Geometry of Feedback Control and Learning

Jingjing Bu

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

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Mehran Mesbahi, Chair

Maryam Fazel, Chair

Lillian J Ratliff

Program Authorized to Offer Degree:
Electrical Engineering

University of Washington

Abstract

Geometry of Feedback Control and Learning

Jingjing Bu

Co-Chairs of the Supervisory Committee:
Professor Mehran Mesbahi
Department of Aeronautics and Astronautics

Professor Maryam Fazel
Department of Electrical & Computer Engineering

In this thesis, we shall study optimal control problems, e.g. linear-quadratic-regulator (LQR), least squares stationary optimal control, linear quadratic (LQ) dynamic games, through the lens of first-order algorithms. The developed theories on these topics are largely derived from model-based dynamic programming. Recently there is a surge of interest in constructing optimal control strategies directly, viewing control synthesis by policy gradient based algorithms. Adopting such a point of view has been partially inspired by the success of learning algorithms, such as Reinforcement Learning (RL), where using principles of Dynamic Programming (DP), one can devise real-time model-free methods for both continuous-time and discrete-time LQR. The direct policy update approach offers advantages in terms of scalability, model-free implementations and richer parameterizations (e.g., structured controller design).

We first study the topological and metrical properties of the set of stabilizing feedback controls. The problem is of interest as this set is the natural domain of the cost functions for optimal problems. We present a complete account of the set-theoretic properties for both single-input-single-out (SISO) and multiple-input-multiple-output (MIMO) systems. We particularly prove an upper bound of number of path-connected components in SISO

systems. An algorithm on how to identify the connected components is proposed as well.

We next move on LQR optimal control. We characterize several analytical properties (smoothness, coerciveness, quadratic growth) that are crucial in the analysis of gradient-based algorithms. We then examine three types of well-posed flows for LQR: gradient flow, natural gradient flow and the quasi-Newton flow. The coercive property suggests that these flows admit unique solutions while gradient dominated property indicates that the corresponding Lyapunov functionals decay at an exponential rate; quadratic growth on the other hand guarantees that the trajectories of these flows are exponentially stable in the sense of Lyapunov. We then discuss the forward Euler discretization of these flows, realized as gradient descent, natural gradient descent and quasi-Newton iteration. We present stepsize criteria for gradient descent and natural gradient descent, guaranteeing that both algorithms converge linearly to the global optima. An optimal stepsize for the quasi-Newton iteration is also proposed, guaranteeing a Q-quadratic convergence rate–and in the meantime–recovering the Hewer algorithm.

We then consider the least squares stationary optimal control, i.e., LQR with indefinite state and input cost matrices. Such a setup has important applications in control design with conflicting objectives, such as linear quadratic dynamic games. We show the global convergence of gradient, natural gradient and quasi-Newton policies for this class of indefinite least squares problems.

Lastly, we study LQ dynamic games, which is closely related to $\mathcal{H}_\infty$ optimal control. We propose projection-free sequential algorithms for linear-quadratic dynamics games. These policy gradient based algorithms are akin to Stackelberg leadership model and can be extended to model-free settings. We show that if the "leader" performs natural gradient descent/ascent, then the proposed algorithm has a global sublinear convergence to the Nash equilibrium. Moreover, if the leader adopts a quasi-Newton policy, the algorithm enjoys a Q-quadratic convergence. Along the way, we examine and clarify the intricacies of adopting sequential policy updates for LQ games, namely, issues pertaining to stabilization, indefinite cost structure, and circumventing projection steps.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

Firstly, the author wishes to express sincere sincere gratitude to my advisor Prof. Mehran Mesbahi for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this proposal. I could not have imagined having a better advisor and mentor for my Ph.D study. The author also would like thank Prof. Maryam Fazel for generous support during my PhD. The insightful discussions have shaped my research directions in so many ways.

DEDICATION

Chapter 1

INTRODUCTION

Linear quadraitic optimal control has been one of the cornerstones of modern control theory [78] since Kalman's seminal work in 1960s [39]. This setting was later extended beyond positive semidefinite cost structure by Willems [78]. LQR, as well as least squares stationary optimal control, is formulated around an optimization problem for determining a sequence of (control) inputs to a linear system in order to minimize a given (integral) quadratic cost over an infinite horizon.[1] From the theoretical point of view, a fundamental property of linear quadratic optimal control synthesis is that the resulting optimal input is in the form of a state feedback; as such, it can be represented as a constant feedback gain on the state of the system [1,39]. The state feedback gain that "solves" the infinite-horizon LQR problem, in turn, can be obtained by solving the algebraic Riccati equation (ARE). That is, in the traditional approach to LQR design, the state feedback gain is revealed after obtaining the "certificate" or "cost-to-go" for the underlying optimal control problem.[2] Historically, a large number of works have studied the solution of ARE, including approaches based on iterative algorithms [31], algebraic solution methods [45], and semidefinite programming [2].

Linear-quadratic (LQ) dynamic and differential games exemplify situations where two players influence an underlying linear dynamics in order to respectively, minimize and maximize a given quadratic cost on the state and the control over an infinite time-horizon.[3] The LQ game setup has a rich history in system and control theory due to its wide range of applications [6, 22, 82]. In particular, $\mathcal{H}_\infty$ optimal control can be interpreted as a two

[1] We shall not delve into the finite-horizon LQR in this paper.

[2] The analogy here would be solving the dual, followed by the recovery of the primal solution.

[3] We will adopt the convention of referring to the continuous time scenario as differential games. Moreover, in this paper, we focus on infinite horizon LQ games without a discount factor.

player zero-sum LQ game [3]. As such, LQ games are generally approached via *generalized algebraic Riccati equation* (GARE) derived from optimal control theory [72]. Adopting a solution approach based on the Riccati equation, in the meantime, has broadly influenced the "data-driven" approaches for solving the generic LQ problem and its extensions. For instance, in the value-iteration for reinforcement learning (RL)—e.g., Q learning—one aims to first estimate the cost-to-go at a given time instance and through this estimate, update the state feedback gain.

