t + \Pi_{\mathcal{R}(U_r)} A \mathbf{z}_t + B(B^\dagger - G_\alpha) A \mathbf{w}_t. \end{aligned}$$

Thus, the first claim follows as  $\mathcal{R}(U_r) = \mathcal{R}(B)$ . For the second claim, note that the definition of  $\mathbf{z}_t$  implies that  $\mathbf{z}_t \perp \mathcal{R}(\mathcal{X}_{t-1})$  for all  $t > 0$ , and  $\mathbf{z}_k \in \mathcal{R}(\mathcal{X}_{t-1})$  for all  $k = 1, \dots, t-1$  and all  $t > 0$ . Hence  $\{\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_t\}$  consists of orthogonal vectors. Finally,  $\Delta_0 = 0$  follows by the definition of  $G_\alpha$  and properties of pseudoinverse.  $\square$

The preceding lemma implies that in the case of  $\alpha = 0$ , the time series generated by Algorithm 4.1 can be considered as the trajectory of a linear system with parameters  $(\tilde{A}, \tilde{B})$  and “input”  $\mathbf{z}_t$  where,

$$\tilde{A} := \Pi_{\mathcal{R}(B)^\perp} A, \quad \tilde{B} := \Pi_{\mathcal{R}(B)} A, \quad (4.7)$$

and  $\mathbf{z}_t = \tilde{K}_t \mathbf{x}_t$ , with the time-varying, state-dependent feedback gain  $\tilde{K}_t = \Pi_{\mathcal{R}(\mathcal{X}_{t-1})^\perp}$ . Note that, in this case, if  $\Pi_{\mathcal{R}(B)}$  and  $A$  commute,<sup>12</sup> then  $\tilde{A}\tilde{B} = 0$  and  $\mathbf{x}_{t+1} = \tilde{A}^{t+1} \mathbf{x}_0 + \tilde{B} \mathbf{z}_t$ .

---

<sup>12</sup>This is the case if (and only if) both matrices are simultaneously diagonalizable (Theorem 1.3.21 in [Horn and Johnson, 2012]). If  $A$  is symmetric, then these matrices commute if (and only if) they are congruent (Theorem 4.5.15 in [Horn and Johnson, 2012]).

Moreover,  $\tilde{A} = 0$  whenever  $\mathcal{R}(A) \subset \mathcal{R}(B)$ , i.e., the system dynamics will only be driven by the feedback signal  $\mathbf{z}_t$ ; these cases will be examined further subsequently. In case of general  $\alpha$ , the system trajectories evolve as,

$$\mathbf{x}_{t+1} = \tilde{A}^{t+1} \mathbf{x}_0 + \sum_{r=0}^t \tilde{A}^{t-r} [\tilde{B} \mathbf{z}_r + \Delta_\alpha \mathbf{w}_r]. \quad (4.8)$$

Finally, an attractive feature of DGR hinges upon the orthogonality of the “hidden” states  $\mathbf{z}_t$  generated during the process.

### *Bounding the State Trajectories Generated by DGR*

In the open loop setting, the generated data from an unstable system can grow exponentially fast with a rate dictated by the largest unstable mode. We show that DGR can prevent this undesirable phenomenon for unstable systems when the system is regularizable. The key property for such an analysis involves the notion of instability number.

**Definition 4.15.** Given the matrix  $A \in \mathbb{R}^{n \times n}$ , for any positive integer  $t \leq n$ , its *instability number of order  $t$*  is defined as,

$$M_t(A) := \sup_{\{\mathbf{v}_1, \dots, \mathbf{v}_t\} \in \mathcal{O}_t^n} \|A\mathbf{v}_1\| \|A\mathbf{v}_2\| \cdots \|A\mathbf{v}_t\|,$$

where  $\mathcal{O}_t^n$  is the collection of all sets of  $t$  “orthonormal” vectors in  $\mathbb{R}^n$ ; for  $t > n$  we define  $M_t(A) = 0$ .

Note that  $M_t(A) \leq \|A\|^t$  for all  $t$ , where  $\|\cdot\|$  denotes the induced operator norm. However the behavior of  $M_t(A)$  is fundamentally distinct from  $\|A\|^t$ . In fact, the instability number of a matrix is distinct from products of any subset of its eigenvalues. Consider for example, a  $t$ -dimensional hypercube with its image under  $A$  as a parallelotope (see Figure 4.1 for a 3D schematic). The instability number is related to the multiplication of the lengths of edges radiating from one vertex of the parallelotope, while  $\det(A^\top A)$  is related to its volume. The instability number of a matrix can in fact be difficult to compute. In what follows, we first provide upper and lower bounds on  $M_t(A)$  characterizing its growth rate with respect to the

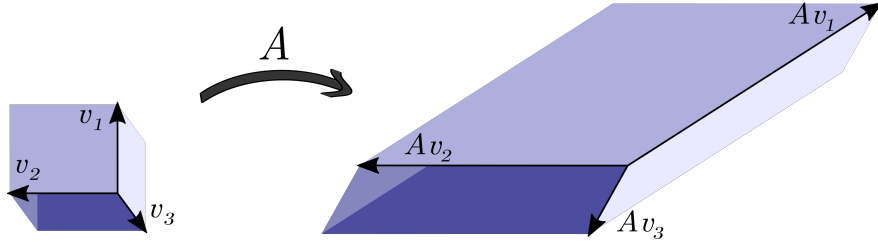


Figure 4.1: A unit cube in the domain of  $A$  that is mapped to a parallelepiped in its range space.

largest singular value of  $A$ . Subsequently, these bounds will be used to provide a bound on the norm of the state trajectory generated by DGR.

**Lemma 4.16.** *Let  $\sigma_1, \dots, \sigma_n$  denote the singular values of  $A \in \mathbb{R}^{n \times n}$  in a descending order. Then for  $t \leq n$ ,*

$$\left[ \frac{\sigma_1^2}{t} \right]^t \leq M_t^2(A) \leq \left[ \frac{\sigma_1^2}{t} \right]^t + \sum_{j=1}^{t-1} \left[ \frac{\sigma_1^2}{t-j} \right]^{t-j} \binom{t}{j} \delta^j + \delta^t,$$

where  $\delta := \sum_{i=2}^t \sigma_i^2$ , with  $M_t(A)$  as defined in Definition 4.15.

**Proof.** Let  $A = W\Sigma U^\top$  be the Singular Value Decomposition (SVD) of  $A$  where  $\Sigma$  is diagonal containing the singular values in descending order and both  $W, U \in \mathbb{R}^{n \times n}$  are unitary. This implies that,

$$\begin{aligned} M_t(A) &= \sup_{\{\mathbf{v}_i\}_1^t \in \mathcal{O}_t^n} \|\Sigma U^\top \mathbf{v}_1\| \|\Sigma U^\top \mathbf{v}_2\| \cdots \|\Sigma U^\top \mathbf{v}_t\| \\ &= \sup_{\{\mathbf{v}_i\}_1^t \in \mathcal{O}_t^n} \|\Sigma \mathbf{v}_1\| \|\Sigma \mathbf{v}_2\| \cdots \|\Sigma \mathbf{v}_t\|, \end{aligned}$$

where the last equality is due to the fact that  $\{U^\top \mathbf{v}_i\}_1^t \in \mathcal{O}_t^n$  only if  $\{\mathbf{v}_i\}_1^t \in \mathcal{O}_t^n$ , since  $U$  is unitary. For the lower-bound, if  $t \leq n$ , we can choose  $\{\mathbf{v}_i\}_1^t \in \mathcal{O}_t^n$  such that  $|\langle \mathbf{e}_1, \mathbf{v}_i \rangle| = 1/\sqrt{t}$  for all  $i = 1, \dots, t$ . This choice is certainly possible as a result of applying Parseval's identity in a  $t$ -dimensional subspace with orthonormal basis  $\{\mathbf{v}_i\}_1^t$  containing the unit vector  $\mathbf{e}_1$ , in

which,  $\mathbf{e}_1$  is represented with all coordinates equal to  $1/\sqrt{t}$  with respect to this basis. We thus conclude that,

$$M_t(A) \geq |\sigma_1 \langle \mathbf{e}_1, \mathbf{v}_1 \rangle| \cdots |\sigma_1 \langle \mathbf{e}_1, \mathbf{v}_t \rangle| \geq (\sigma_1/\sqrt{t})^t,$$

where the left inequality follows from the fact that  $\|\Sigma \mathbf{v}\| \geq |\sigma_1 \langle \mathbf{e}_1, \mathbf{v} \rangle|$  for any  $\mathbf{v} \in \mathbb{R}^n$ . For the upper-bound, define  $\Sigma_t = \text{diag}(\sigma_1, \dots, \sigma_t)$  and since singular values are in descending order we have,

$$\begin{aligned} M_t(A) &\leq \sup_{\{\mathbf{v}_i\}_1^t \in \mathcal{O}_t^t} \prod_{i=1}^t \|\Sigma_t \mathbf{v}_i\| \\ &= \sup_{\{\mathbf{v}_i\}_1^t \in \mathcal{O}_t^t} \prod_{i=1}^t \left[ \sigma_1^2 |\langle \mathbf{e}_1, \mathbf{v}_i \rangle|^2 + \sum_{j=2}^t |\sigma_j \langle \mathbf{e}_j, \mathbf{v}_i \rangle|^2 \right]^{\frac{1}{2}} \\ &\leq \sup_{\{\mathbf{v}_i\}_1^t \in \mathcal{O}_t^t} \prod_{i=1}^t \left[ \sigma_1^2 |\langle \mathbf{e}_1, \mathbf{v}_i \rangle|^2 + \delta \right]^{\frac{1}{2}}. \end{aligned}$$

Define  $\gamma_i = \langle \mathbf{e}_1, \mathbf{v}_i \rangle$ ; then by Bessel's inequality  $\sum_{i=1}^t \gamma_i^2 \leq 1$  whenever  $\{\mathbf{v}_i\}_1^t \in \mathcal{O}_t^t$ . Thereby, by denoting  $\boldsymbol{\gamma} := [\gamma_1 \dots \gamma_t]^\top$ , we can conclude that

$$\begin{aligned} M_t(A) &\leq \sup_{\boldsymbol{\gamma} \in \mathcal{B}_2^t} \prod_{i=1}^t \left[ \sigma_1^2 \gamma_i^2 + \delta \right]^{\frac{1}{2}} \\ &= \sup_{\boldsymbol{\gamma} \in \mathcal{B}_2^t} \left[ \sum_{i=1}^{t+1} \sigma_1^{2(t+1-i)} \delta^{i-1} \sum_{\substack{|\alpha|=t+1-i \\ \alpha \in \{0,1\}^t}} (\gamma_1^2)^{\alpha_1} \cdots (\gamma_t^2)^{\alpha_t} \right]^{\frac{1}{2}}, \end{aligned}$$

where  $\alpha$  is a multi-index of dimension  $t$ , and the last equality follows by direct computation.

Now it is easy to see that for a fixed multi-index  $\alpha$ , if  $|\alpha| = m > 0$  and  $\alpha \in \{0, 1\}^t$  then

$$\sup_{\boldsymbol{\gamma} \in \mathcal{B}_2^t} (\gamma_1^2)^{\alpha_1} \cdots (\gamma_t^2)^{\alpha_t} \leq \left(\frac{1}{m}\right)^m,$$

which follows by the symmetry in optimization variables. Therefore, we can conclude that

$$M_t(A) \leq \left[ \delta^t + \sum_{i=1}^t \left[ \frac{\sigma_1^2}{t+1-i} \right]^{(t+1-i)} \delta^{i-1} \binom{t}{t+1-i} \right]^{\frac{1}{2}},$$

implying the claimed upperbound.  $\square$

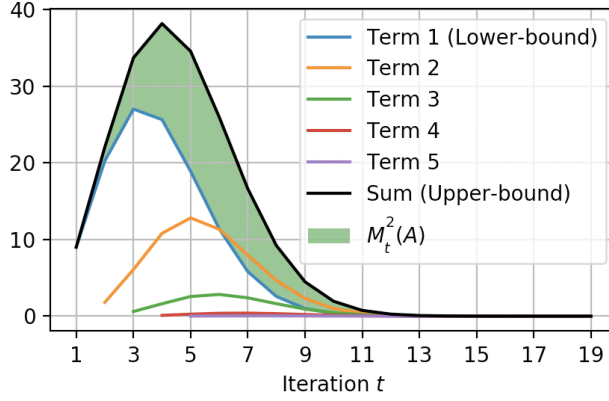


Figure 4.2: Illustration of the upper and lower bounds for instability number of a system with  $\sigma_1 = 3$  and  $\delta = 0.1$  as in Lemma 4.16.

The lower and upper bounds in Lemma 4.16 show that, particularly when  $\delta < 1$ ,  $M_t(A)$  initially grows similar to  $(\sigma_1/\sqrt{t})^t$  for  $t \leq \sigma_1^2$ , in contrast to the exponential growth of  $\|A\|^t = \sigma_1^t$ . This difference becomes more pronounced for  $t > \sigma_1^2$  when  $(\sigma_1/\sqrt{t})^t$  starts *decreasing*. This fact is illustrated via an example in Figure 4.2, where the first five dominant terms of the upper bound are plotted and the green region shows where the actual value of  $M_t^2(A)$  lies. The following result provides an upper bound on the state trajectories for the most general case through the lens of regularizability.

**Theorem 4.17.** *For any regularizable pair  $(A, B)$ , the trajectory generated by Algorithm 4.1 satisfies the following bound for all  $t > 0$ ,*

$$\|\mathbf{x}_{t+1}\| \leq L_{t+1} \|\mathbf{x}_0\|,$$

where  $L_t$  satisfies the recursion,

$$L_{t+1} = a_t + \sum_{r=1}^t b_{t,r} L_r, \quad L_1 = \|A\bar{\mathbf{z}}_0\|,$$

with

$$b_{t,r} = \sqrt{\|\tilde{A}^{t-r} \tilde{B} \bar{\mathbf{z}}_r\|^2 + \|\tilde{A}^{t-r} \Delta_\alpha \bar{\mathbf{w}}_r\|^2},$$

and  $a_t = \|\tilde{A}^t A \bar{\mathbf{z}}_0\|$ , where  $\bar{\mathbf{z}}_r = \mathbf{z}_r / \|\mathbf{z}_r\|$  (if  $\mathbf{z}_r \neq 0$ , otherwise  $\bar{\mathbf{z}}_r = 0$ ), and  $\bar{\mathbf{w}}_r$  is similarly defined.

**Proof.** Knowing that  $\mathbf{x}_1 = A\mathbf{x}_0$ , it follows that  $\|\mathbf{x}_1\| \leq L_1\|\mathbf{x}_0\|$ . Furthermore, for  $t \geq 1$ , (4.8) leads to,

$$\mathbf{x}_{t+1} = \tilde{A}^t A \mathbf{x}_0 + \sum_{r=1}^t \tilde{A}^{t-r} \left[ \tilde{B} \mathbf{z}_r + \Delta_\alpha \mathbf{w}_r \right],$$

since  $\tilde{A} + \tilde{B} = A$  and  $\mathbf{w}_0 = 0$  by definition. This implies that,

$$\begin{aligned} \|\mathbf{x}_{t+1}\| &\leq \|\tilde{A}^t A \mathbf{x}_0\| + \sum_{r=1}^t \|\tilde{A}^{t-r} \tilde{B} \bar{\mathbf{z}}_r\| \|\mathbf{z}_r\| + \|\tilde{A}^{t-r} \Delta_\alpha \bar{\mathbf{w}}_r\| \|\mathbf{w}_r\|, \\ &\leq a_t \|\mathbf{x}_0\| + \sum_{r=1}^t b_{t,r} \|\mathbf{x}_r\|, \end{aligned}$$

where we have used Cauchy–Schwarz inequality in conjunction with the equality  $\|\mathbf{z}_r\|^2 + \|\mathbf{w}_r\|^2 = \|\mathbf{x}_r\|^2$ . Using this recursive bound, the rest of the proof follows by induction.  $\square$

*Remark 4.18.* Note that in the analysis above, when the system is regularizable,  $a_t$  eventually decreases exponentially fast as  $t$  increases. Furthermore, the term  $b_{t,r}$  in the sum increases as  $r$  approaches a fixed  $t$ . Finally, one can show that the obtained upper bound is tight by considering Example 4.1 with  $\lambda_1 > 0$  and  $\lambda_i = 0$  for  $i > 1$ .

Note that computing/estimating the upper bound in Theorem 4.17 requires knowledge on matrix  $A$ , making these estimates more practical for structured systems (for example, see Corollary 4.22 and Remark 4.23). In order to shed light on the intuition behind this upper bound, we next study simpler cases with  $\alpha = 0$ , where there exists small enough  $\kappa$  for which  $b_{t,r} \leq \|\tilde{A}^{t-r} \tilde{B}\| \leq \kappa$  for all  $r < t$ . In particular, we can show that if the system is regularizable and  $\tilde{A}\tilde{B} = 0$ , then the trajectories of the closed loop system will be bounded by a combination of instability number of different orders. This is stated in the following corollary of Theorem 4.17.

**Corollary 4.19.** *For any regularizable pair  $(A, B)$  with  $\tilde{A}\tilde{B} = 0$ , and  $M_t(A)$  as in Definition 4.15, the system trajectory generated by Algorithm 4.1 with  $\alpha = 0$  satisfies the following*

for all  $t > 0$ ,

$$\frac{\|\mathbf{x}_{t+1}\|}{\|\mathbf{x}_0\|} \leq a_t + \sum_{r=1}^{t-1} M_r(A) a_{t-r} + M_{t+1}(A).$$

**Proof.** For brevity, let  $b_t = b_{t,t}$ , then as  $\tilde{A}\tilde{B} = 0$ , the recursion in Theorem 4.17 reduces to  $L_{t+1} = a_t + b_t L_t$  with  $L_1 = \|A\tilde{\mathbf{z}}_0\|$ ; and its solution has the following form for all  $t > 0$ ,

$$L_{t+1} = a_t + b_t \cdots b_2 b_1 L_1 + \sum_{r=1}^{t-1} b_t \cdots b_{t+1-r} a_{t-r}. \quad (4.9)$$

As  $\alpha = 0$  and orthogonal projection is non-expansive, we claim that  $b_r = \|\tilde{B}\tilde{\mathbf{z}}_r\| \leq \|A\tilde{\mathbf{z}}_r\|$  which vanishes whenever  $\mathbf{z}_r = 0$ . In the meantime, by Lemma 4.14,  $\{\mathbf{z}_r\}_0^t$  must be a set of orthogonal vectors for any  $t > 0$ , and thus  $\{\tilde{\mathbf{z}}_r\}_0^t$  is a set of orthogonal vectors that are either normal or zero. Note that if  $t \geq n$ , then  $\{\tilde{\mathbf{z}}_r\}_0^t$  must contain at least one zero vector for dimensional reasons. Therefore, by Definition 4.15, we conclude that  $b_t \cdots b_{t+1-r} \leq M_r(A)$ , for each  $r = 1, \dots, t-1$ . Similarly, as  $L_1 = \|A\tilde{\mathbf{z}}_0\|$ , we have  $b_t \cdots b_2 b_1 L_1 \leq M_{t+1}(A)$ . By using these inequalities in (4.9), the claim follows by Theorem 4.17.  $\square$

The above observation further highlights the importance of the instability number in the context of DGR. Note that the terms in the upper bound involving  $M_r(A)$  vanishes if  $r > n$ .

### *Informativity of the DGR Generated Data*

In the sequel, we show that DGR generates linearly independent state-trajectory data; we refer to this as informativity of data. Then, we proceed to make a connection between this independence structure and the number of excited modes in the system. Before we proceed, let us define  $L_k^t(A)$ , that is based on eigenvalues corresponding to  $k$  modes of a matrix  $A$ ,



as,

$$L_k^t(A) := \begin{pmatrix} 1 & \lambda_1 & \cdots & (\lambda_1)^{t-1} \\ 1 & \lambda_2 & \cdots & (\lambda_2)^{t-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_k & \cdots & (\lambda_k)^{t-1} \end{pmatrix}, \quad 1 \leq t \leq n. \quad (4.10)$$

*Remark 4.20.* Note that  $L_k^t(A)$  has a specific structure that hints at its invertibility. In fact, for  $t = k$ ,  $L_k^k(A)$  is the Vandermonde matrix formed by  $k$  eigenvalues of  $A$  which would be invertible if and only if  $\lambda_1, \dots, \lambda_k$  are distinct. More generally, if  $\{\lambda_1, \dots, \lambda_k\}$  consists of  $r$  distinct eigenvalues (where  $r \leq k$ ), then  $L_k^r(A)$  has full column rank.

Intuitively, informative data—due to its linear independence structure—contain useful information for decision-making purposes. In particular, if the choice of  $\mathbf{x}_0$  results in exciting all modes of  $A$ , one might expect that a useful online regulation algorithm should generate informative data at the same time that it is regulating the state-trajectory. The next theorem formalizes how DGR realizes this expectation depending on what modes of the system are excited by the initial condition.

**Theorem 4.21.** *Let  $\mathbf{x}_0$  excite  $k_1 + k_2$  modes of  $A$ , such that  $k_1$  modes are in  $\mathcal{R}(B)$  and  $k_2$  modes are in  $\mathcal{R}(B)^\perp$ . If the excited modes correspond to distinct eigenvalues, then  $\{\mathbf{x}_0, \dots, \mathbf{x}_{r-1}\}$ , generated by Algorithm 4.1 with  $\alpha = 0$ , is a set of linearly independent vectors for any  $r \leq \max\{k_1, k_2\}$ .*

**Proof.** Without loss of generality, let  $\lambda_1, \dots, \lambda_{k_1}$  be the eigenvalues corresponding to the excited modes  $u_1, \dots, u_{k_1} \in \mathcal{R}(B)$ , and similarly  $\lambda_{k_1+1}, \dots, \lambda_{k_1+k_2}$  be corresponding to  $u_{k_1+1}, \dots, u_{k_1+k_2} \in \mathcal{R}(B)^\perp$ . Recall that  $\mathcal{X}_{t-1} = [\mathbf{x}_0 \ \mathbf{x}_1 \ \dots \ \mathbf{x}_{t-1}]$ ; then by definition of  $\mathbf{z}_t$  in Lemma 4.14, for  $t \geq 1$  there exists scalar coefficients  $\zeta_0^t, \dots, \zeta_{t-1}^t \in \mathbb{R}$  such that  $\mathbf{z}_t = \mathbf{x}_t - \sum_{j=0}^{t-1} \zeta_j^t \mathbf{x}_j$ . This together with the dynamics in Lemma 4.14 imply that  $\mathbf{x}_1 = A\mathbf{x}_0$  and for  $t \geq 2$ ,

$$\mathbf{x}_t = A\mathbf{x}_{t-1} - \Pi_{\mathcal{R}(B)} \sum_{j=0}^{t-2} \zeta_j^t A\mathbf{x}_j. \quad (4.11)$$

Since  $\mathbf{x}_0$  excites  $k_1 + k_2$  modes of the system, we have  $\mathbf{x}_0 = \sum_{\ell=1}^{k_1+k_2} \beta_\ell \mathbf{u}_\ell$ , where  $\beta_\ell$  are some nonzero real coefficients and  $(\lambda_\ell, \mathbf{u}_\ell)$  are eigenpairs of  $A$ . Hence  $\mathbf{x}_1 = A\mathbf{x}_0 = \sum_{\ell=1}^{k_1+k_2} \beta_\ell \lambda_\ell \mathbf{u}_\ell$ , and we claim that for  $t \geq 2$  there exist scalar coefficients  $\xi_1^t, \dots, \xi_{t-1}^t \in \mathbb{R}$  such that,

$$\mathbf{x}_t = \sum_{\ell=1}^{k_1} \beta_\ell \left[ (\lambda_\ell)^t - \sum_{i=1}^{t-1} \xi_i^t (\lambda_\ell)^i \right] \mathbf{u}_\ell + \sum_{\ell=k_1+1}^{k_1+k_2} \beta_\ell (\lambda_\ell)^t \mathbf{u}_\ell. \quad (4.12)$$

The proof of the last claim is by induction. Note that  $A^t \mathbf{x}_0 = \sum_{\ell=1}^{k_1+k_2} \beta_\ell (\lambda_\ell)^t \mathbf{u}_\ell$ , and by substituting this into (4.11) for  $t = 2$  we have that,

$$\begin{aligned} \mathbf{x}_2 &= A\mathbf{x}_1 - \zeta_0^2 \Pi_{\mathcal{R}(B)} A\mathbf{x}_0 \\ &= \Pi_{\mathcal{R}(B)} [A^2 \mathbf{x}_0 - \zeta_0^2 A\mathbf{x}_0] + \Pi_{\mathcal{R}(B)^\perp} A^2 \mathbf{x}_0 \\ &= \sum_{\ell=1}^{k_1} \beta_\ell [(\lambda_\ell)^2 - \zeta_0^2 \lambda_\ell] \mathbf{u}_\ell + \sum_{\ell=k_1+1}^{k_1+k_2} \beta_\ell (\lambda_\ell)^2 \mathbf{u}_\ell, \end{aligned}$$

where the last equality is due to the fact that  $\mathbf{u}_\ell \in \mathcal{R}(B)$  for  $\ell \leq k_1$  and  $\mathbf{u}_\ell \in \mathcal{R}(B)^\perp$  for  $\ell > k_1$ . By choosing  $\xi_1^2 = \zeta_0^2$ , we have shown that (4.12) holds for  $t = 2$ . Now suppose that (4.12) holds for all  $2, \dots, t-1$ ; it now suffices to show that this relation also holds for  $t$ . By substituting the hypothesis for  $2, \dots, t-1$  into (4.11),

$$\begin{aligned} \mathbf{x}_t &= \sum_{\ell=1}^{k_1} \beta_\ell \left[ (\lambda_\ell)^t - \sum_{i=1}^{t-2} \xi_i^{t-1} (\lambda_\ell)^{i+1} \right] \mathbf{u}_\ell + \sum_{\ell=k_1+1}^{k_1+k_2} \beta_\ell (\lambda_\ell)^t \mathbf{u}_\ell - \sum_{\ell=1}^{k_1} \beta_\ell [\zeta_0^t \lambda_\ell + \zeta_1^t (\lambda_\ell)^2] \mathbf{u}_\ell \\ &\quad - \sum_{j=2}^{t-2} \zeta_j^t \sum_{\ell=1}^{k_1} \beta_\ell \left[ (\lambda_\ell)^{j+1} - \sum_{i=1}^{j-1} \xi_i^j (\lambda_\ell)^{i+1} \right] \mathbf{u}_\ell. \end{aligned}$$

Therefore,  $\mathbf{x}_t = \sum_{\ell=1}^{k_1} \beta_\ell [\star] \mathbf{u}_\ell + \sum_{\ell=k_1+1}^{k_1+k_2} \beta_\ell (\lambda_\ell)^t \mathbf{u}_\ell$ , where  $\star$  replaces the expression,

$$(\lambda_\ell)^t - \sum_{i=1}^{t-2} \xi_i^{t-1} (\lambda_\ell)^{i+1} - \sum_{j=0}^{t-2} \zeta_j^t (\lambda_\ell)^{j+1} + \sum_{j=2}^{t-2} \sum_{i=1}^{j-1} \zeta_j^t \xi_i^j (\lambda_\ell)^{i+1}.$$

By appropriate choices of  $\xi_1^t, \dots, \xi_{t-1}^t \in \mathbb{R}$ , we can rewrite  $\star = (\lambda_\ell)^t - \sum_{i=1}^{t-1} \xi_i^t (\lambda_\ell)^i$ . This completes the proof of the claim in (4.12) by induction. Now, let  $\hat{\mathbf{x}} = \sum_{j=0}^{r-1} \gamma_j \mathbf{x}_j$  for some  $\gamma_j \in \mathbb{C}$  and some  $r \leq \max\{k_1, k_2\}$ . Then, by substituting  $\mathbf{x}_j$  from (4.12) and exchanging the

sums over  $j$  and  $\ell$  we have,

$$\widehat{\mathbf{x}} = \sum_{\ell=1}^{k_1} \beta_\ell \left[ \gamma_0 + \gamma_1 \lambda_\ell + \sum_{j=2}^{r-1} \gamma_j [(\lambda_\ell)^j - \sum_{i=1}^{j-1} \xi_i^j (\lambda_\ell)^i] \right] \mathbf{u}_\ell + \sum_{\ell=k_1+1}^{k_1+k_2} \beta_\ell \sum_{j=0}^{r-1} \gamma_j (\lambda_\ell)^j \mathbf{u}_\ell.$$

Now, by exchanging the sums over  $i$  and  $j$ , it follows that,

$$\widehat{\mathbf{x}} = \sum_{\ell=1}^{k_1} \beta_\ell \left[ \gamma_0 + \sum_{i=1}^{r-2} [\gamma_i - \sum_{j=i+1}^{r-1} \gamma_j \xi_i^j] (\lambda_\ell)^i + \gamma_{r-1} (\lambda_\ell)^{r-1} \right] \mathbf{u}_\ell + \sum_{\ell=k_1+1}^{k_1+k_2} \beta_\ell \left[ \sum_{j=0}^{r-1} \gamma_j (\lambda_\ell)^j \right] \mathbf{u}_\ell.$$

Since  $\{\mathbf{u}_\ell\}_1^{k_1+k_2}$  are eigenvectors associated with distinct eigenvalues, they are linearly independent. Thus, noting that  $\beta_\ell \neq 0$  for all  $\ell = 1, \dots, k_1 + k_2$ , then  $\widehat{\mathbf{x}} = 0$  implies that,

$$\gamma_0 + \sum_{i=1}^{r-2} \left[ \gamma_i - \sum_{j=i+1}^{r-1} \gamma_j \xi_i^j \right] (\lambda_\ell)^i + \gamma_{r-1} (\lambda_\ell)^{r-1} = 0,$$

for all  $\ell = 1, \dots, k_1$ ; and  $\sum_{j=0}^{r-1} \gamma_j (\lambda_\ell)^j = 0$ , for all  $\ell = k_1 + 1, \dots, k_1 + k_2$ . By rewriting the last two sets of equations in matrix form we get,

$$\begin{pmatrix} L_{k_1}^r(A)(I - \Xi) \\ \widehat{L}_{k_2}^r(A) \end{pmatrix} \boldsymbol{\gamma} = 0, \quad (4.13)$$

where  $\widehat{L}_{k_2}^r(A)$  is the last  $k_2$  rows of  $L_{k_1+k_2}^r(A)$  and

$$\Xi := \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \xi_1^2 & \xi_1^3 & \cdots & \xi_1^{r-1} \\ 0 & 0 & 0 & \xi_2^3 & \cdots & \xi_2^{r-1} \\ 0 & 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \xi_{r-1}^{r-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \quad \boldsymbol{\gamma} := \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{r-1} \end{pmatrix}. \quad (4.14)$$

Note that  $I - \Xi$  is invertible by construction. Since the excited modes correspond to distinct eigenvalues, if  $r \leq \max\{k_1, k_2\}$ , then either  $L_{k_1}^r(A)$  or  $\widehat{L}_{k_2}^r(A)$  has full column rank. Either way, (4.13) implies that  $\boldsymbol{\gamma} = 0$  and thus  $\{\mathbf{x}_0, \dots, \mathbf{x}_{r-1}\}$  is a set of linearly independent vectors. This observation completes the proof as  $r \leq \max\{k_1, k_2\}$  was chosen arbitrary.  $\square$

The preceding theorem guarantees the linear independence of the state-trajectory generated by DGR whenever the excited modes lie in  $\mathcal{R}(B)$  or  $\mathcal{R}(B)^\perp$ , even though our observations suggest that it must remain valid for arbitrary excitation of the modes. Nonetheless, DGR remains effective in terms of online regulation from an arbitrary choice of  $x_0$  as guaranteed in Theorem 4.17, Corollary 4.19, and subsequently in Corollary 4.22.

#### 4.7 The Special Case of $\mathcal{R}(A) \subset \mathcal{R}(B)$ with $\alpha = 0$

In order to better understand the behavior of DGR, in this subsection, we study the more special case where  $\mathcal{R}(A) \subset \mathcal{R}(B)$ . This includes the case where  $\text{rank}(B) = n$ , i.e., one can directly control each state of the system (e.g. see [Sharf and Zelazo, 2018; Friedkin and Johnsen, 1990]). Note that  $\mathcal{R}(A) \subset \mathcal{R}(B)$  implies that  $\tilde{A} = 0$  which, in turn, results in regularizability of  $(A, B)$ . This, together with Corollary 4.19, results in the following corollary.

**Corollary 4.22.** *For any matrix  $A \in \mathbb{R}^{n \times n}$  and  $B \in \mathbb{R}^{n \times m}$ , where  $\mathcal{R}(A) \subseteq \mathcal{R}(B)$ , the trajectory generated by Algorithm 4.1 with  $\alpha = 0$  satisfies the following for all  $t > 0$ ,*

$$\|\mathbf{x}_{t+1}\| \leq M_{t+1}(A)\|\mathbf{x}_0\|,$$

with  $M_t(A)$  as in Definition 4.15.

**Proof.** Note that  $\tilde{A} = \Pi_{\mathcal{R}(B)^\perp}A = 0$  whenever  $\mathcal{R}(A) \subseteq \mathcal{R}(B)$ . The claim now follows by Corollary 4.19 since  $\tilde{A}\tilde{B} = 0$  and  $a_k = 0$  for all  $k = 1, \dots, t$ .  $\square$

Note that under the hypothesis of Corollary 4.22, in particular  $\mathbf{x}_{t+1} = 0$  for all  $t \geq n$  whenever DGR is in effect for the noiseless dynamics in (4.1) (see Figure 4.3); however, this might happen even before  $t$  reaches  $n$  as will be discussed in Proposition 4.25. Also, the latter bound becomes more structured for a symmetric  $A$  by combining the results from Corollary 4.22 and Lemma 4.16 which we skip for brevity.

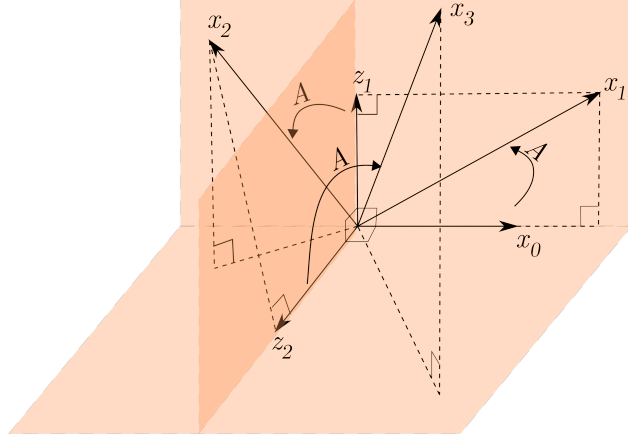


Figure 4.3: A geometric schematic of DGR when  $\mathcal{R}(A) \subseteq \mathcal{R}(B)$ . Since  $\mathbf{z}_0 := \mathbf{x}_0$ ,  $\mathbf{z}_t \perp \mathcal{R}(\mathcal{X}_{t-1})$  and  $\mathbf{z}_t \in \mathcal{R}(\mathcal{X}_t)$  for  $t = 1, 2$ , the set  $\{\mathbf{z}_0, \mathbf{z}_1, \mathbf{z}_2\}$  consists of orthogonal vectors.

*Remark 4.23.* In order to further illustrate the bound stated in Corollary 4.22, assume that  $\delta e \leq 1$ . Then, from Lemma 4.16,

$$\frac{\|\mathbf{x}_t\|^2}{\|\mathbf{x}_0\|^2} \leq \left[\frac{\sigma_1^2}{t}\right]^t + \sum_{j=1}^{\lfloor t/2 \rfloor} \left[\frac{\sigma_1^2}{t-j}\right]^{t-j} \left[\frac{t}{j}\right]^j + \sum_{j=\lfloor t/2 \rfloor + 1}^{t-1} \left[\frac{t\sigma_1^2}{(t-j)^2}\right]^{t-j} + 1,$$

where we have also used  $\binom{t}{j} \leq (et/j)^j$ . This implies that as  $t$  gets larger than  $\sigma_1^2$ , the terms with large powers admit smaller bases and those with large bases will gain smaller powers comparing to  $\sigma_1^{2t}$ . This is despite the fact that for small  $t$ , the relative norm of the state might grow.

In the sequel, as a result of linear independence established in Theorem 4.21 we show how the simplified bounds (derived in Section 4.6) clarify the elimination of the unstable modes in the system.

**Corollary 4.24.** *Suppose  $\mathcal{R}(A) \subseteq \mathcal{R}(B)$  and let  $\mathbf{x}_0$  excite  $k$  modes of  $A$ . If  $r$  eigenvalues corresponding to the  $k$  excited modes are distinct for some  $r \leq k$ , then  $\{\mathbf{x}_0, \dots, \mathbf{x}_{r-1}\}$ , generated by Algorithm 4.1 with  $\alpha = 0$ , is a set of linearly independent vectors.*

**Proof.** Given that  $\mathcal{R}(A) \subseteq \mathcal{R}(B)$ , all the modes of  $A$  are contained in  $\mathcal{R}(B)$ , so without loss

of generality, let  $\lambda_1, \dots, \lambda_k$  be the eigenvalues corresponding to the excited modes  $u_1, \dots, u_k \in \mathcal{R}(B)$ . Then, following the proof of Theorem 4.21, (4.12) reduces to,

$$\mathbf{x}_t = \sum_{\ell=1}^k \beta_\ell \left[ (\lambda_\ell)^t - \sum_{j=1}^{t-1} \xi_j^t (\lambda_\ell)^j \right] \mathbf{u}_\ell.$$

Now, let  $\hat{\mathbf{x}} = \sum_{j=0}^{r-1} \gamma_j \mathbf{x}_j$  for some  $\gamma_j \in \mathbb{C}$  and  $r \leq k$ . Then following the same argument as in the proof of Theorem 4.21 about  $\hat{\mathbf{x}}$ , (4.13) reduces to  $L_k^r(A)(I - \Xi)\boldsymbol{\gamma} = 0$ , with similar definitions of  $\Xi$  and  $\boldsymbol{\gamma}$  as in (4.10), and  $L_k^r(A)$  as in (4.14). Since  $r$  eigenvalues corresponding to  $k$  excited modes are distinct,  $L_k^r(A)$  has full column rank. As  $I - \Xi$  is invertible, we conclude that  $\boldsymbol{\gamma} = 0$  meaning that  $\{\mathbf{x}_0, \dots, \mathbf{x}_{r-1}\}$  are linearly independent.  $\square$

An immediate consequence of the above corollary is that DGR generates data that is effective for simultaneous identification of modes even with multiplicity greater than one.