Recently, there is a surge of interest in constructing optimal control strategies directly, viewing control synthesis by policy gradient based algorithms.[4] Adopting such a point of view has been partially inspired by the success of learning algorithms, such as Reinforcement Learning (RL), where using principles of Dynamic Programming (DP), one can devise real-time model-free methods for both continuous-time and discrete-time LQR [10, 19, 36, 46, 48, 51, 52] (noting that LQR problem can be formulated as an RL problem [8, 52]). However, policy iterations are inherently prohibitive since the cost function has an infinite horizon, is undiscounted and unbounded per stage [8]. The RL perspective for linear quadratic optimal control and LQ dynamic games, particularly direct policy updates, has the merits in terms of extension to model-free setting by means of stochastic (zeroth-order) optimization [20, 57], computational scalability and rich parametrization of the feedback policy. In the meantime, LQR and LQ games have an important role in theoretical RL and multi-agent RL by serving as a benchmark for demonstrating global convergence of policy gradient based algorithms.

The thesis aims to understand first-order algorithms in various optimal control problems. We try to provide a rigorous foundation for optimal control problems when viewed as a direct policy optimization. That is, we first investigate the proper way to define cost function and issues arising in the domain and analytical properties of cost functions. We then study the convergence properties of gradient-base algorithms.

The manuscript is structured as follows: in Chapter 2, we present the set-theoretic prop-

[4] One might as well extrapolate that these methods provide a streamline recipe for learning optimal feedback gains in real-time.

erties stabilizing feedback controls for LTI SISO system; The LQR problem for discrete-time LTI system is treated in Chapter 3. We next study the least squares stationary optimal control problem, i.e., LQR with indefinite cost, in Chapter 4. In Chapter 5, we investigate linear quadratic dynamic games. We conclude the thesis in Chapter 6.

Chapter 2

ON TOPOLOGICAL PROPERTIES OF THE SET OF STABILIZING FEEDBACK GAINS

2.1 Introduction

[1] In classical control, topological and metrical properties of stabilizing feedback gains are of paramount importance for the stability analysis and stabilization of LTI systems [18, 30, 60, 76]. Recently, such properties have received renewed interest in system literature as they have direct implications for adopting learning algorithms for control design. This is particularly the case in the so-called direct policy algorithms, where it is of interest to directly adjust the control gain—without an explicit system identification step—say, using a gradient step. The design objectives in these scenarios are typically functions over direct policies–often desired to be stabilizing feedback gains. For example, in reinforcement learning, the policy gradient updates the feedback iteratively to get desired optimal controller. In this case, the cost functions are defined on the set of stabilizing controllers (assuming $+\infty$ elsewhere).[2] As such, understanding the topological and metrical properties of this set provides valuable insights in designing learning algorithms for dynamic systems. In the meantime, such insights can also reveal fundamental shortcomings in certain optimization algorithms. For example, if the set of stabilizing feedback gains has several path-connected components, the solutions of gradient-type learning algorithms will be highly dependent on the initialization process. It is thus surprising that despite the long historical interest in characterizing the set of stabilizing feedback gains, research works on its set-theoretic and topological properties are rather limited. This is potentially due to significantly more interest in characterizing the set

---

[1]The content of this chapter is published in [14].

[2]In this chapter, we will use "feedback controllers" interchangeably with static feedback gains as "dynamic" controllers are not considered.

of “certificates" for stabilizing controllers, e.g., in terms of linear matrix inequalities.

Of particular relevance to our work in directly characterizing stabilizing feedback gains is that of Ohara and colleagues [60], who examined its differential geometric structure for multiple-input-multiple-output (MIMO) systems. In [63] and [23], an elegant geometric approach has been adopted to parametrize the set of stabilizing feedback gains; in particular, it has been shown that for continuous and discrete single-input-single-output (SISO) and dyadic systems the corresponding sets can be bounded via two and three hyperplanes. Furthermore, the work of Ober [59] has shown that the set of stable SISO systems of order n have $n+1$ connected components in the Euclidean topology while the set of stable MIMO systems is connected. The work reported in [27] focuses on the connectedness of this set for both SISO and MIMO systems. We note that both works [60] and [27] examine continuous-time systems.

This chapter discusses the topological, metrical, and geometric properties of the set of stabilizing controllers for both continuous and discrete-time LTI systems. We show that the set of stabilizing state-feedback gains for a continuous SISO system is regular open, unbounded, in general nonconvex, and path-connected in the Euclidean topology. In the meantime, the set of stabilizing output-feedback controllers is shown to be open but not connected in general, and can be bounded or unbounded. In recent works, based on the implicit assumption that stable and unstable intervals of the feedback gain interlace, it has been stated that the set of stabilizing output feedback controllers for SISO systems can have at most n (in [27]) and $\lceil \frac{n}{2} \rceil$ (in [26]) connected components. If this assumption does not hold, however, the line of reasoning reported in [26, 27] lead to the *upper bounds* of $2n$ and n, respectively.[3] In this work, we prove a tight bound of $\lceil \frac{n}{2} \rceil$ for continuous as well as discrete time LTI systems; all of our results are constructive (they lead to algorithms for characterizing these sets) and rely on basic topology and analytic theory of polynomials [54,

[3]In discussions with authors of [26], it has been pointed out that a perturbation type argument can help address this issue; the approach adopted in this work is direct and leads to a constructive algorithm for characterizing the stabilizing and unstabilizing intervals.