**Proposition 4.25.** *Suppose that  $A$  is diagonalizable with  $\mathcal{R}(A) \subseteq \mathcal{R}(B)$ , and let  $\mathbf{x}_0$  excite  $k$  modes of  $A$  corresponding to  $r$  distinct eigenvalues (where possibly  $r \leq k$ ). Then, in exactly  $r$  iterations of Algorithm 4.1 with  $\alpha = 0$ ,  $\text{span}\{\mathbf{x}_0, \dots, \mathbf{x}_{r-1}\}$  coincides with the subspace containing these excited modes; furthermore,  $\mathbf{x}_{r+1} = 0$ .*

**Proof.** Without loss of generality, let  $\mathbf{x}_0$  excite the  $k$  modes of  $A$  corresponding to  $\lambda_1, \dots, \lambda_r$ . Since  $A$  is diagonalizable, let  $A = U\Lambda U^{-1}$  be its eigen-decomposition and so  $\mathbf{x}_0$  excite  $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ , i.e.,  $\mathbf{x}_0 = \sum_{i=1}^k \beta_i \mathbf{u}_i$ , with  $\beta_i \neq 0$ . Define

$$\mathcal{I}(\lambda_i) = \{j : \mathbf{u}_j \text{ is the eigenvector corresponding to } \lambda_i\},$$

for  $i = 1, \dots, r$ . Furthermore, define the  $r$ -dimensional subspace,

$$S := \text{span} \left\{ \sum_{j \in \mathcal{I}(\lambda_1)} \beta_j \mathbf{u}_j, \dots, \sum_{j \in \mathcal{I}(\lambda_r)} \beta_j \mathbf{u}_j \right\},$$

where the span is taken over the complex field. We prove by induction that  $\mathbf{x}_t \in S$  for all  $t = 1, \dots, r$ . Notice that  $\mathbf{x}_0 \in S$  and suppose that  $\{\mathbf{x}_0, \dots, \mathbf{x}_{t-1}\} \subset S$ ; recall from the proof of Corollary 4.24 that  $\mathbf{x}_t = A\mathbf{z}_{t-1}$ , where  $\mathbf{z}_{t-1} = \Pi_{\mathcal{R}(\mathcal{X}_{t-2})^\perp}(\mathbf{x}_{t-1})$ . Since  $\mathbf{x}_{t-1} \in S$  and

$\text{span}\{\mathbf{x}_0, \dots, \mathbf{x}_{t-2}\} \subset S$ , one can conclude that  $\mathbf{z}_{t-1} \in S$ , and from the definition of  $S$ ,  $\mathbf{x}_t = A\mathbf{z}_{t-1} \in S$ . On the other hand, since  $\lambda_1, \lambda_2, \dots, \lambda_r$  are distinct eigenvalues, by Corollary 4.24,  $\mathbf{dim}(\text{span}\{\mathbf{x}_0, \dots, \mathbf{x}_{r-1}\}) = r$ . By hypothesis of the induction  $\text{span}\{\mathbf{x}_0, \dots, \mathbf{x}_{r-1}\} \subset S$ , and since  $\mathbf{dim}(S) = r$ , we conclude that  $\text{span}\{\mathbf{x}_0, \dots, \mathbf{x}_{r-1}\}$  must be the entire  $S$ , i.e.  $\text{span}\{\mathbf{x}_0, \dots, \mathbf{x}_{r-1}\} = S$ , proving the first claim. Lastly,  $\mathbf{z}_r = \Pi_{\mathcal{R}(\mathcal{X}_{r-1})^\perp}(\mathbf{x}_r) = 0$  since  $\mathbf{x}_r \in S$ , and thus  $\mathbf{x}_{r+1} = A\mathbf{z}_r = 0$ , thereby completing the proof.  $\square$

Following Proposition 4.25, if  $\mathbf{x}_0$  excites  $k$  modes of the system corresponding to distinct eigenvalues with trivial algebraic multiplicities, then Algorithm 4.1 identifies all the excited modes of the system in  $k$  iterations. Furthermore, this implies that  $\mathbf{x}_{k+1} = 0$ , i.e., DGR eliminates the unstable modes and regulates the unknown system in exactly  $k + 1$  iterations. This also implies that, if  $k < n$  then online regulation of the system is achieved, even before enough data is available for full identification of system parameters.

#### 4.7.1 Boosting the Performance of DGR

DGR as introduced in Algorithm 4.1 can become computationally burdensome for large-scale systems. This is mainly due to storing the entire history of data in  $\mathcal{X}_t$  and  $\mathcal{Y}_t$  followed by the update of the controller that finds the pseudoinverse as well as multiplication of these data matrices (steps 7-9). Assuming the SVD-based computation of pseudoinverse, the complexity of the method is<sup>13</sup>  $\mathcal{O}(n^2t)$ . In this section, we show that such computational burden can be circumvented using rank-1 modifications of data matrices as a result of the discrete nature of data collection in our setup. Note that for computing  $K_{t+1}$  from (4.6) we only need to access  $\mathcal{Y}_{t+1}\mathcal{X}_t^\dagger$  (rather than  $\mathcal{X}_t^\dagger$ ). To this end, we leverage the results of [Meyer, 1973] in order to find  $\mathcal{Y}_{t+1}\mathcal{X}_t^\dagger$  recursively as a function of  $\mathcal{Y}_t\mathcal{X}_{t-1}^\dagger$ ,  $\mathcal{X}_{t-1}\mathcal{X}_{t-1}^\dagger$ , and  $\mathbf{x}_t$ .

**Proposition 4.26.** *Let  $\mathcal{X}_{t-1}$  be as in Algorithm 4.1,  $\mathbf{x}_t$  be the state measurement at iteration*

---

<sup>13</sup>The multiplication  $\mathcal{Y}_{t+1}\mathcal{X}_t^\dagger$  enforces another  $\mathcal{O}(n^2t)$  complexity that can be significant for large  $n$ .

$t$  and  $\mathbf{z}_t = \Pi_{\mathcal{R}(\mathcal{X}_{t-1})^\perp} \mathbf{x}_t$ . If  $\mathbf{x}_t \notin \mathcal{R}(\mathcal{X}_{t-1})$  then

$$\mathcal{X}_t^\dagger = \begin{pmatrix} \mathcal{X}_{t-1}^\dagger - \gamma_t \mathbf{z}_t^\dagger \\ \mathbf{z}_t^\dagger \end{pmatrix}; \quad (4.15)$$

otherwise,

$$\mathcal{X}_t^\dagger = \begin{pmatrix} \mathcal{X}_{t-1}^\dagger - \epsilon_t \gamma_t \boldsymbol{\zeta}_t^\top \\ \epsilon_t \boldsymbol{\zeta}_t^\top \end{pmatrix}, \quad (4.16)$$

where  $\epsilon_t \in \mathbb{R}$ ,  $\gamma_t \in \mathbb{R}^t$ , and  $\boldsymbol{\zeta}_t \in \mathbb{R}^n$  are defined as,

$$\epsilon_t := \frac{1}{\|\gamma_t\|^2 + 1}, \quad \gamma_t := \mathcal{X}_{t-1}^\dagger \mathbf{x}_t, \quad \boldsymbol{\zeta}_t := (\mathcal{X}_{t-1}^\dagger)^\top \gamma_t. \quad (4.17)$$

**Proof.** Rearrange  $\mathcal{X}_t$  into,

$$\mathcal{X}_t = \begin{pmatrix} \mathcal{X}_{t-1} & 0 \end{pmatrix} + \mathbf{x}_t \mathbf{e}_{t+1}^\top.$$

Then, it is implied from Theorem 1 in [Meyer, 1973] that

$$\mathcal{X}_t^\dagger = \begin{pmatrix} \mathcal{X}_{t-1} & 0 \end{pmatrix}^\dagger + \left[ \mathbf{e}_{t+1} - \begin{pmatrix} \mathcal{X}_{t-1} & 0 \end{pmatrix}^\dagger \mathbf{x}_t \right] \mathbf{z}_t^\dagger,$$

whenever  $\mathbf{x}_t \notin \mathcal{R}(\mathcal{X}_{t-1})$ . Hence, by leveraging the SVD of  $\mathcal{X}_{t-1}$  and definition of pseudoinverse we get

$$\mathcal{X}_t^\dagger = \begin{pmatrix} \mathcal{X}_{t-1}^\dagger \\ 0 \end{pmatrix} + \left[ \mathbf{e}_{t+1} - \begin{pmatrix} \mathcal{X}_{t-1}^\dagger \\ 0 \end{pmatrix} \mathbf{x}_t \right] \mathbf{z}_t^\dagger = \begin{pmatrix} \mathcal{X}_{t-1}^\dagger - \gamma_t \mathbf{z}_t^\dagger \\ \mathbf{z}_t^\dagger \end{pmatrix}.$$

For the case when  $\mathbf{x}_t \in \mathcal{R}(\mathcal{X}_{t-1})$ , Theorem 3 in [Meyer, 1973] gives,

$$\mathcal{X}_t^\dagger = \begin{pmatrix} \mathcal{X}_{t-1}^\dagger \\ 0 \end{pmatrix} + \mathbf{e}_{t+1} \gamma_t^\top \mathcal{X}_{t-1}^\dagger - \frac{1}{\sigma} \mathbf{p} \mathbf{q}^\top,$$

where,  $\sigma = \|\gamma_t\|^2 + 1$ ,  $\mathbf{p} = -\|\gamma_t\|^2 \mathbf{e}_{t+1} - \begin{pmatrix} \gamma_t \\ 0 \end{pmatrix}$ ,  $\mathbf{q} = -\boldsymbol{\zeta}_t$ . The rest of the proof follows from rearranging the terms and using the definitions in (4.17).  $\square$



As mentioned earlier, the update of the controller requires  $\mathcal{Y}_t \mathcal{X}_{t-1}^\dagger$  that could become prohibitive for large  $n$ . However, we can take advantage of Proposition 4.26 to find this term recursively in order to avoid memory usage as well as computational burden.

**Theorem 4.27.** *Let  $\mathcal{X}_{t-1}$  be as in Algorithm 4.1 and  $\mathbf{x}_t$  be the state measurement collected at  $t$ . For  $t > 0$ , define  $\mathcal{P}_{t-1} := \mathcal{X}_{t-1} \mathcal{X}_{t-1}^\dagger$ ,  $\mathcal{Q}_{t-1} := \mathcal{Y}_t \mathcal{X}_{t-1}^\dagger$  and  $\mathbf{z}_t := \Pi_{\mathcal{R}(\mathcal{X}_{t-1})^\perp} \mathbf{x}_t$ . Then*

$$\begin{aligned} \mathbf{z}_t &= [\mathbf{I} - \mathcal{P}_{t-1}] \mathbf{x}_t, \\ \mathcal{Q}_t &= \mathcal{Q}_{t-1} + [\mathbf{x}_{t+1} - B\mathbf{u}_t - \mathcal{Q}_{t-1} \mathbf{x}_t] \mathbf{z}_t^\dagger, \\ \mathcal{P}_t &= \mathcal{P}_{t-1} + \mathbf{z}_t \mathbf{z}_t^\dagger. \end{aligned}$$

**Proof.** The expression for  $\mathbf{z}_t$  follows directly by properties of pseudoinverse. Next, observing that  $\mathbf{x}_{t+1} - B\mathbf{u}_t = A\mathbf{x}_t$  and so  $\mathcal{Q}_t = A\mathcal{P}_t$ , the recursive relation for  $\mathcal{Q}_t$  can be derived from the one for  $\mathcal{P}_t$ . Finally, Proposition 4.26 implies that, if  $\mathbf{x}_t \notin \mathcal{R}(\mathcal{X}_{t-1})$  then

$$\mathcal{P}_t = \mathcal{X}_t \mathcal{X}_t^\dagger = \begin{pmatrix} \mathcal{X}_{t-1} & \mathbf{x}_t \end{pmatrix} \begin{pmatrix} \mathcal{X}_{t-1}^\dagger - \gamma_t \mathbf{z}_t^\dagger \\ \mathbf{z}_t^\dagger \end{pmatrix} = \mathcal{P}_{t-1} + \mathbf{z}_t \mathbf{z}_t^\dagger;$$

otherwise,  $\mathcal{P}_t = \mathcal{P}_{t-1}$  because  $\mathcal{X}_{t-1} \gamma_t = \mathcal{X}_{t-1} \mathcal{X}_{t-1}^\dagger \mathbf{x}_t = \mathbf{x}_t$ . But in this case,  $\mathbf{z}_t = (\mathbf{I} - \mathcal{P}_{t-1}) \mathbf{x}_t = 0$  and therefore, the same recursion holds.  $\square$

Given the recursions introduced in Theorem 4.27, the refined (fast) version of DGR is displayed in Algorithm 4.2. At each iteration  $t$ , we update  $\mathcal{Q}_t$  based on the information from the new data and the projection  $\mathcal{P}_{t-1}$  (hidden in  $\mathbf{z}_t^\dagger$ ), which itself gets updated as a part of the recursion. The  $n \times n$  matrix  $\mathcal{Q}_t$  is then employed for the controller's update. Notice that  $\mathbf{z}_t$  is the same as in Lemma 4.14, however, here we compute it using  $\mathcal{P}_{t-1}$  which is obtained recursively.

Next, we discuss the convergence of DGR algorithm in the following. Note that, by Theorem 4.27, DGR and F-DGR are equivalent, and the following corollary provides a useful necessary and sufficient conditions for the convergence of  $\mathcal{Q}_t$ .

---

**Algorithm 4.2:** Fast Data-Guided Regulation (F-DGR)

---

- 1: **Initialization**
  - 2: Measure  $\mathbf{x}_0$ , set  $K_0 = 0$  and  $G_\alpha = [\alpha I + B^\top B]^\dagger B^\top$
  - 3: Set  $\mathcal{P}_{-1} = \mathcal{Q}_{-1} = \mathbf{0}$  and  $t = 0$
  - 4: **While stopping criterion not met**
  - 5: Compute  $\mathbf{u}_t = -K_t \mathbf{x}_t$
  - 6: Run system (4.1) and measure  $\mathbf{x}_{t+1}$
  - 7: Set  $\mathbf{z}_t = [\mathbf{I} - \mathcal{P}_{t-1}] \mathbf{x}_t$
  - 8:  $\mathcal{Q}_t = \mathcal{Q}_{t-1} + [\mathbf{x}_{t+1} - B \mathbf{u}_t - \mathcal{Q}_{t-1} \mathbf{x}_t] \mathbf{z}_t^\dagger$
  - 9:  $\mathcal{P}_t = \mathcal{P}_{t-1} + \mathbf{z}_t \mathbf{z}_t^\dagger$
  - 10:  $K_{t+1} = G_\alpha \mathcal{Q}_t$
  - 11:  $t = t + 1$
- 

**Corollary 4.28.** *Suppose that  $A$  is full-rank. Let  $\mathcal{Q}_t$  be as defined in Algorithm 4.2 and  $\mathcal{X}_{t-1}$  as in Algorithm 4.1. Then  $\mathcal{Q}_{t+k} = \mathcal{Q}_{t-1}$  for all  $k \geq 0$  if and only if  $\mathbf{x}_{t+k} \in \mathcal{R}(\mathcal{X}_{t-1})$  for all  $k \geq 0$ .*

**Proof.** By Theorem 4.27, we know that

$$\mathcal{Q}_t - \mathcal{Q}_{t-1} = A(\mathcal{P}_t - \mathcal{P}_{t-1}) = A \mathbf{z}_t \mathbf{z}_t^\dagger.$$

So, if  $\mathbf{x}_t \in \mathcal{R}(\mathcal{X}_{t-1})$  then  $\mathbf{z}_t = 0$  and thus  $\mathcal{Q}_t - \mathcal{Q}_{t-1} = 0$ ; otherwise  $\mathcal{Q}_t - \mathcal{Q}_{t-1} = A \mathbf{z}_t \mathbf{z}_t^\top / \mathbf{z}_t^\top \mathbf{z}_t$ , which does not vanish as  $A$  is assumed to be full-rank and  $\mathbf{z}_t \neq 0$ . The rest of the proof follows from this observation.  $\square$

A simple implication of the preceding corollary is that, in the worst case scenario—when  $A \in \mathbb{R}^{n \times n}$  has no zero eigenvalues and all of its modes are excited—the algorithm converges as soon as  $n$  linearly independent state measurements have been collected from the noise-less dynamics in (4.1). From this point on,  $\mathbf{u}_t$ 's as computed in DGR and F-DGR coincide with the solution of (4.5). However, the convergence of the controller in DGR and F-DGR may

happen earlier in the process whenever the future state measurements lie in the range of previous ones. For example, under hypothesis of Proposition 4.25, when  $\mathbf{x}_0$  excites  $k$  modes of  $A$  corresponding to  $r \leq k < n$  distinct eigenvalues, then both DGR and F-DGR converge in  $r$  iterations. In this case, the proposed controller may not coincide with the actual solution of the optimization in (4.5). Nonetheless, the online regulation of the system is guaranteed in general by Theorem 4.17. Note that, in presence of process noise in the dynamics (4.1), Corollary 4.28 is not valid and the convergence behavior of DGR (and F-DGR) will be dictated by the noise stochastics.

Finally,  $\mathbf{z}_t$  reflects the informativity of the newly generated data  $\mathbf{x}_t$ . In fact, based on its definition,  $\mathcal{Q}_t$  provides an estimate of  $A$  up to iteration  $t$ . Hence, the update of  $\mathcal{Q}_t$  as in Algorithm 4.2 essentially adjusts the prior estimate of  $A$  based on the new information encoded in the term  $A\mathbf{z}_t\mathbf{z}_t^\dagger$ . All in all, the machinery provided in this section circumvents the computational load of finding pseudoinverses by leveraging the recursive nature of the solution methodology.

#### 4.8 Numerical Simulations

In order to showcase the advantages of the proposed method in practical settings, we have implemented DGR on data collected from the X-29A aircraft. The Grumman X-29A is an experimental aircraft initially tested for its forward-swept wing; it was designed with a high degree of longitudinal static instability (due to the location of the aerodynamic center on the wings) for maneuverability, where linear models were leveraged to determine the closed-loop stability (Figure 4.4). The primary task of the control laws is to stabilize the longitudinal motion of the aircraft. To this end, the dynamic elements of the flight control system is designed for two general modes: 1) the Normal Digital Powered Approach (ND-PA) used in the takeoff and landing phase of the flight, and 2) the Normal Digital Up-and-Away (ND-UA) when otherwise.

For both flight modes, we study the case where the dynamics of the aircraft has been perturbed and unknown. This can be due to a mis-estimation of system parameters and/or

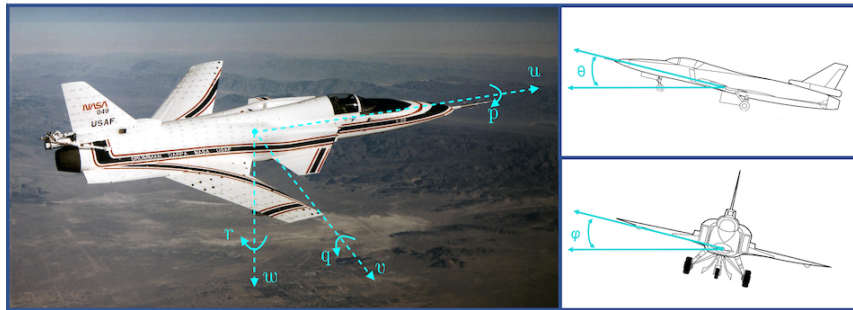


Figure 4.4: Grumman X-29A (*Credits: NASA Photo*), mainly known for its extreme instability while providing high-quality maneuverability; the longitudinal and lateral-directional states are illustrated.

any unpredicted flaw in the flight dynamics due to malfunction/damage. In this setting, the control laws designed for the original system fail and the system can become highly unstable. We then let DGR regulate the system; in this case, since the aircraft continues to operate safely, one can use any data-driven identification, stabilization, or robust control methods once enough data has been collected.

The longitudinal and lateral-directional dynamics each contains 4 states (see Figure 4.4). The nominal system parameters in each operating mode are obtained from Tables 9-10 and 13-14 in [Bosworth, 1992] (with fixed discretization step-size 0.05), whereas perturbation  $\Delta A$  is assumed to shift the dynamics to,

$$\mathbf{x}_{t+1} = (A + \Delta A)\mathbf{x}_t + B\mathbf{u}_t + \boldsymbol{\omega}_t,$$

where the elements of  $\Delta A$  are sampled from a normal distribution  $\mathcal{N}(0, 0.05)$ , and  $\boldsymbol{\omega}_t \sim \mathcal{N}(0, 0.01)$  denotes the process noise. Note that even though the nominal dynamics is known in this example, the proposed machinery makes no such *a priori* estimate, and assumes a completely unknown dynamics  $A_{\text{new}} := A + \Delta A$ . The original controller for the unperturbed system in each mode is assumed to be a closed-loop infinite horizon LQR with state and input weights  $Q = I$  and  $R = 10^{-7}$ .

We now aim to regulate the unstable system  $A_{\text{new}}$  from random initial states (where each state is sampled from  $\mathcal{N}(0, 10.0)$ ). Note that both the original system and the perturbed system have effective input characteristics that make them regularizable (with  $\rho(\tilde{A}) = 0.998$  and  $\rho(\tilde{A}_{\text{new}}) = 0.927$  for ND-PA mode, and  $\rho(\tilde{A}) = 0.998$  and  $\rho(\tilde{A}_{\text{new}}) = 0.932$  for ND-UA mode). The resulting state trajectories for ND-PA and ND-UA modes are demonstrated in Figures 4.5 and 4.6, and Figures 4.7 and 4.8, respectively. Without DGR, the norm of the state  $\|\mathbf{x}_t\|$  would grow rapidly (red curve) as the unknown system is unstable and the original control laws fail.<sup>14</sup> As the plots suggest, with DGR in the feedback loop (with the choice of  $\alpha = 5 \times 10^{-7}$ ),<sup>15</sup> the unstable modes can be suppressed resulting in stabilization of the system (the norm of the states in this case is demonstrated in black and each state is depicted in faded color).

Up to iteration  $t = 36$  for longitudinal and  $t = 30$  for lateral directional dynamics (shown with vertical dashed-line), enough data is generated in order to estimate the new system dynamics, or apply any other data-driven control using the data, (safely) generated by DGR up to this point. In what follows, we first showcase the complementary utility of DGR for identification-and-control; we then illustrate how it can also be incorporated for data-driven control.

In particular, the data is informative enough to identify system parameters through least squares denoted by  $\hat{A}$ . Therefore, one stopping criterion—which is also used here—is the point where the estimate of system parameters  $\hat{A}$  has converged. Then, one can replace DGR with a closed-loop infinite horizon LQR controller with some cost-weights  $Q$  and  $R$  which is obtained using the new estimate of the system dynamics. Here we set  $Q = I$  and  $R = 10^{-7}$  in order to make it comparable to the one-step quadratic cost used for DGR. In contrast to the original unstable LQR controller (red curve), it is shown that the new LQR controller for  $\hat{A}$  (blue curve) is stabilizing since we now have a more accurate estimate of the (perturbed)

---

<sup>14</sup>Since the LQR solution, in general, may have small stability margins for general parameter perturbations [Zhou et al., 1996].

<sup>15</sup>The positive choice for  $\alpha$  adjusts the compromise between the regulation of states in order to decrease 2-norm of the input. This may lead to larger upper bound specially when the system is unstable.

system parameters using the data generated safely by DGR in the loop.

Next, while the DGR is still in effect, the generated data matrix is not ill-conditioned and thus can be utilized to implement a data-driven control algorithm from that point onward. Due to the presence of noise and uncertainty, we have implemented the regularized version of Data-driven Model Predictive Control (MPC) as in [Coulson et al., 2019; Berberich et al., 2020b] with parameters  $T_{\text{ini}} = 1$ ,  $N = 4$ ,  $Q = 400I$ ,  $R = 0.05I$ ,  $\lambda_\sigma = 10^4$  and  $\lambda_g = 1$  for both dynamics, where the input is persistently exciting. With DGR, after enough data has been generated for each dynamics, the data-driven MPC algorithm is initiated; the norm of the corresponding state vector is depicted in yellow dash-dotted line labeled as “DGR+DeePC.”

On the other hand, one could consider implementing the data-driven MPC without DGR. However, this would require offline data which is not available a priori. Nonetheless, just for the purpose of comparison, this has been implemented based on *offline data* obtained from the original unstable plant. The resulting norm of the state has been depicted in cyan labeled as “Offline data+DeePC”. Due to the ill-posedness of the data matrix resulting from an unstable plant, it is observed that the practical tuning of the parameters could be problematic. This is due to the fact that maintaining the stability and feasibility of the resulting convex optimization problems is challenging due to conditioning in the dataset (for similar observations see for example [De Persis and Tesi, 2019]). These trajectories are terminated whenever the corresponding optimization problem was not numerically solvable/feasible. We have used the CVXPY package for solving the convex programs derived in all these cases [Diamond and Boyd, 2016].

Finally, the convergence of DGR algorithm in terms of the designed controller  $K_t$  is illustrated in Figure 4.9, where the values are seen to be well behaved after 30 to 40 iterations from the noisy dynamics. Furthermore, we note that as the process noise  $\omega_t$  decreases, the error for  $K_t$  tends to zero; in the absence of noise, the limiting error is in fact negligible.

In these examples, the bound derived in Theorem 4.17—that requires the knowledge of  $A_{\text{new}}$ —is plotted in green for comparison. The behavior of the bound follows our observations in Remark 4.18; the bound increases as the algorithm initially tries to “detect” the unstable

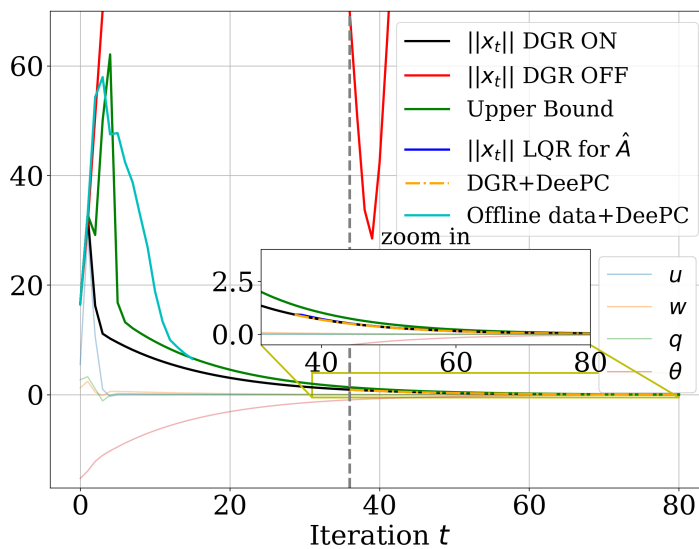


Figure 4.5: The state trajectory of X-29 in ND-PA mode with and without DGR for longitudinal control.

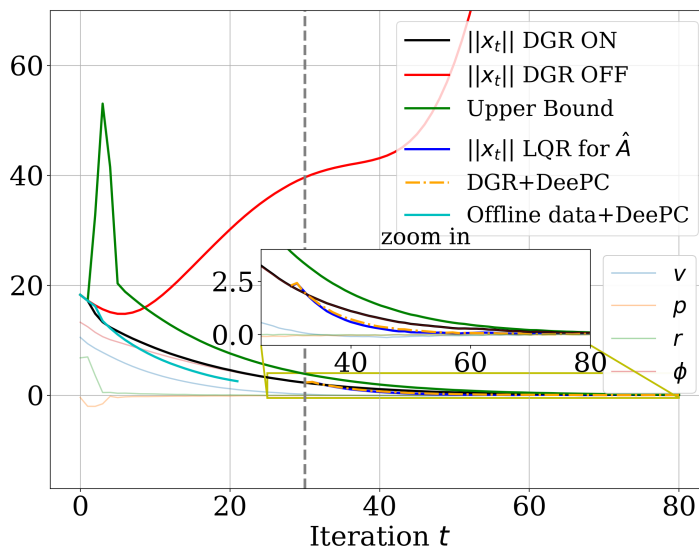


Figure 4.6: The state trajectory of X-29 in ND-PA mode with and without DGR for lateral-directional control.

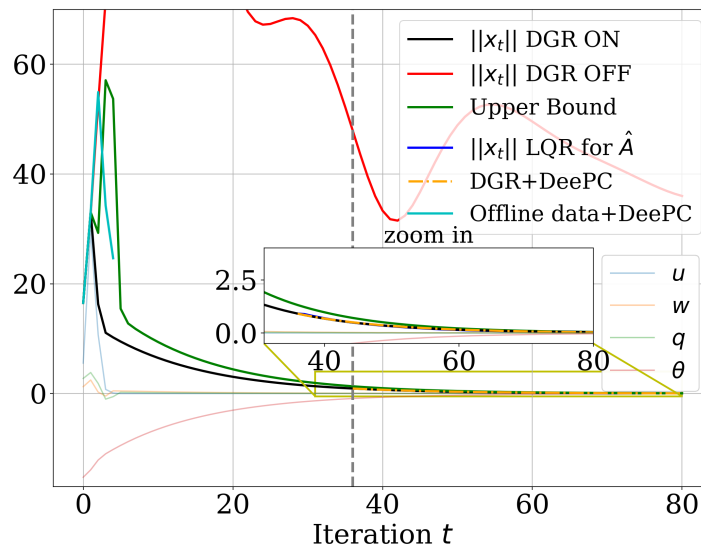


Figure 4.7: The state trajectory of X-29 in ND-UA mode with and without DGR for longitudinal control.

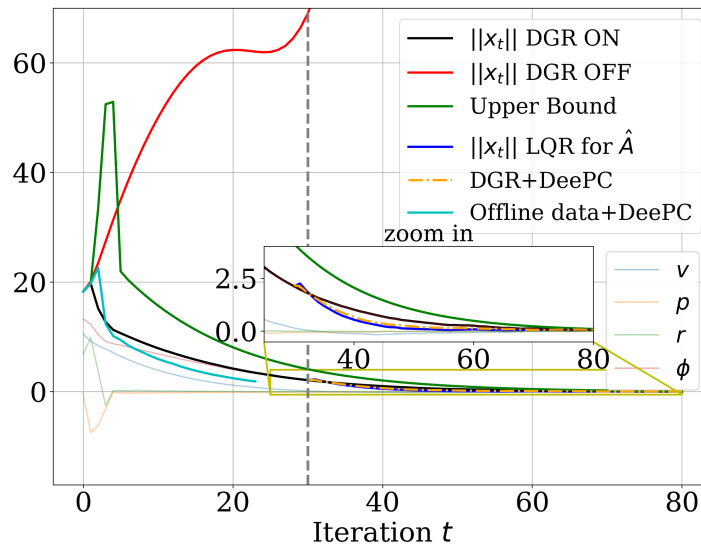


Figure 4.8: The state trajectory of X-29 in ND-UA mode with and without DGR for lateral-directional control.



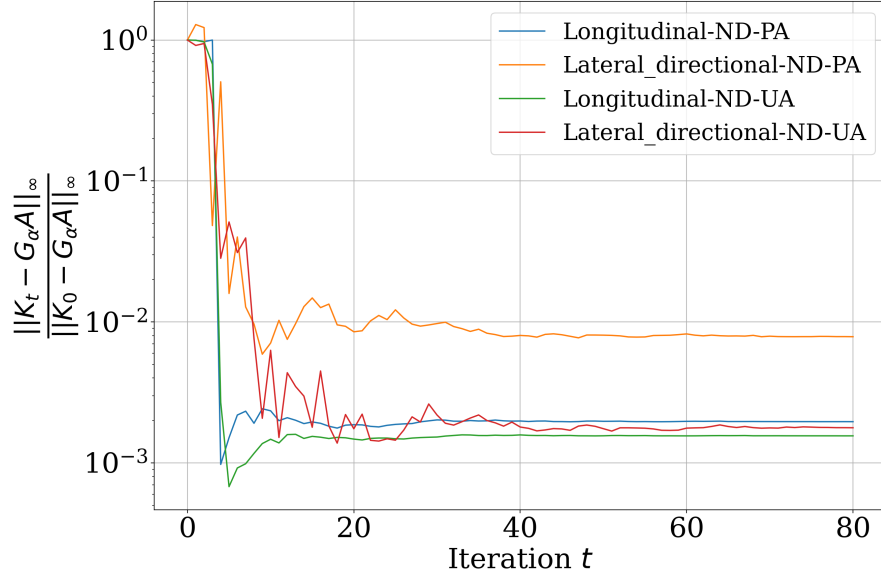


Figure 4.9: The convergence behavior of DGR algorithm for longitudinal and lateral-directional dynamics for both of flight modes.

modes, followed by suppressing these modes for regulation. We finally note that for large enough iterations, the rate of change of the upper bound is dictated by  $\rho(\tilde{A}_{\text{new}})$ , which in this case, is slightly less than one.<sup>16</sup>

#### 4.9 Remarks and Future Directions

In this chapter, we have introduced and characterized “regularizability,” a novel system theoretic notion that quantifies the ability of data-driven finite-time regulation for a linear system. This notion is in contrast to asymptotic behavior of the system, which is commonly characterized by stabilizability/controllability. Furthermore, we have proposed DGR, an online iterative feedback regulator for a partially unknown, potentially unstable, linear system using streaming data from a single trajectory. In addition to regulation of an unknown system, DGR leads to informative data that can subsequently be used for data-guided stabilization

---

<sup>16</sup>The code for these simulations can be found at <https://github.com/shahriarta/Data-Guided-Regulation>.

or system identification.<sup>17</sup> Along the way, we have provided another system theoretic notion referred to as “instability number” in order to analyze the performance of DGR and derive bounds on the trajectory of the system over a finite-time interval. Subsequently, we presented the application of the proposed online synthesis procedure on a highly maneuverable unstable aircraft. This example underscores how DGR can be integrated with other state-of-the-art data driven methods to achieve better performance through improved numerical conditioning.

The extensions of the results presented in this chapter to noisy dynamics as well as considering an unknown input matrix—in price of the guaranteed performance from the onset—are deferred to our future work. Furthermore, state regulation becomes more challenging when one only relies on partial observation of system’s trajectory or when the system is known to have multi-scale dynamics. Finally, our setup would be more practical considering input constraints, e.g., rate limits. While it is straightforward to address such extensions via convex constraints in the proposed DGR procedure, analysis of the resulting closed loop trajectory is more involved. The result of this chapter has been mainly adapted from [Talebi et al., 2021b].

Finally, recalling the result of Chapter 3, we considered the problem of collectively optimizing a distributed cost over a network of agents. That can be viewed as a cooperative game among these agents and a natural question in this setting would be what if the agents do not cooperate; rather, they tend to act only in their own favor—knowing that other agents might do the same. This viewpoint then motivates the problem of learning the Nash equilibrium for network of agents with “distributed cost”. This is the topic of next chapter, where we provide algorithms for learning the Nash equilibrium among (static) distributed agents with regret-type guarantees.

---

<sup>17</sup> This is guaranteed for example when  $\alpha = 0$ .

## Chapter 5

# DISTRIBUTED LEARNING IN NETWORK GAMES: A DUAL AVERAGING APPROACH

In this chapter, we propose a distributed no-regret learning algorithm for network games using a primal-dual method, i.e., dual averaging. With only locally available observations, we consider the scenario where each player optimizes a global objective, formed by local objective functions on the nodes of a given communication graph. Our learning algorithm for each player involves taking steps along their individual pay-off gradients, dictated by local observations of the other player's actions—where the local nature of information exchange is encoded by the network. The output is then projected back—again locally—to the set of admissible actions for each player. We provide the regret analysis of this distributed learning algorithm for the case of a deterministic network that is subjected to two teams with distinct objectives, and obtain  $O(\sqrt{T} \log(T))$  regret bound. Our analysis indicates the key correlation between the rate of convergence and network structure/connectivity that also appears in the distributed optimization setup via dual averaging. Finally, illustrative examples showcase the performance of our algorithm in relation to the size and connectivity of the communication network.

### **5.1 Introduction**

Networked systems analysis has been on the cutting edge of multi-disciplinary research over the past few years with applications spanning from robotic swarms to biological networks. Common to all of these systems, a global objective is achieved based on local interactions—which in turn—require local decision-making that is inherently restricted by limited information exchange and prescribed set of admissible policies [Mesbahi and Egerstedt, 2010]. As

such, many computationally efficient optimization algorithms [Nemirovsky and Yudin, 1983; Nesterov, 2009; Xiao, 2010; Sedghi et al., 2019] when implemented on a network have scaling issues when the network size grows. Not surprisingly, there has been an extensive literature on exploiting the structure of a information-exchange graph in order to reduce the complexity of distributed methods built upon subgradient optimization [Nedić and Ozdaglar, 2009], mirror decent [Doan et al., 2019], and dual averaging [Duchi et al., 2012], just to name a few. In the meantime, while these works mainly focus on a cooperative information sharing, many of the real-world systems exhibit non-cooperative behaviors due to a plethora of reasons including intrusions/attacks, greedy agents, competition for constrained resources, misaligned incentives, or the adversarial nature of environments that necessitate a game-theoretical model.

Game theory has been successfully employed in non-cooperative decision-making [Basar and Olsder, 1995; Engwerda, 2005a; Lambertini, 2018]. A variety of dynamic games including zero-sum, non-zero-sum [Starr and Ho, 1969] and Stackelberg games [Simaan and Cruz, 1973] have been studied in the literature for decades that model various non-cooperative interactions among decision makers, in scenarios such as network security [Manshaei et al., 2013], wireless networks [Srivastava et al., 2005; Han et al., 2012], and multi-agent systems such as pursuit evasion games [Vidal et al., 2002; Talebi et al., 2017],

Conventional centralized game-theoretic machinery are barely leveraged in real-world decision-making due to poor computational performance, lack of information, and policy limitations. In this direction, well-established algorithms such as primal-dual methods have been used for convergence analysis of learning and optimization in the game setup. For instance, by introducing *variational stability* condition, [Mertikopoulos and Zhou, 2018] considers a centralized multiagent decision process where each agent employs an online mirror descent learning algorithm. Also, [Bravo et al., 2018] studies the behaviour of learning with “bandit” feedback (a framework for low-information environments) in non-cooperative concave games. However, in spite of the improved convergence rates, scalability and limited communications are inherent source of intractability in these learning algorithms.

As the key connection, Nash Equilibrium (NE) plays a central role in game theory roughly defined as a strategy from which no player has any incentive to deviate unilaterally. Upon the convexity and continuous differentiability, the NE can be characterized as the solution of a variational inequality (VI) problem. Distributed Nash seeking refers to the class of problems that aims for learning a global NE with only local information [Li and Marden, 2013; Salehisadaghiani and Pavel, 2018; Frihauf et al., 2011; Alpcan and Başar, 2005]. In this direction, each node in the network usually represents an individual player, e.g., [Tatarenko et al., 2018] proposes an accelerated gradient decent for monotone games, [Gharesifard and Cortés, 2013] considers a zero-sum game between two networks characterizing conditions for convergence to NE, and [Grammatico, 2018] leverages proximal map composed with an averaging map where, under specific assumptions on the topology of the graph connecting agents, the overall algorithmic map becomes non-expansive. Additionally, the existence of a unique Nash equilibrium can be guaranteed under assumptions on the admissible action sets (like compactness) and the game Jacobian (like monotonicity and its variants) [Rosen, 1965; Facchinei and Pang, 2003; Parise and Ozdaglar, 2019].