64].[4] The separate treatment for continuous and discrete time systems is warranted; in fact, in contrast to the folklore expectation of unified properties for continuous and discrete time systems, there are counterexamples to show that the analogies between the two are far from complete [37, 79]. The distinct difference between continuous and discrete LTI systems might be due to the fact that the generalized bilinear transform has poles and is thus not continuous [34, 56]. Therefore, generalizing the proposed topological properties of the set of stabilizing feedback gains from continuous LTI systems to discrete ones is not straightforward. Nevertheless, in this chapter we show that the set of stabilizing state feedback gains for discrete-time LTI SISO systems enjoys some of the topological properties as its continuous counterpart, i.e., open and path-connected in Euclidean topology and nonconvex if states are greater than two. But in contrast to the continuous case, the set of stabilizing state feedback gains is *bounded*. For output feedback SISO systems, the corresponding set of stabilizing gains is open, bounded and in general nonconvex, but is no longer path-connected. Accordingly, we prove that the set can have at most $\lceil \frac{n}{2} \rceil$ path-connected components, which is a tight bound supported by simulation results. The present work also proposes an algorithm for determining the intervals of stabilizing feedback gains for general continuous and discrete LTI systems. This algorithm also computes the number of unstable roots in each unstable interval.

The chapter is organized as follows: in §5.2, we introduce the notations and the preliminary background; §2.3 and §2.4 are devoted to set-theoretic properties of Hurwitz and Schur stabilizing feedback gains, respectively, followed by numerical examples. Our results are intermingled with observations and remarks that further provide insights into some of the geometric and topological intricacies of feedback stabilization.

[4]Thus, the emphasis on the SISO case; some of these results have been extended to MIMO case in [12].

2.2 Notation and Preliminaries

We denote by $\mathbb{M}_n(\mathbb{R})$ the set of $n \times n$ real matrices and $\mathbb{GL}_n(\mathbb{R})$ as its subset of invertible matrices. The determinant of a matrix A is denoted $\mathbf{det}(A)$; $\mathbb{R}^n$ and $\mathbb{C}^n$ denote the n-dimensional real and complex Euclidean spaces with $n = 1$ identified with real and complex numbers. $\mathcal{A}^c$ denotes the complement of the set $\mathcal{A}$. A subset $\mathcal{A}$ of $\mathcal{T}$ is called regular open if $\mathcal{A}$ is equal to the interior of its closure. For a vector $v \in \mathbb{R}^n$, we use v_j to denote the jth entry of v, where $v = (v_1, \dots, v_n)^\top$. We denote the open unit disk of $\mathbb{C}$ by $\mathbb{D} = \{\lambda \in \mathbb{C} : |\lambda| < 1\}$ and the left-half plane by $\mathbb{H}_- = \{\lambda \in \mathbb{C} : \mathbf{Re}(\lambda) < 0\}$;[5] $\mathbb{H}_-^n$ will be the n-dimensional version of $\mathbb{H}_-$. The notation $|\lambda|$ denotes the modulus of the complex number $\lambda \in \mathbb{C}$ and $\bar{\lambda}$ denotes its complex conjugate. We use $\mathbb{C}[\lambda]$ and $\mathbb{R}[\lambda]$ to denote polynomials with complex and real coefficients, respectively, where λ is the corresponding indeterminant of the polynomial. For a polynomial p over $\mathbb{C}$ or $\mathbb{R}$, we use p' to denote its derivative with respect to the indeterminant, unless noted otherwise. A monic polynomial is a univariate polynomial with leading coefficient 1; as such, it has the form $\lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_0$. By Fundamental Theorem of Algebra, a monic polynomial $p(\lambda) = \lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_0 \in \mathbb{C}[\lambda]$ has n roots (or zeros) counting multiplicities; we let $\mathcal{Z}_p$ to denote this set (each zero is repeated according to its multiplicity). Mind that $\mathcal{Z}_p$ is not a well-defined object in $\mathbb{C}^n$ as we do not impose a natural ordering amongst the roots. Thus, if $\mathcal{Z}_p = \{\lambda_1, \dots, \lambda_n\}$ with each $\lambda_j \in \mathbb{C}$, $\sigma\mathcal{Z}_p = \{\lambda_{\sigma(1)}, \dots, \lambda_{\sigma(n)}\}$ denotes the same set of roots for every permutation σ in the permutation group S_n. Hence $\mathcal{Z}_p$ is more naturally viewed as an element of the quotient space $\mathbb{C}^n/S_n$, where the underlying equivalence relation $u \sim v$ is via,

$$u = (u_1, \dots, u_n)^\top = (v_{\sigma(1)}, \dots, v_{\sigma(n)})^\top,$$

for some $\sigma \in S_n$; endow this quotient space $\mathbb{C}^n/S_n$ with a quotient topology induced by the canonical projection $\pi : \mathbb{C}^n \to \mathbb{C}^n/S_n$.

The following result will subsequently be used in our analysis.

[5] **Re** and **Im** refer to the real and imaginary parts of the complex number; when applied to a set of complex numbers, these operations are naturally extended to each element of that set.

Theorem 2.2.1. *([28]) There is a homeomorphism* $h\colon \mathbb{C}^n \to \mathbb{C}^n/S_n$*, mapping the coefficients of a monic complex polynomial to its zeros.*

In this manuscript, we are concerned with "real" linear time-invariant (LTI) systems, i.e., systems with real parameters and feedback gains. Hence the roots of the corresponding characteristic polynomial $p(x) \in \mathbb{R}[x]$ will be invariant under complex conjugation, namely if $\mathcal{Z}_p = \{z_1, \dots, z_n\} \in \mathbb{C}^n/S_n$, then[6]

$$\bar{\mathcal{Z}}_p = \{\bar{z}_1, \dots, \bar{z}_n\} = \mathcal{Z}_p;$$

denote by $\mathbb{C}^n_*$ as the set of vectors in $\mathbb{C}^n$ that are invariant under entry-wise conjugation.