In this chapter, we introduce a non-cooperative game between two teams, each consisting of players that interact over a network. While the goal of each team is to learn the global NE, players have no information about the other side, neither do they enjoy a global decision-making capability. Thus the equilibrium learning is only based on local observations of the opponent actions and distributed decision-making of nodes within their team. This setup resembles a scenario where each node in the network is subjected to actions that contribute to distinct objectives—in this sense, there is duality in the interactions between nodes in the network. For example, each node can represent a socio-economic entity, with objectives that are not completely aligned with each other (altruistic vs. profit-seeking). Financial networks consisting of nodes that are subjected to political influence from two opposing parties yet provide another scenario of interest. The network in these scenarios provides the backbone for information exchange and coordination amongst the teams. As such, it becomes imperative to characterize the role of the network in the evolution of team’s strategies, potentially

towards NE. Of prime interest in this chapter is how algebraic and combinatorial properties of the network, such as its connectivity, contribute to the convergence of distributed learning in the context of games. The choice of a primal/dual approach that is built around dual averaging [Nesterov, 2009], and its online implementation [Xiao, 2010], is primarily motivated by this overarching objective; this choice is also consistent with how dual averaging has been used in the context of distributed optimization to underscore the “network-effect” on the convergence and optimality properties of distributed optimization, where there is no “duality” in the nodes’ operation. However, in order to extend the work of [Duchi et al., 2012] to the game setting, one has to pay a particular attention to the information structure, as for example, it would be unreasonable to assume information sharing with the opponents in the game setting.

Accordingly, in our contribution: (1) We suppose players have neither *a priori* information about their adversary, nor global coordination over the network. Instead, they have a chance to choose a local strategy at each node and receive a local cost (reward), resulting in learning the global NE in a distributed fashion, (2) It is shown that the algorithm enjoys a similar regret bound as dual averaging methods which is well-known to be tight in black-box setting [Nemirovsky and Yudin, 1983]. We also discuss the convergence of the running average of actions to NE at each node under stronger regularity assumptions.

The rest of the chapter is organized as follows: In §5.2 we provide a quick overview of mathematical tools that are used in the chapter. In §5.3 we introduce the problem setup and propose our method. In §5.5 the required foundations to ultimately prove  $O(\sqrt{T} \log(T))$  regret bound associated with our algorithm are built up. We provide an illustrative example in §5.6.

## 5.2 Mathematical Preliminaries

We denote by  $\mathbb{R}$  the set of real numbers. A column vector with  $n$  elements is referred to as  $v \in \mathbb{R}^n$ , where  $v_i$  represents the  $i$ th element in  $v$ . The matrix  $M \in \mathbb{R}^{p \times q}$  contains  $p$  rows and  $q$  columns with  $M_{ij}$  denoting the element in the  $i$ th row and  $j$ th column of  $M$ . The square

matrix  $N \in \mathbb{R}^{n \times n}$  is *symmetric* if  $N^\top = N$ , where  $N^\top$  denotes the *transpose* of the matrix  $N$ . The  $n \times 1$  vector of all ones is denoted by  $\mathbf{1}$ . A doubly stochastic matrix  $P \in \mathbb{R}^{n \times n}$  is defined as a non-negative square matrix such that  $\sum_j P_{ij} = \sum_j P_{jk} = 1$  for all  $i$  and  $k$ , and  $\sigma_2(P)$  indicates the second largest singular value of  $P$ . We define  $[n] = \{1, \dots, n\}$ . The *Euclidean norm* of a vector  $x \in \mathbb{R}^n$  is defined as  $\|x\| = (x^\top x)^{1/2} = (\sum_{i=1}^n x_i^2)^{1/2}$  and the *dual norm* to  $\|x\|$  is denoted by  $\|x\|_* := \sup_{\|u\|=1} \langle x, u \rangle$ . Also, the *1-norm* is defined as  $\|x\|_1 = \sum_{i=1}^n |x_i|$ . A function  $f$  is convex if  $f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$  for all  $\theta \in (0, 1)$  and for all  $x, y$  in its convex domain, and  $g$  is a subgradient of  $f$  at point  $z$  if  $f(y) \geq f(z) + g^\top(y - z)$  for all  $y$ . If  $f$  is convex and differentiable, the gradient of  $f$ ,  $\nabla f(x)$ , is also a subgradient of  $f$  at  $x$ . The set of all subgradients of  $f$  at  $x$  is called subdifferential and denoted by  $\partial f(x)$ . A graph is characterized by the 2-tuple  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V}$  is the set of nodes and  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$  denotes the set of edges. An edge exists from node  $i$  to  $j$  if  $(i, j) \in \mathcal{E}$  which can also be shown by  $j \in \mathcal{N}_i$  where  $\mathcal{N}_i$  is the set of neighbors of node  $i$ . Then we say a graph is *complete* if  $(i, j) \in \mathcal{E}$  for all nodes  $i, j$ , and is *random  $k$ -regular* if  $|\mathcal{N}_i| = k$  for all  $i$  in some random order, and is called a *cycle* in case  $j \in \mathcal{N}_i$  if and only if  $j = i \pm 1$  with a cyclic order, i.e.,  $(1, n) \in \mathcal{E}$  in case there are  $n$  nodes in the graph.

### 5.3 Background and Problem Formulation

In this section we introduce the main framework of our analysis. Herein, we briefly mention some background material on dual averaging which is the workhorse of our methodology. We then continue by proposing the distributed setup followed by a two-player game-theoretic framework, which can be generalized to multi-player setting with minimum extra effort.

#### 5.3.1 Standard Dual Averaging

The dual averaging algorithm proposed by Nesterov [Nesterov, 2009] is a subgradient scheme for non-smooth convex problems. The primal-dual nature of this method generates two sequence of iterates  $\{x(t), z(t)\}_{t=0}^\infty$  contained within  $\mathcal{X} \times \mathbb{R}^d$  such that the updates of  $z(t)$  is responsible for averaging the support functions in the dual space, while the updates of  $x(t)$

establishes a dynamically updated scale between the primal and dual spaces. More precisely, after receiving the subgradient  $g(t) \in \partial f(x(t))$  at iteration  $t$ , the algorithm is updated as follows,

$$\begin{aligned} z(t+1) &= z(t) + \gamma(t)g(t), \\ x(t+1) &= \Pi_{\mathcal{X}}^{\psi}(-z(t+1), \alpha(t)), \end{aligned} \tag{5.1}$$

where  $\gamma(t) > 0$ ,  $\{\alpha(t)\}_{t=0}^{\infty}$  is a positive non-increasing sequence, and

$$\Pi_{\mathcal{X}}^{\psi}(z, \alpha) := \operatorname{argmin}_{x \in \mathcal{X}} \left\{ \langle -z, x \rangle + \frac{1}{\alpha} \psi(x) \right\} \tag{5.2}$$

is a generalized projection according to a strongly convex *prox-function*  $\psi(\cdot)$ .

### 5.3.2 Our Model

We consider the distributed learning problem for a game between two teams (players), both playing on a network consisting of  $n$  nodes connected via a communication graph  $\mathcal{G}$ <sup>1</sup>. To this end, each team has a representative on each node, hence  $2n$  members in total (Figure 5.1). The teams are grouped within the sets  $\mathcal{I}_{\ell} = \{(\ell, 1), \dots, (\ell, n)\}$  for  $\ell \in \{A, B\}$  and each team has a choice of action  $x_{\ell} \in \mathcal{X}_{\ell} \subset \mathbb{R}^{d_{\ell}}$  that minimizes the following global cost:

$$f_{\ell}(x_A, x_B) = \frac{1}{n} \sum_{i=1}^n f_{\ell,i}(x_A, x_B), \tag{5.3}$$

which is the average of its members' costs at each node  $i$ , denoted by  $f_{\ell,i}$ . The goal is to learn a global NE while players have no global decision-making capability. Instead,  $\mathcal{G}$  is assumed to be connected and players can communicate within their own team according to the network structure.

To learn a global NE, each team updates the state of its nodes using a distributed dual averaging type method. The network-based information flow of our algorithm is related to [Duchi et al., 2012] in the sense of consensus error of the dual variable, however, in a

---

<sup>1</sup>It is worth noting that different networks could be associated with each team where the communication graph would be their overlaps, which is not considered here due to brevity.



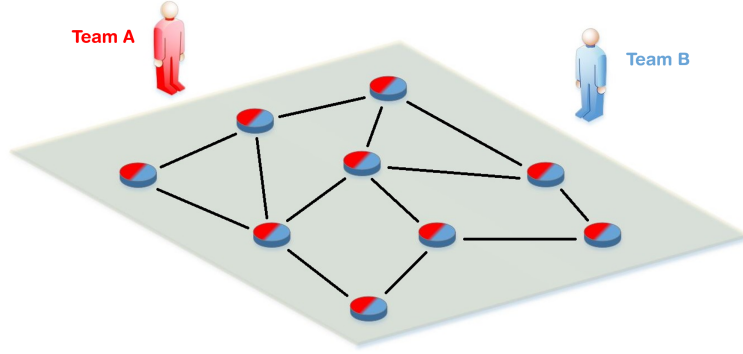


Figure 5.1: Schematic of our game setup including two teams (red A and blue B) with each team having a representative at each node.

non-cooperative game-theoretic setup, convergence of such iterative methods to NE is non-trivial due to nature of equilibria, limited information exchange, etc. Sufficient conditions of convergence is further discussed in §5.5.

In our proposed algorithm, at each node  $i$  and iteration  $t$ , player  $\ell \in \{A, B\}$  maintains an estimate of its team's action as  $x_{\ell,i}(t)$ . A communication protocol is designed for sharing dual variables among the nodes of each team, where node  $i$  updates its dual variable  $z_{\ell,i}(t)$  using a convex combination of those of its neighboring teammates. Then it maps  $z_{\ell,i}(t)$  back to the set of admissible actions  $\mathcal{X}_\ell$  followed by taking its local action  $x_{\ell,i}(t)$ . Subsequently, players observe the actions of the opponent at each node and locally obtain an estimate of the subgradient of their distributed cost (or reward). We show that under some regularity assumptions, this process provides players with enough local information to decide about the next step and eventually learn the global NE.

#### 5.4 Algorithm: TDA

We assume that the structure of  $\mathcal{G}$  induces a doubly stochastic matrix  $P_\ell$  available to each team where  $P_{\ell,ij} > 0$  if and only if nodes  $i$  and  $j$  are connected. Then each player  $\ell \in \{A, B\}$

at iteration  $t$  and each node  $i \in V$  computes updates by the following,

$$\begin{aligned} z_{\ell,i}(t+1) &= \sum_{j \in \mathcal{N}_{\ell,i}} P_{\ell,ij} z_{\ell,j}(t) + \gamma_{\ell}(t) g_{\ell,i}(t), \\ x_{\ell,i}(t+1) &= \Pi_{\mathcal{X}_{\ell}}^{\psi}(-z_{\ell,i}(t+1), \alpha_{\ell}(t)), \end{aligned} \quad (5.4)$$

where  $z_{\ell,i}$  and  $x_{\ell,i}$  are the dual variable and the local action of player  $\ell$  at node  $i$  respectively,  $g_{\ell,i}$  is a subgradient of the local cost  $f_{\ell,i}$  at the local actions  $x_{\ell,i}(t)$  as,

$$g_{\ell,i}(t) \in \partial_{\ell} f_{\ell,i}(x_{A,i}(t), x_{B,i}(t)), \quad (5.5)$$

where  $\partial_{\ell}$  is the differential w.r.t. the action of player  $\ell$ . Finally,  $\alpha_{\ell}(t)$  and  $\gamma_{\ell}(t)$  are sequences of positive stepsize with  $\alpha_{\ell}(t)$  being non-increasing. Note that  $x_{\ell,i}$  can be viewed as the local copy of  $x_{\ell}$  at node  $i$ , and its updates require access to only the  $i$ th row of the matrix  $P_{\ell}$ . We refer to the updates in (5.4) as *TDA*. The proposed methodology is summarized in Algorithm 5.1. We define the *running local average* at node  $i$ , for player  $\ell \in \{A, B\}$  as,

$$\hat{x}_{\ell,i}(t) := \frac{1}{t} \sum_{s=0}^{t-1} x_{\ell,i}(s). \quad (5.6)$$

In the sequel, we first show the sub-linear regret bound for TDA algorithm with appropriate choices of stepsizes. Finally, we discuss future directions regarding convergence of action iterates to NE under more regularity conditions.

## 5.5 Analysis of TDA

In this section, we present the main results of the chapter and the corresponding analysis. We first make a few key assumptions that are used in our subsequent analysis.

**Assumption 5.1.** The undirected communication graph  $\mathcal{G}$  is connected.

**Assumption 5.2.** The cost functions  $f_A$  and  $f_B$  satisfy the following:

1. The cost  $f_A(\cdot, x_2) : \mathcal{X}_A \mapsto \mathbb{R}$  is convex for any  $x_2$ , and similarly, the cost  $f_B(x_1, \cdot) : \mathcal{X}_B \mapsto \mathbb{R}$  is convex for any  $x_1$ .

---

**Algorithm 5.1:** Team-based Dual Averaging (TDA)

---

- 1: **Inputs:** For player  $\ell$
  - 2:     Local black-box oracle at node  $i$  to compute a subgradient of the local cost at any test point
  - 3:     Doubly stochastic matrix  $P_\ell$  induced by the network structure
  - 4: **Outputs:**
  - 5:     Estimates of NE  $\hat{x}_{\ell,i}(t)$  at node  $i$  for player  $\ell$
  - 6: **Initialize:**
  - 7:      $t = 0$
  - 8:     **For** Player  $\ell \in \{A, B\}$  at node  $i \in [n]$
  - 9:          $z_{\ell,i}(0) = 0$
  - 10:         Take random action  $x_{\ell,i}(0)$
  - 11: **while convergence**
  - 12:     **For** Player  $\ell \in \{A, B\}$  at node  $i \in [n]$
  - 13:         Observe the opponents local action and get
  - 14:          $g_{\ell,i}(t) \in \partial_\ell f_{\ell,i}(x_{A,i}(t), x_{B,i}(t))$
  - 15:         **For** Player  $\ell \in \{A, B\}$  at node  $i \in [n]$
  - 16:             Update the dual variable  $z_{\ell,i}(t+1)$  by (5.4)
  - 17:             Calculate and take action  $x_{\ell,i}(t+1)$  by (5.4)
  - 18:      $t = t + 1$
  - 19: **Return:** the *running average*  $\hat{x}_{\ell,i}$  by (5.6)
-

2. The cost  $f_A(\cdot, x_2)$  is  $L_A$ -Lipschitz continuous for any  $x_2$  such that  $|f_A(x_1, x_2) - f_A(x'_1, x_2)| \leq L_A \|x_1 - x'_1\|$ ,  $\forall x_1, x'_1 \in \mathcal{X}_A$ . Similarly, the cost  $f_B(x_1, \cdot)$  is  $L_B$ -Lipschitz continuous for any  $x_1$ , i.e.,  $|f_B(x_1, x_2) - f_B(x_1, x'_2)| \leq L_B \|x_2 - x'_2\|$ ,  $\forall x_2, x'_2 \in \mathcal{X}_B$ .

In contrast with optimization problems where the notion of “optimality” plays a central role, in a game setup, the objective is seeking an equilibrium rather than an “optimal” solution. In order to incorporate the interactions of players in convergence analysis of distributed algorithms, monotonicity has been known as a useful sufficient condition for problem “regularity,” originally due to the seminal work by Rosen [Rosen, 1965]. The monotonicity as defined by Rosen, and subsequently used in the game literature (e.g., [Facchinei and Pang, 2003; Parise and Ozdaglar, 2019] and references therein), refers to the property of the operator generated by the pseudo-gradient map of the game. The monotonicity condition provides sufficient condition for convergence of the action iterates which is discussed in Section §5.5.5.

### 5.5.1 Regret Analysis

First, we introduce two lemmas that equip us with the required tools to prove the basic convergence. Herein, we borrow some tools from prior works, as such, we only mention the key steps that distinguishes our contribution in the game setup. Also, it is worth mentioning that even though we prove convergence of TDA with  $\gamma_\ell(t) = 1$ , this choice can improve the convergence rates numerically and its characteristics will be analyzed in a followup work. For that, let us define  $\tilde{f}_A(\cdot; t)$  and  $\tilde{f}_B(\cdot; t)$  as follows

$$\begin{aligned}\tilde{f}_A(\cdot; t) &:= \frac{1}{n} \sum_{j=1}^n f_{A,j}(\cdot, x_{B,j}(t)), \\ \tilde{f}_B(\cdot; t) &:= \frac{1}{n} \sum_{j=1}^n f_{B,j}(x_{A,j}(t), \cdot),\end{aligned}$$

where  $x_{A,i}(t)$  and  $x_{B,j}(t)$  are the sequence generated by TDA algorithm. Now, for  $\ell \in \{A, B\}$  we define  $\mathcal{R}_\ell^{(i)}$  to be the regret of player  $\ell$  for implementation of TDA algorithm at node  $i$

versus any fixed action  $x_\ell^*$  while the other player is also implementing TDA, i.e.

$$\mathcal{R}_\ell^{(i)}(T) := \sum_{t=1}^T \tilde{f}_\ell(x_{\ell,i}(t); t) - \sum_{t=1}^T \tilde{f}_\ell(x_\ell^*; t). \quad (5.7)$$

**Lemma 5.3.** *Following Algorithm 5.1, suppose that player  $\ell \in \{A, B\}$  has access to  $g_{\ell,i}$  for  $(\ell, i) \in \mathcal{I}_\ell$  at each node  $i \in [n]$ , and  $\{\alpha_\ell(t)\}_{t=0}^\infty$  is a non-increasing sequence of positive stepsizes and  $\gamma_\ell(t) = 1$ . Then under Assumption 5.2 at each node, for any fixed  $x^* = (x_A^*, x_B^*) \in \mathcal{X}_A \times \mathcal{X}_B$  at node  $i$ ,*

$$\mathcal{R}_\ell^{(i)}(T) \leq \frac{\psi(x_\ell^*)}{\alpha_\ell(T)} + \frac{L_\ell^2}{2} \sum_{t=1}^T \alpha_\ell(t-1) + L_\ell \sum_{t=1}^T \alpha_\ell(t) \left( \mathcal{Z}_\ell^{(i)}(t) + \frac{2}{n} \sum_{j=1}^n \mathcal{Z}_\ell^{(j)}(t) \right)$$

where  $\mathcal{Z}_\ell^{(i)}(t) = \|\bar{z}_\ell(t) - z_{\ell,i}(t)\|_*$  is the consensus error at each time with  $\bar{z}_\ell(t) = \frac{1}{n} \sum_{i=1}^n z_{\ell,i}(t)$  as the averaging factor in the dual space.

**Proof.** For each team as the generalized projection, define

$$y_\ell(t) := \Pi_{\mathcal{X}_\ell}^\psi(-\bar{z}_\ell(t), \alpha(t)) = \operatorname{argmin}_{x \in \mathcal{X}_\ell} \left\{ \sum_{s=1}^{t-1} \left\langle \frac{1}{n} \sum_{i=1}^n g_{\ell,i}(s), x \right\rangle + \frac{1}{\alpha_\ell(t)} \psi(x) \right\}, \quad (5.8)$$

which follows from the iterative form of  $\bar{z}_\ell(t)$ ,

$$\bar{z}_\ell(t+1) = \bar{z}_\ell(t) + \frac{1}{n} \sum_{j=1}^n g_{\ell,j}(t),$$

and the zero choice of the dual initialization, where the above iterative form is implied by the doubly stochastic nature of  $P_\ell$ .

Now using the  $L_A$ -Lipschitz property of  $f_A$  we can show that,

$$\mathcal{R}_A^{(i)}(T) \leq \sum_{t=1}^T \left[ \tilde{f}_A(y_A(t)) - \tilde{f}_A(x_A^*) \right] + \sum_{t=1}^T L_A \|y_A(t) - x_{A,i}(t)\|.$$

Also, by adding and subtracting  $\sum_j f_{A,j}(x_{A,j}(t), x_{B,j}(t))$  to the first term and using convexity we have,

$$\mathcal{R}_A^{(i)}(T) \leq \sum_{t=1}^T \frac{1}{n} \sum_{j=1}^n \langle g_{A,j}(t), x_{A,j}(t) - x_A^* \rangle + \sum_{t=1}^T \frac{L_A}{n} \sum_{j=1}^n [\|y_A(t) - x_{A,j}(t)\| + \|y_A(t) - x_{A,i}(t)\|].$$

where  $g_{A,j}(t) \in \partial_A f_{A,j}(x_{A,j}(t), x_{B,j}(t))$  and the  $L_A$ -Lipschitz condition is leveraged again. Then by adding and subtracting  $y_A(t)$  in the inner-product we get that

$$\begin{aligned} \mathcal{R}_A^{(i)}(T) &\leq \sum_{t=1}^T \frac{1}{n} \sum_{j=1}^n \langle g_{A,j}(t), y_A(t) - x_A^* \rangle + \sum_{t=1}^T \frac{L_A}{n} \sum_{j=1}^n [2\|y_A(t) - x_{A,j}(t)\| + \|y_A(t) - x_{A,i}(t)\|] \\ &\leq \sum_{t=1}^T \left\langle \frac{1}{n} \sum_{j=1}^n g_{A,j}(t), y_A(t) - x_A^* \right\rangle + \frac{L_A}{n} \sum_{t=1}^T \alpha_A(t) \sum_{j=1}^n [2\mathcal{Z}_A^{(j)}(t) + \mathcal{Z}_A^{(i)}(t)], \end{aligned}$$

resulted from the  $L_A$ -Lipschitz property of  $f_A$  and  $\alpha$ -Lipschitz continuity of the generalized projection  $\Pi_{\mathcal{X}}^\psi(\cdot, \alpha)$  (which is a direct consequence of Lemma 1 in [Nesterov, 2009]). Finally, the results follows by applying Theorem 2 and Equation (3.3) in [Nesterov, 2009] (also restated as Lemma 3 in [Duchi et al., 2012]) to the first term in the inequality above and using the fact that  $\|g_{A,j}\|_* \leq L_A$ . Similar analysis results in the bound for  $\mathcal{R}_B^{(i)}(T)$ .  $\square$

### 5.5.2 Choice of the learning rate $\alpha_\ell(t)$

In order to achieve convergence, an appropriate choice of stepsize (learning rate)  $\alpha_\ell(t)$  is required. Next lemma shows how specific choices of  $\alpha_\ell(t)$  result in practical bounds on  $\mathcal{R}_\ell^{(i)}$  by getting rid of  $\mathcal{Z}_\ell^{(k)}$  terms.

**Theorem 5.4.** *Under Assumption 5.1 and 5.2, and notation of Lemma 5.3, suppose that  $\psi(x_\ell^*) \leq R_\ell^2$ . By choosing the stepsize,*

$$\alpha_\ell(t) = \frac{R_\ell \sqrt{1 - \sigma_2(P_\ell)}}{\sqrt{13} L_\ell \sqrt{t}},$$

for  $\ell \in \{A, B\}$  and  $\gamma_\ell(t) = 1$ , following Algorithm 5.1 at each node  $i$  we get,

$$\mathcal{R}_\ell^{(i)}(T) \leq \sqrt{T} \log(T\sqrt{n}) \frac{2\sqrt{13} R_\ell L_\ell}{\sqrt{1 - \sigma_2(P_\ell)}}.$$

**Proof.** Stacking the updates of dual variables in (5.4) into a matrix form  $Z_\ell = [z_{\ell,1} \ \dots \ z_{\ell,n}]$  and similarly for  $G_\ell$ . Then for an undirected graph ( $P_\ell^\top = P_\ell$ ), we get

$$Z_\ell(t+1) = Z_\ell(t)P_\ell + G_\ell(t).$$

Define,

$$\Phi_\ell(t + 1, s) = P_\ell^{t+1-s},$$

where  $\Phi$  can be a model of some transition matrix for a discrete-time linear system with states  $Z$ . This results in,

$$Z_\ell(t + 1) = Z(s)\Phi_\ell(t + 1, s) + \sum_{r=s+1}^{t+1} G_\ell(r - 1)\Phi_\ell(t + 1, r)$$

for  $0 \leq s \leq t$ . Noting  $\Phi_\ell(t + 1, s)\mathbb{1} = \mathbb{1}$  and from definitions  $\bar{z}_\ell(t) = Z_\ell(t)\mathbb{1}/n$  and  $z_{\ell,i}(t) = Z_\ell(t)e_i$ , it is straightforward to show,

$$\bar{z}_\ell(t) - z_{\ell,i}(t) = \bar{z}_\ell(s) - Z(s)\Phi_\ell(t, s)e_i + \sum_{r=s+1}^t G_\ell(r - 1)[\mathbb{1}/n - \Phi_\ell(t, r)e_i].$$

For simplicity we assume  $z_{\ell,i}(0) = \bar{z}_\ell(0)$  (say by choosing  $z_{\ell,i}(0) = 0$ ), then we have  $\bar{z}_\ell(s) - Z(s)\Phi_\ell(t, s)e_i = 0$  at  $s = 0$ . This implies,

$$\bar{z}_\ell(t) - z_{\ell,i}(t) = \sum_{r=1}^t G_\ell(r - 1)[\mathbb{1}/n - \Phi_\ell(t, r)e_i].$$

We can then proceed to bound this error,

$$\|\bar{z}_\ell(t) - z_{\ell,i}(t)\|_* \leq L_\ell \sum_{r=1}^t \|\mathbb{1}/n - \Phi_\ell(t, r)e_i\|_1,$$

where we used  $\|g_{\ell,i}\|_* \leq L_\ell$  and norm inequalities. Consider the following standard inequality ([Diaconis and Stroock, 1991]),

$$\|\mathbb{1}/n - \Phi_\ell(t, r)e_i\|_1 \leq \sqrt{n}\sigma_2(P_\ell)^{t+1-r}.$$

Now we say  $r$  is small if  $r \leq t + 1 + \frac{\log(T\sqrt{n})}{\log \sigma_2(P_\ell)}$ , otherwise it is large. Then by splitting the sum, one would note that for the small  $r$ ,

$$\|\mathbb{1}/n - \Phi_\ell(t, r)e_i\|_1 \leq \frac{1}{T},$$

since  $\sigma_2(P_\ell) < 1$ . Otherwise, for the large  $r$ ,

$$\|\mathbb{1}/n - \Phi_\ell(t, r)e_i\|_1 \leq 2.$$

Then it can be shown that,

$$\mathcal{Z}_\ell^{(i)} = \|\bar{z}_\ell(t) - z_{\ell,i}(t)\|_* \leq 2L_\ell \frac{\log(T\sqrt{n})}{1 - \sigma_2(P_\ell)}, \quad (5.9)$$

using  $\log \sigma_2(P_\ell)^{-1} \geq 1 - \sigma_2(P_\ell)$ . Then from Lemma 5.3,

$$\mathcal{R}_\ell^{(i)}(T) \leq \frac{\psi(x_\ell^*)}{\alpha_\ell(T)} + \frac{L_\ell^2}{2} \sum_{t=1}^T \alpha_\ell(t-1) + \frac{6L_\ell^2 \log(T\sqrt{n})}{1 - \sigma_2(P_\ell)} \sum_{t=1}^T \alpha_\ell(t).$$

Define the sequence  $\{\alpha(t)\}_{t=0}^\infty$  as,

$$\alpha_\ell(t) = \frac{K_\ell}{\sqrt{t}}, \quad \alpha_\ell(0) = 1.$$

Since  $\psi(x_\ell^*) \leq R_\ell^2$  and  $\sum_{t=1}^T t^{-1/2} \leq 2\sqrt{T} - 1$ ,

$$\mathcal{R}_\ell^{(i)}(T) \leq \sqrt{T} \left[ \frac{R_\ell^2}{K_\ell} + K_\ell L_\ell^2 + 12K_\ell L_\ell^2 \left( \frac{\log(T\sqrt{n})}{1 - \sigma_2(P_\ell)} \right) \right].$$

Choosing  $K_\ell = \frac{R_\ell \sqrt{1 - \sigma_2(P_\ell)}}{\sqrt{13}L_\ell}$  proves the claim.  $\square$

### 5.5.3 Convergence of TDA Algorithm

A natural question on the performance of the TDA algorithm pertains to the convergence of the corresponding action iterates. However, it is known that the convergence of action iterates for this general class of algorithms cannot be guaranteed without further regularity assumptions. Nevertheless, with minimal continuity assumptions, our next result ensures that the point of convergence of TDA algorithm is in fact a NE. First a relevant definition.

**Definition 5.5.** A network game is called *continuous* if the cost of player  $\ell \in \{A, B\}$  at node  $i \in [n]$  satisfies,

- $f_{\ell,i}(x_\ell, x_{-\ell})$  is continuously differentiable in  $x_\ell$ ,
- $f_{\ell,i}(x_\ell, x_{-\ell})$  and  $\nabla f_{\ell,i}(x_\ell, x_{-\ell})$  are both continuous in the joint variable  $(x_\ell, x_{-\ell})$ .



**Theorem 5.6.** *Under the notation adopted in Theorem 5.4, for a continuous network game, if Algorithm 5.1 converges, i.e.,  $x_{\ell,i}(t) \rightarrow x_{\ell,i}^*$  as  $t \rightarrow \infty$ , then  $x_{\ell,i}^* = x_{\ell}^*$  for all  $i \in [n]$  and  $(x_{\ell}^*, x_{-\ell}^*)$  is a NE.*

**Proof.** By continuous differentiability, we have that  $g_{\ell,i} = \nabla f_{\ell,i}$  for all  $i$  and  $g_{\ell,i} \rightarrow g_{\ell,i}^*$  due to the joint continuity. Now by  $\alpha_{\ell}$ -Lipschitz continuity of the generalized projection we have,

$$\|x_{\ell,i}(t) - x_{\ell,j}(t)\| \leq \alpha_{\ell}(t) [\mathcal{Z}_{\ell,i}(t) + \mathcal{Z}_{\ell,j}(t)].$$

From (5.9) and the choice of  $\alpha_{\ell}(t) = K_{\ell}/\sqrt{t}$  we conclude that for all  $i, j$ ,  $\|x_{\ell,i}(t) - x_{\ell,j}(t)\| \rightarrow 0$  as  $t \rightarrow \infty$ , and thus  $x_{\ell,i}^* = x_{\ell}^*$  for all  $i$ . Similarly, this implies  $y_{\ell}(t) \rightarrow x_{\ell}^*$  since,

$$\|y_{\ell}(t) - x_{\ell}^*\| \leq \alpha_{\ell}(t)\mathcal{Z}_{\ell,i}(t) + \|x_{\ell,i}(t) - x_{\ell}^*\|.$$

The rest of the proof is by contradiction. Suppose  $x^* = (x_{\ell}^*, x_{-\ell}^*)$  is not a NE and define  $G_{\ell}^* = G_{\ell}(x^*)$ . Then by definition of NE in our setup and noting that  $\nabla_{\ell} f_{\ell}(x^*) = G_{\ell}^* \mathbb{1}/n$ , at least for one of the players (say player  $\ell$ ) we have the following (Proposition 1.4.2 in [Facchinei and Pang, 2003]),

$$\exists q_{\ell} \in \mathcal{X}_{\ell} \quad \text{s.t.} \quad \langle G_{\ell}^* \mathbb{1}/n, q_{\ell} - x_{\ell}^* \rangle < 0.$$

By continuity there exist a constant  $c > 0$  and neighborhoods  $U, V$  of points  $x_{\ell}^*, G_{\ell}^* \mathbb{1}/n$ , respectively, such that,

$$\langle G'_{\ell} \mathbb{1}/n, q_{\ell} - x'_{\ell} \rangle \leq -c, \quad \forall x'_{\ell} \in U, \quad \forall G'_{\ell} \text{ s.t. } G'_{\ell} \mathbb{1}/n \in V.$$

On the other hand, by definition of  $\Pi_{\mathcal{X}}^{\psi}$  and  $y_{\ell}$  as in (5.8), and strong convexity of  $\psi$  we can conclude that (Theorem 23.5 in [Rockafellar, 2015])  $-\alpha_{\ell}(t)\bar{z}_{\ell}(t) \in \partial\psi(y_{\ell}(t))$ , and therefore,

$$\psi(q_{\ell}) - \psi(y_{\ell}(t)) \geq -\alpha_{\ell}(t)\langle \bar{z}_{\ell}(t), q_{\ell} - y_{\ell}(t) \rangle.$$

Note that by convergence of  $y_{\ell}(t)$ , there exists  $N$  such that  $y_{\ell}(t) \in U$  and  $G_{\ell}(t)\mathbb{1}/n \in V$ ,  $\forall t \geq N$ . Furthermore,  $\bar{z}_{\ell}(t) = \sum_{r=1}^{t-1} G_{\ell}(r)\mathbb{1}/n$  since  $z_{\ell,i}(0) = \bar{z}_{\ell}(0)$ ; thus we can conclude

that

$$\begin{aligned} & \psi(q_\ell) - \psi(y_\ell(t)) \\ & \geq \alpha_\ell(t) \sum_{r=N}^{t-1} c - \alpha_\ell(t) \langle \sum_{r=1}^{N-1} G_\ell(r) \mathbb{1}/n, q_\ell - y_\ell(t) \rangle. \end{aligned}$$

Now as  $t \rightarrow \infty$  the right hand side of the above inequality approaches positive infinity, which is a contradiction.  $\square$

#### 5.5.4 Convergence of Function Values using Cross-monotonicity assumption

Here, we propose *cross-monotonicity* as a regularity assumption that refers to the operator  $\Omega$  generated by the cross-pseudo-gradient map defined below. The reason for introducing this new regularity condition in the context of network games is to achieve convergence of function values, even though the action iterates might not converge necessarily. Later in Section §5.5.5, the conventional monotonicity is discussed that can ensure the convergence of action iterates.

**Definition 5.7.** (*Cross-monotonicity*) A two-player game setup with cost functions  $f_1(x_1, x_2)$  and  $f_2(x_1, x_2)$  is *cross-monotone* if

$$\left\langle \Omega(x_1, x_2) - \Omega(x_1^*, x_2^*), (x_1, x_2) - (x_1^*, x_2^*) \right\rangle \geq 0, \quad (5.10)$$

for all  $(x_1, x_2), (x_1^*, x_2^*) \in \mathcal{X}_1 \times \mathcal{X}_2$  where  $\Omega$  is a 2-tuple defined as  $\Omega(x_1, x_2) = (\xi(x_1, x_2), \zeta(x_1, x_2))$  with any

$$\xi(x_1, x_2) \in \partial_1 f_2(x_1, x_2), \quad \zeta(x_1, x_2) \in \partial_2 f_1(x_1, x_2).$$

The above inner-product is defined over the product space  $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ .

**Assumption 5.8.** The cost  $f_A(x_1, \cdot) : \mathcal{X}_B \mapsto \mathbb{R}$  is concave for any  $x_1$ , and similarly, the cost  $f_B(\cdot, x_2) : \mathcal{X}_A \mapsto \mathbb{R}$  is concave for any  $x_2$ .

**Proposition 5.9.** *Under Assumptions 5.1, 5.2 and 5.8 at each node and notations in Lemmas 5.3 and Theorem 5.4, if  $x^*$  is a global NE and the game is cross-monotone, then following Algorithm 5.1 at each node  $i$  we get,*

$$\lim_{T \rightarrow \infty} \mathcal{F}_A^{(i)}(T) = \lim_{T \rightarrow \infty} \mathcal{F}_B^{(i)}(T) = 0,$$

where

$$\begin{aligned} \mathcal{F}_A^{(i)}(t) &= f_A(\hat{x}_{A,i}(t), x_B^*) - f_A(x_A^*, x_B^*), \\ \mathcal{F}_B^{(i)}(t) &= f_B(x_A^*, \hat{x}_{B,i}(t)) - f_B(x_A^*, x_B^*), \\ \mathcal{H}_A^{(i)}(t) &= \langle h_{A,i}^*(t) - h_{A,i}(t), x_{B,i}(t) - x_B^* \rangle, \\ \mathcal{H}_B^{(i)}(t) &= \langle h_{B,i}^*(t) - h_{B,i}(t), x_{A,i}(t) - x_A^* \rangle. \end{aligned}$$

with  $\hat{x}_{\ell,i}(t)$  defined as in (5.6), and  $h_{\ell,i}$  and  $h_{\ell,i}^*$  as the sub-gradients of  $f_{\ell,i}$  detailed below,

$$\begin{aligned} h_{A,i}(t) &\in \partial_B f_{A,i}(x_{A,i}(t), x_{B,i}(t)), & h_{A,i}^*(t) &\in \partial_B f_{A,i}(x^*), \\ h_{B,i}(t) &\in \partial_A f_{B,i}(x_{A,i}(t), x_{B,i}(t)), & h_{B,i}^*(t) &\in \partial_A f_{B,i}(x^*). \end{aligned}$$

**Proof.** Convexity of  $f_A$  (in its first element) results in,

$$\mathcal{F}_A^{(i)}(T) \leq \frac{1}{T} \sum_{t=1}^T f_A(x_{A,i}(t), x_B^*) - f_A(x^*),$$

Using the  $L_A$ -Lipschitz property of  $f_A$  and  $\alpha$ -Lipchitz continuity of the generalized projection  $\Pi_{\mathcal{X}}^\psi(\cdot, \alpha)$  (similar as in lemma 5.3) we can show that,

$$\frac{1}{T} \sum_{t=1}^T f_A(x_{A,i}(t), x_B^*) - f_A(x^*) \leq \frac{1}{T} \sum_{t=1}^T f_A(y_A(t), x_B^*) - f_A(x^*) + \frac{L_A}{T} \sum_{t=1}^T \alpha_A(t) \mathcal{Z}_A^{(i)}(t).$$

By adding and subtracting  $\sum_j f_{A,j}(x_{A,j}(t), x_B^*)$ , we have the following bound,

$$f_A(y_A(t), x_B^*) - f_A(x^*) \leq \frac{L_A}{n} \sum_{j=1}^n \|y_A(t) - x_{A,j}(t)\| + \frac{1}{n} \sum_{j=1}^n f_{A,j}(x_{A,j}(t), x_B^*) - f_{A,j}(x^*).$$

Also, by adding and subtracting we get,

$$\begin{aligned} f_{A,j}(x_{A,j}(t), x_B^*) - f_{A,j}(x_A^*, x_B^*) &= \left( f_{A,j}(x_{A,j}(t), x_B^*) - f_{A,j}(x_{A,j}(t), x_{B,j}(t)) \right) \\ &\quad + \left( f_{A,j}(x_{A,j}(t), x_{B,j}(t)) - f_{A,j}(x_A^*, x_{B,j}(t)) \right) \\ &\quad + \left( f_{A,j}(x_A^*, x_{B,j}(t)) - f_{A,j}(x_A^*, x_B^*) \right). \end{aligned}$$

Now by Assumption 5.8, we can determine the following bound,

$$f_A(y_A(t), x_B^*) - f_A(x^*) \leq \frac{L_A}{n} \sum_{j=1}^n \alpha_A(t) \mathcal{Z}_A^{(j)}(t) + \frac{1}{n} \sum_{j=1}^n \langle g_{A,j}(t), x_{A,j}(t) - x_A^* \rangle + \frac{1}{n} \sum_{j=1}^n \mathcal{H}_A^{(j)}(t).$$

Also similar to the proof of lemma 5.3, by using Theorem 2 and Equation (3.3) in [Nesterov, 2009] (or Lemma 3 in [Duchi et al., 2012]) we have,

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^T \sum_{j=1}^n \langle g_{A,j}(t), x_{A,j}(t) - x_A^* \rangle &\leq \sum_{t=1}^T \left\langle \frac{1}{n} \sum_{j=1}^n g_{A,j}(t), y_A(t) - x_A^* \right\rangle + \frac{L_A}{n} \sum_{t=1}^T \sum_{j=1}^n \alpha_A(t) \mathcal{Z}_A^{(j)}(t) \\ &\leq \frac{14R_\ell L_\ell \sqrt{T}}{\sqrt{13} \sqrt{1 - \sigma_2(P_\ell)}} + \frac{L_A}{n} \sum_{t=1}^T \sum_{j=1}^n \alpha_A(t) \mathcal{Z}_A^{(j)}(t) \end{aligned}$$

By combining all the preceding upper bounds we have that,

$$\mathcal{F}_A^{(i)}(T) \leq \frac{L_A}{nT} \sum_{t=1}^T \sum_{j=1}^n \alpha_A(t) [2\mathcal{Z}_A^{(j)}(t) + \mathcal{Z}_A^{(i)}(t)] + \frac{14R_\ell L_\ell}{\sqrt{13} \sqrt{1 - \sigma_2(P_\ell)} \sqrt{T}} + \frac{1}{nT} \sum_{t=1}^T \sum_{j=1}^n \mathcal{H}_A^{(j)}(t).$$

By doing similar analysis for  $\mathcal{F}_B^{(i)}(T)$ , using (5.9), and the step size in theorem 5.4 we get that,

$$\mathcal{F}_A^{(i)}(T) + \mathcal{F}_B^{(i)}(T) \leq \frac{\log(T\sqrt{n})}{\sqrt{T}} \sum_{\ell \in \{A,B\}} \frac{2\sqrt{13}R_\ell L_\ell}{\sqrt{1 - \sigma_2(P_\ell)}} + \frac{1}{T} \frac{1}{n} \sum_{t=1}^T \sum_{j=1}^n \mathcal{H}_A^{(j)}(t) + \mathcal{H}_B^{(j)}(t).$$

Expanding the last term, and defining the operator  $\Omega_j$  (as in (5.10) for  $f_1 = f_{A,j}$  and  $f_2 = f_{B,j}$ ) we get,

$$\mathcal{H}_A^{(j)}(t) + \mathcal{H}_B^{(j)}(t) = -\langle \Omega_j(x_{A,j}(t), x_{B,j}(t)) - \Omega_j(x_A^*, x_B^*), (x_{A,j}(t), x_{B,j}(t)) - x^* \rangle$$

Since the game is cross-monotone at each node, i.e., the operator  $S_j$  is monotone (specifically at  $\xi = h_{B,j}$  and  $\zeta = h_{A,j}$ ) and we get that  $\mathcal{H}_A^{(j)}(t) + \mathcal{H}_B^{(j)}(t) \leq 0$ . Hence,

$$\mathcal{F}_A^{(i)}(T) + \mathcal{F}_B^{(i)}(T) \leq \frac{\log(T\sqrt{n})}{\sqrt{T}} \sum_{\ell \in \{A,B\}} \frac{2\sqrt{13}R_\ell L_\ell}{\sqrt{1 - \sigma_2(P_\ell)}}$$

Finally, since  $x^*$  is a global NE,  $\mathcal{F}_A^{(i)}(T), \mathcal{F}_B^{(i)}(T) \geq 0$  and the convergence will subsequently be guaranteed.  $\square$

### 5.5.5 Convergence of Action Iterates

For simplicity, suppose that  $f_A$  and  $f_B$  are differentiable, then under Assumption 5.2.1 a NE exists [Facchinei and Pang, 2003]. Define the operator  $\mathcal{Q}$  generated as the pseudo-gradient map as  $\mathcal{Q}(x_1, x_2) = (\nabla_A f_A(x_1, x_2), \nabla_B f_B(x_1, x_2))$ . Additionally, suppose that  $\mathcal{Q}$  is (strictly) monotone, i.e.

$$\left\langle \mathcal{Q}(x_1, x_2) - \mathcal{Q}(x_1^*, x_2^*), (x_1, x_2) - (x_1^*, x_2^*) \right\rangle \geq 0,$$

for all  $(x_1, x_2), (x_1^*, x_2^*) \in \mathcal{X}_A \times \mathcal{X}_B$  with equality if and only if  $(x_1, x_2) = (x_1^*, x_2^*)$  (where inner-product is defined over the product space  $\mathbb{R}^{d_A} \times \mathbb{R}^{d_B}$ ). Then it follows that there exists a unique NE [Facchinei and Pang, 2003] (see also Theorem 2 in [Rosen, 1965], where also a sufficient condition for this monotonicity condition is given in terms of the game Jacobin). Finally, combining this regularity assumption with the convergence of TDA algorithm in function values can facilitate establishing of convergence of the running average of action iterates to the unique NE, and will be investigated in a followup work.