The relation between the coefficients and roots for real polynomials follows from Theorem 2.2.1, as the restriction of a homeomorphism is again a homeomorphism.

Corollary 2.2.1.1. *Suppose that* $p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0 \in \mathbb{R}[\lambda]$ *is a real polynomial. Then there is a homeomorphism* $\hat{h} : \mathbb{R}^n \to \mathbb{C}^n_*/S_n$*, mapping coefficients of* $p(\lambda)$ *to its roots.*

Consider now the continuous LTI SISO system,

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + b\mathbf{u}(t), \quad \mathbf{y}(t) = c^\top \mathbf{x}(t), \tag{2.1}$$

and its discrete time variant,

$$\mathbf{x}(k+1) = A\mathbf{x}(k) + b\mathbf{u}(k), \quad \mathbf{y}(k) = c^\top \mathbf{x}(k), \tag{2.2}$$

where $A \in \mathbb{M}_n(\mathbb{R})$ and $b, c \in \mathbb{R}^n$; such systems will be abbreviated in terms of the triplet $(A, b, c^\top)$. We say that the system $(A, b, c^\top)$ is controllable and observable if it satisfies the Kalman Rank Condition [80], namely, $\mathbf{rank}([b, Ab, \dots, A^{n-1}b]) = n$ and $\mathbf{rank}([c^\top, c^\top A, \dots, c^\top A^{n-1}]^\top) = n$. For synthesizing state feedback gains for SISO systems with dynamics of order n, we are interested in identifying $k \in \mathbb{R}^n$ to synthesize the control signal $\mathbf{u}(t) = k^\top \mathbf{x}(t)$; if output-feedback is of interest, the feedback gain is a scalar $k \in \mathbb{R}$ and $\mathbf{u}(t) = k\mathbf{y}(t) = kc^\top \mathbf{x}(t)$. For a

[6] Note the equivalence relation on $\mathbb{C}^n/S_n$.

controllable and observable triplet $(A, b, c^\top)$, we denote the set of *Hurwitz stabilizing output feedback gains* as,[7]

$$\mathcal{H} = \{k \in \mathbb{R} : \max\{\mathbf{Re}(\mathcal{Z}_{p(\lambda, A-kbc^\top)})\} < 0\}, \tag{2.3}$$

and the set of *Schur stabilizing output-feedback gains* by

$$\mathcal{S} = \{k \in \mathbb{R} : \max\{|\mathcal{Z}_{p(\lambda, A-kbc^\top)}|\} < 1\}, \tag{2.4}$$

where $p(\lambda, A-kbc^\top) = \mathbf{det}(\lambda I - A + kbc^\top)$ denotes the characteristic polynomial of the closed-loop system with feedback gain $k \in \mathbb{R}$. Naturally, we could have defined the sets $\mathcal{H}$ and $\mathcal{S}$ in terms of the eigenvalues of $A - kbc^\top$; thus $\max\{|\mathcal{Z}_{p(\lambda,A)}|\}$ is simply the spectral radius of the matrix A, that we denote by $\rho(A)$.

When we examine state feedback with the controllable system parameters (A, b), the sets $\mathcal{H}_\mathbf{x}$ and $\mathcal{S}_\mathbf{x}$ are defined as,

$$\mathcal{H}_\mathbf{x} = \{k \in \mathbb{R}^n : \max\{\mathbf{Re}(\mathcal{Z}_{p(lambda, A-bk^\top)})\} < 0\}, \tag{2.5}$$

$$\mathcal{S}_\mathbf{x} = \{k \in \mathbb{R}^n : \max\{|\mathcal{Z}_{p(\lambda, A-bk^\top)}|\} < 1\} = \{k \in \mathbb{R}^n : \rho(A - bk^\top) < 1\}, \tag{2.6}$$

where $p(\lambda, A - bk^\top) = \mathbf{det}(\lambda I - A + bk^\top)$ and we have used the subscript $\mathbf{x}$ to denote state feedback.[8]

We now observe a relation between $\mathcal{H}$ and $\mathcal{H}_\mathbf{x}$; a similar relation holds between $\mathcal{S}$ and $\mathcal{S}_\mathbf{x}$. This relation will be used in our subsequent analysis.

Observation 2.2.2. *For a controllable and observable system* $(A, b, c^\top)$*,*

$$\mathcal{H} = \{k \in \mathbb{R} : (kc) \cap \mathcal{H}_\mathbf{x} \neq \varnothing\}.$$

Proof. We only need to observe that $k \in \mathcal{H}$ is equivalent to having $kc \in \mathcal{H}_\mathbf{x}$. $\square$

---

[7]The max operation on a set identifies the maximum element of that set.

[8]That is, the state "$\mathbf{x}$" is available for feedback.

Denote by $(A^\flat, b^\flat)$ the controllable canonical form [38] of the pair (A, b); $\mathcal{H}_{\mathbf{x}}^\flat, \mathcal{S}_{\mathbf{x}}^\flat$ then denote the corresponding set of Hurwitz/Schur stabilizing state feedback gains [38]. We now observe that the sets $\mathcal{H}_{\mathbf{x}}^\flat$ and $\mathcal{S}_{\mathbf{x}}^\flat$ are related to $\mathcal{H}_{\mathbf{x}}$ and $\mathcal{S}_{\mathbf{x}}$ through a change of coordinates.

Observation 2.2.3. *Let (A, b) be a controllable pair and $(A^\flat, b^\flat)$ be its corresponding controllable canonical form. Then $\mathcal{S}_{\mathbf{x}} = T^\top \mathcal{S}_{\mathbf{x}}^\flat := \{T^\top k : k \in \mathcal{S}_{\mathbf{x}}^\flat\}$ and $\mathcal{H}_{\mathbf{x}} = T^\top \mathcal{H}_{\mathbf{x}}^\flat := \{T^\top k : k \in \mathcal{H}_{\mathbf{x}}^\flat\}$, where $T \in \mathbb{GL}_n(\mathbb{R})$ and $(A^\flat, b^\flat) = (TAT^{-1}, Tb)$.*

Proof. We only prove the relation between $\mathcal{S}_{\mathbf{x}}^\flat$ and $\mathcal{S}_{\mathbf{x}}$; the proof for $\mathcal{H}_{\mathbf{x}}^\flat$ and $\mathcal{H}_{\mathbf{x}}$ follows analogously. When $k \in \mathcal{S}_{\mathbf{x}}^\flat$,

$$\rho(A^\flat - b^\flat k^\top) = \rho(T(A - bk^\top T)T^{-1}) = \rho(A - bk^\top T) < 1;$$

hence, $T^\top k \in \mathcal{S}_{\mathbf{x}}$ and $T^\top \mathcal{S}_{\mathbf{x}}^\flat \subseteq \mathcal{S}_{\mathbf{x}}$. The set inclusion in the other direction follows analogously; as such, $T^\top \mathcal{S}_{\mathbf{x}}^\flat = \mathcal{S}_{\mathbf{x}}$. $\square$

Remark. *Since $k \mapsto T^\top k$ is a diffeomorphism on $\mathbb{R}^n$, topological properties of the set of stabilizing feedback gains, such as connectedness, can be studied for the controllable canonical form instead. Given Observation 2.2.3, one we may assume, without loss of generality, that (A, b) is in the controllable canonical form in order to characterize the stabilizing output feedback gains. In fact, since $(A^\flat, b^\flat) = (TAT^{-1}, Tb)$, the set of stabilizing gains for $(TAT^{-1}, Tb, c^\top T^{-1})$ will coincide with $\mathcal{H}$ and $\mathcal{S}$ for the triple $(A, b, c^\top)$. This observation is useful as the characteristic polynomial of the closed-loop system for $(A^\flat, b^\flat, c^\top)$ admits a rather simple form.*

2.3 Properties of Hurwitz Stabilizing Feedback Gains

Consider again the continuous LTI SISO system (2.1) in relation to the sets $\mathcal{H}$ (2.3) and $\mathcal{H}_{\mathbf{x}}$ (2.5). The diffeomorphism between the set of stabilizing feedback gains for a system and its controllable canonical form (Observation 2.2.3) allows us to prove a number of topological and metrical properties for $\mathcal{H}$ and $\mathcal{H}_{\mathbf{x}}$. In fact, Observation 2.2.3 leads to the following properties through the application of theory of polynomials: (a) $\mathcal{H}$ is open in the Euclidean topology for both state-feedback and output-feedback systems, (b) $\mathcal{H}_{\mathbf{x}}$ is unbounded but $\mathcal{H}$ can be either bounded or unbounded, (c) the sets $\mathcal{H}$ and $\mathcal{H}_{\mathbf{x}}$ are both convex when the corresponding system has order two, (d) $\mathcal{H}_{\mathbf{x}}$ is connected and $\mathcal{H}$ can have at most $\lceil \frac{n}{2} \rceil$ connected components. We now provide the proofs for these observations.

Lemma 2.3.1. *The set $\mathcal{H}$ is open in $\mathbb{R}$ and $\mathcal{H}_{\mathbf{x}}$ is open in $\mathbb{R}^n$.*

Proof. By Observation 2.2.3, without loss of generality, we shall assume that the system $(A, b, c^\top)$ is in the controllable canonical form. Let $a = (a_0, \ldots, a_{n-1})$ be the last row of A and $c = (c_0, \ldots, c_{n-1})^\top$. For any $k \in \mathbb{R}$, the characteristic polynomial of this system is given by,

$$p(\lambda, k) = \lambda^n + (a_{n-1} - kc_{n-1})\lambda^{n-1} + \cdots + (a_0 - kc_0).$$

We note that for a fixed $k \in \mathbb{R}$, the map $\tilde{\upsilon} : \mathbb{C}^n_*/S_n \to \mathbb{R}$ given by

$$\mathcal{Z}_{p(\lambda,k)} \mapsto \max\{\mathbf{Re}(\mathcal{Z}_{p(\lambda,k)})\},$$

is continuous since $\upsilon := \max \circ \mathbf{Re} : \mathbb{C}^n_* \to \mathbb{R}$ is continuous and there is a unique continuous map $\tilde{\upsilon} : \mathbb{C}^n/S_n \to \mathbb{R}$ such that $\upsilon = \tilde{\upsilon} \circ \pi$:

$$\begin{tikzcd} \mathbb{C}^n \arrow[d, "\pi"] \arrow[dr, "\upsilon"] & \\ \mathbb{C}^n_*/S_n \arrow[r, "\tilde{\upsilon}"] & \mathbb{R}; \end{tikzcd}$$

this follows from the properties of the quotient topology (see Theorem 3.73 in [47]). Thus the map,

$$g : k \mapsto \mathcal{Z}_{p(\lambda,k)} \mapsto \max\{\mathbf{Re}(\mathcal{Z}_{p(\lambda,k)})\},$$

is continuous as it is a composition of continuous maps. Thus as the pre-image of the open interval $(-\infty, 0)$ under the continuous map g, the set $\mathcal{H}$ is an open subset of $\mathbb{R}$ and as such,

$\mathcal{H}$ is a union of disjoint open intervals.[9] Following a similar argument, the map $g_{\mathbf{x}} : \mathbb{R}^n \to \mathbb{R}$ defined via the composition,

$$g_{\mathbf{x}} : k \mapsto \mathcal{Z}_{p(\lambda,k)} \mapsto \max\{\mathbf{Re}(\mathcal{Z}_{p(\lambda,k)})\},$$

is continuous and hence $\mathcal{H}_{\mathbf{x}}$ is open in $\mathbb{R}^n$. □

We shall point out another favorable property of $\mathcal{H}_{\mathbf{x}}$, namely that it is *regular* open. In other words, the closure of the set of Hurwitz stabilizing controllers is the set of marginally stabilizing controllers and $\mathcal{H}_{\mathbf{x}}$ is precisely the interior of the set of marginally stabilizing controllers.