## 5.6 Numerical Simulations

In this section we illustrate the performance of our method for a two-team game on a network where each player has an objective,

$$f_{\ell,i}(x_\ell, x_{-\ell}) = \frac{1}{2} \|x_\ell - a_{\ell,i}\|^2 + \frac{1}{4} \langle x_\ell, x_{-\ell} - b_{\ell,i} \rangle$$

where  $a_{\ell,i}$  and  $b_{\ell,i}$  are arbitrary prescribed parameters. The information about each player's cost is only known locally (at each node) and unknown to the opponent. Although the cost functions are locally Lipschitz, they can be treated as a Lipschitz continuous function over any bounded (potentially large) domain. We have simulated the performance of TDA over complete, random 6-regular, and cycle graphs each consisting of 50 nodes. Figure 5.3 shows the random 6-regular network simulated here. To illustrate the advantage of dual averaging in our algorithm, we compare the results with another distributed algorithm — referred to as *TMD* — which is an extension of *Distributed Mirror Descent* algorithm [Doan et al., 2019]

implemented on our game model, where  $\psi(\cdot) = \frac{1}{2} \|\cdot\|^2$ . Also, as in standard form, the stepsize of this algorithm  $\beta_\ell(t) = \frac{K_\ell}{t^{0.8}}$  is chosen to be not summable but square summable where  $K_\ell$  is the constant as in Theorem 5.4. The TMD algorithm is detailed in Algorithm 5.2.

We define NAE as the normalized mean of the running average error from NE over all nodes of the network as follows,

$$\text{NAE}(t) = \frac{\sum_{i=1}^n \|\hat{x}_{A,i}(t) - x_A^*\| + \|\hat{x}_{B,i}(t) - x_B^*\|}{\sum_{i=1}^n \|\hat{x}_{A,i}(0) - x_A^*\| + \|\hat{x}_{B,i}(0) - x_B^*\|} \quad (5.11)$$

Figure 5.2 shows the NAE at each iteration for both of TDA and TMD algorithms. It can be noted that on the networks with the same structure, the TDA method has a convergence rate much faster than the TMD, which is due to dual averaging nature of TDA. Clearly, as the connectivity of the network decreases from complete to cycle, convergence rate becomes slower. This is captured in Proposition 5.9 where lower connectivity in the network leads to an increase of  $\sigma_2(P_\ell)$  closer to 1, which results in a larger bound for each time horizon  $T$ .

Next, we examine the performance of TDA in terms of the required iteration for a given error (from the NE). To do so, we consider four complete graphs with 25, 50, 80, and 120 nodes and require the algorithm to achieve a NAE of less than 0.1. Figure 5.4 illustrates the number of iterations  $T$  needed for each network to achieve  $\text{NAE}(T) < 0.1$ . Since, the initialization of the algorithm is random, for each network we have illustrated the mean and variance of the number of iterations required by 30 realizations. It is clear that, the number of required iterations to achieve the same error-bound, increases exponentially in this simulation. Part of this dynamic is due to the fact that TDA is a first-order method in game setting and it is converging towards an equilibrium point, unlike distributed optimization case where the convergence is towards an attractive optimal point.

## 5.7 Remarks and Future Directions

In this chapter, we proposed a new model for network games where each player's cost is distributed over a network. Without global information and only with distributed decision-making for players, we have shown that dual averaging can be applied to the scenario where

---

**Algorithm 5.2:** Team-based Mirror Descent
 

---

- 1: **Inputs:** For player  $\ell \in \{A, B\}$
  - 2:     Local black-box oracle at node  $i$  to compute a subgradient of the local cost at any test point
  - 3:     Doubly stochastic matrix  $P_\ell$  induced by the network structure
  - 4: **Outputs:**
  - 5:     Estimates of NE as  $x_{\ell,i}(t)$  at node  $i$  for player  $\ell$
  - 6: **Initialize:**
  - 7:      $t = 0$
  - 8:     **For** Player  $\ell \in \{A, B\}$  at node  $i \in [n]$
  - 9:         Take random action  $x_{\ell,i}(0)$
  - 10: **while convergence**
  - 11:     **For** Player  $\ell \in \{A, B\}$  at node  $i \in [n]$
  - 12:         Observe the opponents action:  $x_{-\ell,i}(t)$
  - 13:         Calculate the average estimate:
  - 14:              $v_{\ell,i}(t) = \sum_{j \in \mathcal{N}_{\ell,i}} P_{\ell,ij} x_{\ell,j}(t)$
  - 15:             Get:  $g_{\ell,i}(t) \in \partial_\ell f_{\ell,i}(v_{\ell,i}(t), x_{-\ell,i}(t))$
  - 16:     **For** Player  $\ell \in \{A, B\}$  at node  $i \in [n]$
  - 17:         Calculate and take action
  - 18:              $x_{\ell,i}(t+1) = \text{Proj}_{\mathcal{X}_\ell}[v_{\ell,i}(t) - \beta_\ell(t)g_{\ell,i}(t)]$
  - 19:      $t = t + 1$
  - 20: **Return**  $x_{\ell,i}(t)$  at node  $i$  for player  $\ell \in \{A, B\}$
-

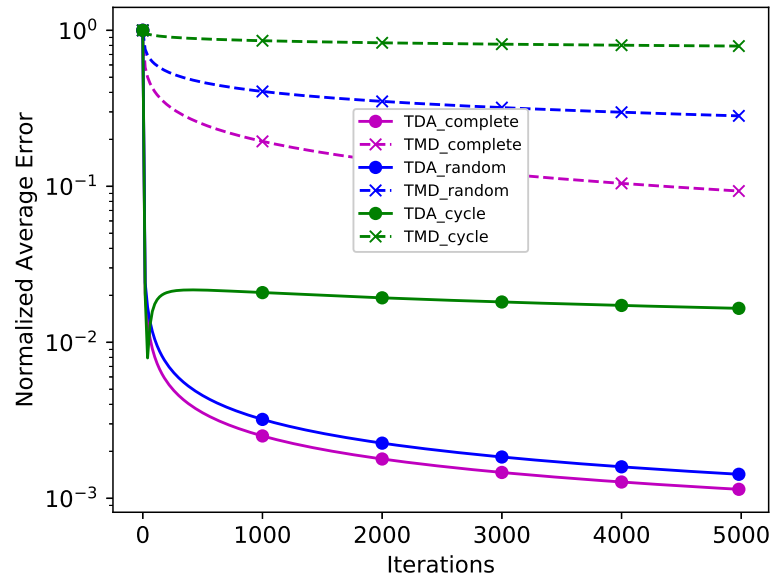


Figure 5.2: NAE at each iteration for both TDA and TMD algorithms in complete, random 6-regular, and cycle graphs with 50 nodes.

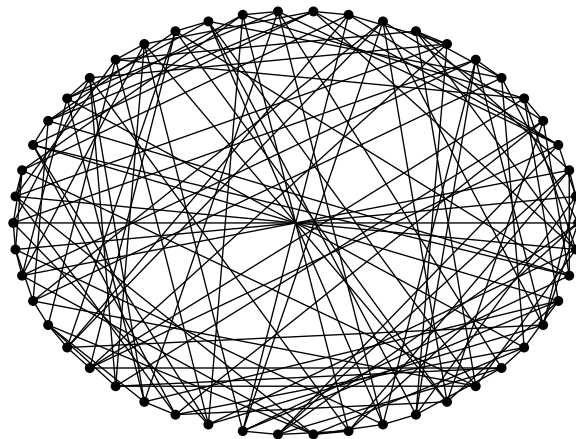


Figure 5.3: The random 6-regular networks with 50 nodes simulated in our setup.



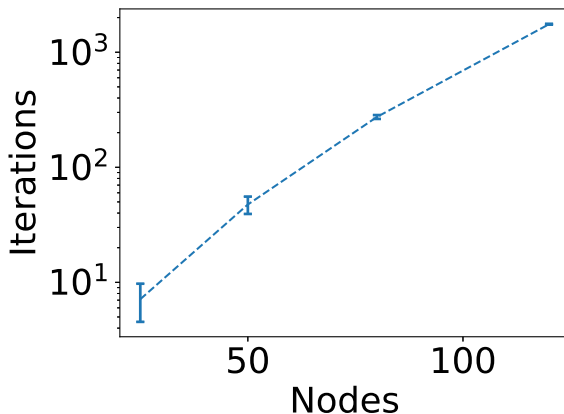


Figure 5.4: Number of iterations needed for each network so that NAE is less than 0.1.

two teams, with distinct objectives, can coordinate with their respective team members, over the network. In particular, using assumptions adopted in distributed optimization, we have shown that with proper choice of step-size the regret associated with our algorithm is sub-linear. Also, with cross-monotonicity condition on the game structure, the convergence to the function values are guaranteed.

Stochastic access to subgradient of cost functions as well as noisy observations of opponent’s action is considered as the next immediate extension of this work. Additionally, analysis of the stepsize for the dual variables and equipping the algorithm with feasible second-order information are yet other achievable furtherance that can improve the convergence behavior of this method close to equilibrium.

Next, the convergence of action iterates is still under investigation. Furthermore, the practicality of cross-monotonicity property needs to be further examined. Additionally, the result of this chapter has been mainly adapted from [Talebi et al., 2019].

Finally, in all the chapters so far, we have discussed learning constrained policies in various control and game settings. Yet, it is known that the estimation problem has various interesting connections to optimal control problems including “duality”. In other words, one

can show that the optimal mean-squared estimation problem has a dual reformulation in the form of an optimal LQR. Then, through this duality, one can learn the optimal estimation policy directly from output measurement data. This is the subject of the following final chapter in this dissertation.

## Chapter 6

# OPTIMAL FILTERING FOR LINEAR SYSTEMS WITH UNKNOWN NOISE COVARIANCES

This chapter examines learning the optimal filtering policy, known as the Kalman gain, for a linear system with unknown noise covariance matrices using noisy output data. The learning problem is formulated as a stochastic policy optimization, aiming to minimize the output prediction error. This formulation provides a direct bridge between data-driven optimal control and, its dual, optimal filtering. Our contributions are twofold. Firstly, we conduct a thorough convergence analysis of the stochastic gradient descent algorithm, adopted for the filtering problem, accounting for biased gradients and stability constraints. Secondly, we carefully leverage a combination of tools from linear system theory and high-dimensional statistics to derive bias-variance error bounds that scale logarithmically with problem dimension, and, in contrast to subspace methods, the length of output trajectories only affects the bias term.

### **6.1 Introduction**

The duality of control and estimation plays a crucial role in system theory, linking two distinct synthesis problems [Kalman, 1960a,b; Pearson, 1966; Mortensen, 1968; Bensoussan, 2018; Fleming and Mitter, 1982]. This duality is an effective bridge between two distinct disciplines, facilitating development of theoretical and computational techniques in one domain and then adopting them for use in the other. For example, the stability proof of the Kalman filter relies on the stabilizing characteristic of the optimal feedback gain in the dual LQR optimal control problem [Xiong, 2008, Ch.9]. In this chapter, we leverage this duality to learn optimal filtering policies using recent advances in data-driven algorithms for optimal control.

We consider the estimation problem for a system with a known linear dynamic and observation model, but unknown process and measurement noise covariances. Our objective is to learn the optimal steady-state Kalman gain using a training dataset comprising independent realizations of the observation signal. This problem has a rich history in system theory, often explored within the context of adaptive Kalman filtering [Mehra, 1970, 1972; Carew and Belanger, 1973; Belanger, 1974; Myers and Tapley, 1976; Tajima, 1978]. A comprehensive summary of four solution approaches to this problem can be found in the classical reference [Mehra, 1972]. These approaches include Bayesian inference [Magill, 1965; Hilborn and Lainiotis, 1969; Matisko and Havlena, 2010], Maximum likelihood [Kashyap, 1970; Shumway and Stoffer, 1982], covariance matching [Myers and Tapley, 1976], and innovation correlation methods [Mehra, 1970; Carew and Belanger, 1973]. While Bayesian and maximum likelihood approaches are computationally intensive, and covariance matching introduces biases in practice, the innovation correlation-based approaches have gained popularity and have been the subject of recent research [Odelson et al., 2006; Åkesson et al., 2008; Dunik et al., 2009]. For an excellent survey on this topic, refer to the article [Zhang et al., 2020]. However, it is important to note that these approaches often lack non-asymptotic guarantees and heavily depend on statistical assumptions about the underlying model.

In the realm of optimal control, significant progress has been made in the development of data-driven synthesis methods. Notably, recent advances have focused on the adoption of first-order methods for state-feedback LQR problems [Bu et al., 2019a, 2020a]. The direct optimization of policies from a gradient-dominant perspective has first proven in [Fazel et al., 2018] to be remarkably effective with global convergence despite non-convex optimization landscape. It has been demonstrated that despite the non-convex nature of the cost function, when expressed directly in terms of the policy, first-order methods exhibit global convergence to the optimal policy. Building upon this line of work, the use of first-order methods for policy optimization has been explored in variants of the LQR problem. These include OLQR [Fatkhullin and Polyak, 2020], model-free setup [Mohammadi et al., 2021a], risk-constrained setup [Zhao et al., 2023], LQG [Tang et al., 2021], and most recently, Riemannian constrained

LQR [Talebi and Mesbahi, 2022]. These investigations have expanded the scope of data-driven optimal control, demonstrating the versatility and applicability of first-order methods for a wide range of synthesis problems.

The objective of this chapter is to provide fresh insights into the classical estimation problem by leveraging the duality between control and estimation and incorporating recent advances in data-driven optimal control. Specifically, building on the fundamental connection between the optimal mean-squared error estimation problem and the LQR problem (Prop. 6.3), we reformulate determining the optimal Kalman gain as a problem of synthesizing an optimal policy for the adjoint system, under conditions that differ from those explored in the existing literature (see (6.10) and Remark 6.5). Upon utilizing this relationship, we propose a SGD algorithm for learning the optimal Kalman gain, accompanied by novel non-asymptotic error guarantees in presence of biased gradient and stability constraint. Our approach opens up promising avenues for addressing the estimation problem with robust and efficient data-driven techniques. The following is an informal statement of our main results (combination of Thm. 6.17 and Thm. 6.24), and missing proofs appear in the supplementary materials.

**Theorem 6.1** (Informal). *Suppose the system is observable and both dynamic and measurement noise are bounded. Then, with high probability, direct policy updates using stochastic gradient descent with small stepsize converges linearly and globally (from any initial stabilizing policy) to the optimal steady-state Kalman gain.*

More recently, the problem of learning the Kalman gain has been considered from a system identification perspective, for completely *unknown* linear systems [Lale et al., 2020a; Tsiamis and Pappas, 2019, 2023; Umenberger et al., 2022]. In [Tsiamis and Pappas, 2019] and [Tsiamis and Pappas, 2023], subspace system identification methods are used to obtain error bounds for learning the Markov parameters of the model over a time horizon and establish logarithmic regret guarantee for output prediction error. Due to the inherent difficulty of learning a completely unknown stochastic system from partial observations, subspace

methods assume marginal *stability of the unknown system*, and lead to sub-optimal sample complexity bounds that grow with the number of Markov parameters, instead of the number of unknowns [Tsiamis et al., 2022, pp. 14]. Alternatively, [Umenberger et al., 2022] considers minimizing the output prediction error and introduces a model-free policy gradient approach, under the same stability assumptions, that achieves *sublinear convergence rate*. This work provides a middle ground between completely known and completely unknown systems, for a learning scenario that not only has relevant practical implications, but also utilizes the duality relationship to LQR to establish *linear convergence rates* even for *unstable systems* as long as they are *observable*.

## 6.2 Background and Problem Formulation

Herein, first we propose the model setup in detail and discuss the Kalman filter as the estimation strategy. Consider the discrete-time filtering problem given by the stochastic difference equations,

$$x(t+1) = Ax(t) + \xi(t), \quad \text{and} \quad y(t) = Hx(t) + \omega(t), \quad (6.1)$$

where  $x(t) \in \mathbb{R}^n$  is the state of the system,  $y(t) \in \mathbb{R}^m$  is the observation signal, and  $\{\xi(t)\}_{t \in \mathbb{Z}}$  and  $\{\omega(t)\}_{t \in \mathbb{Z}}$  are the uncorrelated zero-mean random vectors, that represent the process and measurement noise respectively, with the following covariances,

$$\mathbb{E}[\xi(t)\xi(t)^\top] = Q \in \mathbb{R}^{n \times n}, \quad \mathbb{E}[\omega(t)\omega(t)^\top] = R \in \mathbb{R}^{m \times m},$$

for some positive (semi-)definite matrices  $Q, R \succeq 0$ . Let  $m_0$  and  $P_0 \succeq 0$  denote the mean and covariance of the initial condition  $x_0$ .

In the filtering setup, the state  $x(t)$  is hidden, and the objective is to estimate it given the history of the observation signal  $\mathcal{Y}(t) = \{y(0), y(1), \dots, y(t-1)\}$ . The best mean-squared error (MSE) estimate of  $x(t)$  is defined according to

$$\hat{x}(t) = \arg \min_{\hat{x} \in \sigma(\mathcal{Y}(t))} \mathbb{E} [\|x(t) - \hat{x}\|^2] \quad (6.2)$$

where  $\sigma(\mathcal{Y}(t))$  denotes the  $\sigma$ -algebra generated by the history of the observation signal  $\mathcal{Y}(t)$ . If the model parameters  $(A, H, Q, R)$  are known, the optimal MSE estimate  $\hat{x}(t)$  can be recursively computed by the Kalman filter algorithm [Kalman, 1960b]:

$$\hat{x}(t+1) = A\hat{x}(t) + L(t)(y(t) - H\hat{x}(t)), \quad \hat{x}(0) = m_0, \quad (6.3a)$$

$$P(t+1) = AP(t)A^\top + Q - AP(t)H^\top S(t)^{-1}HP(t)A^\top, \quad P(0) = P_0, \quad (6.3b)$$

where  $S(t) = HP(t)H^\top + R$ ,  $L(t) := AP(t)H^\top S(t)^{-1}$  is the Kalman gain, and  $P(t) := \mathbb{E}[(x(t) - \hat{x}(t))(x(t) - \hat{x}(t))^\top]$  is the error covariance matrix.

**Assumption 6.2.** The pair  $(A, H)$  is detectable, and the pair  $(A, B)$ , where  $Q = BB^\top$ , is stabilizable.

Under this assumption, the error covariance  $P(t)$  converges to a steady-state value  $P_\infty$ , resulting in a steady-state Kalman gain  $L_\infty = AP_\infty H^\top (HP_\infty H^\top + R)^{-1}$  [Kwakernaak and Sivan, 1969; Lewis, 1986]. It is common to evaluate the steady-state Kalman gain  $L_\infty$  offline and use it, instead of  $L(t)$ , to update the estimate in real-time.

### 6.2.1 Learning problem

Now, we describe our learning setup:

- 1) The system matrices  $A$  and  $H$  are known, but the process and the measurement noise covariances,  $Q$  and  $R$ , are *not* available.
- 2) We have access to an oracle that generates independent realizations of the observation signal for given length  $T$ :  $\{y(t)\}_{t=0}^T$ . However, ground-truth measurements of the state  $x(t)$  is *not* available.

This setting arises in various important engineering applications, such as aircraft wing dynamics, where merely approximate or reduced-order linear models are available due to difficulty in analytically capturing the effect of complex dynamics or disturbances, hence represented by noise with unknown covariance matrices [Zhang et al., 2020].

Inspired by the structure of the Kalman filter, our goal is to learn the steady-state Kalman gain  $L_\infty$  from the data described in the learning setup:

Given: independent random realizations of  $\{y(0), \dots, y(T)\}$  with the parameters  $A, H$

Learn: steady-state Kalman gain  $L_\infty$

For that, we formulate the learning problem as a stochastic optimization described next.

### 6.2.2 Stochastic optimization formulation

Define  $\hat{x}_L(T)$  to be the estimate given by the Kalman filter at time  $T$  realized by the constant gain  $L$ . Rolling out the update law (6.3a) for  $t = 0$  to  $t = T - 1$ , and replacing  $L(t)$  with  $L$ , leads to the following expression for the estimate  $\hat{x}_L(T)$  as a function of  $L$ ,

$$\hat{x}_L(T) = A_L^T m_0 + \sum_{t=0}^{T-1} A_L^{T-t-1} L y(t), \quad (6.4)$$

where  $A_L := A - LH$ . Note that evaluating this estimate does not require knowledge of  $Q$  or  $R$ . However, it is not possible to directly aim to learn the gain  $L$  by minimizing the MSE (6.2) because the ground-truth measurement of the state  $x(T)$  is not available. Instead, we propose to minimize the MSE in predicting the observation  $y(T)$  as a surrogate objective function:

$$\min_L J_T^{\text{est}}(L) := \mathbb{E} [\|y(T) - \hat{y}_L(T)\|^2] \quad (6.5)$$

where  $\hat{y}_L(T) := H\hat{x}_L(T)$ . Note that while the objective function involves finite time horizon  $T$ , our goal is to learn the steady-state Kalman gain  $L_\infty$ .

Numerically, this problem falls into the category of stochastic optimization and can be solved by algorithms such as Stochastic Gradient Descent (SGD). Such an algorithm would need access to independent realizations of the observation signal which are available. Theoretically however, it is not yet clear if this optimization problem is well-posed and admits a unique minimizer. This is the subject of following section where certain properties of the objective function, such as its gradient dominance and smoothness, are established. These



theoretical results are then used to analyze first-order optimization algorithms and provide stability guarantees of the estimation policy iterates. The results are based on the duality relationship between estimation and control that is presented next.

### 6.3 Estimation-Control Duality Relationship

The stochastic optimization problem (6.5) is related to an LQR problem through the application of the classical duality relationship between estimation and control [Åström, 2012, Ch.7.5]. In order to do so, we introduce the adjoint system (dual to (6.1)) according to:

$$z(t) = A^\top z(t+1) - H^\top u(t+1), \quad z(T) = a \quad (6.6)$$

where  $z(t) \in \mathbb{R}^n$  is the adjoint state and  $\mathcal{U}(T) := \{u(1), \dots, u(T)\} \in \mathbb{R}^{mT}$  are the control variables (dual to the observation signal  $\mathcal{Y}(T)$ ). The adjoint state is initialized at  $z(T) = a \in \mathbb{R}^n$  and simulated *backward in time* starting with  $t = T - 1$ . We introduce an LQR cost for the adjoint system:

$$J_T^{\text{LQR}}(a, \mathcal{U}_T) := z^\top(0)P_0z(0) + \sum_{t=1}^T [z^\top(t)Qz(t) + u^\top(t)Ru(t)], \quad (6.7)$$

and formalize a relationship between linear estimation policies for the system (6.1) and linear control policies for the adjoint system (6.6). A linear estimation policy takes the observation history  $\mathcal{Y}_T \in \mathbb{R}^{mT}$  and outputs an estimate  $\hat{x}_{\mathcal{L}}(T) := \mathcal{L}(\mathcal{Y}_T)$  where  $\mathcal{L} : \mathbb{R}^{mT} \rightarrow \mathbb{R}^n$  is a linear map. The adjoint of this linear map, denoted by  $\mathcal{L}^\dagger : \mathbb{R}^n \rightarrow \mathbb{R}^{mT}$ , is used to define a control policy for the adjoint system (6.6) which takes the initial condition  $a \in \mathbb{R}^n$  and outputs the control signal  $\mathcal{U}_{\mathcal{L}^\dagger} = \mathcal{L}^\dagger(a)$ .

$$\begin{array}{ccc} \{y(0), \dots, y(T-1)\} & \xrightarrow{\mathcal{L}} & \hat{x}_{\mathcal{L}}(T) \\ \{u(1), \dots, u(T)\} & \xleftarrow{\mathcal{L}^\dagger} & a \end{array}$$

The duality relationship between optimal MSE estimation and LQR control is summarized in the following proposition. The proof is presented in the supplementary material.

**Proposition 6.3.** *Consider the estimation problem for the system (6.1) and the LQR problem (6.7) subject to the adjoint dynamics (6.6).*

1. *For any linear estimation policy  $\hat{x}_{\mathcal{L}}(T) = \mathcal{L}(\mathcal{Y}_T)$ , and for any  $a \in \mathbb{R}^n$ , we have the identity*

$$\mathbb{E} [|a^\top x(T) - a^\top \hat{x}_{\mathcal{L}}(T)|^2] = J_T^{LQR}(a, \mathcal{U}_{\mathcal{L}^\dagger}(T)), \quad (6.8)$$

where  $\mathcal{U}_{\mathcal{L}^\dagger}(T) = \mathcal{L}^\dagger(a)$ .

2. *In particular, for a Kalman filter with constant gain  $L$ , the output prediction error (6.5)*

$$\mathbb{E} [\|y(T) - \hat{y}_L(T)\|^2] = \sum_{i=1}^m J_T^{LQR}(H_i, \mathcal{U}_{L^\top}(T)) + \text{tr}[R], \quad (6.9)$$

where  $\mathcal{U}_{L^\top}(T) = \{L^\top z(1), L^\top z(2), \dots, L^\top z(T)\}$ , i.e., the feedback control policy with constant gain  $L^\top$ , and  $H_i^\top \in \mathbb{R}^n$  is the  $i$ -th row of the  $m \times n$  matrix  $H$  for  $i = 1, \dots, m$ .

*Remark 6.4.* The duality is also true in the continuous-time setting where the estimation problem is related to a continuous-time LQR. Recent extensions to the nonlinear setting appears in [Kim et al., 2019] with a comprehensive study in [Kim, 2022]. This duality is distinct from the maximum likelihood approach which involves an optimal control problem over the original dynamics instead of the adjoint system [Bensoussan, 2018].

### 6.3.1 Duality in steady-state regime

Using the duality relationship (6.9), the MSE in prediction (6.5) is expressed as:

$$J_T^{\text{est}}(L) = \text{tr}[X_T(L)H^\top H] + \text{tr}[R],$$

where  $X_T(L) := A_L^T P_0 (A_L^\top)^T + \sum_{t=0}^{T-1} A_L^t (Q + LRL^\top) (A_L^\top)^t$ . Define the set of Schur stabilizing gains

$$\mathcal{S} := \{L \in \mathbb{R}^{n \times m} : \rho(A_L) < 1\}.$$

For any  $L \in \mathcal{S}$ , in the steady-state, the mean-squared prediction error assumes the form,

$$\lim_{T \rightarrow \infty} J_T^{\text{est}}(L) = \text{tr}[X_\infty(L)H^\top H] + \text{tr}[R],$$

where  $X_\infty(L) := \lim_{T \rightarrow \infty} X_T(L)$  and coincides with the unique solution  $X$  of the discrete Lyapunov equation  $X = A_L X A_L^\top + Q + L R L^\top$  (existence of unique solution follows from  $\rho(A_L) < 1$ ). Given the steady-state limit, we formally analyze the following constrained optimization problem:

$$\begin{aligned} \min_{L \in \mathcal{S}} \quad & \leftarrow J(L) := \text{tr}[X_{(L)}H^\top H], \\ \text{s.t.} \quad & X_{(L)} = A_L X_{(L)} A_L^\top + Q + L R L^\top. \end{aligned} \tag{6.10}$$

*Remark 6.5.* Note that the latter problem is technically the dual of the optimal LQR problem as formulated in [Bu et al., 2019a] by relating  $A \leftrightarrow A^\top$ ,  $-H \leftrightarrow B^\top$ ,  $L \leftrightarrow K^\top$ , and  $H^\top H \leftrightarrow \Sigma$ . However, the main difference here is that the product  $H^\top H$  may *not* be positive definite, for example, due to rank deficiency in  $H$  specially when  $m < n$ . Thus, in general, the cost function  $J(L)$  is not necessarily coercive in  $L$ , which can drastically effect the optimization landscape. For the same reason, in contrast to the LQR case [Fazel et al., 2018; Bu et al., 2019a], the gradient dominant property of  $J(L)$  is not clear in the filtering setup. In the next section, we show that such issues can be avoided as long as the pair  $(A, H)$  is observable. Also, the learning problem posed here is distinct from its LQR counterpart (see Table 6.1).

### 6.3.2 Optimization landscape

The first result is concerned with the behaviour of the objective function at the boundary of the optimization domain. It is known [Bu et al., 2019b] that the set of Schur stabilizing gains  $\mathcal{S}$  is regular open, contractible, and unbounded when  $m \geq 2$  and the boundary  $\partial\mathcal{S}$  coincides with the set  $\{L \in \mathbb{R}^{n \times m} : \rho(A - LH) = 1\}$ . For simplicity of presentation, we consider a slightly stronger assumption:

**Assumption 6.6.** The pair  $(A, H)$  is observable, and the noise covariances  $Q \succ 0$  and  $R \succ 0$ .

Table 6.1: Differences between SGD algorithms for optimal LQR and optimal estimation problems

Problem	Parameters			Constraints	Gradient Oracle	
	cost value	$Q$ and $R$	$A$ and $H$	stability $\mathcal{S}$	model	biased
LQR [Fazel et al., 2018]	known	known	unknown	yes	$\mathbb{E} [J(L + r\Delta)\Delta]$ $\Delta \sim U(\mathbb{S}^{mn})$	yes
Estimation (this work)	unknown	unknown	known	yes	$\mathbb{E} [\nabla\varepsilon(L, \mathcal{Y})]$ $\mathcal{Y} \sim \text{output data}$	yes
Vanila SGD	*	*	*	no	$\mathbb{E} [\nabla\varepsilon(L, \mathcal{Y})]$ $\mathcal{Y} \sim \text{data dist.}$	no

**Lemma 6.7.** *The function  $J(\cdot) : \mathcal{S} \rightarrow \mathbb{R}$  is real-analytic and coercive with compact sublevel sets; i.e.,*

$$L \rightarrow \partial\mathcal{S} \text{ or } \|L\| \rightarrow \infty \text{ each implies } J(L) \rightarrow \infty,$$

and  $\mathcal{S}_\alpha := \{L \in \mathbb{R}^{n \times m} : J(L) \leq \alpha\}$  is compact and contained in  $\mathcal{S}$  for any finite  $\alpha > 0$ .

The next result establishes the gradient dominance property of the objective function. While this result is known in the LQR setting ([Fazel et al., 2018; Bu et al., 2019a]), the extension to the estimation setup is not trivial as  $H^\top H$ , which takes the role of the covariance matrix of the initial state in LQR, may not be positive definite (instead, we only assume  $(A, H)$  is observable). This, apparently minor issue, hinders establishing the gradient dominated property globally. However, we recover this property on every sublevel sets of  $J$  which is sufficient for the subsequent convergence analysis.

**Lemma 6.8.** *Consider the constrained optimization problem (6.10). Then,*

- *The explicit formula for the gradient of  $J$  is:  $\nabla J(L) = 2Y_{(L)}(-LR + A_L X_{(L)} H^\top)$ ,*

where  $Y_{(L)} = Y$  is the unique solution of  $Y = A_L^\top Y A_L + H^\top H$ .

- The global minimizer  $L^* = \arg \min_{L \in \mathcal{S}} J(L)$  satisfies  $L^* = AX^*H^\top (R + HX^*H^\top)^{-1}$ , with  $X^*$  being the unique solution of  $X^* = A_{L^*}X^*A_{L^*}^\top + Q + L^*R(L^*)^\top$ .
- The function  $J(\cdot) : \mathcal{S}_\alpha \rightarrow \mathbb{R}$ , for any non-empty sublevel set  $\mathcal{S}_\alpha$  for some  $\alpha > 0$ , satisfies the following inequalities; for all  $L, L' \in \mathcal{S}_\alpha$ :

$$c_1[J(L) - J(L^*)] + c_2\|L - L^*\|_F^2 \leq \langle \nabla J(L), \nabla J(L) \rangle, \quad (6.11a)$$

$$c_3\|L - L^*\|_F^2 \leq J(L) - J(L^*), \quad (6.11b)$$

$$\|\nabla J(L) - \nabla J(L')\|_F \leq \ell \|L - L'\|_F, \quad (6.11c)$$

for positive constants  $c_1, c_2, c_3$  and  $\ell$  that are only a function of  $\alpha$  and independent of  $L$ .

Note that the expression for the gradient is consistent with Proposition 3.8 in [Bu et al., 2019a] after applying the duality relationship explained in Remark 6.5.

*Remark 6.9.* The proposition above implies that  $J(\cdot)$  has the Polyak-Łojasiewicz (PL) property (aka gradient dominance) on every  $\mathcal{S}_\alpha$ ; i.e., for any  $L \in \mathcal{S}_\alpha$  we have  $J(L) - J(L^*) \leq \frac{1}{c_1(\alpha)} \langle \nabla J(L), \nabla J(L) \rangle$ . The inequality (6.11a) is more general as it characterizes the dominance gap in terms of the iterate error from the optimality. This is useful in obtaining the iterate convergence results in the next section. Also, the Lipschitz bound resembles its “dual” counterpart in [Bu et al., 2019a, Lemma 7.9], however, it is *not* implied as a simple consequence of duality because  $H^\top H$  may not be positive definite.

#### 6.4 Algorithm: SGD for Learning the Kalman Gain

In order to emphasize on the estimation time horizon  $T$  for various measurement sequences, we use  $\mathcal{Y}_T := \{y(t)\}_{t=0}^T$  to denote the measurement time-span. Note that, any choice of  $L \in \mathcal{S}$  corresponds to a filtering strategy that outputs the following prediction,

$$\hat{y}_L(T) = HA_L^T m_0 + \sum_{t=0}^{T-1} HA_L^{T-t-1} Ly(t).$$

We denote the squared-norm of the estimation error for this filtering strategy as,

$$\varepsilon(L, \mathcal{Y}_T) := \|e_T(L)\|^2,$$

where  $e_T(L) := y(T) - \hat{y}_L(T)$ . We also define the *truncated* objective function as

$$J_T(L) := \mathbb{E} [\varepsilon(L, \mathcal{Y}_T)],$$

where the expectation is taken over all possible random measurement sequences, and note that, at the steady-state limit, we obtain  $\lim_{T \rightarrow \infty} J_T(L) = J(L)$ .

The SGD algorithm aims to solve this optimization problem by replacing the gradient, in the Gradient Descent (GD) update, with an unbiased estimate of the gradient in terms of samples from the measurement sequence. In particular, with access to an oracle that produces independent realization of the measurement sequence, say  $M$  random independent measurements sequences  $\{\mathcal{Y}_T^i\}_{i=1}^M$ , the gradient can be approximated as follows: denote the approximated cost value

$$\hat{J}_T(L) := \frac{1}{M} \sum_{i=1}^M \varepsilon(L, \mathcal{Y}_T^i),$$

then the approximate gradient with batch-size  $M$  is  $\nabla \hat{J}_T(L) = \frac{1}{M} \sum_{i=1}^M \nabla_L \varepsilon(L, \mathcal{Y}_T^i)$ . This forms an unbiased estimate of the gradient of the “truncated objective”, i.e.,  $\mathbb{E} [\nabla \hat{J}_T(L)] = \nabla J_T(L)$ . Next, for implementation purposes, we compute the gradient estimate explicitly in terms of the measurement sequence and the filtering policy  $L$ .