Lemma 2.3.2. *Let*

$$\mathcal{B}_{\mathbf{x}} = \{k \in \mathbb{R}^n : \max\{\mathbf{Re}(\mathcal{Z}_{p(\lambda,A-bk^\top)})\} = 0\}.$$

Then $\mathcal{B}_{\mathbf{x}}$ is the boundary of $\mathcal{H}_{\mathbf{x}}$ (that is, $\partial\mathcal{H}_{\mathbf{x}}$) and the closure of $\mathcal{H}$ is,

$$\bar{\mathcal{H}}_{\mathbf{x}} = \mathcal{H}_{\mathbf{x}} \cup \mathcal{B}_{\mathbf{x}} = \{k \in \mathbb{R}^n : \max\{\mathbf{Re}(\mathcal{Z}_{p(\lambda,A-bk^\top)})\} \leq 0\}. \tag{2.7}$$

Proof. Note that for any $k \in \mathcal{B}_{\mathbf{x}}$, we can perturb the marginally stable eigenvalues to be stable, i.e., having negative real parts. This means that there exists a sequence $\{k_n\} \subseteq \mathcal{H}_{\mathbf{x}}$ such that $k_n \to k$. On the other hand, we may as well perturb the marginally stable eigenvalues to become unstable, i.e., having positive real parts, suggesting that there is a sequence $\{k'_n\} \subseteq \mathcal{H}_{\mathbf{x}}^c$ such that $k'_n \to k$. Hence $\partial H_{\mathbf{x}} = \mathcal{B}_{\mathbf{x}}$ and (2.7) follows. □

It is also immediate to deduce that the interior of $\bar{\mathcal{H}}_{\mathbf{x}}$, namely $\left(\bar{\mathcal{H}}_{\mathbf{x}}\right)^\circ$, is $\mathcal{H}_{\mathbf{x}}$; thus $\mathcal{H}_{\mathbf{x}}$ is regular open.

Let us now examine the boundedness of the sets $\mathcal{H}$ and $\mathcal{H}_{\mathbf{x}}$.

Observation 2.3.3. *The set $\mathcal{H}_{\mathbf{x}}$ is unbounded.*

[9]That is, it can always be represented as such.

Proof. By Observation 2.2.3, it suffices to assume that the pair (A, b) is in controllable canonical form. For any $k = (k_0, \dots, k_n)^\top \in \mathbb{R}^n$, the characteristic polynomial of the corresponding closed-loop system is,

$$p(\lambda, k) = \lambda^n + (a_{n-1} - k_{n-1})\lambda^{n-1} + \cdots + (a_0 - k_0). \tag{2.8}$$

By Pole-shifting theorem [65], for every n-tuple $(-M, -M, \dots, -M)$, with $M \in \mathbb{R}$, there is some $k(M) \in \mathbb{R}^n$ such that $k(M) \in \mathcal{H}_{\mathbf{x}}$ and the zeros of $p(\lambda, A - bk(M)^\top)$ are exactly $(-M, \dots, -M)$.[10] But $a_0 - k_0(M) = (-M)^n$ by Vieta's formula [77].[11] Hence $\mathcal{H}_{\mathbf{x}}$ is not bounded. □

For an output feedback system, the set $\mathcal{H}$ can either be bounded or unbounded depending on the properties of the system $(A, b, c^\top)$; this is demonstrated in the following example.

Example 1. *Let the triplet $(A^\flat, b^\flat, c^\top)$, in controllable canonical form, be controllable.*[12] *Let $a = (a_0, a_1, \dots, a_{n-1})$ be the last row of $A^\flat$. Then,*

a. *If for some $i, j \in \{0, \dots, n-1\}$, $c_i > 0$ and $c_j < 0$, then $\mathcal{H}$ is bounded.*

b. *Suppose that $n = 4$ and the entries of c are positive. If $c_3c_2c_1 < c_3^2c_0$ then $\mathcal{H}$ is unbounded; when $c_3c_2c_1 > c_3^2c_0$, $\mathcal{H}$ is bounded.*

The assertions in this example are consequences of the Routh-Hurwitz Criterion. If k is stabilizing, then

$$a_{n-1} - kc_{n-1} > 0, \dots, a_0 - kc_0 > 0.$$

If there exists a k that satisfies the above inequalities, and for some i, j, $c_i > 0$ and $c_j < 0$, then k satisfies,

$$\frac{a_j}{c_j} < k < \frac{a_i}{c_i}.$$

[10]The notation $k(M)$ underscores the dependence of k on M.

[11]The subscript indexing of k_i is with reference to (2.8).

[12]Of course, the controllability of the triplet is independent of c.

Hence, $\mathcal{H}$ must be bounded. For part (b), the Routh-Hurwitz Criterion states that,

$$a_3 - kc_3 > 0, a_2 - kc_2 > 0, a_1 - kc_1 > 0, a_0 - kc_0 > 0,$$
$$(a_3 - kc_3)(a_2 - kc_2)(a_1 - kc_1) > (a_1 - kc_1)^2 + (a_3 - kc_3)^2(a_0 - kc_0).$$

We note that for sufficiently negative k, the last inequality holds since $c_1c_2c_3 > c_3^2c_0$; hence $\mathcal{H}$ is not bounded. On the other hand, if $c_1c_2c_3 < c_3^2c_0$, then k must be bounded from below; thus $\mathcal{H}$ is bounded.