**Lemma 6.10.** *Given  $L \in \mathcal{S}$  and a sequence of measurements  $\mathcal{Y} = \{y(t)\}_{t=0}^T$ , we have,*

$$\begin{aligned} \nabla_L \varepsilon(L, \mathcal{Y}) &= -2H^\top e_T(L) y(T-1)^\top \\ &+ 2 \sum_{t=1}^{T-1} -(A_L^\top)^t H^\top e_T(L) y(T-t-1)^\top + \sum_{k=1}^t (A_L^\top)^{t-k} H^\top e_T(L) y(T-t-1)^\top L^\top (A_L^\top)^{k-1} H^\top. \end{aligned}$$

Finally, using this approximate gradient, the so-called SGD update proceeds as,

$$L_{k+1} = L_k - \eta_k \nabla_L \hat{J}_T(L),$$

for  $k \in \mathbb{Z}$ , where  $\eta_k > 0$  is the step-size.

*Remark 6.11.* Computing this approximate gradient only requires the knowledge of the system matrices  $A$  and  $H$ , and does *not* require the noise covariance information  $Q$  and  $R$ . Simulation results for the SGD algorithm are provided in the supplementary material.

Although the convergence of the SGD algorithm is expected to follow similar to the GD algorithm under the gradient dominance condition and Lipschitz property, the analysis becomes complicated due to the possibility of the iterated gain  $L_k$  leaving the sub-level sets. It is expected that a convergence guarantee would hold under high-probability due to concentration of the gradient estimate around the true gradient. The complete analysis in this direction is provided in the subsequent sections.

We first provide sample complexity and convergence guarantees for SGD with a biased estimation of gradient for locally Lipschitz objective functions and in presence of stability constraint  $\mathcal{S}$ . Subsequently, we study the stochastic problem of estimating the gradient for the estimation problem. Distinct features of our approach as compared with similar formulations in the literature are highlighted in Table 6.1.

#### 6.4.1 SGD with biased gradient and stability constraint

First, we characterize the “robustness” of a policy at which we aim to estimate the gradient. This is formalized in the following lemma which is a consequence of [Talebi and Mesbahi, 2022, Lemma IV.1].

**Lemma 6.12.** *Consider any  $L \in \mathcal{S}$  and let  $Z$  be the unique solution of  $Z = A_L Z A_L^\top + \Lambda$  for any  $\Lambda \succ 0$ . Then,  $L + \Delta \in \mathcal{S}$  for any  $\Delta \in \mathbb{R}^{n \times m}$  satisfying*

$$0 \leq \|\Delta\|_F \leq \underline{\lambda}(\Lambda) / [2 \bar{\lambda}(Z) \|H\|].$$

Second, we provide a uniform lowerbound for the stepsize of gradient descent for an approximated direction “close” to the true gradient direction.

**Lemma 6.13** (Uniform Lower Bound on Stepsize). *Let  $L_0 \in \mathcal{S}_\alpha$  for some  $\alpha \geq \alpha^* := J(L^*)$ , and choose any finite  $\beta \geq \alpha$ . Consider any direction  $E$  such that  $\|E - \nabla J(L_0)\|_F \leq$*

$\gamma\|\nabla J(L_0)\|_F$  for some  $\gamma \in [0, 1]$ , then we have  $J(L_0 - \eta E) \leq \beta$  for any  $\eta$  satisfying:

$$0 \leq \eta \leq \frac{1 - \gamma}{(\gamma + 1)^2} \cdot \frac{1}{\ell(\beta)} + \frac{c_3(\alpha)}{\ell(\alpha)[\alpha - \alpha^*]} \sqrt{\frac{\beta - \alpha}{2\ell(\beta)}}.$$

*Remark 6.14.* Note that for the case of exact gradient direction, i.e. when  $E = \nabla J(L_0)$ , we have  $\gamma = 0$  and choosing  $\beta = \alpha$  implies the known uniform bound of  $\eta \leq \frac{1}{\ell(\beta)}$  for feasible stepsizes as expected. Also, by this choice of  $\beta$ , this guarantees that the next iterate remains in sublevel set  $\mathcal{S}_\alpha$ . This lemma generalizes this uniform bound for general directions and (potentially) larger sublevel set.

The next lemma provides a decay guarantee for one iteration of gradient descent with an approximate direction which will be used later for convergence of SGD with a biased gradient estimate.

**Proposition 6.15** (Linear Decay in Cost Value). *Suppose  $L_0 \in \mathcal{S}_\alpha$  for some  $\alpha > 0$  and a direction  $E \neq 0$  is given such that  $\|E - \nabla J(L)\|_F \leq t\|\nabla J(L)\|_F$  for some  $\gamma < 1$ . Let  $\bar{\eta}_0 := (1-\gamma)/(\gamma+1)^2\ell(\alpha)$ . Then,  $L_1 := L_0 - \bar{\eta}_0 E$  remains in the same sublevel set, i.e.,  $L_1 \in \mathcal{S}_\alpha$ . Furthermore, we obtain the following linear decay of the cost value:*

$$J(L_1) - J(L^*) \leq [1 - c_1(\alpha)\bar{\eta}(1 - \gamma)/2] [J(L_0) - J(L^*)].$$

The next result is a direct consequence of this proposition. Roughly speaking, it guarantees that SGD algorithm with this biased estimation of gradient obtains a linear convergence rate outside a small set  $\mathcal{C}_\tau$  around optimality defined as

$$\mathcal{C}_\tau := \{L \in \mathcal{S} \mid \|\nabla J(L)\|_F \leq s_0/\tau\},$$

for some  $\tau \in (0, 1)$  and arbitrarily small  $s_0 > 0$ . First, we assume access to the following oracle that provides a biased estimation of the true gradient.

**Assumption 6.16.** Suppose, for some  $\alpha > 0$ , we have access to a biased estimation of the gradient  $\widehat{\nabla J}(L)$  such that, there exists constants  $s, s_0 > 0$  implying  $\|\widehat{\nabla J}(L) - \nabla J(L)\|_F \leq s\|\nabla J(L)\|_F + s_0$  for all  $L \in \mathcal{S}_\alpha \setminus \mathcal{C}_\tau$ .



**Theorem 6.17** (Convergence). *Suppose Assumption 6.16 holds with small enough  $s$  and  $s_0$  such that  $s \leq \epsilon\gamma$  and  $\mathcal{S}_\alpha \setminus \mathcal{C}_\epsilon$  is non-empty for some  $\epsilon, \gamma \in (0, 1)$ . Then, SGD algorithm starting from any  $L_0 \in \mathcal{S}_\alpha \setminus \mathcal{C}_{\gamma(1-\epsilon)}$  with fixed stepsize  $\bar{\eta} := \frac{(1-\gamma)}{(\gamma+1)^2\ell(\alpha)}$  generates a sequence of policies  $\{L_k\}$  that are stable (i.e. each  $L_k \in \mathcal{S}_\alpha$ ) and cost values decay linearly before entering  $\mathcal{C}_{\gamma(1-\epsilon)}$ ; i.e.,*

$$J(L_k) - J(L^*) \leq [1 - c_1(\alpha)\bar{\eta}(1-\gamma)/2]^k [J(L_0) - J(L^*)],$$

for each  $k \geq 0$  unless  $L_j \in \mathcal{C}_{\gamma(1-\epsilon)}$  for some  $j \leq k$ .

**Proof.** It suffices to note that for any  $L \in \mathcal{S}_\alpha \setminus \mathcal{C}_\epsilon$  it follows  $s\|\nabla J(L)\|_F + s_0 < \epsilon\gamma + (1-\epsilon)\gamma = \gamma$ . The claim then follows by applying Proposition 6.15 inductively.  $\square$

#### 6.4.2 The observation model of the estimation problem

Herein, first we show that the estimation error and its differential can be characterized as a “simple norm” of the concatenated noise (Proposition 6.19). This norm is induced by a metric that encapsulates the system dynamics which is explained below. Before proceeding to the results of this section, we assume that both the process and measurement noise are bounded:

**Assumption 6.18.** Assume that (almost surely)  $\|x_0\|, \|\xi(t)\| \leq \kappa_\xi$  and  $\|\omega(t)\| \leq \kappa_\omega$  for all  $t$ . Also, for simplicity, suppose the initial state has zero mean, i.e.,  $m_0 = 0_n$ .

For two vectors  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^{(T+1)n}$ , we define

$$\langle \mathbf{v}, \mathbf{w} \rangle_{\mathcal{A}_L} := \text{tr}[\mathbf{v}\mathbf{w}^\top \mathcal{A}_L^\top H^\top H \mathcal{A}_L],$$

where  $\mathcal{A}_L := \begin{pmatrix} A_L^0 & A_L^1 & \dots & A_L^T \end{pmatrix}$ . Also, define  $\mathcal{M}_L[E] := \begin{pmatrix} M_0[E] & M_1[E] & \dots & M_T[E] \end{pmatrix}$  with  $M_0[E] = 0$ ,  $M_1[E] = EH$  and  $M_{i+1}[E] = \sum_{k=0}^i A_L^{i-k} E H A_L^k$  for  $i = 1, \dots, T-2$ .

**Proposition 6.19.** *The estimation error  $\varepsilon(L, \mathcal{Y}_T)$  takes the following form*

$$\varepsilon(L, \mathcal{Y}_T) = \|\eta_L\|_{\mathcal{A}_L}^2,$$

where  $\eta_L := \xi - (I \otimes L)\omega$  with

$$\begin{aligned}\xi^\top &= \left( \xi(T-1)^\top \quad \dots \quad \xi(0)^\top \quad x(0)^\top \right), \\ \omega^\top &= \left( \omega(T-1)^\top \quad \dots \quad \omega(0)^\top \quad 0_m^\top \right).\end{aligned}$$

Furthermore, its differential acts on small enough  $E \in \mathbb{R}^{n \times m}$  as,

$$d\varepsilon(\cdot, \mathcal{Y}_T)|_L(E) = -2 \langle \eta_L, (I \otimes E)\omega \rangle_{\mathcal{A}_L} + \text{tr}[\mathcal{X}_L \mathcal{N}_L[E]],$$

where  $\mathcal{X}_L := \eta_L \eta_L^\top$  and  $\mathcal{N}_L[E] := \mathcal{M}_L[E]^\top H^\top H \mathcal{A}_L + \mathcal{A}_L^\top H^\top H \mathcal{M}_L[E]$ .

Now, we want to bound the error in the estimated gradient  $\nabla \widehat{J}_T(L)$  by considering the concentration error (on length  $T$  trajectories) and truncation error separately as follows:

$$\|\nabla \widehat{J}_T(L) - \nabla J(L)\| \leq \|\nabla \widehat{J}_T(L) - \nabla J_T(L)\| + \|\nabla J_T(L) - \nabla J(L)\|,$$

recalling that  $J_T(L) = \mathbb{E}[\varepsilon(L, \mathcal{Y}_T)]$  by definition.

Next, we aim to provide the analysis of concentration error on trajectories of length  $T$  with probability bounds. However, for any pair of real  $(T \times T)$ -matrices  $M$  and  $N$ , by Cauchy Schwartz inequality we obtain that  $|\text{tr}[MN]| \leq \|M\|_F \|N\|_F \leq \sqrt{T} \|M\| \|N\|_F$ . This bound becomes loose (in terms of dimension  $T$ ) as the condition number of  $N$  increases<sup>1</sup>.

Nonetheless, we are able to provide concentration error bounds that “scale well with respect to the length  $T$ ” which hinges upon the following idea: from von Neumann Trace Inequality [Horn and Johnson, 2012, Theorem 8.7.6] one obtains that

$$|\text{tr}[MN]| \leq \sum_{i=1}^T \sigma_i(M) \sigma_i(N) \leq \|M\| \|N\|_*, \quad (6.12)$$

where  $\|N\|_* := \text{tr}[\sqrt{N^\top N}] = \sum_i \sigma_i(N)$  is the *nuclear norm* with  $\sigma_i(N)$  denoting the  $i$ -th largest singular value of  $N$ . Additionally, the same inequality holds for non-square matrices of appropriate dimension which is tight in terms of dimension.

---

<sup>1</sup>The reason is that the first equality is sharp whenever  $M$  is in the “direction” of  $N$ , while the second inequality is sharp whenever condition number of  $M$  is close to one.

**Proposition 6.20** (Concentration independent of length  $T$ ). *Consider  $M$  independent length  $T$  trajectories  $\{\mathcal{Y}_T^i\}_{i=1}^M$  and suppose Assumption 6.18 holds. Let  $\nabla \widehat{J}_T(L) := \frac{1}{M} \sum_{i=1}^M \nabla \varepsilon(L, \mathcal{Y}_T^i)$ , then for any  $s > 0$ ,*

$$\mathbb{P} \left[ \|\nabla \widehat{J}_T(L) - \nabla J_T(L)\| \geq s \right] \leq 2n \exp \left[ \frac{-Ms^2/2}{\nu_L^2 + 2\nu_L s/3} \right],$$

where  $\nu_L := 4\kappa_L^2 C_L^3 \|H\|^2 \|H\|_* / [1 - \sqrt{\rho(A_L)}]^3$  with  $\kappa_L = \kappa_\xi + \|L\| \kappa_\omega$ .

We can also show how the truncation error decays linearly as  $T$  grows, with constants that are independent of the system dimension  $n$ :

**Proposition 6.21** (Truncation Error Bound). *Under Assumption 6.18, the truncation error is bounded:*

$$\|\nabla J(L) - \nabla J_T(L)\| \leq \bar{\gamma}_L \sqrt{\rho(A_L)}^{T+1},$$

where  $\bar{\gamma}_L := 10\kappa_L^4 C_L^6 \|H\|^2 \|H\|_* / [1 - \rho(A_L)]^2$ .

*Remark 6.22.* Notice how the trajectory length  $T$  determines the bias in the estimated gradient. However, the concentration error bound is independent of  $T$  and only depends on the noise bounds proportionate to  $\kappa_\xi^4$ ,  $\kappa_\omega^4$  and the stability margin of  $A_L$  proportionate to  $C_L^6 / (1 - \sqrt{\rho(A_L)})^6$ .

Finally, by combining the truncation bound in Proposition 6.21 with concentration bounds in Proposition 6.20 we can provide probabilistic bounds on the “estimated cost”  $\widehat{J}_T(L)$  and the “estimated gradient”  $\nabla \widehat{J}_T(L)$ . Its precise statement is deferred to the supplementary materials (Theorem 6.32).

### 6.4.3 Sample complexity of SGD for Kalman Gain

Note that the open-loop system may be unstable. Often in learning literature, it is assumed that the closed-loop system can be contractible ( i.e., the spectral norm  $\|A_L\| < 1$ ) which is quiet convenient for analysis, however, it is not a reasonable system theoretic assumption. Herein, we emphasize that we only require the close-loop system to be Schur stable,

meaning that  $\rho(A_L) < 1$ ; yet, it is very well possible that the system is not contractible. Handling systems that are merely stable requires more involved system theoretic tools that are established in the following lemma.

**Lemma 6.23** (Uniform Bounds for Stable Systems). *Suppose  $L \in \mathcal{S}$ , then there exist a constant  $C_L > 0$  such that*

$$\|A_L^k\| \leq C_L \sqrt{\rho(A_L)^{k+1}}, \quad \forall k \geq 0.$$

*Furthermore, consider  $\mathcal{S}_\alpha$  for some  $\alpha > 0$ , then there exist constants  $D_\alpha > 0$ ,  $C_\alpha > 0$  and  $\rho_\alpha \in (0, 1)$  such that  $\|L\| \leq D_\alpha$ ,  $C_L \leq C_\alpha$ , and  $\rho(A_L) \leq \rho_\alpha$  for all  $L \in \mathcal{S}_\alpha$ .*

The following result provides sample complexity bounds for this stochastic oracle to provide a biased estimation of the gradient that satisfies our oracle model of SGD analysis in Assumption 6.16.

**Theorem 6.24.** *Under the premise of Proposition 6.20, consider  $\mathcal{S}_\alpha$  for some  $\alpha > 0$  and choose any  $s, s_0 > 0$  and  $\tau \in (0, 1)$ . Suppose the trajectory length  $T \geq \ln \left( \bar{\gamma}_\alpha \sqrt{\min(n, m)} / s_0 \right) / \ln (1 / \sqrt{\rho_\alpha})$  and the batch size  $M \geq 4\nu_\alpha^2 \min(n, m) \ln(2n/\delta) / (s s_0)^2$ , where  $\bar{\gamma}_\alpha := 10(\kappa_\xi + D_\alpha \kappa_\omega)^4 C_\alpha^6 \|H\|^2 \|H\|_*$  and  $\nu_\alpha := 5C_\alpha^3 \|H\|^2 \|H\|_* (\kappa_\xi + D_\alpha \kappa_\omega)^2 / [1 - \sqrt{\rho_\alpha}]^3$ . Then, with probability no less than  $1 - \delta$ , Assumption 6.16 holds.*

**Proof.** First, note that for any  $L \in \mathcal{S}_\alpha$  by Lemma 6.23 we have that  $\bar{\gamma}_L \leq \bar{\gamma}_\alpha$  and  $\nu_L \leq \nu_\alpha$ . Then, note that the lower bound on  $T$  implies that  $\bar{\gamma}_\alpha \sqrt{\min(n, m)} \sqrt{\rho_\alpha}^{T+1} \leq s_0$ . The claim then follows by applying Theorem 6.32 and noting that for any  $L \notin \mathcal{C}_\tau$  the gradient is lowerbound as  $\|\nabla J(L)\| > s_0/\tau$ .  $\square$

*Remark 6.25.* By combining Theorem 6.17 and Theorem 6.24 we claim that in order to obtain  $\epsilon$  error on the cost value we need to run  $O(\ln(1/\epsilon))$  steps of SGD on data trajectories of length at least  $O(\ln(1/\epsilon))$ . And, this is achieved with probability at least  $\delta$  if the number of these trajectories are  $O(\ln(1/\delta)/\epsilon^2)$ .

## 6.5 Analysis and Extended Results

In this section we present the remaining analysis of the main results of this chapter and provide extended versions in some cases.

### 6.5.1 Proof of the duality relationship: Proposition 6.3

1. By pairing the original state dynamics (6.1) and its dual (6.6):

$$z(t+1)^\top x(t+1) - z(t)^\top x(t) = z(t+1)^\top \xi(t) + u(t+1)^\top Hx(t).$$

Summing this relationship from  $t = 0$  to  $t = T - 1$  yields,

$$z(T)^\top x(T) = z(0)^\top x(0) + \sum_{t=0}^{T-1} z(t+1)^\top \xi(t) + u(t+1)^\top Hx(t).$$

Upon subtracting the estimate  $a^\top \hat{x}_{\mathcal{L}}(T)$ , using the adjoint relationship

$$\sum_{t=0}^{T-1} u(t+1)^\top y(t) = a^\top \hat{x}_{\mathcal{L}}(T). \quad (6.13)$$

and  $z(T) = a$ , lead to

$$a^\top x(T) - a^\top \hat{x}_{\mathcal{L}}(T) = z(0)^\top x(0) + \sum_{t=0}^{T-1} z(t+1)^\top \xi(t) - u(t+1)^\top w(t).$$

Squaring both sides and taking the expectation concludes the duality result in (6.8).

2. Consider the adjoint system (6.6) with the linear feedback law  $u(t) = L^\top z(t)$ . Then,

$$z(t) = (A_L^\top)^{T-t} a, \quad \text{for } t = 0, 1, \dots, T. \quad (6.14)$$

Therefore, as a function of  $a$ ,  $u(t) = L^\top (A_L^\top)^{T-t} a$ . These relationships are used to identify the control policy

$$\mathcal{L}^\dagger(a) = (u(1), \dots, u(T)) = (L^\top (A_L^\top)^{T-1} a, \dots, L^\top a).$$

This control policy corresponds to an estimation policy by the adjoint relationship (6.13):

$$a^\top \hat{x}_{\mathcal{L}}(T) = \sum_{t=0}^{T-1} a^\top A_L^{T-t-1} Ly(t), \quad \forall a \in \mathbb{R}^n.$$

This relationship holds for all  $a \in \mathbb{R}^n$ . Therefore,

$$\hat{x}_{\mathcal{L}}(T) = \sum_{t=0}^{T-1} A_L^{T-t-1} L y(t),$$

which coincides with the Kalman filter estimate with constant gain  $L$  given by the formula (6.4) (with  $m_0 = 0$ ). Therefore, the adjoint relationship (6.13) relates the control policy with constant gain  $L^\top$  to the Kalman filter with constant gain  $L$ . The result (6.9) follows from the identity

$$\begin{aligned} \mathbb{E} [\|y(T) - \hat{y}_L(T)\|^2] &= \mathbb{E} [\|Hx(T) - H\hat{x}_L(T)\|^2 + \|w(T)\|^2] \\ &= \sum_{i=1}^m \mathbb{E} [|H_i^\top x(T) - H_i^\top \hat{x}_L(T)|^2] + \text{tr}[R], \end{aligned}$$

and the application of the first result (6.9) with  $a = H_i$ .

### 6.5.2 Proofs for the results for the analysis of the optimization landscape

#### Preliminary lemma

The following lemmas are a direct consequence of duality and useful for our subsequent analysis.

**Lemma 6.26.** *The set of Schur stabilizing gains  $\mathcal{S}$  is regular open, contractible, and unbounded when  $m \geq 2$  and the boundary  $\partial\mathcal{S}$  coincides with the set  $\{L \in \mathbb{R}^{n \times m} : \rho(A - LH) = 1\}$ . Furthermore,  $J(\cdot)$  is real analytic on  $\mathcal{S}$ .*

**Proof.** Consider the duality described in Remark 6.5. The proof then follows identical to [Bu et al., 2019a, Lemmas 3.5 and 3.6] by noting that the spectrum of a matrix is identical to the spectrum of its transpose.  $\square$

Next, we present the proof of Lemma 6.7 that provides sufficient conditions to recover the coercive property of  $J(\cdot)$  which resembles Lemma 3.7 in [Bu et al., 2019a] (but extended for the time-varying parameters). The proof is removed on the account of space and appears in the extended version of this work.

*Remark 6.27.* This approach recovers the claimed coercivity also in the control setting with weaker assumptions. In particular, using this result, one can replace the positive definite condition on the covariance of the initial condition in [Bu et al., 2019a], i.e.,  $\Sigma \succ 0$ , with just the controllability of  $(A, \Sigma^{1/2})$ .

*Proof of Lemma 6.7*

**Proof.** Consider any  $L \in \mathcal{S}$  and note that the right eigenvectors of  $A$  and  $A_L$  that are annihilated by  $H$  are identical. Thus, by PBH test, observability of  $(A, H)$  is equivalent to observability of  $(A_L, H)$ . Therefore, there exists a positive integer  $n_0 \leq n$  such that

$$H_{n_0}^\top(L) := \begin{pmatrix} H^\top & A_L^\top H^\top & \dots & (A_L^\top)^{n_0-1} H^\top \end{pmatrix}$$

is full-rank, implying that  $H_{n_0}^\top(L)H_{n_0}(L)$  is positive definite. Now, recall that for any such stabilizing gain  $L$ , we compute

$$\begin{aligned} J(L) &= \text{tr} \left[ \sum_{t=0}^{\infty} (A_L)^t (Q + LRL^\top) (A_L^\top)^t H^\top H \right] \\ &= \text{tr} \left[ \sum_{t=0}^{\infty} \sum_{k=0}^{n_0-1} (A_L)^{n_0t+k} (Q + LRL^\top) (A_L^\top)^{n_0t+k} H^\top H \right] \\ &= \text{tr} \left[ \sum_{t=0}^{\infty} (A_L)^{n_0t} (Q + LRL^\top) (A_L^\top)^{n_0t} H_{n_0}^\top(L) H_{n_0}(L) \right] \\ &=: \text{tr} [X_{n_0}(L) H_{n_0}^\top(L) H_{n_0}(L)], \end{aligned}$$

where we used the cyclic property of trace and the inequality follows because for any PSD matrices  $P_1, P_2 \succeq 0$  we have

$$\text{tr} [P_1 P_2] = \text{tr} \left[ P_2^{\frac{1}{2}} P_1 P_2^{\frac{1}{2}} \right] \geq 0. \quad (6.15)$$

Also,  $X_{n_0}(L)$  is well defined because  $A_L$  is Schur stable if and only if  $(A_L)^{n_0}$  is. Moreover,  $X_{n_0}(L)$  coincides with the unique solution to the following Lyapunov equation

$$X_{n_0}(L) = (A_L)^{n_0} X_{n_0}(L) (A_L^\top)^{n_0} + Q + LRL^\top.$$

Next, as  $R \succeq 0$ ,

$$\begin{aligned}
J(L) &\geq \underline{\lambda}(H_{n_0}^\top(L)H_{n_0}(L))\text{tr}[X_{n_0}(L)] \\
&\geq \underline{\lambda}(H_{n_0}^\top(L)H_{n_0}(L))\text{tr}\left[\sum_{t=0}^{\infty}(A_L)^{n_0t}Q(A_L^\top)^{n_0t}\right] \\
&\geq \underline{\lambda}(H_{n_0}^\top(L)H_{n_0}(L))\underline{\lambda}(Q)\sum_{t=0}^{\infty}\text{tr}\left[(A_L^\top)^{n_0t}(A_L)^{n_0t}\right] \\
&\geq \underline{\lambda}(H_{n_0}^\top(L)H_{n_0}(L))\underline{\lambda}(Q)\sum_{t=0}^{\infty}\rho(A_L)^{2n_0t}, \tag{6.16}
\end{aligned}$$

where the last inequality follows by the fact that

$$\text{tr}\left[(A_L^\top)^{n_0t}(A_L)^{n_0t}\right] = \|(A_L)^{n_0t}\|_F^2 \geq \|(A_L)^{n_0t}\|_{\text{op}}^2 \geq \rho((A_L)^{n_0t})^2 = \rho(A_L)^{2n_0t},$$

with  $\|\cdot\|_{\text{op}}$  denoting the operator norm induced by 2-norm. Now, by Lemma 6.26 and continuity of the spectral radius, as  $L_k \rightarrow \partial\mathcal{S}$  we observe that  $\rho(A_{L_k}) \rightarrow 1$ . But then, the obtained lowerbound implies that  $J(L_k) \rightarrow \infty$ . On the other hand, as  $Q \succ 0$ ,  $R \succ 0$  are both time-independent, by using a similar technique we also provide the following lowerbound

$$\begin{aligned}
J(L) &\geq \text{tr}\left[(Q + LRL^\top)\sum_{t=0}^{\infty}(A_L^\top)^{n_0t}H_{n_0}^\top(L)H_{n_0}(L)(A_L)^{n_0t}\right] \\
&\geq \underline{\lambda}(H_{n_0}^\top(L)H_{n_0}(L))\text{tr}[Q + LRL^\top] \\
&\geq \underline{\lambda}(H_{n_0}^\top(L)H_{n_0}(L))\text{tr}[RLL^\top] \\
&\geq \underline{\lambda}(H_{n_0}^\top(L)H_{n_0}(L))\underline{\lambda}(R)\|L\|_F^2,
\end{aligned}$$

where  $\|\cdot\|_F$  denotes the Frobenius norm. Therefore, by equivalency of norms on finite dimensional spaces,  $\|L_k\| \rightarrow \infty$  implies that  $J(L_k) \rightarrow \infty$  which concludes that  $J(\cdot)$  is coercive on  $\mathcal{S}$ . Finally, note that for any  $L \notin \mathcal{S}$ , by (6.16) we can argue that  $J(L) = \infty$ , therefore the sublevel sets  $\mathcal{S}_\alpha \subset \mathcal{S}$  whenever  $\alpha$  is finite. The compactness of  $\mathcal{S}_\alpha$  is then a direct consequence of the coercive property and continuity of  $J(\cdot)$  (Lemma 6.26).  $\square$



*Derivation of the gradient formula in Lemma 6.8*

Next, we aim to compute the gradient of  $J$  in a more general format. We do the derivation for time-varying  $R$  and  $Q$  with specialization to time-invariant setting at the end. For any admissible  $\Delta$ , we have

$$\begin{aligned} X_\infty(L + \Delta) - X_\infty(L) &= \sum_{t=1}^{\infty} (A_L)^t (Q_t + LR_t L^\top) (\star)^\top + (\star) (Q_t + LR_t L^\top) (A_L^\top)^t \\ &\quad - \sum_{t=0}^{\infty} (A_L)^t (\Delta R_t L^\top + LR_t \Delta^\top) (A_L^\top)^t + o(\|\Delta\|), \end{aligned}$$

where the  $\star$  is hiding the following term

$$\sum_{k=1}^t (A_L)^{t-k} \Delta H (A_L)^{k-1}.$$

Therefore, by linearity and cyclic permutation property of trace, we get that

$$\begin{aligned} J(L + \Delta) - J(L) &= \text{tr} \left[ \Delta H \sum_{t=1}^{\infty} \sum_{k=1}^t 2 (A_L)^{k-1} (Q_t + LR_t L^\top) (A_L^\top)^t H^\top H (A_L)^{t-k} \right] \\ &\quad - \text{tr} \left[ \Delta \sum_{t=0}^{\infty} 2 R_t L^\top (A_L^\top)^t H^\top H (A_L)^t \right] + o(\|\Delta\|). \end{aligned}$$

Finally, by considering the Euclidean metric on real matrices induced by the inner product  $\langle Q, P \rangle = \text{tr}[Q^\top P]$ , we obtain the gradient of  $J$  as follows

$$\nabla J(L) = -2 \sum_{t=0}^{\infty} (A_L^\top)^t H^\top H (A_L)^t L R_t - 2 \sum_{t=1}^{\infty} \sum_{k=1}^t (A_L^\top)^{t-k} H^\top H (A_L)^t (Q_t + LR_t L^\top) (A_L^\top)^{k-1} H^\top,$$

whenever the series are convergent! And, by switching the order of the sums it simplifies to

$$\begin{aligned} \nabla J(L) &= -2 \sum_{t=0}^{\infty} (A_L^\top)^t H^\top H (A_L)^t L R_t \\ &\quad + 2 \sum_{k=1}^{\infty} \sum_{t=k}^{\infty} \left[ (A_L^\top)^{t-k} H^\top H (A_L)^{t-k} \right] A_L \left[ (A_L)^{k-1} (Q_t + LR_t L^\top) (A_L^\top)^{k-1} \right] H^\top \\ &= -2 \sum_{t=0}^{\infty} (A_L^\top)^t H^\top H (A_L)^t L R_t + 2 \sum_{t=0}^{\infty} \left[ (A_L^\top)^t H^\top H (A_L)^t \right] \\ &\quad \cdot A_L \left[ \sum_{k=0}^{\infty} (A_L)^k (Q_{t+k+1} + LR_{t+k+1} L^\top) (A_L^\top)^k \right] H^\top. \end{aligned}$$

For the case of time-independent  $Q$  and  $R$ , this reduces to

$$\begin{aligned}\nabla J(L) &= -2Y_{(L)}LR + 2Y_{(L)}A_L \left[ \sum_{k=0}^{\infty} (A_L)^k (Q + LRL^\top) (A_L^\top)^k \right] H^\top \\ &= 2Y_{(L)} [-LR + A_L X_{(L)} H^\top].\end{aligned}$$

where  $Y_{(L)} = Y$  is the unique solution of

$$Y = A_L^\top Y A_L + H^\top H.$$

*Proof of existence of global minimizer in Lemma 6.8*

The domain  $\mathcal{S}$  is non-empty whenever  $(A, H)$  is observable. Thus, by continuity of  $L \rightarrow J(L)$ , there exists some finite  $\alpha > 0$  such that the sublevel set  $\mathcal{S}_\alpha$  is non-empty and compact. Therefore, the minimizer is an interior point and thus must satisfy the first-order optimality condition  $\nabla J(L^*) = 0$ . Therefore, by coercivity, it is stabilizing and unique which satisfies

$$L^* = AX^*H^\top (R + HX^*H^\top)^{-1},$$

with  $X^*$  being the unique solution of

$$X^* = A_{L^*} X^* A_{L^*}^\top + Q + L^* R (L^*)^\top. \quad (6.17)$$

As expected, the global minimizer  $L^*$  is equal to the steady-state Kalman gain, but explicitly dependent on the noise covariances  $Q$  and  $R$ .

*Proof of gradient dominance property in Lemma 6.8*

Note that  $X = X_{(L)}$  satisfies

$$X = A_L X A_L^\top + Q + L R L^\top. \quad (6.18)$$

Then, by combining (6.17) and (6.18), and some algebraic manipulation, we recover part of the gradient information, i.e.,  $(-LR + A_L X H^\top)$ , in the gap of cost matrices by arriving at

the following identity

$$\begin{aligned}
& X - X^* - A_{L^*}(X - X^*)A_{L^*}^\top \\
&= (LR - A_L X H^\top)(L - L^*)^\top + (L - L^*)(RL^\top - H X A_L^\top) \\
&\quad - (L - L^*)R(L - L^*)^\top - (L - L^*)H X H^\top(L - L^*)^\top \\
&\preceq \frac{1}{a}(LR - A_L X H^\top)(RL^\top - H X A_L^\top) + a(L - L^*)(L - L^*)^\top \\
&\quad - (L - L^*)(R + H X H^\top)(L - L^*)^\top
\end{aligned} \tag{6.19}$$

where the upperbound is valid for any choice of  $a > 0$ . Now, as  $R \succ 0$ , we choose  $a = \underline{\lambda}(R)/2$ .

As  $X \succeq 0$ , it further upperbounds

$$\begin{aligned}
X - X^* - A_{L^*}(X - X^*)A_{L^*}^\top &\preceq \frac{2}{\underline{\lambda}(R)}(-LR + A_L X H^\top)(-RL^\top + H X A_L^\top) \\
&\quad - \frac{\underline{\lambda}(R)}{2}(L - L^*)(L - L^*)^\top.
\end{aligned}$$

Now, let  $\tilde{X}$  and  $\hat{X}$  be, respectively, the unique solution of the following Lyapunov equations

$$\begin{aligned}
\tilde{X} &= A_{L^*} \tilde{X} A_{L^*}^\top + (-LR + A_L X H^\top)(-RL^\top + H X A_L^\top), \\
\hat{X} &= A_{L^*} \hat{X} A_{L^*}^\top + (L - L^*)(L - L^*)^\top.
\end{aligned}$$

Then by comparison, we conclude that

$$X - X^* \preceq \frac{2}{\underline{\lambda}(R)} \tilde{X} - \frac{\underline{\lambda}(R)}{2} \hat{X}.$$

Recall that by the fact in (6.15),

$$J(L) - J(L^*) = \text{tr}[(X - X^*)H^\top H] \leq \frac{2}{\underline{\lambda}(R)} \text{tr}[\tilde{X} H^\top H] - \frac{\underline{\lambda}(R)}{2} \text{tr}[\hat{X} H^\top H]. \tag{6.20}$$

Let  $Y^* \succ 0$  be the unique solution of

$$Y^* = A_{L^*}^\top Y^* A_{L^*} + H^\top H,$$

then, by cyclic permutation property

$$\begin{aligned}
\operatorname{tr} \left[ \tilde{X} H^\top H \right] &= \operatorname{tr} [(-LR + A_L X H^\top)(-RL^\top + H X A_L^\top) Y^*] \\
&\leq \frac{\bar{\lambda}(Y^*)}{\underline{\lambda}^2(Y_{(L)})} \operatorname{tr} [Y_{(L)}(-LR + A_L X H^\top)(-RL^\top + H X A_L^\top) Y_{(L)}] \\
&= \frac{\bar{\lambda}(Y^*)}{4 \underline{\lambda}^2(Y_{(L)})} \langle \nabla J(L), \nabla J(L) \rangle
\end{aligned} \tag{6.21}$$

where the last equality follows by the obtained formula for the gradient  $\nabla J(L)$ . Similarly, we obtain that

$$\operatorname{tr} \left[ \hat{X} H^\top H \right] = \operatorname{tr} [(L - L^*)(L - L^*)^\top Y^*] \geq \underline{\lambda}(Y^*) \|L - L^*\|_F^2. \tag{6.22}$$

Notice that the mapping  $L \rightarrow Y_{(L)}$  is continuous on  $\mathcal{S} \supset \mathcal{S}_\alpha$ , and also by observability of  $(A, H)$ ,  $Y_{(L)} \succ 0$  for any  $L \in \mathcal{S}$ . To see this, let  $H_{n_0}(L) \succ 0$  be as defined in Lemma 6.7.

Then,

$$\begin{aligned}
Y_{(L)} &= \sum_{t=0}^{\infty} (A_L^\top)^t (H^\top H) (A_L)^t \\
&= \sum_{t=0}^{\infty} \sum_{k=0}^{n_0-1} (A_L^\top)^{n_0 t+k} (H^\top H) (A_L)^{n_0 t+k} \\
&= \sum_{t=0}^{\infty} (A_L^\top)^{n_0 t} H_{n_0}^\top(L) H_{n_0}(L) (A_L)^{n_0 t} \\
&\succeq H_{n_0}^\top(L) H_{n_0}(L) \succ 0.
\end{aligned}$$

Now, by Lemma 6.7,  $\mathcal{S}_\alpha$  is compact and therefore we claim that the following infimum is attained with some positive value  $\kappa_\alpha$ :

$$\inf_{L \in \mathcal{S}_\alpha} \underline{\lambda}(Y_{(L)}) =: \kappa_\alpha > 0. \tag{6.23}$$

Finally, the first claimed inequality follows by combining the inequalities (6.20), (6.21) and (6.22), with the following choice of

$$c_1(\alpha) := \frac{2 \underline{\lambda}(R)}{\bar{\lambda}(Y^*)} \kappa_\alpha^2, \quad \text{and} \quad c_2(\alpha) := \frac{\underline{\lambda}(Y^*) \underline{\lambda}(R)^2}{\bar{\lambda}(Y^*)} \kappa_\alpha^2.$$

For the second claimed inequality, one arrive at the following identity by similar computation to (6.19):

$$\begin{aligned} X - X^* - A_L(X - X^*)A_L^\top &= (L^*R - A_{L^*}X^*H^\top)(L - L^*)^\top + (L - L^*)(RL^{*\top} - HX^*A_{L^*}^\top) \\ &\quad + (L - L^*)R(L - L^*)^\top + (L - L^*)HX^*H^\top(L - L^*)^\top \\ &= (L - L^*)(R + HX^*H^\top)(L - L^*)^\top \end{aligned}$$

where the second equality follows because  $Y_{(L)} \succ 0$  and thus

$$L^*R - A_{L^*}X^*H^\top = -Y_{(L)}^{-1}\nabla J(L^*) = 0.$$

Recall that

$$J(L) - J(L^*) = \text{tr}[(X - X^*)H^\top H],$$

then by the equality in (6.19) and cyclic property of trace we obtain

$$J(L) - J(L^*) = \text{tr}[ZY_{(L)}],$$

where

$$\begin{aligned} Z &:= (L - L^*)(R + HX^*H^\top)(L - L^*)^\top \\ &\succeq \underline{\lambda}(R)(L - L^*)(L - L^*)^\top. \end{aligned}$$

Therefore, for any  $L \in \mathcal{S}_\alpha$ , we have

$$J(L) - J(L^*) \geq \underline{\lambda}(Y_{(L)})\text{tr}[Z] \geq \underline{\lambda}(R)\kappa_\alpha\|L - L^*\|_F^2,$$

and thus, we complete the proof of first part by the following choice of

$$c_3(\alpha) = \underline{\lambda}(R)\kappa_\alpha.$$

*Proof of Lipschitz property in Lemma 6.8*

Next, we provide the proof of locally Lipschitz property. Notice that the mappings  $L \rightarrow X_{(L)}$ ,  $L \rightarrow Y_{(L)}$  and  $L \rightarrow A_L$  are all real-analytic on the open set  $\mathcal{S} \supset \mathcal{S}_\alpha$ , and thus so is the mapping

$L \rightarrow \nabla J(L) = 2Y_{(L)} [-LR + A_L X_{(L)} H^\top]$ . Also, by Lemma 6.7,  $\mathcal{S}_\alpha$  is compact and therefore the mapping  $L \rightarrow \nabla J(L)$  is  $\ell$ -Lipschitz continuous on  $\mathcal{S}_\alpha$  for some  $\ell = \ell(\alpha) > 0$ . In the rest of the proof, we attempt to characterize  $\ell(\alpha)$  in terms of the problem parameters. By direct computation we obtain

$$\begin{aligned} \nabla J(L_1) - \nabla J(L_2) &= (2Y_{(L_1)} - 2Y_{(L_2)}) [-L_1 R + A_{L_1} X_{(L_1)} H^\top] \\ &\quad + 2Y_{(L_2)} ([-L_1 R + A_{L_1} X_{(L_1)} H^\top] - [-L_2 R + A_{L_2} X_{(L_2)} H^\top]) \\ &= 2(Y_{(L_1)} - Y_{(L_2)}) [-L_1(R + HX_{(L_1)}H^\top) + AX_{(L_1)}H^\top] \\ &\quad + 2Y_{(L_2)} [(L_2 - L_1)(R + HX_{(L_1)}H^\top) + A_{L_2}(X_{(L_1)} - X_{(L_2)})H^\top]. \end{aligned}$$

Therefore,

$$\|\nabla J(L_1) - \nabla J(L_2)\|_F^2 \leq \ell_1^2 \|Y_{(L_1)} - Y_{(L_2)}\|_F^2 + \ell_2^2 \|L_1 - L_2\|_F^2 + \ell_3^2 \|X_{(L_1)} - X_{(L_2)}\|_F^2 \quad (6.24)$$

where

$$\begin{aligned} \ell_1 &= \ell_1(L_1) := 2\| -L_1(R + HX_{(L_1)}H^\top) + AX_{(L_1)}H^\top \|_{\text{op}}, \\ \ell_2 &= \ell_2(L_1, L_2) := 2\|Y_{(L_2)}\|_{\text{op}} \|R + HX_{(L_1)}H^\top\|_{\text{op}}, \\ \ell_3 &= \ell_3(L_2) := 2\|Y_{(L_2)}\|_{\text{op}} \|A_{L_2}\|_{\text{op}} \|H^\top\|_{\text{op}}. \end{aligned}$$

On the other hand, by direct computation we obtain

$$\begin{aligned} Y_{(L_1)} - Y_{(L_2)} - A_{L_1}^\top (Y_{(L_1)} - Y_{(L_2)}) A_{L_1} &= (L_2 - L_1)^\top H^\top Y_{(L_2)} A_{L_2} + A_{L_2}^\top Y_{(L_2)} H (L_2 - L_1) \\ &\quad + (L_1 - L_2)^\top H^\top Y_{(L_2)} H (L_1 - L_2) \\ &\leq \|L_1 - L_2\|_F \ell_4 I \end{aligned} \quad (6.25)$$

where

$$\ell_4 = \ell_4(L_1, L_2) := 2\|H^\top Y_{(L_2)} A_{L_2}\|_{\text{op}} + \|H^\top Y_{(L_2)} H (L_1 - L_2)\|_{\text{op}}.$$

Now, consider the mapping  $L \rightarrow Z_{(L)}$  where  $Z_{(L)} = Z$  is the unique solution of the following Lyapunov equation:

$$Z = A_L^\top Z A_L + I,$$

which is well-defined and continuous on  $\mathcal{S} \supset \mathcal{S}_\alpha$ . Therefore, by comparison, we claim that

$$\|Y_{(L_1)} - Y_{(L_2)}\|_F \preceq \|L_1 - L_2\|_F \ell_4 \|Z_{(L_1)}\|_F.$$

By a similar computation to that of (6.19), we obtain that

$$\begin{aligned} & X_{(L_1)} - X_{(L_2)} - A_{L_2}(X_{(L_1)} - X_{(L_2)})A_{L_2}^\top \\ &= (L_1R - A_{L_1}X_{(L_1)}H^\top)(L_1 - L_2)^\top + (L_1 - L_2)(RL_1^\top - HX_{(L_1)}A_{L_1}^\top) \\ &\quad - (L_1 - L_2)R(L_1 - L_2)^\top - (L_1 - L_2)HX_{(L_1)}H^\top(L_1 - L_2)^\top \\ &\preceq \|L_1 - L_2\|_F \ell_5 (Q + L_2RL_2^\top) \end{aligned} \tag{6.26}$$

where

$$\ell_5 = \ell_5(L_1) := 2\| -L_1R + A_{L_1}X_{(L_1)}H^\top \|_{\text{op}} / \underline{\lambda}(Q).$$

Therefore, by comparison, we claim that

$$\|X_{(L_1)} - X_{(L_2)}\|_F \preceq \|L_1 - L_2\|_F \ell_5 \|X_{(L_2)}\|_F.$$

Finally, by compactness of  $\mathcal{S}_\alpha$ , we claim that the following supremums are attained and thus, are achieved with some *finite* positive values:

$$\begin{aligned} \bar{\ell}_1(\alpha) &:= \sup_{L_1, L_2 \in \mathcal{S}_\alpha} \ell_1(L_1) \ell_4(L_1, L_2) \|Z_{(L_1)}\|_F, \\ \bar{\ell}_2(\alpha) &:= \sup_{L_1, L_2 \in \mathcal{S}_\alpha} \ell_2(L_1, L_2), \\ \bar{\ell}_3(\alpha) &:= \sup_{L_1, L_2 \in \mathcal{S}_\alpha} \ell_3(L_2) \ell_5(L_1) \|X_{(L_2)}\|_F. \end{aligned}$$

Then, the claim follows by combining the bound in (6.24) with (6.25) and (6.26), and the following choice of

$$\ell(\alpha) := \sqrt{\bar{\ell}_1^2(\alpha) + \bar{\ell}_2^2(\alpha) + \bar{\ell}_3^2(\alpha)}.$$

### 6.5.3 First order methods for optimal filtering

Herein, we show how the optimization landscape allows for finding globally optimum filtering policy by carefully adopting a gradient descent algorithm despite non-convexity of both the

cost function and the feasible domain. This intuition serves as a baseline in analyzing the SGD algorithm where a gradient oracle is inexact and also biased.

### *Gradient Flow (GF)*

For completeness, in the next two sections, we analyze first-order methods in order to solve the minimization problem (6.10). In this section, we consider a policy update according to the the GF dynamics:

$$[\text{GF}] \quad \dot{L}_s = -\nabla J(L_s).$$

We summarize the convergence result in the following Proposition which is a direct consequence of Lemma 6.8.

**Proposition 6.28.** *Consider any sublevel set  $\mathcal{S}_\alpha$  for some  $\alpha > 0$ . Then, for any initial policy  $L_0 \in \mathcal{S}_\alpha$ , the GF updates converges to optimality at a linear rate of  $\exp(-c_1(\alpha))$  (in both the function value and the policy iterate). In particular, we have*

$$J(L_s) - J(L^*) \leq (\alpha - J(L^*)) \exp(-c_1(\alpha)s),$$

and

$$\|L_s - L^*\|_F^2 \leq \frac{\alpha - J(L^*)}{c_3(\alpha)} \exp(-c_1(\alpha)s).$$

**Proof.** Consider a Lyapunov candidate function  $V(L) := J(L) - J(L^*)$ . Under the GF dynamics

$$\dot{V}(L_s) = -\langle \nabla J(L_s), \nabla J(L_s) \rangle \leq 0.$$

Therefore,  $L_s \in \mathcal{S}_\alpha$  for all  $s > 0$ . But then, by Lemma 6.8, we can also show that

$$\dot{V}(L_s) \leq -c_1(\alpha)V(L_s) - c_2(\alpha)\|L_s - L^*\|_F^2, \quad \text{for } s > 0.$$

By recalling that  $c_1(\alpha) > 0$  is a positive constant independent of  $L$ , we conclude the following exponential stability of the GF:

$$V(L_s) \leq V(L_0) \exp(-c_1(\alpha)s),$$



for any  $L_0 \in \mathcal{S}_\alpha$  which, in turn, guarantees convergence of  $J(L_s) \rightarrow J(L^*)$  at the linear rate of  $\exp(-c_1(\alpha)s)$ . Finally, the linear convergence of the policy iterates follows directly from the second bound in Lemma 6.8:

$$\|L_s - L^*\|_F^2 \leq \frac{1}{c_3(\alpha)} V(L_s) \leq \frac{V(L_0)}{c_3(\alpha)} \exp(-c_1(\alpha)s).$$

The proof concludes by noting that  $V(L_0) \leq \alpha - J(L^*)$  for any such initial value  $L_0 \in \mathcal{S}_\alpha$ .  $\square$

### *Gradient Descent (GD)*

Here, we consider the GD policy update:

$$[\text{GD}] \quad L_{k+1} = L_k - \eta_k \nabla J(L_k),$$

for  $k \in \mathbb{Z}$  and a positive stepsize  $\eta_k$ . Given the convergence result for the GF, establishing convergence for GD relies on carefully choosing the stepsize  $\eta_k$ , and bounding the rate of change of  $\nabla J(L)$ —at least on each sublevel set. This is achieved by the Lipschitz bound for  $\nabla J(L)$  on any sublevel set.

In what follows, we establish linear convergence of the GD update. Our convergence result only depends on the value of  $\alpha$  for the initial sublevel set  $\mathcal{S}_\alpha$  that contains  $L_0$ . Note that our proof technique is distinct from those in [Bu et al., 2019a] and [Mohammadi et al., 2021b]; nonetheless, it involves a similar argument using the gradient dominance property of  $J$ .

**Theorem 6.29.** *Consider any sublevel set  $\mathcal{S}_\alpha$  for some  $\alpha > 0$ . Then, for any initial policy  $L_0 \in \mathcal{S}_\alpha$ , the GD updates with any fixed stepsize  $\eta_k = \eta \in (0, 1/\ell(\alpha)]$  converges to optimality at a linear rate of  $1 - \eta c_1(\alpha)/2$  (in both the function value and the policy iterate). In particular, we have*

$$J(L_k) - J(L^*) \leq [\alpha - J(L^*)](1 - \eta c_1(\alpha)/2)^k,$$

and

$$\|L_k - L^*\|_F^2 \leq \left[ \frac{\alpha - J(L^*)}{c_3(\alpha)} \right] (1 - \eta c_1(\alpha)/2)^k,$$

with  $c_1(\alpha)$  and  $c_3(\alpha)$  as defined in Lemma 6.8.

**Proof.** First, we argue that the GD update with such a step size does not leave the initial sublevel set  $\mathcal{S}_\alpha$  for any initial  $L_0 \in \mathcal{S}_\alpha$ . In this direction, consider  $L(\eta) = L_0 - \eta \nabla J(L_0)$  for  $\eta \geq 0$  where  $L_0 \neq L^*$ . Then, by compactness of  $\mathcal{S}_\alpha$  and continuity of the mapping  $\eta \rightarrow J(L(\eta))$  on  $\mathcal{S} \supset \mathcal{S}_\alpha$ , the following supremum is attained with a positive value  $\bar{\eta}_0$ :

$$\bar{\eta}_0 := \sup\{\eta : J(L(\zeta)) \leq \alpha, \forall \zeta \in [0, \eta]\},$$

where positivity of  $\bar{\eta}_0$  is a direct consequence of the strict decay of  $J(L(\eta))$  for sufficiently small  $\eta$  as  $\nabla J(L_0) \neq 0$ . This implies that  $L(\eta) \in \mathcal{S}_\alpha \subset \mathcal{S}$  for all  $\eta \in [0, \bar{\eta}_0]$  and  $J(L(\bar{\eta}_0)) = \alpha$ . Next, by the Fundamental Theorem of Calculus and smoothness of  $J(\cdot)$  (Lemma 6.26), for any  $\eta \in [0, \bar{\eta}_0]$  we have that,

$$\begin{aligned} J(L(\eta)) - J(L_0) - \langle \nabla J(L_0), L(\eta) - L_0 \rangle &= \int_0^1 \langle \nabla J(L(\eta s)) - \nabla J(L_0), L(\eta) - L_0 \rangle ds \\ &\leq \|L(\eta) - L_0\|_F \int_0^1 \|\nabla J(L(\eta s)) - \nabla J(L_0)\|_F ds \\ &\leq \ell(\alpha) \|L(\eta) - L_0\|_F \int_0^1 \|L(\eta s) - L_0\|_F ds \\ &= \frac{1}{2} \ell(\alpha) \eta \|L(\eta) - L_0\|_F \|\nabla J(L_0)\|_F, \end{aligned}$$

where  $\|\cdot\|_F$  denotes the Frobenius norm, the first inequality is a consequence of Cauchy-Schwartz, and the second one is due to (6.11c) and the fact that  $L(\eta s)$  remains in  $\mathcal{S}_\alpha$  for all  $s \in [0, 1]$ .<sup>2</sup> By the definition of  $L(\eta)$ , it now follows that,

$$J(L(\eta)) - J(L_0) \leq \eta \|\nabla J(L_0)\|_F^2 \left( \frac{\ell(\alpha)\eta}{2} - 1 \right). \quad (6.27)$$

This implies  $J(L(\eta)) \leq J(L_0)$  for all  $\eta \leq 2/\ell(\alpha)$ , and thus concluding that  $\bar{\eta}_0 \geq 2/\ell(\alpha)$ . This justifies that  $L(\eta) \in \mathcal{S}_\alpha$  for all  $\eta \in [0, 2/\ell(\alpha)]$ . Next, if we consider the GD update with

---

<sup>2</sup>Note that a direct application of Descent Lemma [Beck, 2017, Lemma 5.7] may not be justified as one has to argue about the uniform bound for the Hessian of  $J$  over the non-convex set  $\mathcal{S}_\alpha$  where  $J$  is  $\ell(\alpha)$ -Lipschitz only on  $\mathcal{S}_\alpha$ . Also see the proof of [Mohammadi et al., 2021b, Theorem 2].

any fixed stepsize  $\eta \in (0, 1/\ell(\alpha)]$  and apply the bound in (6.27) and the gradient dominance property in Lemma 6.8, we obtain

$$J(L_1) - J(L_0) \leq \eta c_1 \left( \frac{\ell(\alpha)\eta}{2} - 1 \right) [J(L_0) - J(L^*)],$$

which by subtracting  $J(L^*)$  results in

$$J(L_1) - J(L^*) \leq \left( 1 - \frac{\eta c_1}{2} \right) [J(L_0) - J(L^*)],$$

as  $\eta c_1(\ell(\alpha)\eta/2 - 1) \leq -\eta c_1/2$  for all  $\eta \in (0, 1/\ell(\alpha)]$ . By induction, and the fact that both  $c_1(\alpha)$  and the choice of  $\eta$  only depends on the value of  $\alpha$ , we conclude the convergence in the function value at a linear rate of  $1 - (\eta c_1/2)$  and the constant coefficient of  $\alpha - J(L^*) \geq J(L_0) - J(L^*)$ . To complete the proof, the linear convergence of the policy iterates follows directly from the second bound in Lemma 6.8.  $\square$

#### 6.5.4 Proofs for the analysis of the constrained SGD algorithm

*lemma 6.10: Derivation of stochastic gradient formula*

**Proof.** For small enough  $\Delta \in \mathbb{R}^{n \times m}$ ,

$$\begin{aligned} \varepsilon(L + \Delta, \mathcal{Y}) - \varepsilon(L, \mathcal{Y}) &= \|e_T(L + \Delta)\|^2 - \|e_T(L)\|^2 \\ &= 2\text{tr} [(e_T(L + \Delta) - e_T(L))e_T^\top(L)] + o(\|\Delta\|). \end{aligned}$$

The difference

$$e_T(L + \Delta) - e_T(L) = E_1(\Delta) + E_2(\Delta) + o(\|\Delta\|),$$

with the following terms that are linear in  $\Delta$ :

$$\begin{aligned} E_1(\Delta) &:= -\sum_{t=0}^{T-1} H(A_L)^t \Delta y(T-t-1), \\ E_2(\Delta) &:= \sum_{t=1}^{T-1} \sum_{k=1}^t H(A_L)^{t-k} \Delta H(A_L)^{k-1} L y(T-t-1). \end{aligned}$$

Therefore, combining the two identities, the definition of gradient under the inner product  $\langle A, B \rangle := \text{tr}[AB^\top]$ , and ignoring the higher order terms in  $\Delta$  yields,

$$\langle \nabla_L \varepsilon(L, y), \Delta \rangle = 2\text{tr}[(E_1(\Delta) + E_2(\Delta))e_T^\top(L)],$$

which by linearity and cyclic permutation property of trace reduces to:

$$\begin{aligned} \langle \nabla_L \varepsilon(L, y), \Delta \rangle = & -2\text{tr} \left[ \Delta \left( \sum_{t=0}^{T-1} y(T-t-1)e_T^\top(L)H(A_L)^t \right) \right] \\ & + 2\text{tr} \left[ \Delta \left( \sum_{t=1}^{T-1} \sum_{k=1}^t H(A_L)^{k-1}Ly(T-t-1)e_T^\top(L)H(A_L)^{t-k} \right) \right]. \end{aligned}$$

This holds for all admissible  $\Delta$ , concluding the formula for the gradient.  $\square$

*Proof of Lemma 6.12: Robustness of the policy with respect to perturbation*

**Proof.** Recall the stability certificate  $s_K$  as proposed in [Talebi and Mesbahi, 2022, Lemma IV.1] for a choice of constant mapping  $\mathcal{Q} : K \rightarrow \Lambda \succ 0$  and dual problem parameters as discussed in Remark 6.5. Then, we arrive at  $s_K = \underline{\lambda}(\Lambda)/(2\bar{\lambda}(Z)\|H^\top\Delta^\top\|)$ , for which  $\rho(A^\top + H^\top(L^\top + \eta\Delta^\top)) < 1$  for any  $\eta \in [0, s_K]$ . But, the spectrum of a square matrix and its transpose are identical, thus  $L + \eta\Delta \in \mathcal{S}$  for any such  $\eta$ . The claim then follows by noting that the operator norm of a matrix and its transpose are identical, and the resulting lowerbound as follows  $s_K \geq \underline{\lambda}(\Lambda)/(2\bar{\lambda}(Z)\|H\|\|\Delta\|_F)$ .  $\square$

*6.5.5 Proof of Lemma 6.13: Uniform lower-bound on stepsize*

**Proof.** Without loss of generality, suppose  $L_0 \neq L^*$  and  $E \neq 0$ , and let  $L(\eta) := L_0 - \eta E$ . By compactness of  $\mathcal{S}_\beta$  and continuity of the mapping  $\eta \rightarrow J(L(\eta))$  on  $\mathcal{S} \supset \mathcal{S}_\beta$ , the following supremum is attained by  $\eta_\beta$ :

$$\eta_\beta := \sup\{\eta : J(L(\zeta)) \leq \beta, \forall \zeta \in [0, \eta]\}. \quad (6.28)$$

Note that  $\eta_\beta$  is strictly positive for  $\beta > \alpha$  because  $L_0 \in \mathcal{S}_\alpha \subset \mathcal{S}_\beta$  and  $J(\cdot)$  is coercive (Lemma 6.7) and its domain is open (Lemma 6.12). This implies that  $L(\eta) \in \mathcal{S}_\beta \subset \mathcal{S}$  for all  $\eta \in [0, \eta_\beta]$  and  $J(L(\eta_\beta)) = \beta$ .

Next, we want to show that  $\eta_\beta$  is uniformly lower bounded with high probability. By the Fundamental Theorem of Calculus, for any  $\eta \in [0, \eta_\beta]$  we have

$$\begin{aligned}
J(L(\eta)) - J(L_0) - \langle \nabla J(L_0), L(\eta) - L_0 \rangle &= \int_0^1 \langle \nabla J(L(\eta s)) - \nabla J(L_0), L(\eta) - L_0 \rangle ds \\
&\leq \|L(\eta) - L_0\|_F \int_0^1 \|\nabla J(L(\eta s)) - \nabla J(L_0)\|_F ds \\
&\leq \ell(\beta) \|L(\eta) - L_0\|_F \int_0^1 \|L(\eta s) - L_0\|_F ds \\
&= \frac{1}{2} \ell(\beta) \eta^2 \|E\|_F^2, \tag{6.29}
\end{aligned}$$

where the first inequality is a consequence of Cauchy-Schwartz, and the second one is due to (6.11c) and the fact that  $L(\eta s)$  remains in  $\mathcal{S}_\beta$  for all  $s \in [0, 1]$  by definition of  $\eta_\beta$ . Note that the assumption implies  $\|E\|_F \leq (\gamma + 1) \|\nabla J(L_0)\|_F$ . Thus, (6.29) implies that

$$\begin{aligned}
J(L(\eta)) &\leq J(L_0) - \eta \langle \nabla J(L_0), E \rangle + \frac{1}{2} \ell(\beta) \eta^2 \|E\|_F^2 \\
&\leq J(L_0) - \eta \langle \nabla J(L_0), E - \nabla J(L_0) \rangle - \eta \|\nabla J(L_0)\|_F^2 + \frac{1}{2} \ell(\beta) \eta^2 \|E\|_F^2 \tag{6.30} \\
&\leq J(L_0) + \|\nabla J(L_0)\|_F^2 \left[ \frac{1}{2} (\gamma + 1)^2 \ell(\beta) \eta^2 + (\gamma - 1) \eta \right].
\end{aligned}$$

Therefore, for  $\eta$  to be a feasible point in the supremum in (6.28), it suffices to satisfy:

$$\frac{1}{2} \|\nabla J(L_0)\|_F^2 [(\gamma + 1)^2 \ell(\beta) \eta^2 + 2(\gamma - 1) \eta] \leq \beta - \alpha,$$

or equivalently,

$$\left[ (\gamma + 1) \ell(\beta) \eta + \frac{\gamma - 1}{\gamma + 1} \right]^2 \leq \left( \frac{\gamma - 1}{\gamma + 1} \right)^2 + \frac{2\ell(\beta)[\beta - \alpha]}{\|\nabla J(L_0)\|_F^2}.$$

But then it suffices to have

$$(\gamma + 1) \ell(\beta) \eta + \frac{\gamma - 1}{\gamma + 1} \leq \frac{\sqrt{2\ell(\beta)[\beta - \alpha]}}{\|\nabla J(L_0)\|_F}.$$

Finally, note that by Lemma 6.8 and (6.11c) we have

$$\|\nabla J(L_0)\|_F \leq \frac{\ell(\alpha)}{c_3(\alpha)} [J(L_0) - J(L^*)] \leq \frac{\ell(\alpha)}{c_3(\alpha)} [\alpha - \alpha^*]. \tag{6.31}$$

Using this uniform bound of gradient on sublevel set  $\mathcal{S}_\alpha$  and noting that  $\gamma \in [0, 1]$ , we can obtain the sufficient condition for  $\eta$  to be feasible. This completes the proof.  $\square$

### 6.5.6 Proof of Proposition 6.15: Linear decay in cost value

**Proof.** Suppose  $L_0 \neq L^*$  and let  $L(\eta) := L_0 - \eta E$ . Note that  $E$  may not be necessarily in the direction of decay in  $J(L)$ , however, we can argue the following:

Choose  $\beta = \alpha$  in Lemma 6.13 and note that  $\eta_\beta$  as defined in (6.28) will be lower bounded as  $\eta_\beta \geq \bar{\eta}_0$ , and thus  $\bar{\eta}_0$  is feasible. Recall that  $L(\eta) \in \mathcal{S}_\beta \subset \mathcal{S}$  for all  $\eta \in [0, \eta_\beta]$ . Also, for any  $\eta \in [0, \bar{\eta}_0]$  and  $\gamma \in [0, 1)$ , from (6.30) we obtain that:

$$\begin{aligned} J(L(\eta)) - J(L_0) &\leq \|\nabla J(L_0)\|_F^2 \left[ \frac{1}{2}(\gamma + 1)^2 \ell(\alpha) \eta^2 + (\gamma - 1)\eta \right] \\ &\leq c_1(\alpha) [J(L_0) - J(L^*)] \left[ \frac{1}{2}(\gamma + 1)^2 \ell(\alpha) \eta^2 + (\gamma - 1)\eta \right] \end{aligned}$$

where, as  $\gamma < 1$ , the last inequality follows by (6.11a) for any  $\eta \leq \min\{2\bar{\eta}_0, \eta_\beta\}$ . By the choice of  $\bar{\eta}_0$ , then we obtain that

$$J(L(\bar{\eta}_0)) - J(L_0) \leq -c_1(\alpha) \left[ \frac{(\gamma - 1)^2}{2(\gamma + 1)^2 \ell(\alpha)} \right] [J(L_0) - J(L^*)]$$

This implies that

$$J(L(\bar{\eta}_0)) - J(L^*) \leq \left( 1 - c_1(\alpha) \left[ \frac{(\gamma - 1)^2}{2(\gamma + 1)^2 \ell(\alpha)} \right] \right) [J(L_0) - J(L^*)].$$

□

### 6.5.7 Proofs of the result for observation model and sample complexity

#### Preliminary lemmas and their proofs

First, we provide the proof for the complete version of Lemma 6.23:

*Lemma 6.23'* (Uniform Bounds for Stable Systems). Suppose  $L \in \mathcal{S}$ , then there exists a constant  $C_L > 0$  such that

$$\|A_L^k\| \leq C_L \left( \sqrt{\rho(A_L)} \right)^{k+1}, \quad \forall k \geq 0,$$

whenever  $\rho(A_L) > 0$ , and otherwise  $\sqrt{\rho(A_L)}$  is replaced with any arbitrarily small  $r \in (0, 1)$ . Additionally,

$$\begin{aligned} \sum_{i=0}^{\infty} \|A_L^i\| &\leq \frac{C_L}{1 - \sqrt{\rho(A_L)}} \\ \sum_{i=0}^{\infty} \|M_i[E]\| &\leq \frac{1 + 2C_L^2 \rho(A_L)^{3/2}}{[1 - \sqrt{\rho(A_L)}]^2} \|EH\| \end{aligned}$$

Furthermore, consider  $\mathcal{S}_\alpha$  for some  $\alpha > 0$ , then there exist constants  $D_\alpha > 0$ ,  $C_\alpha > 0$  and  $\rho_\alpha \in (0, 1)$  such that  $\|L\| \leq D_\alpha$ ,  $C_L \leq C_\alpha$  and  $\rho(A_L) \leq \rho_\alpha$ ,  $\forall L \in \mathcal{S}_\alpha$ .

**Proof of Lemma 6.23.** Recall the Cauchy Integral formula for matrix functions [Higham, 2008, Theorem 1.12]: for any matrix  $M \in \mathbb{C}^{n \times n}$ ,

$$f(M) = \frac{1}{2\pi i} \oint_{\Gamma} f(z)(zI - M)^{-1} dz,$$

whenever  $f$  is real analytic on and inside a closed contour  $\Gamma$  that encloses spectrum of  $M$ . Note that  $L \in \mathcal{S}$  implying that  $\rho(A_L) < 1$ . Now, fix some  $r \in (\rho(A_L), 1)$  and define  $\Gamma(\theta) = re^{i\theta}$  with  $\theta$  ranging on  $[0, 2\pi]$ . Therefore, Cauchy Integral formula applies to  $f(z) = z^k$  for any positive integer  $k$  and the contour  $\Gamma$  defined above. So, for matrix  $A_L$ , we obtain

$$A_L^k = \frac{1}{2\pi i} \oint_{\Gamma} z^k (zI - A_L)^{-1} dz = \frac{1}{2\pi i} \int_0^{2\pi} r^k e^{ik\theta} (re^{i\theta} I - A)^{-1} r de^{i\theta},$$

implying that

$$\|A_L^k\| \leq \frac{r^{k+1}}{2\pi} \int_0^{2\pi} \|(re^{i\theta} I - A_L)^{-1}\| d\theta \leq r^{k+1} \max_{\theta \in [0, 2\pi]} \|(re^{i\theta} I - A_L)^{-1}\|.$$

Finally, the first claim follows by choosing  $r = \sqrt{\rho(A_L)}$  (whenever  $\rho(A_L) > 0$ , otherwise  $r \in (0, 1)$  can be chosen arbitrarily small) and defining

$$C_L := \max_{\theta \in [0, 2\pi]} \|(\sqrt{\rho(A_L)} e^{i\theta} I - A_L)^{-1}\|$$

which is attained and bounded.

Next, by applying the first claim, we have

$$\sum_{i=0}^T \|A_L^i\| \leq C_L \frac{1 - \sqrt{\rho(A_L)}^T}{1 - \sqrt{\rho(A_L)}},$$

implying the second bound. For the third claim, note that for  $i = 1, 2, \dots$  we obtain

$$\begin{aligned} \|M_{i+1}[E]\| &\leq \sum_{k=0}^i \|A_L^{i-k}\| \|A_L^k\| \|EH\| \leq \|EH\| C_L^2 \sum_{k=0}^i \left[ \sqrt{\rho(A_L)} \right]^{(i-k+1)+(k+1)} \\ &\leq \|EH\| C_L^2 \sqrt{\rho(A_L)} \left[ (i+1) \cdot \rho(A_L)^{(i+1)/2} \right]. \end{aligned} \quad (6.32)$$

But, then by recalling that  $M_0[E] = 0$  and  $\|M_1[E]\| = \|EH\|$  we have

$$\sum_{i=0}^{\infty} \|M_i[E]\| \leq \|EH\| + \|EH\| \left[ \frac{2C_L^2 \rho(A_L)^{3/2}}{[1 - \sqrt{\rho(A_L)}]^2} \right]$$

where we used the following convergent sum for any  $\rho \in (0, 1)$ :

$$\sum_{i=1}^{\infty} (i+1) \cdot \rho^{i+1} = \frac{(2-\rho)\rho^2}{(1-\rho)^2} \leq \frac{2\rho^2}{(1-\rho)^2}.$$

This implies the third bound.

The final claim follows directly from compactness of sublevel set  $\mathcal{S}_\alpha$  (Lemma 6.7) and continuity of the mappings  $(L, \theta) \mapsto (\rho(A_L), \theta) \mapsto \|(\sqrt{\rho(A_L)}e^{i\theta}I - A_L)^{-1}\|$  on  $\mathcal{S}_\alpha \times [0, 2\pi]$  whenever  $\rho(A_L) > 0$  (and otherwise considering the mapping  $(L, \theta) \mapsto \|(re^{i\theta}I - A_L)^{-1}\|$  for arbitrarily small and fixed  $r \in (0, 1)$ ).  $\square$

As mentioned in Section 6.4.2, a key idea behind these error bounds that scale well with respect to the length  $T$  is the following consequence of von Neumann Trace Inequality [Horn and Johnson, 2012, Theorem 8.7.6]:

$$|\text{tr}[MN]| \leq \sum_{i=1}^T \sigma_i(M)\sigma_i(N) \leq \|M\| \|N\|_*,$$

with  $\|\cdot\|_*$  denoting the nuclear norm. Additionally, as a direct consequence of Courant-Fischer Theorem, one can also show that nuclear norm is sub-multiplicative. More precisely,

$$\|AB\|_* \leq \|A\| \|B\|_* \leq \|A\|_* \|B\|_*.$$

Next, we require the following lemma to bound these errors.



**Lemma 6.30.** For any  $L \in \mathcal{S}_\alpha$ , we have

$$\begin{aligned} \|\mathcal{A}_L^\top H^\top H \mathcal{A}_L\|_* &\leq \frac{C_L^2 \|H^\top H\|_*}{1 - \rho(A_L)}, \\ \|\mathcal{N}_L[E]\|_* &\leq \frac{[2C_L + 4C_L^3 \rho(A_L)^{3/2}] \|H\| \|H^\top H\|_*}{[1 - \rho(A_L)]^2} \|E\|. \end{aligned}$$

*Proof of Lemma 6.30.* For the first claim, note that  $\mathcal{A}_L^\top H^\top H \mathcal{A}_L$  is positive semi-definite, so

$$\begin{aligned} \|\mathcal{A}_L^\top H^\top H \mathcal{A}_L\|_* &= \text{tr}[\mathcal{A}_L^\top H^\top H \mathcal{A}_L] \\ &\leq \text{tr}[H^\top H] \|\mathcal{A}_L \mathcal{A}_L^\top\| \\ &\leq \|H^\top H\|_* \left\| \sum_{i=0}^T A_L^i (A_L^\top)^i \right\| \\ &\leq \|H^\top H\|_* \sum_{i=0}^T \|A_L^i\|^2 \\ &\leq \|H^\top H\|_* \frac{C_L^2}{1 - \rho(A_L)} \end{aligned}$$

where the last inequality follows by Lemma 6.23. Next, we have

$$\begin{aligned} \|\mathcal{M}_L[E]\| &= \|\mathcal{M}_L[E] \mathcal{M}_L[E]^\top\|^{1/2} \\ &\leq \left[ \sum_{i=0}^T \|M_i[E]\|^2 \right]^{1/2} \\ &\leq \|EH\| + \|EH\| C_L^2 \sqrt{\rho(A_L)} \left[ \sum_{i=0}^T (i+1)^2 \cdot \rho(A_L)^{(i+1)} \right]^{1/2} \\ &\leq \|EH\| + \|EH\| C_L^2 \sqrt{\rho(A_L)} \frac{2\rho(A_L)}{[1 - \rho(A_L)]^{3/2}} \\ &\leq \|EH\| \left[ \frac{1 + 2C_L^2 \rho(A_L)^{3/2}}{[1 - \rho(A_L)]^{3/2}} \right] \end{aligned}$$

where the second inequality follows by (6.32) and the third one by the following convergent sum for any  $\rho \in (0, 1)$ :

$$\sum_{i=1}^{\infty} (i+1)^2 \cdot \rho^{i+1} = \frac{\rho^2(\rho^2 - 3\rho + 4)}{(1 - \rho)^3} \leq \frac{4\rho^2}{(1 - \rho)^3}.$$

Also, by the properties of nuclear norm

$$\begin{aligned}
\|H^\top H \mathcal{A}_L\|_* &= \text{tr} \left[ \sqrt{H^\top H \mathcal{A}_L \mathcal{A}_L^\top H^\top H} \right] \\
&\leq \|\mathcal{A}_L \mathcal{A}_L^\top\|^{1/2} \|H^\top H\|_* \\
&\leq \left[ \sum_{i=0}^{\infty} \|A_L^i\|^2 \right]^{1/2} \|H^\top H\|_* \\
&\leq \left[ \frac{C_L^2}{1 - \rho(A_L)} \right]^{1/2} \|H^\top H\|_*,
\end{aligned}$$

where the last inequality follows by Lemma 6.23. Finally, notice that

$$\begin{aligned}
\|\mathcal{N}_L[E]\|_* &\leq 2\|\mathcal{A}_L^\top H^\top H \mathcal{M}_L[E]\|_* \\
&\leq 2\|\mathcal{M}_L[E]\| \|\|H^\top H \mathcal{A}_L\|_*
\end{aligned}$$

and thus combining the last three bounds implies the second claim. This completes the proof.  $\square$

The next tool we will be using is the following famous bound on random matrices which is a variant of Bernstein inequality:

**Lemma 6.31** (Matrix Bernstein Inequality [Tropp, 2015, Corollary 6.2.1]). *Let  $Z$  be a  $d_1 \times d_2$  random matrices such that  $\mathbb{E}[Z] = \bar{Z}$  and  $\|Z\| \leq K$  almost surely. Consider  $M$  independent copy of  $Z$  as  $Z_1, \dots, Z_M$ , then for every  $t \geq 0$ , we have*

$$\mathbb{P} \left[ \left\| \frac{1}{M} \sum_i Z_i - \bar{Z} \right\| \geq t \right] \leq (d_1 + d_2) \exp \left\{ \frac{-Mt^2/2}{\sigma^2 + 2Kt/3} \right\}$$

where  $\sigma^2 = \max\{\|\mathbb{E}[ZZ^\top]\|, \|\mathbb{E}[Z^\top Z]\|\}$  is the per-sample second moment. This bound can be expressed as the mixture of sub-gaussian and sub-exponential tail as  $(d_1 + d_2) \exp \left\{ -c \min \left\{ \frac{t^2}{\sigma^2}, \frac{t}{2K} \right\} \right\}$  for some  $c$ .

We are now well-equipped to provide the main proofs.