We further make an observation on a necessary condition for the unboundedness of the set $\mathcal{H}$; see also [26].

Observation 2.3.4. *If $(A, b, c^\top)$ is controllable and observable, a necessary condition for $\mathcal{H}$ to be unbounded is that the nonzero entries of c have the same sign. Moreover, if $\mathcal{H}$ is unbounded, then $\mathcal{H}$ must only include one of the two intervals: $(-\infty, M)$ and (M', ∞) for some $M, M' \in \mathbb{R}$.*

Proof. If a monic polynomial is Hurwitz stable, all of its coefficients are positive (Theorem 2.4 in [80]). Hence, we have $a_j - kc_j > 0$ for every j. If the nonzero entries of c do not have the same sign, k should be bounded. Since the system is observable, $c \neq 0$. Thereby, either $k < M$ or $k > M'$, for some $M, M' \in \mathbb{R}$. $\square$

We now make a few observations on the convexity of the sets $\mathcal{H}$ and $\mathcal{H}_{\mathbf{x}}$; needless to say, these observations have direct algorithmic implications. It is known that a convex combination of stable polynomials is not necessarily convex [9]. However, a convex combination of stable monic polynomials with degree 2 is convex; this follows from the Routh-Hurwitz criterion.

Observation 2.3.5. *For the state feedback system (A, b), the set $\mathcal{H}_{\mathbf{x}}$ is convex if $n = 2$.*

Proof. It suffices to show this for $(A^\flat, b^\flat)$. Let $k = (k_0, k_1)^\top$ and $k' = (k_0', k_1')^\top$ be two stabilizing feedback gains. Then the characteristic polynomial of the corresponding closed-loop systems are,

$$p_1(\lambda) = \lambda^2 + (a_1 - k_1)\lambda + (a_0 - k_0),$$
$$p_2(\lambda) = \lambda^2 + (a_1 - k_1')\lambda + (a_0 - k_0').$$

Note that p_1 and p_2 are stable if and only if the coefficients are positive.[13] Hence, if $\hat{k} = (1-\delta)k + \delta k'$, for $\delta \in (0,1)$, then $p_{\hat{k}}(x) = (1-\delta)p_1(x) + \delta p_2(x)$ and $p_{\hat{k}}$ is stable by the positivity of its coefficients. □

Observation 2.3.6. *For the output feedback system $(A, b, c^\top)$ with $n = 2$, the set $\mathcal{H}$ is convex.*

Proof. Recall that by Observation 2.2.2,

$$\mathcal{H} = \{k \in \mathbb{R} : (kc) \cap \mathcal{H}_{\mathbf{x}} \neq \varnothing\}.$$

Noting that $\mathbf{span}(c) \cap H_{\mathbf{x}}$ is a convex subset of $\mathbb{R}^2$; by Observation 2.3.5, the proof follows. □

We can also use the Routh-Hurwitz stability criteria to show the nonconvexity of the set of stabilizing feedback gains when $n > 2$. For example, let $n = 3$ and consider the controllable canonical form $(A^\flat, b^\flat)$ with the last row of $A^\flat$ set to 0. Then the stabilizing feedback gain is parametrized by three parameters $k = (k_1, k_2, k_3)$ and the characteristic polynomial of $A - bk^\top$ is

$$p(\lambda, A - bk^\top) = \lambda^3 + k_3\lambda^2 + k_2\lambda + k_1.$$

The following example essentially shows that convex combinations of stable polynomials are not necessarily stable.

Example 2. *Consider the system,*

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \; b = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

[13] For convexity analysis, this is in fact the key property for systems of order 2.

For $k_1 = (-24, -5, -5)^\intercal$ and $k_2 = (-0.9, -1, -1)^\intercal$, the characteristic polynomials of the corresponding closed-loop systems are given by $p_1(\lambda) = \lambda^3 + 5\lambda^2 + 5\lambda + 24$ and $p_2(\lambda) = \lambda^3 + \lambda^2 + \lambda + 0.9$. Both polynomials are stable and as such, $k_1, k_2 \in \mathcal{H}_\mathbf{x}$; however, $k' = 0.5k_1 + 0.5k_2$ yields an unstable characteristic polynomial $p(\lambda) = \lambda^3 + 3\lambda^2 + 3\lambda + 12.45$.

2.3.1 Connectedness properties of $\mathcal{H}_\mathbf{x}$ and $\mathcal{H}$

We will now delve into connectedness of the sets $\mathcal{H}$ and $\mathcal{H}_\mathbf{x}$, requiring more delicate arguments as compared with their boundedness and convexity properties. For the state feedback system,

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + b\mathbf{u}(t),$$

by Corollary 2.2.1.1, $\mathcal{H}_\mathbf{x}$ is connected in $\mathbb{R}^n$. We now show that this set is in fact contractible, i.e., it can be continuously deformed to a point [47].

Lemma 2.3.7. *When (A, b) is controllable, the set of stabilizing feedback controllers $\mathcal{H}_\mathbf{x} \subseteq \mathbb{R}^n$ is connected and contractible.*

Proof. Let $\{\mathbb{H}^n_-\}_*$ denote the set of n-tuples $v \in \mathbb{H}^n_-$ invariant under (entry-wise) complex conjugation. We note that $\{\mathbb{H}^n_-\}_*/S_n$ is connected in $\mathbb{C}^n/S_n$ by noting that every $v \in \{\mathbb{H}^n_-\}_*/S_n$ is path-connected to $(-1, \dots, -1)^\intercal$. This path in fact defines a homotopy between the identity and the constant map $(-1, \dots, -1)$. Mind that

$$\mathcal{H} = (a_{n-1}, \dots, a_0)^\intercal - \hat{\sigma}^{-1}(\{\mathbb{H}^n_-\}_*/S_n),$$

i.e., an affine translation in $\mathbb{R}^n$. By Corollary 2.2.1.1, it now follows that $\mathcal{H}_\mathbf{x}$ is connected and contractible. $\square$

Remark. *Note that even though the set $\mathcal{H}_\mathbf{x}$ is contractible to a point, it is not necessarily star-convex. This is due to nonlinearity of the Vieta's map $\hat{\sigma}^{-1}$.*

Connected Components of $\mathcal{H}$

We now develop bounds on the number of connected components of $\mathcal{H}$ as it is not necessary connected. Lemma 2.3.9 provides a bound of n and Lemma 2.3.12 will tighten the bound to

$\lceil \frac{n}{2} \rceil$. We have chosen to present the two lemmas in sequence, since the proof of Lemma 2.3.9 is straightforward but tightening the bound to $\lceil \frac{n}{2} \rceil$ requires more delicate analysis.