*Proof of Proposition 6.19*

**Proof.** Recall that

$$\varepsilon(L, \mathcal{Y}_T) = \|Hx(T) - H\hat{x}(T)\|^2 = \sum_{i=1}^m |H_i^\top x(T) - H_i^\top \hat{x}(T)|^2$$

where  $H_i^\top$  is the  $i$ -th row of  $H$ . Also, by duality, if  $z(t) = (A_L^\top)^{T-t} H_i$  is the adjoint dynamics' closed-loop trajectory with control signal  $u(t) = L^\top z(t)$  then

$$H_i^\top x(T) - H_i^\top \hat{x}(T) = \mathbf{z}_i^\top \xi - \mathbf{u}^\top \omega$$

where

$$\begin{aligned} \mathbf{z}_i^\top &= \left( z(T)^\top \quad z(T-1)^\top \quad \dots \quad z(1)^\top \quad z(0)^\top \right), \\ \mathbf{u}^\top &= \left( u(T)^\top \quad u(T-1)^\top \quad \dots \quad u(1)^\top \quad 0_m^\top \right). \end{aligned}$$

But  $\mathbf{u} = (I \otimes L^\top) \mathbf{z}_i$  and then  $\mathbf{z}_i = \mathcal{A}_L^\top H_i$ . Therefore,

$$\begin{aligned} \varepsilon(L, \mathcal{Y}_T) &= \sum_{i=1}^m |H_i^\top x(T) - H_i^\top \hat{x}(T)|^2 \\ &= \sum_{i=1}^m \text{tr} [\xi \xi^\top \mathbf{z}_i \mathbf{z}_i^\top] - \text{tr} [(\xi \omega^\top (I \otimes L^\top) + (I \otimes L) \omega \xi^\top) \mathbf{z}_i \mathbf{z}_i^\top] \\ &\quad + \text{tr} [\omega \omega^\top (I \otimes L^\top) \mathbf{z}_i \mathbf{z}_i^\top (I \otimes L)] \end{aligned}$$

Then, by using the fact that  $\sum_{i=1}^m H_i H_i^\top = H^\top H$ , we obtain that

$$\varepsilon(L, \mathcal{Y}_T) = \text{tr} [\mathcal{X}_L \mathcal{A}_L^\top H^\top H \mathcal{A}_L].$$

Thus, we can rewrite the estimation error as

$$\varepsilon(L, \mathcal{Y}_T) = \langle \xi, \xi \rangle_{\mathcal{A}_L} + \langle (I \otimes L) \omega, (I \otimes L) \omega \rangle_{\mathcal{A}_L} - 2 \langle \xi, (I \otimes L) \omega \rangle_{\mathcal{A}_L} = \|\eta_L\|_{\mathcal{A}_L}^2$$

Next, we can compute that for small enough  $E$

$$\mathcal{A}_{L+E} - \mathcal{A}_L = \mathcal{M}_L[E] + o(\|E\|).$$

This implies that

$$d(\mathcal{A}_L^\top H^\top H \mathcal{A}_L)|_L[E] = \mathcal{M}_L[E]^\top H^\top H \mathcal{A}_L + \mathcal{A}_L^\top H^\top H \mathcal{M}_L[E]$$

On the other hand,

$$\begin{aligned} \mathcal{X}_{L+E} - \mathcal{X}_L &= (I \otimes E)\omega\omega^\top(I \otimes L^\top) + (I \otimes L)\omega\omega^\top(I \otimes E^\top) - \xi\omega^\top(I \otimes E^\top) + (I \otimes E)\omega\xi^\top \\ &\quad + o(\|E\|). \end{aligned}$$

Therefore, the second claim follows by the chain rule.  $\square$

*Proof of the Proposition 6.20*

We provide the proof for a detailed version of Proposition 6.20:

*Proposition 6.20'* (Concentration independent of length  $T$ ). Consider length  $T$  trajectories  $\{\mathcal{Y}_{[t_0, t_0+T]}^i\}_{i=1}^M$  and let  $\widehat{J}_T(L) := \frac{1}{M} \sum_{i=1}^M \varepsilon(L, \mathcal{Y}_T^i)$ . Then, under Assumption 6.18, for any  $s > 0$

$$\mathbb{P} \left[ |\widehat{J}_T(L) - J_T(L)| \leq s \right] \geq 1 - 2n \exp \left[ \frac{-Ms^2/2}{\mu_L^2 + 2\mu_L s/3} \right],$$

$$\mathbb{P} \left[ \|\nabla \widehat{J}_T(L) - \nabla J_T(L)\| \leq s \right] \geq 1 - 2n \exp \left[ \frac{-Ms^2/2}{\nu_L^2 + 2\nu_L s/3} \right]$$

where  $\kappa_L = \kappa_\xi + \|L\|\kappa_\omega$  and

$$\begin{aligned} \mu_L &:= \frac{\kappa_L^2 C_L^2}{[1 - \sqrt{\rho(A_L)}]^2} \|H^\top H\|_* \\ \nu_L &:= \frac{2\kappa_L \kappa_\omega C_L^2 + [C_L + 2C_L^3 \rho(A_L)^{3/2}] \|H\| \kappa_L^2}{[1 - \sqrt{\rho(A_L)}]^3} \|H^\top H\|_*. \end{aligned}$$

***Proof of Proposition 6.20.*** Note that

$$\begin{aligned} \mathbb{E} [\xi \xi^\top] &= \begin{pmatrix} I \otimes Q & 0 \\ 0 & P_0 \end{pmatrix} =: \mathcal{Q}, \\ \mathbb{E} [\omega \omega^\top] &= \begin{pmatrix} I \otimes R & 0 \\ 0 & 0_m \end{pmatrix} =: \mathcal{R}, \end{aligned}$$

and  $\mathbb{E}[\xi \omega^\top] = 0$ . Assume  $m_0 = 0$  and recall that  $\langle \nabla \varepsilon(L, \mathcal{Y}_T), E \rangle = d\varepsilon(\cdot, \mathcal{Y}_T)|_L(E)$  thus, using Proposition 6.19, we can rewrite the  $J_T(L)$  and its gradient as

$$J_T(L) = \mathbb{E}[\varepsilon(L, \mathcal{Y}_T)] = \mathbb{E}[\|\eta_L\|_{\mathcal{A}_L}^2] = \text{tr}[\mathbb{E}[\mathcal{X}_L] \mathcal{A}_L^\top H^\top H \mathcal{A}_L]$$

where  $\mathbb{E}[\mathcal{X}_L] = \mathcal{Q} + (I \otimes L)\mathcal{R}(I \otimes L^\top)$ . Therefore, by definition of  $\widehat{J}_T(L)$  we obtain

$$\widehat{J}_T(L) = \text{tr}[\mathcal{Z}_L \mathcal{A}_L^\top H^\top H \mathcal{A}_L]$$

with  $\mathcal{Z}_L = \frac{1}{M} \sum_{i=1}^M \mathcal{X}_L(\mathcal{Y}^i)$  which can be expanded as

$$\mathcal{Z}_L = \mathcal{Z}_1 + (I \otimes L)\mathcal{Z}_2(I \otimes L^\top) - \mathcal{Z}_3(I \otimes L^\top) - (I \otimes L)\mathcal{Z}_3^\top,$$

where

$$\mathcal{Z}_1 = \frac{1}{M} \sum_{i=1}^M \xi_i \xi_i^\top, \quad \mathcal{Z}_2 = \frac{1}{M} \sum_{i=1}^M \omega_i \omega_i^\top, \quad \mathcal{Z}_3 = \frac{1}{M} \sum_{i=1}^M \xi_i \omega_i^\top,$$

and  $\mathbb{E}[\mathcal{Z}_L] = \mathcal{Q} + (I \otimes L)\mathcal{R}(I \otimes L^\top)$ . Therefore,

$$\widehat{J}_T(L) - J_T(L) = \text{tr}[(\mathcal{Z}_L - \mathbb{E}[\mathcal{Z}_L]) \mathcal{A}_L^\top H^\top H \mathcal{A}_L].$$

Thus, by cyclic permutation property of trace and (6.12) we obtain

$$|\widehat{J}_T(L) - J_T(L)| \leq \|\mathcal{A}_L (\mathcal{Z}_L - \mathbb{E}[\mathcal{Z}_L]) \mathcal{A}_L^\top\| \|H^\top H\|_*. \quad (6.33)$$

Next, we consider the symmetric random matrix  $\mathcal{A}_L (\mathcal{Z}_L - \mathbb{E}[\mathcal{Z}_L]) \mathcal{A}_L^\top$ . Note that  $\|\xi(t) - L\omega(t)\| \leq \kappa_L$  almost surely and thus

$$\|\mathcal{A}_L \mathcal{X}_L \mathcal{A}_L^\top\| = \|\mathcal{A}_L \eta_L\|^2 \leq \kappa_L^2 \left[ \sum_{i=0}^{\infty} \|A_L^i\| \right]^2 \leq \mu_L / \|H^\top H\|_*.$$

It then follows that

$$\|\mathbb{E}[(\mathcal{A}_L \mathcal{X}_L \mathcal{A}_L^\top)^2]\| \leq \mathbb{E}[\|\mathcal{A}_L \mathcal{X}_L \mathcal{A}_L^\top\|^2] \leq \mu_L^2 / \|H^\top H\|_*^2.$$

Therefore, by Lemma 6.31 we obtain that

$$\mathbb{P}[\|\mathcal{A}_L (\mathcal{Z}_L - \mathbb{E}[\mathcal{Z}_L]) \mathcal{A}_L^\top\| \geq t] \leq 2n \exp \left[ \frac{-M \|H^\top H\|_*^2 t^2 / 2}{\mu_L^2 + 2\mu_L \|H^\top H\|_* t / 3} \right].$$

Substituting  $t$  with  $t/\|H^\top H\|_*$  together with (6.33) implies the first claim.

Similarly, we can compute that

$$\begin{aligned} \langle \nabla J_T(L), E \rangle &= \langle \mathbb{E} [\nabla \varepsilon(L, \mathcal{Y}_T)], E \rangle \\ &= 2\text{tr} [(I \otimes L)\mathcal{R}(I \otimes E^\top)\mathcal{A}_L^\top H^\top H \mathcal{A}_L] + \text{tr} [(\mathcal{Q} + (I \otimes L)\mathcal{R}(I \otimes L^\top))\mathcal{N}_L[E]], \end{aligned}$$

and thus

$$\begin{aligned} \langle \nabla \widehat{J}_T(L) - \nabla J_T(L), E \rangle &= -2\text{tr} [\mathcal{Z}_3(I \otimes E^\top)\mathcal{A}_L^\top H^\top H \mathcal{A}_L] \\ &\quad + 2\text{tr} [(I \otimes L)(\mathcal{Z}_2 - \mathcal{R})(I \otimes E^\top)\mathcal{A}_L^\top H^\top H \mathcal{A}_L] \\ &\quad + \text{tr} [(\mathcal{Z}_L - \mathbb{E}[\mathcal{Z}_L])\mathcal{N}_L[E]]. \end{aligned}$$

Thus, by cyclic permutation property of trace and (6.12) we obtain that

$$|\langle \nabla \widehat{J}_T(L) - \nabla J_T(L), E \rangle| \leq \left\| \frac{1}{M} \sum_{i=1}^M S_L(E, \mathcal{Y}_T^i) - \mathbb{E} [S_L(E, \mathcal{Y}_T^i)] \right\| \|H^\top H\|_* \quad (6.34)$$

where  $S_L(E, \mathcal{Y})$  is the symmetric part of the following random matrix

$$-2\mathcal{A}_L \xi \omega^\top (I \otimes E^\top) \mathcal{A}_L^\top + 2\mathcal{A}_L (I \otimes L) \omega \omega^\top (I \otimes E^\top) \mathcal{A}_L^\top + 2\mathcal{A}_L \mathcal{X}_L \mathcal{M}_L[E]^\top.$$

Next, we provide the following almost sure bounds for each term: first,

$$\begin{aligned} \|\mathcal{A}_L \mathcal{Z}_3(I \otimes E^\top) \mathcal{A}_L^\top\| &\leq \|\mathcal{A}_L \xi\| \|\mathcal{A}_L (I \otimes E) \omega\| \\ &\leq \kappa_\xi \left[ \sum_{i=0}^T \|A_L^i\| \right] \kappa_\omega \|E\| \left[ \sum_{i=0}^T \|A_L^i\| \right] \\ &\leq \kappa_\xi \kappa_\omega \frac{C_L^2}{[1 - \sqrt{\rho(A_L)}]^2} \|E\| \end{aligned}$$

where the last equality follows by Lemma 6.23; second, similarly

$$\begin{aligned} \|\mathcal{A}_L (I \otimes L) \omega \omega^\top (I \otimes E^\top) \mathcal{A}_L^\top\| &\leq \|\mathcal{A}_L (I \otimes L) \omega\| \|\mathcal{A}_L (I \otimes E) \omega\| \\ &\leq \kappa_\omega^2 \|L\| \|E\| \left[ \sum_{i=0}^T \|A_L^i\| \right]^2 \\ &\leq \kappa_\omega^2 \frac{C_L^2}{[1 - \sqrt{\rho(A_L)}]^2} \|L\| \|E\|; \end{aligned}$$

and finally

$$\begin{aligned}
\|\mathcal{A}_L \mathcal{X}_L \mathcal{M}_L [E]^\top\| &\leq \|\mathcal{A}_L \eta_L\| \|\mathcal{M}_L [E] \eta_L\| \\
&\leq \kappa_L^2 \left[ \sum_{i=0}^T \|A_L^i\| \right] \left[ \sum_{i=0}^T \|M_i [E]\| \right] \\
&\leq \kappa_L^2 \left[ \frac{C_L + 2C_L^3 \rho(A_L)^{3/2}}{[1 - \sqrt{\rho(A_L)}]^3} \right] \|EH\|
\end{aligned}$$

where the last inequality follows by Lemma 6.23. Now, by combining the last three bounds we can claim that almost surely

$$\|S_L(E, \mathcal{Y})\| \leq \frac{\nu_L}{\|H^\top H\|_*} \|E\|.$$

This also implies that

$$\|\mathbb{E} [S_L(E, \mathcal{Y})^2]\| \leq \mathbb{E} [\|S_L(E, \mathcal{Y})\|^2] \leq \frac{\nu_L^2}{\|H^\top H\|_*^2} \|E\|^2.$$

Therefore, by Lemma 6.31 we obtain that

$$\mathbb{P} \left[ \left\| \frac{1}{M} \sum_{i=1}^M S_L(E, \mathcal{Y}_T^i) - \mathbb{E} [S_L(E, \mathcal{Y}_T)] \right\| \geq t \right] \leq 2n \exp \left[ \frac{-M \|H^\top H\|_*^2 t^2 / 2}{\nu_L^2 \|E\|^2 + 2\nu_L \|H^\top H\|_* \|E\| t / 3} \right]$$

Thus, by substituting  $t$  with  $t\|E\|/\|H^\top H\|_*$  and applying this bound to (6.34) we obtain that

$$\mathbb{P} \left[ |\langle \nabla \widehat{J}_T(L) - \nabla J_T(L), E \rangle| \leq t\|E\| \right] \geq 1 - 2n \exp \left[ \frac{-Mt^2/2}{\nu_L^2 + 2\nu_L t/3} \right]$$

Finally, choosing  $E = \nabla \widehat{J}_T(L) - \nabla J_T(L)$  proves the second claim.  $\square$

### *Proof of the Proposition 6.21*

We provide the proof for a detailed version of Proposition 6.21:

*Proposition 6.21'* (Truncation Error Bound). Suppose  $m_0 = 0$ , then under Assumption 6.18 we have

$$|J(L) - J_T(L)| \leq \bar{\xi}_L \frac{\rho(A_L)^{T+1}}{1 - \rho(A_L)},$$

and

$$\|\nabla J(L) - \nabla J_T(L)\| \leq \bar{\gamma}_L \frac{\sqrt{\rho(A_L)}^{T+1}}{[1 - \rho(A_L)]^2}$$

where

$$\begin{aligned} \bar{\xi}_L &:= [\kappa_\xi^2 + (\kappa_\xi^2 + \kappa_\omega^2 \|L\|^2)C_L^2] \|H^\top H\|_* C_L^2, \\ \bar{\gamma}_L &:= 2 [\kappa_\xi^2 + C_L^2(\kappa_\xi^2 + \kappa_\omega^2 \|L\|^2)] C_L^2 \|H\| \|H^\top H\|_* \\ &\quad + 2\kappa_\omega^2(\kappa_\xi^2 + \kappa_\omega^2 \|L\|^2) \|L\| \|H\| \|H^\top H\|_* (C_L + 2C_L^3 \rho(A_L)^{3/2}) C_L^3 \sqrt{\rho(A_L)}^{T+1}. \end{aligned}$$

**Proof of Proposition 6.21.** For the purpose of this proof, we denote the same matrices by  $\mathcal{A}_{L,T}$  and  $\mathcal{M}_{L,T}[E]$  in order to emphasize on length  $T$ . Recall that

$$J_T(L) = \mathbb{E} [\varepsilon(L, \mathcal{Y})] = \text{tr} [\mathbb{E} [\mathcal{X}_L] \mathcal{A}_{L,T}^\top H^\top H \mathcal{A}_{L,T}],$$

where  $\mathbb{E} [\mathcal{X}_L] = \mathcal{Q} + (I \otimes L) \mathcal{R} (I \otimes L^\top)$ , which implies

$$J_T(L) = \text{tr} [[I \otimes (Q + LRL^\top)] \mathcal{A}_{L,T-1}^\top H^\top H \mathcal{A}_{L,T-1}] + \text{tr} [P_0(A_L^\top)^T H^\top H A_L^T],$$

On the other hand,

$$J(L) = \lim_{t \rightarrow \infty} \text{tr} [[I \otimes (Q + LRL^\top)] \mathcal{A}_{L,t}^\top H^\top H \mathcal{A}_{L,t}],$$

and thus

$$\begin{aligned} J(L) - J_T(L) &= - \text{tr} [P_0(A_L^\top)^T H^\top H A_L^T] \\ &\quad + \lim_{t \rightarrow \infty} \text{tr} [[I \otimes (Q + LRL^\top)] [I \otimes (A_L^\top)^T] \mathcal{A}_{L,t}^\top H^\top H \mathcal{A}_{L,t} [I \otimes A_L^T]] \\ &= - \text{tr} [A_L^T P_0(A_L^\top)^T H^\top H] \\ &\quad + \lim_{t \rightarrow \infty} \text{tr} [[I \otimes A_L^T (Q + LRL^\top) (A_L^\top)^T] \mathcal{A}_{L,t}^\top H^\top H \mathcal{A}_{L,t}]. \end{aligned}$$

Therefore, by the properties of trace and Lemma 6.23 we obtain that

$$\begin{aligned} |J(L) - J_T(L)| &\leq \|P_0\| \|A_L^T\|^2 \text{tr} [H^\top H] + \|Q + LRL^\top\| \|(A_L)^\top\|^2 \lim_{t \rightarrow \infty} \text{tr} [\mathcal{A}_{L,t}^\top H^\top H \mathcal{A}_{L,t}] \\ &\leq \|P_0\| C_L^2 \rho(A_L)^{T+1} \|H^\top H\|_* + (\|Q\| + \|R\| \|L\|^2) C_L^2 \rho(A_L)^{T+1} \frac{C_L^2 \|H^\top H\|_*}{1 - \rho(A_L)}, \end{aligned}$$



where the last line follows by Lemma 6.30. So,

$$|J(L) - J_T(L)| \leq \left[ \|P_0\| + \frac{(\|Q\| + \|R\| \|L\|^2) C_L^2}{1 - \rho(A_L)} \right] \|H^\top H\|_* C_L^2 \rho(A_L)^{T+1}$$

This, together with Assumption 6.18 imply the first claim.

Next, for simplicity we adopt the notation  $\mathcal{A}_{L,\infty}$  to interpret the limit as  $t \rightarrow \infty$ , then similar to the proof of Proposition 6.20 we can compute that

$$\begin{aligned} \langle \nabla J(L) - \nabla J_T(L), E \rangle &= -2\text{tr} [M_T[E] P_0 (A_L^\top)^T H^\top H] \\ &\quad + \text{tr} [I \otimes A_L^\top (Q + LRL^\top) (A_L^\top)^T \mathcal{N}_{L,\infty}[E]] \\ &\quad + 2\text{tr} [I \otimes M_T[E] (Q + LRL^\top) (A_L^\top)^T \mathcal{A}_{L,\infty}^\top H^\top H \mathcal{A}_{L,\infty}] \\ &\quad + 2\text{tr} [I \otimes A_L^\top ERL^\top (A_L^\top)^T \mathcal{A}_{L,\infty}^\top H^\top H \mathcal{A}_{L,\infty}]. \end{aligned}$$

Therefore, using (6.12) and Lemma 6.23 we have the following bound

$$\begin{aligned} |\langle \nabla J(L) - \nabla J_T(L), E \rangle| &\leq 2\|EH\| C_L^2 (T+1) \rho(A_L)^{T+1} \|P_0\| \|H^\top H\|_* \\ &\quad + \|Q + LRL^\top\| C_L^2 \rho(A_L)^{T+1} \|\mathcal{N}_{L,\infty}[E]\|_* \\ &\quad + 2\|Q + LRL^\top\| \|EH\| C_L^2 (T+1) \rho(A_L)^{T+1} \|\mathcal{A}_{L,\infty}^\top H^\top H \mathcal{A}_{L,\infty}\|_* \\ &\quad + 2\|ERL^\top\| C_L^2 \rho(A_L)^{T+1} \|\mathcal{A}_{L,\infty}^\top H^\top H \mathcal{A}_{L,\infty}\|_* \end{aligned}$$

which by Lemma 6.30 is bounded as follows

$$\begin{aligned} |\langle \nabla J(L) - \nabla J_T(L), E/\|E\| \rangle| &\leq 2\|P_0\| \|H\| \|H^\top H\|_* C_L^2 (T+1) \rho(A_L)^{T+1} \\ &\quad + \|Q + LRL^\top\| \|H\| \|H^\top H\|_* \frac{[2C_L^3 + 4C_L^5 \rho(A_L)^{3/2}] \rho(A_L)^{T+1}}{[1 - \rho(A_L)]^2} \\ &\quad + 2\|Q + LRL^\top\| \|H\| \|H^\top H\|_* \frac{C_L^4 (T+1) \rho(A_L)^{T+1}}{1 - \rho(A_L)} \\ &\quad + 2\|R\| \|L\| \|H^\top H\|_* \frac{C_L^4 \rho(A_L)^{T+1}}{1 - \rho(A_L)}. \end{aligned}$$

Finally, choosing  $E = \nabla J(L) - \nabla J_T(L)$  together with Assumption 6.18 implies

$$\begin{aligned} \|\nabla J(L) - \nabla J_T(L)\| &\leq 2 \left[ \frac{\kappa_\xi^2 + C_L^2 (\kappa_\xi^2 + \kappa_\omega^2 \|L\|^2)}{1 - \rho(A_L)} \right] C_L^2 \|H\| \|H^\top H\|_* (T+1) \rho(A_L)^{T+1} \\ &\quad + 2 \left[ \frac{\kappa_\omega^2 (\kappa_\xi^2 + \kappa_\omega^2 \|L\|^2) \|L\| \|H\| (C_L + 2C_L^3 \rho(A_L)^{3/2})}{1 - \rho(A_L)} \right] \|H^\top H\|_* \frac{C_L^3 \rho(A_L)^{T+1}}{1 - \rho(A_L)}. \end{aligned}$$

Finally, the second claim follows by the following simple facts:

$$(T+1)\rho(A_L)^{T+1} \leq \frac{\sqrt{\rho(A_L)}^{T+1}}{1-\rho(A_L)}, \quad \forall T > 0,$$

as  $\max_{t \geq 0} t\rho^t = \frac{2}{e \ln 1/\rho} \leq \frac{1}{\ln 1/\rho} \leq \frac{1}{1-\rho}$  for any  $\rho \in (0, 1)$ . This completes the proof.  $\square$

### *Complete description of Theorem 6.24*

The following is a detailed version of Theorem 6.24:

*Theorem 6.24'.* Suppose  $m_0 = 0_n$  and Assumption 6.18 holds for a data-set  $\{\mathcal{Y}_T^i\}_{i=1}^M$ . Define  $\nabla \widehat{J}_T(L) := \frac{1}{M} \sum_{i=1}^M \nabla \varepsilon(L, \mathcal{Y}_T^i)$ , where  $\nabla_L \varepsilon(L, \mathcal{Y})$  is obtained in Lemma 6.10. Consider  $\mathcal{S}_\alpha$  for some  $\alpha > 0$  and any  $s, s_0 > 0$  and  $\tau \in (0, 1)$ . Suppose the trajectory length

$$T \geq \ln \left( \frac{\bar{\gamma}_\alpha \sqrt{\min(n, m)}}{s_0} \right) / \ln \left( \frac{1}{\sqrt{\rho_\alpha}} \right)$$

and the batch size

$$M \geq \left[ 2 \left( \frac{\nu_\alpha \sqrt{\min(n, m)}}{s s_0 / \tau} \right)^2 + \frac{4}{3} \left( \frac{\nu_\alpha \sqrt{\min(n, m)}}{s s_0 / \tau} \right) \right] \ln(2n/\delta),$$

where

$$\begin{aligned} \bar{\gamma}_\alpha &:= 2 \left[ \kappa_\xi^2 + C_\alpha^2 (\kappa_\xi^2 + \kappa_\omega^2 D_\alpha^2) \right] C_\alpha^2 \|H\| \|H^\top H\|_* \\ &\quad + 2\kappa_\omega^2 (\kappa_\xi^2 + \kappa_\omega^2 D_\alpha^2) D_\alpha \|H\| \|H^\top H\|_* (C_\alpha + 2C_\alpha^3 \rho_\alpha^{3/2}) C_\alpha^3 \sqrt{\rho_\alpha}^{T+1}, \\ \nu_\alpha &:= \frac{2(\kappa_\xi + D_\alpha \kappa_\omega) \kappa_\omega C_\alpha^2 + \left[ C_\alpha + 2C_\alpha^3 \rho_\alpha^{3/2} \right] \|H\| (\kappa_\xi + D_\alpha \kappa_\omega)^2}{[1 - \sqrt{\rho_\alpha}]^3 \|H^\top H\|_*}, \end{aligned}$$

with  $\rho_\alpha$ ,  $C_\alpha$  and  $D_\alpha$  defined in Lemma 6.23. Then, with probability no less than  $1 - \delta$ , Assumption 6.16 holds.

### *Additional concentration bound results*

Combining the truncation bound in Proposition 6.21 with concentration bounds in Proposition 6.20 we can provide probabilistic bounds on the “estimated cost”  $\widehat{J}_T(L)$  and the

“estimated gradient”  $\nabla \widehat{J}_T(L)$ . The result involves the bound for the Frobenius norm of the error with probabilities independent of  $T$ .

Theorem 6.24, can be viewed as a simplified application of Theorem 6.32 to characterize the required minimum trajectory length and minimum batch so that the approximate gradient satisfies Assumption 6.16, with a specific  $s$  and  $s_0$ .

**Theorem 6.32.** *Suppose Assumption 6.18 holds. For any  $s > 0$  and  $L \in \mathcal{S}_\alpha$ , if*

$$M \geq \left[ 2 \left[ \frac{\nu_L \sqrt{\min(n, m)}}{s \|\nabla J(L)\|_F} \right]^2 + \frac{4}{3} \left[ \frac{\nu_L \sqrt{\min(n, m)}}{s \|\nabla J(L)\|_F} \right] \right] \ln(2n/\delta),$$

then with probability no less than  $1 - \delta$ ,

$$\|\nabla \widehat{J}_T(L) - \nabla J(L)\|_F \leq s \|\nabla J(L)\|_F + \bar{\gamma}_L \sqrt{\min(n, m)} \sqrt{\rho(A_L)}^{T+1},$$

with  $\nu_L$  and  $\bar{\gamma}_L$  defined in Proposition 6.20 and Proposition 6.21, respectively.

**Proof of Theorem 6.32.** Recall that for any  $L \in \mathcal{S}_\alpha$  for some  $\alpha > 0$  we have

$$\|\nabla \widehat{J}_T(L) - \nabla J(L)\| \leq \|\nabla \widehat{J}_T(L) - \nabla J_T(L)\| + \|\nabla J_T(L) - \nabla J(L)\|.$$

Thus, by Proposition 6.20 with  $s$  replaced by  $s \|\nabla J(L)\|_F / \sqrt{\min(n, m)}$  and applying Proposition 6.21 to the second term, we obtain that with probability at least  $1 - \delta$ :

$$\|\nabla \widehat{J}_T(L) - \nabla J(L)\| \leq \frac{s \|\nabla J(L)\|_F}{\sqrt{\min(n, m)}} + \bar{\gamma}_L \frac{\sqrt{\rho(A_L)}^{T+1}}{[1 - \rho(A_L)]^2},$$

where

$$\delta \geq 2n \exp \left[ \frac{-Ms^2/2}{\left[ \frac{\nu_L \sqrt{\min(n, m)}}{\|\nabla J(L)\|_F} \right]^2 + 2 \left[ \frac{\nu_L \sqrt{\min(n, m)}}{\|\nabla J(L)\|_F} \right] s/3} \right].$$

Noticing  $\|\nabla \widehat{J}_T(L) - \nabla J(L)\|_F \leq \sqrt{\min(n, m)} \|\nabla \widehat{J}_T(L) - \nabla J(L)\|$  and rearranging terms will complete the proof.  $\square$

One can also provide the analogous concentration error bounds where the probabilities are independent of the system dimension  $n$ .

**Proposition 6.33** (Concentration independent of system dimension  $n$ ). *Under the same hypothesis, we have*

$$\mathbb{P} \left[ |\widehat{J}_T(L) - J_T(L)| \leq s \right] \geq 1 - 2T \exp \left[ \frac{-Ms^2/2}{\bar{\mu}_L^2 T^2 + 2\bar{\mu}_L T s/3} \right],$$

and

$$\begin{aligned} \mathbb{P} \left[ \|\nabla \widehat{J}_T(L) - \nabla J_T(L)\| \leq s \right] &\geq 1 - 2T \exp \left[ \frac{-Ms^2/2}{\bar{\nu}_L^2 T^2 + 2\bar{\nu}_L T s/3} \right] \\ &\quad - 2T \exp \left[ \frac{-Ms^2/2}{\kappa_\omega^2 \bar{\mu}_L^2 T^2 + 2\kappa_\omega \bar{\mu}_L T s/3} \right] \end{aligned}$$

where

$$\begin{aligned} \bar{\mu}_L &:= \frac{C_L^2 \|H^\top H\|_* \kappa_L^2}{1 - \rho(A_L)} \\ \bar{\nu}_L &:= \frac{[2C_L + 4C_L^3 \rho(A_L)^{3/2}] \|H\| \|H^\top H\|_* \kappa_L^2}{[1 - \rho(A_L)]^2} \end{aligned}$$

**Proof of Proposition 6.33.** Similar to the previous proof, we have

$$\widehat{J}_T(L) - J_T(L) = \text{tr} [(\mathcal{Z}_L - \mathbb{E}[\mathcal{Z}_L]) \mathcal{A}_L^\top H^\top H \mathcal{A}_L],$$

and thus, by (6.12) we obtain

$$|\widehat{J}_T(L) - J_T(L)| \leq \|(\mathcal{Z}_L - \mathbb{E}[\mathcal{Z}_L])\| \|\mathcal{A}_L^\top H^\top H \mathcal{A}_L\|_*. \quad (6.35)$$

Next, we consider the symmetric random matrix  $(\mathcal{Z}_L - \mathbb{E}[\mathcal{Z}_L])$  and recall that  $\|\xi(t) - L\omega(t)\| \leq \kappa_L$  almost surely; thus

$$\|\mathcal{X}_L\| = \|\eta_L\|^2 \leq \kappa_L^2 T^2.$$

It then follows that

$$\|\mathbb{E}[\mathcal{X}_L^2]\| \leq \mathbb{E}[\|\mathcal{X}_L\|^2] \leq \kappa_L^4 T^4.$$

Therefore, by Lemma 6.31 we obtain that

$$\mathbb{P} [\|(\mathcal{Z}_L - \mathbb{E}[\mathcal{Z}_L])\| \geq t] \leq 2T \exp \left[ \frac{-Mt^2/2}{\kappa_L^4 T^4 + 2\kappa_L^2 T^2 t/3} \right] \quad (6.36)$$

Substituting  $t$  with  $t/\|\mathcal{A}_L^\top H^\top H \mathcal{A}_L\|_*$  together with (6.35) implies the first claim because by Lemma 6.30  $\kappa_L^2 T^2 \|\mathcal{A}_L^\top H^\top H \mathcal{A}_L\|_* \leq \bar{\mu}_L T$ .

Again, similar to (6.34) in the previous proof, by (6.12) we obtain that

$$\begin{aligned} |\langle \nabla \widehat{J}_T(L) - \nabla J_T(L), E \rangle| &\leq \left\| \frac{1}{M} \sum_{i=1}^M S_L(E, \mathcal{Y}_T^i) - \mathbb{E} [S_L(E, \mathcal{Y}_T^i)] \right\| \|\mathcal{A}_L^\top H^\top H \mathcal{A}_L\|_* \\ &\quad + \left\| \frac{1}{M} \sum_{i=1}^M \mathcal{X}_L(\mathcal{Y}^i) - \mathbb{E} [\mathcal{X}_L(\mathcal{Y}^i)] \right\| \|\mathcal{N}_L[E]\|_* \end{aligned} \quad (6.37)$$

where  $S_L(E, \mathcal{Y})$  is the symmetric part of the following random matrix

$$-2\xi \omega^\top (I \otimes E^\top) + 2(I \otimes L) \omega \omega^\top (I \otimes E^\top).$$

So, we claim that almost surely

$$\begin{aligned} \|S_L(E, \mathcal{Y})\| &\leq 2\|(I \otimes E)\omega\|(\|\xi\| + \|(I \otimes L)\omega\|) \\ &\leq 2\|E\| \kappa_\omega T (\kappa_\xi T + \|L\| \kappa_\omega T) \\ &= \kappa_L \kappa_\omega T^2 \|E\|, \end{aligned}$$

and thus

$$\|\mathbb{E} [S_L(E, \mathcal{Y})^2]\| \leq \mathbb{E} [\|S_L(E, \mathcal{Y})\|^2] \leq \kappa_L^2 \kappa_\omega^2 T^4 \|E\|^2.$$

Therefore, by Lemma 6.31 we obtain that

$$\mathbb{P} \left[ \left\| \frac{1}{M} \sum_{i=1}^M S_L(E, \mathcal{Y}_T^i) - \mathbb{E} [S_L(E, \mathcal{Y}_T^i)] \right\| \geq t \right] \leq 2T \exp \left[ \frac{-Mt^2/2}{\kappa_\omega^2 \kappa_L^2 T^4 \|E\|^2 + 2\kappa_L \kappa_\omega T^2 \|E\| t/3} \right]$$

Substituting  $t$  with  $t/\|\mathcal{A}_L^\top H^\top H \mathcal{A}_L\|_*$  implies that

$$\begin{aligned} \mathbb{P} \left[ \left\| \frac{1}{M} \sum_{i=1}^M S_L(E, \mathcal{Y}_T^i) - \mathbb{E} [S_L(E, \mathcal{Y}_T^i)] \right\| \|\mathcal{A}_L^\top H^\top H \mathcal{A}_L\|_* \geq t \right] \\ \leq 2T \exp \left[ \frac{-Mt^2/2}{\kappa_\omega^2 \bar{\mu}_L^2 T^2 \|E\|^2 + 2\kappa_\omega \bar{\mu}_L T \|E\| t/3} \right] \end{aligned}$$

because by Lemma 6.30 we have  $\kappa_L^2 T^2 \|\mathcal{A}_L^\top H^\top H \mathcal{A}_L\|_* \leq \bar{\mu}_L T$ .

Next, by substituting  $t$  with  $t/\|\mathcal{N}_L[E]\|_*$  in (6.36) we have

$$\mathbb{P} [\|\mathcal{Z}_L - \mathbb{E} [\mathcal{Z}_L]\| \|\mathcal{N}_L[E]\|_* \geq t] \leq 2T \exp \left[ \frac{-Mt^2/2}{\bar{\nu}_L^2 T^2 \|E\|^2 + 2\bar{\nu}_L T \|E\| t/3} \right]$$

because Lemma 6.30 implies that  $\kappa_L^2 T^2 \|\mathcal{N}_L[E]\|_* \leq \bar{\nu}_L T \|E\|$ . Thus, by combining the last two inequalities and using the union bound for (6.37) we obtain that

$$\begin{aligned} \mathbb{P} \left[ |\langle \nabla \hat{J}_T(L) - \nabla J_T(L), E \rangle| \geq t \right] &\leq 2T \exp \left[ \frac{-Mt^2/2}{\bar{\nu}_L^2 T^2 \|E\|^2 + 2\bar{\nu}_L T \|E\| t/3} \right] \\ &\quad + 2T \exp \left[ \frac{-Mt^2/2}{\kappa_\omega^2 \bar{\mu}_L^2 T^2 \|E\|^2 + 2\kappa_\omega \bar{\mu}_L T \|E\| t/3} \right] \end{aligned}$$

Finally, substituting  $t$  with  $t\|E\|$  and choosing  $E = \nabla \hat{J}_T(L) - \nabla J_T(L)$  proves the second claim.  $\square$

*Remark 6.34.* We obtained a better bound for truncation of the gradient as

$$\|\nabla J(L) - \nabla J_T(L)\| \leq \bar{\gamma}_1(L) \frac{\sqrt{\rho(A_L)}^{T+1}}{[1 - \rho(A_L)] \ln(1/\rho(A_L))} + \bar{\gamma}_2(L) \frac{\rho(A_L)^{T+1}}{[1 - \rho(A_L)]^2},$$

which has been simplified for now.

### 6.5.8 Innovation correlation based approach

Before finishing this section and moving to numerical results, we provide a few remarks on the innovation correlation approach as an alternative direction. It is based on inspecting the autocovariance of the innovation error process

$$e_t(L) = y(t) - H\hat{x}_L(t),$$

where  $\hat{x}_L(t)$  is the estimate obtained by a Kalman filter with a sub-optimal, but stabilizing, Kalman gain  $L$ ,  $y(t)$  is the noisy output measurement, and  $H$  is the observation matrix. The  $k$ -step autocovariance matrix  $[e_{t-k}(L)e_t(L)^\top]$ , for  $k = 0, 1, \dots, \ell$ , is approximated empirically using batch of independent observation signals. This is then compared to the its asymptotic analytical expression to approximate the noise covariance matrices [Odelson et al., 2006]. Alternatively, the autocovariance matrices can be used to directly estimate the optimal Kalman

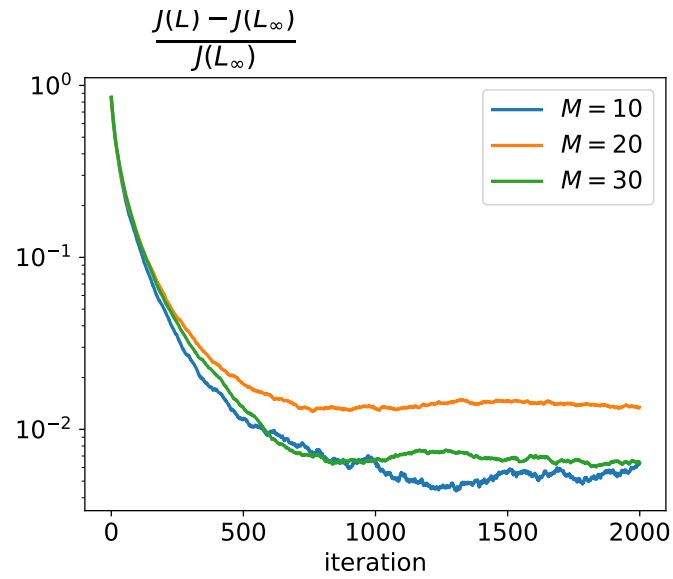
gain  $L_\infty$ . In particular, it is known that, with the optimal Kalman gain, the innovation process becomes independent (or white), i.e.,  $[e_{t-k}(L_\infty)e_t(L_\infty)^\top] = 0$  for  $k > 0$ . This criteria is used in [Zhang et al., 2020] to obtain the optimal Kalman gain by minimizing the normalized autocovariance matrices with a gradient-based optimization algorithm.

## 6.6 Numerical Simulations

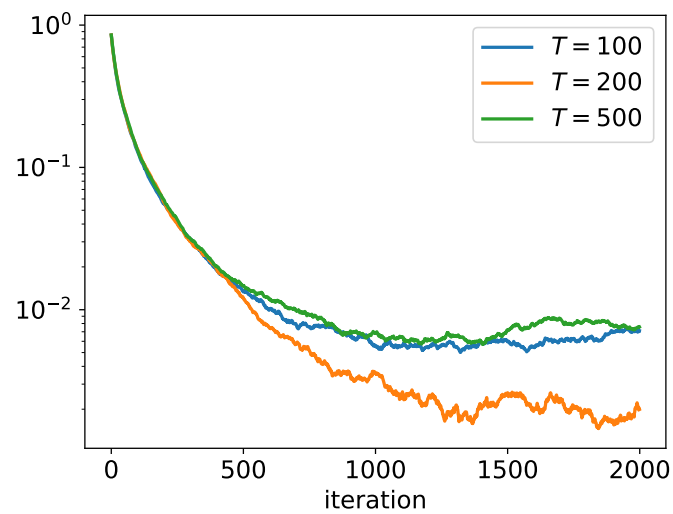
Herein, we showcase the application of the developed theory for improving the estimation policy for an LTI system. Specifically, we consider an undamped mass-spring system with known parameters  $(A, H)$  with  $n = 2$  and  $m = 1$ . In the hindsight, we consider a variance of 0.1 for each state dynamic noise, a state covariance of 0.05 and a variance of 0.1 for the observation noise. Assuming a trajectory of length  $T$  at every iteration, the approximate gradient is obtained as in Lemma 6.10, only requiring an output data sequence collected from the system in (6.1). Then, the progress of policy updates using the SGD algorithm for different values of trajectory length  $T$  and batch size  $M$  are depicted in Figure 6.1 where each figure shows *average progress* over 20 rounds of simulation. The figure demonstrates a linear convergence outside of a neighborhood of global optimum that depends on the bias term in the approximate gradient (due to truncated data trajectories). The rate then drops when the policy iterates enter into this neighborhood which is expected as every update only relies on a *biased gradient*—in contrast to the linear convergence established for deterministic GD (to the exact optimum) using the true gradient.

## 6.7 Remarks and Future Directions

In this chapter, we considered the problem of learning the optimal (steady-state) Kalman gain for linear systems with unknown process and measurement noise covariances. Our approach builds on the duality between optimal control and estimation, resulting in a direct stochastic PO algorithm for learning the optimal filter gain. We also provided convergence guarantees and finite sample complexity with bias and variance error bounds that scale well with problem parameters. In particular, the variance is independent of the length of



(a)



(b)

Figure 6.1: SGD directly from output data and without prior knowledge of the noise covariances or state information. Mean progress of the normalized estimation error over 20 simulations obtained from data trajectories of a) different batch size  $M$  and b) different length  $T$ .



data trajectories and scales logarithmically with problem dimension, and bias term decreases exponentially with the length.

This work contributes a generic optimization algorithm and introduces a filtering strategy for estimating dynamical system states. While theoretical, it raises privacy concerns similar to the model-based Kalman filter. Limitations include the need for prior knowledge of system parameters, nonetheless, parameter uncertainties can be treated practically as process and measurement noise. Finally, sample complexities depend on the stability margin  $1 - \sqrt{\rho(A_L)}$ , inherent to the system generating the data. The result of this chapter has been mainly adapted from [Talebi et al., 2023].

## BIBLIOGRAPHY

- Pierre-Antoine Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2009.
- Juergen Ackermann. Parameter space design of robust control systems. *IEEE Transactions on Automatic Control*, 25(6):1058–1072, 1980.
- Naman Agarwal, Brian Bullins, Elad Hazan, Sham Kakade, and Karan Singh. Online control with adversarial disturbances. In *International Conference on Machine Learning*, pages 111–119, 2019.
- Bernt M Åkesson, John Bagterp Jørgensen, Niels Kjølstad Poulsen, and Sten Bay Jørgensen. A generalized autocovariance least-squares method for Kalman filter tuning. *Journal of Process Control*, 18(7-8):769–779, 2008.
- Atiye Alaeddini, Siavash Alemzadeh, Afshin Mesbahi, and Mehran Mesbahi. Linear model regression on time-series data: non-asymptotic error bounds and applications. In *IEEE Conference on Decision and Control*, pages 2259–2264, 2018.
- Siavash Alemzadeh and Mehran Mesbahi. Influence models on layered uncertain networks: A guaranteed-cost design perspective. In *IEEE Conference on Decision and Control*, pages 5251–5256, 2018.
- Siavash Alemzadeh and Mehran Mesbahi. Distributed Q-learning for dynamically decoupled systems. In *American Control Conference*, pages 772–777, 2019.
- Siavash Alemzadeh, Shahriar Talebi, and Mehran Mesbahi. D3PI: Data-driven distributed policy iteration for homogeneous interconnected systems. *arXiv preprint arXiv:2103.11572*, 2021.

- Tansu Alpcan and Tamer Başar. *Distributed Algorithms for Nash Equilibria of Flow Control Games*, pages 473–498. Birkhäuser Boston, Boston, MA, In *Advances in Dynamic Games: Applications to Economics, Finance, Optimization, and Stochastic Control*, 2005.
- Brian DO Anderson and John B Moore. *Optimal control: linear quadratic methods*. Courier Corporation, 2007.
- Karl J Åström. *Introduction to Stochastic Control Theory*. Courier Corporation, 2012.
- Karl J Åström and Björn Wittenmark. *Adaptive control*. Courier Corporation, 2013.
- Bassam Bamieh, Fernando Paganini, and Munther A Dahleh. Distributed control of spatially invariant systems. *IEEE Transactions on Automatic Control*, 47(7):1091–1107, 2002.
- Stefanos Baros, Chin-Yao Chang, Gabriel E. Colon-Reyes, and Andrey Bernstein. Online data-enabled predictive control. *arXiv preprint arXiv:2003.03866*, 2020.
- Tamer Basar and Geert J Olsder. *Dynamic Noncooperative Game Theory*. Academic Press, 1995.
- Amir Beck. *First-Order Methods in Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2017.
- Pierre R Belanger. Estimation of noise covariance matrices for a linear time-varying stochastic process. *Automatica*, 10(3):267–275, 1974.
- Richard Bellman. Dynamic programming. *Science*, 153(3731):34–37, 1966.
- Alain Bensoussan. *Estimation and Control of Dynamical Systems*, volume 48. Springer, 2018.
- Julian Berberich, Anne Koch, Carsten W. Scherer, and Frank Allgower. Robust data-driven state-feedback design. In *American Control Conference (ACC)*, pages 1532–1538, 2020a.

- Julian Berberich, Johannes Köhler, Matthias A Müller, and Frank Allgöwer. Data-driven model predictive control with stability and robustness guarantees. *IEEE Transactions on Automatic Control*, 66(4):1702–1717, 2020b.
- Dario A Bini, Bruno Iannazzo, and Beatrice Meini. *Numerical Solution of Algebraic Riccati Equations*. Society for Industrial and Applied Mathematics (SIAM), 2011.
- Béla Bollobás. *Modern Graph Theory*, volume 184. Springer Science & Business Media, 2013.
- Francesco Borrelli and Tamás Keviczky. Distributed LQR design for identical dynamically decoupled systems. *IEEE Transactions on Automatic Control*, 53(8):1901–1912, 2008.
- Francesco Borrelli, Tamás Keviczky, Gary J Balas, Greg Stewart, Kingsley Fregene, and Datta Godbole. Hybrid decentralized control of large scale systems. In *International Workshop on Hybrid Systems: Computation and Control*, pages 168–183. Springer, 2005.
- John T Bosworth. *Linearized aerodynamic and control law models of X-29A airplane and comparison with flight data*, volume 4356. NASA, 1992.
- Steven J Bradtke, B Erik Ydstie, and Andrew G Barto. Adaptive linear quadratic control using policy iteration. In *American Control Conference*, volume 3, pages 3475–3479, 1994.
- Mario Bravo, David Leslie, and Panayotis Mertikopoulos. Bandit learning in concave n-person games. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 5661–5671. Curran Associates, Inc., 2018.
- Jingjing Bu, Afshin Mesbahi, Maryam Fazel, and Mehran Mesbahi. LQR through the lens of first order methods: Discrete-time case. *arXiv preprint arXiv:1907.08921*, 2019a.
- Jingjing Bu, Afshin Mesbahi, and Mehran Mesbahi. On topological and metrical properties of stabilizing feedback gains: the MIMO case. *arXiv preprint arXiv:1904.02737*, 2019b.

- Jingjing Bu, Afshin Mesbahi, and Mehran Mesbahi. Policy gradient-based algorithms for continuous-time linear quadratic control. *arXiv preprint arXiv:2006.09178*, 2020a.
- Jingjing Bu, Afshin Mesbahi, and Mehran Mesbahi. On topological properties of the set of stabilizing feedback gain. *IEEE Transactions on Automatic Control*, 2020b.
- Eduardo Camponogara, Dong Jia, Bruce H Krogh, and Sarosh Talukdar. Distributed model predictive control. *IEEE control systems magazine*, 22(1):44–52, 2002.
- Burian Carew and P Belanger. Identification of optimum filter steady-state gain for systems with unknown noise covariances. *IEEE Transactions on Automatic Control*, 18(6):582–587, 1973.
- Ting-Jui Chang and Shahin Shahrampour. Distributed online linear quadratic control for linear time-invariant systems. In *2021 American Control Conference (ACC)*, pages 923–928. IEEE, 2021.
- Airlie Chapman, Andrés D González, Mehran Mesbahi, Leonardo Dueñas-Osorio, and Raissa M D’Souza. Data-guided control: Clustering, graph products, and decentralized control. In *IEEE Conference on Decision and Control*, pages 493–498. IEEE, 2017.
- Xinyi Chen and Elad Hazan. Black-box control for linear dynamical systems. *arXiv preprint arXiv:2007.06650*, 2020.
- Pierre Comon and Gene H Golub. Tracking a few extreme singular values and vectors in signal processing. *Proceedings of the IEEE*, 78(8):1327–1343, 1990.
- Jeremy Coulson, John Lygeros, and Florian Dörfler. Data-enabled predictive control: In the shallows of the deepc. In *2019 18th European Control Conference (ECC)*, pages 307–312. IEEE, 2019.
- Maurício C De Oliveira, Jacques Bernussou, and José C Geromel. A new discrete-time robust stability condition. *Systems & control letters*, 37(4):261–265, 1999a.

- Maurício C De Oliveira, Jacques Geromel, and Liu Hsu. LMI characterization of structural and robust stability: the discrete-time case. *Linear Algebra and its applications*, 296(1-3): 27–38, 1999b.
- Claudio De Persis and Pietro Tesi. Formulas for data-driven control: Stabilization, optimality, and robustness. *IEEE Transactions on Automatic Control*, 65(3):909–924, 2019.
- Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. *arXiv preprint arXiv:1710.01688*, 2017.
- Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. *Foundations of Computational Mathematics*, pages 1–47, 2019a.
- Sarah Dean, Stephen Tu, Nikolai Matni, and Benjamin Recht. Safely learning to control the constrained linear quadratic regulator. In *American Control Conference*, pages 5582–5588, 2019b.
- John E Dennis Jr and Robert B Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Society for Industrial and Applied Mathematics, 1996.
- Paresh Deshpande, Prathyush P Menon, Christopher Edwards, and Ian Postlethwaite. A distributed control law with guaranteed LQR cost for identical dynamically coupled linear systems. In *American Control Conference (ACC)*, pages 5342–5347. IEEE, 2011.
- Paresh Deshpande, PP Menon, Christopher Edwards, and Ian Postlethwaite. Sub-optimal distributed control law with  $\mathcal{H}_2$  performance for identical dynamically coupled linear systems. *IET Control Theory & Applications*, 6(16):2509–2517, 2012.
- Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of Markov chains. *The Annals of Applied Probability*, 1(1):36–61, 2 1991.

- Steven Diamond and Stephen Boyd. CVXPY: A python-embedded modeling language for convex optimization. *The Journal of Machine Learning Research*, 17(1):2909–2913, 2016.
- Thinh T. Doan, Subhonmesh Bose, D. Hoa Nguyen, and Carolyn L. Beck. Convergence of the iterates in mirror descent methods. *IEEE Control Systems Letters*, 3(1):114–119, 2019.
- Peter Dorato, Vito Cerone, and Chaouki Abdallah. *Linear Quadratic Control: An introduction*. Simon & Schuster, Inc., 1994.
- John C. Duchi, Alekh Agarwal, and Martin J. Wainwright. Dual averaging for distributed optimization: Convergence analysis and network scaling. *IEEE Transactions on Automatic Control*, 57(3):592–606, 2012.
- Jindrich Dunik, Miroslav Simandl, and Ondrej Straka. Methods for estimating state and measurement noise covariance matrices: Aspects and comparison. *IFAC Proceedings Volumes*, 42(10):372–377, 2009.
- Jacob Engwerda. *LQ Dynamic Optimization and Differential Games*. John Wiley & Sons, 2005a.
- Jacob Engwerda. *LQ Dynamic Optimization and Differential Games*. John Wiley & Sons, 2005b.
- Francisco. Facchinei and Jong-Shi. Pang. *Finite-dimensional Variational Inequalities and Complementarity Problems*. Springer, 2003.
- Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. Finite-time adaptive stabilization of linear systems. *IEEE Transactions on Automatic Control*, 64(8), 2019.
- Ilyas Fatkhullin and Boris Polyak. Optimizing static linear feedback: Gradient method. *arXiv preprint arXiv:2004.09875*, 2020.

- Salar Fattahi, Nikolai Matni, and Somayeh Sojoudi. Learning sparse dynamical systems from a single sample trajectory. In *IEEE Conference on Decision and Control*, pages 2682–2689, 2019.
- J Alexander Fax and Richard M Murray. Information flow and cooperative control of vehicle formations. *IEEE Transactions on Automatic Control*, 49(9):1465–1476, 2004.
- Maryam Fazel, Rong Ge, Sham Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1467–1476. PMLR, 10–15 Jul 2018.
- Han Feng and Javad Lavaei. On the exponential number of connected components for the feasible set of optimal decentralized control problems. In *2019 American Control Conference (ACC)*, pages 1430–1437, 2019.
- Wendell H Fleming and Sanjoy K Mitter. Optimal control and nonlinear filtering for non-degenerate diffusion processes. *Stochastics: An International Journal of Probability and Stochastic Processes*, 8(1):63–77, 1982.
- Noah E Friedkin and Eugene C Johnsen. Social influence and opinions. *Journal of Math. Sociology*, 15(3-4):193–206, 1990.
- Paul Frihauf, Miroslav Krstic, and Tamer Basar. Nash equilibrium seeking in noncooperative games. *IEEE Transactions on Automatic Control*, 57(5):1192–1207, 2011.
- Daniel Gabay. Minimizing a differentiable function over a differential manifold. *Journal of Optimization Theory and Applications*, 37(2):177–219, 1982.
- Zoran Gajic and Muhammad Tahir Javed Qureshi. *Lyapunov Matrix Equation in System Stability and Control*. Courier Corporation, 2008.



- Bahman Ghahesifard and Jorge Cortés. Distributed convergence to Nash equilibria in two-network zero-sum games. *Automatica*, 49(6):1683–1692, 2013.
- Andrés D González, Airlie Chapman, Leonardo Dueñas-Osorio, Mehran Mesbahi, and Raissa M D’Souza. Efficient infrastructure restoration strategies using the recovery operator. *Computer-Aided Civil and Infrastructure Engineering*, 32(12):991–1006, 2017.
- Graham C Goodwin, Stefan F Graebe, Mario E Salgado, et al. *Control System Design*. Upper Saddle River, NJ: Prentice Hall,, 2001.
- Sergio Grammatico. Proximal dynamics in multiagent network games. *IEEE Transactions on Control of Network Systems*, 5(4):1707–1716, Dec 2018.
- Vijay Gupta, Babak Hassibi, and Richard M Murray. A sub-optimal algorithm to synthesize control laws for a network of dynamic agents. *International Journal of Control*, 78(16):1302–1313, 2005.
- Zhu Han, Dusit Niyato, Walid Saad, Tamer Başar, and Are Hjørungnes. *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*. Cambridge university press, 2012.
- Simon Haykin. Adaptive filter theory. *Information and System Science*. Prentice Hall, 2002.
- João P. Hespanha. *Linear Systems Theory*. Princeton University Press, 2018.
- Gary Hewer. An iterative technique for the computation of the steady state gains for the discrete optimal regulator. *IEEE Transactions on Automatic Control*, 16(4):382–384, 1971.
- Nicholas J Higham. *Functions of Matrices: Theory and computation*. SIAM, 2008.
- Charles G Hilborn and Demetrios G Lainiotis. Optimal estimation in the presence of unknown parameters. *IEEE Transactions on Systems Science and Cybernetics*, 5(1):38–43, 1969.

- Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge University Press, 2012.
- Zhongsheng Hou, Huijun Gao, and Frank L Lewis. Data-driven control and learning systems. *IEEE Transactions on Industrial Electronics*, 64(5):4070–4075, 2017.
- Yeung Sam Hung and AG MacFarlane. *Multivariable Feedback: A quasi-classical approach*. Springer-Verlag New York, Inc., 1982.
- P Ioannou. Decentralized adaptive control of interconnected systems. *IEEE Transactions on Automatic Control*, 31(4):291–298, 1986.
- Tetsuya Iwasaki, Robert E. Skelton, and José C. Geromel. Linear quadratic suboptimal control with static output feedback. *Systems and Control Letters*, 23(6):421–430, 1994.
- Pushpak Jagtap, George J. Pappas, and Majid Zamani. Control barrier functions for unknown nonlinear systems using gaussian processes. *arXiv preprint arXiv:2010.05818*, 2020.
- Yu Jiang and Zhong-Ping Jiang. Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics. *Automatica*, 48(10):2699–2704, 2012.
- Rudolf E Kalman. On the general theory of control systems. In *Proceedings First International Conference on Automatic Control, Moscow, USSR*, pages 481–492, 1960a.
- Rudolf E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 03 1960b.
- Rangasami L. Kashyap. Maximum likelihood identification of stochastic linear systems. *IEEE Transactions on Automatic Control*, 15(1):25–34, 1970.
- Tosio Kato. *Perturbation Theory for Linear Operators*, volume 132. Springer Science & Business Media, 2013.

- Pramod P Khargonekar, Ian R Petersen, and Kemin Zhou. Robust stabilization of uncertain linear systems: quadratic stabilizability and  $H^\infty$  control theory. *IEEE Transactions on Automatic control*, 35(3):356–361, 1990.
- H. Jin Kim and Andrew Y Ng. Stable adaptive control with online learning. In *Advances in Neural Info. Processing Systems*, pages 977–984, 2005.
- Jin Won Kim. Duality for nonlinear filtering. *arXiv preprint arXiv:2207.07709*, 2022.
- Jin Won Kim, Prashant G Mehta, and Sean P Meyn. What is the Lagrangian for nonlinear filtering? In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 1607–1614. IEEE, 2019.
- Steven G Krantz and Harold R Parks. *A Primer of Real Analytic Functions*. Springer Science & Business Media, 2002.
- Huibert Kwakernaak and Raphael Sivan. *Linear Optimal Control Systems*, volume 1072. Wiley-interscience, 1969.
- Sahin Lale, Kamyar Azizzadenesheli, Babak Hassibi, and Anima Anandkumar. Logarithmic regret bound in partially observable linear dynamical systems. *Advances in Neural Information Processing Systems*, 33:20876–20888, 2020a.
- Sahin Lale, Kamyar Azizzadenesheli, Babak Hassibi, and Anima Anandkumar. Regret bound of adaptive control in linear quadratic gaussian (LQG) systems. *arXiv preprint arXiv:2003.05999*, 2020b.
- Luca Lambertini. *Differential Games in Industrial Economics*. Cambridge University Press, 2018.
- Peter Lancaster. Explicit solutions of linear matrix equations. *SIAM review*, 12(4):544–566, 1970.
- John M Lee. *Introduction to Smooth Manifolds*. Springer, second edition, 2013.

- John M Lee. *Introduction to Riemannian Manifolds*. Cham, Switzerland: Springer Nature, second edition, 2018.
- W. Levine, T. Johnson, and M. Athans. Optimal limited state variable feedback controllers for linear systems. *IEEE Transactions on Automatic Control*, 16(6):785–793, 1971.
- William Levine and Michael Athans. On the determination of the optimal constant output feedback gains for linear multivariable systems. *IEEE Transactions on Automatic control*, 15(1):44–48, 1970.
- Frank Lewis. *Optimal Estimation with an Introduction to Stochastic Control Theory*. New York, Wiley-Interscience, 1986.
- Frank L Lewis and Kyriakos G Vamvoudakis. Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(1):14–25, 2010.
- Frank L Lewis, Draguna Vrabić, and Kyriakos G Vamvoudakis. Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers. *IEEE Control Systems Magazine*, 32(6):76–105, 2012.
- Jinna Li, Hamidreza Modares, Tianyou Chai, Frank L Lewis, and Lihua Xie. Off-policy reinforcement learning for synchronization in multiagent graphical games. *IEEE transactions on neural networks and learning systems*, 28(10):2434–2445, 2017.
- Na Li and Jason R. Marden. Designing games for distributed optimization. *IEEE Journal on Selected Topics in Signal Processing*, 7(2):230–242, 2013.
- Zhongkui Li, Zhisheng Duan, Lihua Xie, and Xiangdong Liu. Distributed robust control of linear multi-agent systems with parameter uncertainties. *International Journal of Control*, 85(8):1039–1050, 2012.

- Lennart Ljung. System identification. *Wiley encyclopedia of electrical and electronics engineering*, pages 1–19, 1999.
- Lennart Ljung. *System Identification*. Wiley Online Library, 2001.
- Rogelio Lozano, Pedro Castillo, Pedro Garcia, and Alejandro Dzul. Robust prediction-based control for unstable delay systems: Application to the yaw control of a mini-helicopter. *Automatica*, 40(4):603–612, 2004.
- Yuwei Luo, Zhuoran Yang, Zhaoran Wang, and Mladen Kolar. Natural actor-critic converges globally for hierarchical linear quadratic regulator. *arXiv preprint arXiv:1912.06875*, 2019.
- David Magill. Optimal adaptive estimation of sampled stochastic processes. *IEEE Transactions on Automatic Control*, 10(4):434–439, 1965.
- Pertti M. Mäkilä and Hannu T. Toivonen. Computational methods for parametric LQ problems—a survey. *IEEE Transactions on Automatic Control*, 32(8):658–671, 1987.
- Krithika Manohar, Eurika Kaiser, Steven L Brunton, and J Nathan Kutz. Optimized sampling for multiscale dynamics. *Multiscale Modeling & Simulation*, 17(1):117–136, 2019.
- Mohammad Hossein Manshaei, Quanyan Zhu, Tansu Alpcan, Tamer Başçar, and Jean-Pierre Hubaux. Game theory meets network security and privacy. *ACM Computing Surveys (CSUR)*, 45(3):25, 2013.
- Karl Mårtensson and Anders Rantzer. Gradient methods for iterative distributed control synthesis. In *Proceedings of the 48th IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 549–554. IEEE, 2009.
- Paolo Massioni and Michel Verhaegen. Distributed control for identical dynamically coupled systems: A decomposition approach. *IEEE Transactions on Automatic Control*, 54(1):124–135, 2009.

- Peter Matisko and Vladimír Havlena. Noise covariances estimation for Kalman filter tuning. *IFAC Proceedings Volumes*, 43(10):31–36, 2010.
- Daniel McFadden. On the controllability of decentralized macroeconomic systems: The assignment problem. In *Mathematical Systems Theory and Economics I/II*, pages 221–239. Springer, 1969.
- Raman Mehra. On the identification of variances and adaptive Kalman filtering. *IEEE Transactions on Automatic Control*, 15(2):175–184, 1970.
- Raman Mehra. Approaches to adaptive filtering. *IEEE Transactions on Automatic Control*, 17(5):693–698, 1972.
- Panayotis Mertikopoulos and Zhengyuan Zhou. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, pages 1–43, 2018.
- Mehran Mesbahi and Magnus Egerstedt. *Graph Theoretic Methods in Multiagent Networks*. Princeton University Press, 2010.
- Carl D Meyer, Jr. Generalized inversion of modified matrices. *SIAM Journal on Applied Mathematics*, 24(3):315–323, 1973.
- Daniel D Moerder and Anthony J. Calise. Convergence of a numerical algorithm for calculating optimal output feedback gains. *IEEE Transactions on Automatic Control*, 30(9):900–903, 1985.
- Hesameddin Mohammadi, Mahdi Soltanolkotabi, and Mihailo R. Jovanović. On the linear convergence of random search for discrete-time LQR. *IEEE Control Systems Letters*, 5(3):989–994, 2021a.
- Hesameddin Mohammadi, Armin Zare, Mahdi Soltanolkotabi, and Mihailo R. Jovanović. Convergence and sample complexity of gradient methods for the model-free linear quadratic regulator problem. *arXiv preprint arXiv:1912.11899*, 2021b.

- Takehiro Mori, Norio Fukuma, and Michiyoshi Kuwahara. On the discrete lyapunov matrix equation. *IEEE Transactions on Automatic Control*, 27(2):463–464, 1982.
- Richard E. Mortensen. Maximum-likelihood recursive nonlinear filtering. *Journal of Optimization Theory and Applications*, 2(6):386–394, 1968.
- Nader Motee and Ali Jadbabaie. Optimal control of spatially distributed systems. *IEEE Transactions on Automatic Control*, 53(7):1616–1629, 2008.
- Kenneth Myers and BD Tapley. Adaptive sequential estimation with unknown noise statistics. *IEEE Transactions on Automatic Control*, 21(4):520–523, 1976.
- Karl Mårtensson. *Gradient Methods for Large-Scale and Distributed Linear Quadratic Control*. PhD thesis, Lund University, 2012.
- D Subbaram Naidu and Anthony J Calise. Singular perturbations and time scales in guidance and control of aerospace systems: A survey. *Journal of Guidance, Control, and Dynamics*, 24(6):1057–1078, 2001.
- Kumpati S Narendra and Anuradha M Annaswamy. *Stable Adaptive Systems*. Courier Corporation, 2012.
- Angelina Nedić and Asuman Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009.
- Arkadij S Nemirovsky and David B Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley, New York, 1983.
- Yurii Nesterov. Primal-dual Subgradient Methods for Convex Problems. *Mathematical Programming*, 120(1 SPEC. ISS.):221–259, 2009.
- Jorge Nocedal and Stephen Wright. *Numerical Optimization*. Springer Science & Business Media, 2006.

- Ann Nowé, Peter Vrancx, and Yann-Michaël De Hauwere. Game theory and multi-agent reinforcement learning. In *Reinforcement Learning*, pages 441–470. Springer, 2012.
- Erfan Nozari, Yingbo Zhao, and Jorge Cortés. Network identification with latent nodes via autoregressive models. *IEEE Transactions on Control of Network Systems*, 5(2):722–736, 2017.
- Brian J Odelson, Murali R Rajamani, and James B Rawlings. A new autocovariance least-squares method for estimating noise covariances. *Automatica*, 42(2):303–308, 2006.
- Atsumi Ohara and Shun-Ichi Amari. Differential geometric structures of stable state feedback systems with dual connections. *IFAC Proceedings Volumes*, 25(21):176–179, 1992.
- Samet Oymak and Necmiye Ozay. Non-asymptotic identification of LTI systems from a single trajectory. In *American Control Conference*, pages 5655–5661, 2019.
- Kadri Ozcaldiran and Frank L. Lewis. On the regularizability of singular systems. *IEEE Transactions on Automatic Control*, 35(10):1156–1160, 1990.
- Christos H Papadimitriou and John Tsitsiklis. Intractable problems in control theory. *SIAM journal on control and optimization*, 24(4):639–654, 1986.
- Francesca Parise and Asuman Ozdaglar. A variational inequality framework for network games: Existence, uniqueness, convergence and sensitivity analysis. *Games and Economic Behavior*, (2010):1–43, 2019.
- Youngsuk Park, Ryan A. Rossi, Zheng Wen, Gang Wu, and Handong Zhao. Structured policy iteration for linear quadratic regulator. *arXiv preprint arXiv:2007.06202*, 2020.
- J.D. Pearson. On the duality between estimation and control. *SIAM Journal on Control*, 4(4):594–600, 1966.
- Joshua L Proctor, Steven L Brunton, and J Nathan Kutz. Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, 15(1):142–161, 2016.



- Tankred Rautert and Ekkehard W Sachs. Computational design of optimal output feedback controllers. *SIAM Journal on Optimization*, 7(3):837–852, 1997.
- Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.
- J. Ben Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 33(3):520–534, 1965.
- M. Rotkowitz and S. Lall. A characterization of convex problems in decentralized control\*. *IEEE Transactions on Automatic Control*, 51(2):274–286, 2006.
- Michael Rotkowitz and Sanjay Lall. A characterization of convex problems in decentralized control. *IEEE transactions on Automatic Control*, 50(12):1984–1996, 2005.
- Farzad Salehisadaghiani and Lacra Pavel. Distributed Nash equilibrium seeking in networked graphical games. *Automatica*, 87:17–24, 1 2018.
- Nils Sandell, Pravin Varaiya, Michael Athans, and Michael Safonov. Survey of decentralized control methods for large scale systems. *IEEE Transactions on Automatic Control*, 23(2):108–128, 1978.
- Tuhin Sarkar, Alexander Rakhlin, and Munther A Dahleh. Finite-time system identification for partially observed LTI systems of unknown order. *arXiv preprint arXiv:1902.01848*, 2019.
- Mahlagha Sedghi, George Atia, and Michael Georgiopoulos. Robust manifold learning via conformity pursuit. *IEEE Signal Processing Letters*, 26(3):425–429, March 2019.
- Mahlagha Sedghi, George Atia, and Michael Georgiopoulos. A multi-criteria approach for fast and robust representative selection from manifolds. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1, 2020.
- Denis Serre. *Matrices: Theory and Applications*. Springer Science and Media, 2nd edition, 2010.

- Miel Sharf and Daniel Zelazo. Network identification: A passivity and network optimization approach. In *IEEE Conference on Decision and Control*, pages 2107–2113. IEEE, 2018.
- Robert H Shumway and David S Stoffer. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3(4):253–264, 1982.
- Marwaan Simaan and Jose B Cruz. On the Stackelberg strategy in nonzero-sum games. *Journal of Optimization Theory and Applications*, 11(5):533–555, 1973.
- Jerome D. Simon and Sanjoy K Mitter. A theory of modal control. *Information and Control*, 13(4):316–353, 1968.
- Sigurd Skogestad, Kjetil Havre, and Truls Larsson. Control limitations for unstable plants. *IFAC Proceedings*, 35(1):485–490, 2002.
- Steven T Smith. Optimization techniques on Riemannian manifolds. *Fields Institute Communications*, 3(3):113–135, 1994.
- R Padma Sree and M Chidambaram. *Control of unstable systems*. Alpha Science Int’l Ltd., 2006.
- Vivek Srivastava, James Neel, Allen B MacKenzie, Rekha Menon, Luiz A DaSilva, James Edward Hicks, Jeffrey H Reed, and Robert P Gilles. Using game theory to analyze wireless ad hoc networks. *IEEE Communications Surveys & Tutorials*, 7(4):46–56, 2005.
- Alan Wilbor Starr and Yu-Chi Ho. Nonzero-sum differential games. *Journal of Optimization Theory and Applications*, 3(3):184–206, 1969.
- Gunter Stein. Respect the unstable. *IEEE Control Systems Magazine*, 23(4):12–25, 2003.
- Dusan M Stipanovic, Gökhan Inalhan, Rodney Teo, and Claire J Tomlin. Decentralized overlapping control of a formation of unmanned aerial vehicles. *Automatica*, 40(8):1285–1296, 2004.

- Yvonne R. Stürz, Annika Eichler, and Roy S. Smith. Distributed control design for heterogeneous interconnected systems. *IEEE Transactions on Automatic Control*, 66(11):5112–5127, 2021.
- Koji Tajima. Estimation of steady-state Kalman filter gain. *IEEE Transactions on Automatic Control*, 23(5):944–945, 1978.
- Shahriar Talebi and Mehran Mesbahi. Policy optimization over submanifolds for constrained feedback synthesis. *IEEE Transactions on Automatic Control (to appear)*, *arXiv preprint arXiv:2201.11157*, 2022.
- Shahriar Talebi, Marwan A. Simaan, and Zhihua Qu. Cooperative, non-cooperative and greedy pursuers strategies in multi-player pursuit-evasion games. In *2017 IEEE Conference on Control Technology and Applications (CCTA)*, pages 2049–2056, Aug 2017.
- Shahriar Talebi, Siavash Alemzadeh, Lillian J Ratliff, and Mehran Mesbahi. Distributed learning in network games: A dual averaging approach. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 5544–5549. IEEE, 2019.
- Shahriar Talebi, Siavash Alemzadeh, Niyousha Rahimi, and Mehran Mesbahi. Online regulation of unstable LTI systems from a single trajectory. *arXiv preprint arXiv:2006.00125*, 2020.
- Shahriar Talebi, Siavash Alemzadeh, and Mehran Mesbahi. Distributed model-free policy iteration for networks of homogeneous systems. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 6970–6975, 2021a.
- Shahriar Talebi, Siavash Alemzadeh, Niyousha Rahimi, and Mehran Mesbahi. On regularizability and its application to online control of unstable LTI systems. *IEEE Transactions on Automatic Control*, 67(12):6413–6428, 2021b.
- Shahriar Talebi, Amirhossein Taghvaei, and Mehran Mesbahi. Data-driven optimal filtering for linear systems with unknown noise covariances. *arXiv preprint arXiv:2305.17836*, 2023.

- Yujie Tang, Yang Zheng, and Na Li. Analysis of the optimization landscape of linear quadratic gaussian (LQG) control. In *Proceedings of the 3rd Conference on Learning for Dynamics and Control*, volume 144, pages 599–610. PMLR, June 2021.
- Tatiana Tatarenko, Wei Shi, and Angelia Nedic. Accelerated gradient play algorithm for distributed nash equilibrium seeking. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 3561–3566. IEEE, 12 2018.
- Hannu T Toivonen. A globally convergent algorithm for the optimal constant output feedback problem. *International Journal of Control*, 41(6):1589–1599, 1985.
- Hannu T. Toivonen and Pertti M. Mäkilä. Newton’s method for solving parametric linear quadratic control problems. *International Journal of Control*, 46(3):897–911, 1987.
- Joel A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.
- Anastasios Tsiamis and George J Pappas. Finite sample analysis of stochastic system identification. *arXiv preprint arXiv:1903.09122*, 2019.
- Anastasios Tsiamis and George J. Pappas. Online learning of the kalman filter with logarithmic regret. *IEEE Transactions on Automatic Control*, 68(5):2774–2789, 2023.
- Anastasios Tsiamis, Ingvar Ziemann, Nikolai Matni, and George J Pappas. Statistical learning theory for control: A finite sample perspective. *arXiv preprint arXiv:2209.05423*, 2022.
- Stephen Tu and Benjamin Recht. The gap between model-based and model-free methods on the linear quadratic regulator: An asymptotic viewpoint. In *Proceedings of the 32nd Conference on Learning Theory*, volume 99, pages 3036–3083. PMLR, 25–28 Jun 2019.
- Jack Umenberger, Max Simchowitz, Juan Perdomo, Kaiqing Zhang, and Russ Tedrake. Globally convergent policy search for output estimation. In S. Koyejo, S. Mohamed, A. Agar-

- wal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 22778–22790. Curran Associates, Inc., 2022.
- Henk J. Van Waarde, Jaap Eising, Harry L. Trentelman, and M. Kanat Camlibel. Data informativity: a new perspective on data-driven analysis and control. *IEEE Transactions on Automatic Control*, pages 1–1, 2020.
- Henk J Van Waarde, Jaap Eising, Harry L Trentelman, and M Kanat Camlibel. Data informativity: a new perspective on data-driven analysis and control. *IEEE Transactions on Automatic Control*, 2020.
- Rene Vidal, Omid Shakernia, H Jin Kim, David Hyunchul Shim, and Shankar Sastry. Probabilistic pursuit-evasion games: Theory, implementation, and experimental evaluation. *IEEE Transactions on Robotics and Automation*, 18(5):662–669, 2002.
- Draguna Vrabie, O Pastravanu, Murad Abu-Khalaf, and Frank L Lewis. Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica*, 45(2):477–484, 2009.
- Andrew Wagenmaker and Kevin Jamieson. Active learning for identification of linear dynamical systems. In *Conference on Learning Theory*, pages 3487–3582. PMLR, 2020.
- Shih Ho Wang and EJ Davison. On the stabilization of decentralized control systems. *IEEE Transactions on Automatic Control*, 18(5):473–478, 1973.
- Wei Wang, Fangfang Zhang, and Chunyan Han. Distributed linear quadratic regulator control for discrete-time multi-agent systems. *IET Control Theory & Applications*, 11(14):2279–2287, 2017.
- Jan C. Willems, Paolo Rapisarda, Ivan Markovsky, and Bart L.M. De Moor. A note on persistency of excitation. *Systems & Control Letters*, 54(4):325–329, 2005.
- Max A Woodbury. Inverting modified matrices. *Memorandum report*, 42(106):336, 1950.

- Lin Xiao. Dual averaging methods for regularized stochastic learning and online optimization. *The Journal of Machine Learning Research*, 11(1):2543–2596, 2010.
- Jie Xiong. *An Introduction to Stochastic Filtering Theory*, volume 18. OUP Oxford, 2008.
- Yue Yu, Shahriar Talebi, Henk J van Waarde, Ufuk Topcu, Mehran Mesbahi, and Behçet Açıkmeşe. On controllability and persistency of excitation in data-driven control: Extensions of willems’ fundamental lemma. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 6485–6490. IEEE, 2021.
- Lingyi Zhang, David Sidoti, Adam Bienkowski, Krishna R Pattipati, Yaakov Bar-Shalom, and David L Kleinman. On the identification of noise covariances and adaptive Kalman filtering: A new look at a 50 year-old problem. *IEEE Access*, 8:59362–59388, 2020.
- Feiran Zhao, Keyou You, and Tamer Başar. Global convergence of policy gradient primal-dual methods for risk-constrained LQRs. *IEEE Transactions on Automatic Control*, 2023.
- Kemin Zhou, John C Doyle, and Keith Glover. *Robust and Optimal Control*. Prentice-Hall, Inc., 1996.

## VITA

Shahriar Talebi is a researcher specializing in control theory, game theory, differential geometry, and learning for control. He is currently pursuing a Ph.D. degree in Aeronautics and Astronautics under the supervision of Prof. Mehran Mesbahi. In 2014, he obtained his B.Sc. degree in Electrical Engineering from Sharif University of Technology, Tehran, Iran, where he focused on control theory. Building upon this foundation, he pursued an M.Sc. degree in Electrical Engineering at the University of Central Florida (UCF), Orlando, FL, graduating in 2017. During this time, his studies, supervised by Prof. Marwan Simaan, centered around game theory and its applications in (non-)cooperative dynamical systems. In Spring 2023, Shahriar further enriched his academic background by earning an M.S. degree in Mathematics from the University of Washington (UW), Seattle, WA, with a specialization in differential geometry. This additional expertise, supervised by Prof. John M. Lee, enhanced his ability to tackle complex problems that require a deep understanding of geometric structures and mathematical principles.

His research interests span learning for control and estimation, game theory, variational analysis, data-driven control, and networked dynamical systems. By developing innovative techniques and methodologies, Shahriar aims to analyze distributed systems in both cooperative and non-cooperative environments, advancing our understanding of these complex systems and facilitating practical applications across diverse fields.

During his academic journey, Shahriar has been recognized for his outstanding achievements. In 2022, he was honored with the Excellence in Teaching Award at UW, a prestigious recognition that acknowledges his unwavering dedication and exceptional performance in classroom teaching. In addition, he has received several esteemed scholarships, including the William E. Boeing Endowed Fellowship, the Paul A. Carlstedt Endowment, and the Latvian

Arctic Pilot–A. Vagners Memorial Scholarship during his time at UW in 2018-19. Prior to that, in 2017, he was awarded the Frank Hubbard Engineering Scholarship at UCF.