Let us start our analysis by recalling the smooth dependence of simple roots of a polynomial on its coefficients, subsequently used in Lemma 2.3.12.

Lemma 2.3.8. *Let $a_j : \mathcal{I} \to \mathbb{R}$ be C^∞ functions for $j = \{1, \ldots, n\}$, where $\mathcal{I} \subseteq \mathbb{R}$ is an open interval. If $t_0 \in \mathcal{I}$ and λ_0 is a simple root of the polynomial $f(\lambda) = a_n(t_0)\lambda^n + a_{n-1}(t_0)\lambda^{n-1} + \cdots + a_0(t_0) \in \mathbb{R}[\lambda]$ with $t_0 \in \mathcal{I}$, then there exists a C^∞ function $\eta : \mathcal{J} \to \mathbb{C}$ over an open interval $\mathcal{J} \subseteq \mathcal{I}$ such that $t_0 \in \mathcal{J}$, $\eta(t_0) = \lambda_0$ and $\eta(t)$ is a zero of*

$$f(\lambda, t) = a_n(t)\lambda^n + a_{n-1}(t)\lambda^{n-1} + \cdots + a_0(t),$$

for every $t \in \mathcal{J}$.

Proof. This follows from Implicit Function Theorem [44]. First note that $f(\lambda, t)$ is C^∞ in both λ and t; we note that at (λ_0, t_0), $f'(\lambda_0, t_0) \neq 0$ since λ_0 is a simple zero. □

We now prove an upper bound of n on the number of connected components of $\mathcal{H}$.

Lemma 2.3.9. *If $\mathcal{H} \neq \varnothing$, it has at most n connected components.*

Proof. Note that in the SISO case, $\mathcal{H}$ is a subset of $\mathbb{R}$; furthermore, it suffices to show that $(A^\flat, b^\flat, c^\top)$ has at most n connected components. Recall that for $k \in \mathbb{R}$, the characteristic polynomial of a closed-loop system $A^\flat - kb^\flat c^\top$ is given by

$$p(\lambda, k) = \lambda^n + (a_{n-1} - kc_{n-1})\lambda^{n-1} + \cdots + a_0 - kc_0,$$

where $a := (a_0, a_1, \ldots, a_{n-1})$ is the last row of $A^\flat$ and c_j's are components of c. Let $\zeta : \mathbb{R}^n \to \mathcal{P}_n(\lambda)$ denote the natural bijection, assigning coefficients to monic polynomials. We denote by,

$$\Gamma = \{a \in \mathbb{R}^n : \zeta(a) \text{ has at least one zero on imaginary axis}\},$$

and $\ell(k) = a - kc$, i.e., a parametrized line in $\mathbb{R}^n$. Suppose that $\ell(k)$ intersects Γ for finitely many k's, listed in an increasing order $k_1, \dots, k_q$; this fact will be proved subsequently. Let $n_{p(\lambda,k)}(\mathbb{H}_-)$ denote the number of roots of $p(\lambda, k)$ in $\mathbb{H}_-$. Moreover, let $\gamma(r)$ be a counterclockwise oriented curve in $\mathbb{C}$ consisting the line segment $[-ir, ir]$ and the semicircle $S(r,\theta) = re^{i\theta}$, with $\theta \in [\pi/2, 3\pi/2]$. For each $k \in (k_j, k_{j+1})$, we define

$$m_r(k) = \frac{1}{2\pi i}\int_{\gamma(r)} \frac{p'(\lambda,k)}{p(\lambda,k)}\, d\lambda, \text{ and } m(k) = \lim_{r\to\infty} m_r(k).$$

Note that if $p(\lambda, k)$ does not vanish on $\gamma(r)$, by Cauchy's Argument Principle [67], $m_r(k)$ is the number of zeros of $p(\lambda, k)$ inside the curve $\gamma(r)$. However, since $p(\lambda, k)$ has at most n roots, the integral is well-defined except at finitely many r's. Hence $m(k)$ is well-defined and $m(k) = n_{p(\lambda,k)}(\mathbb{H}_-)$. In the meantime, the function $m(k)$ is continuous in k and integer-valued, and thereby, $m(k) = n_{p(\lambda,k)}$ is constant on each interval (k_j, k_{j+1}). That is, either $n_{p(\lambda,k)} = n$ or $n_{p(\lambda,k)} < n$, corresponding to either stabilizing or non-stabilizing gains, respectively. So by inspecting the number of intersections between $\ell(k)$ and Γ, one can derive an upper bound on the number of connected components of $\mathcal{H}$. Let

$$r(\lambda) = c_{n-1}\lambda^{n-1} + \cdots + c_0, \text{ and } s(\lambda) = a_{n-1}\lambda^{n-1} + \cdots + a_0. \tag{2.9}$$

Consider an intersection of $\ell(k)$ and Γ that is, when for some $k \in \mathbb{R}$, there exists $\lambda = i\beta$, $\beta \in \mathbb{R}$, for which $p(i\beta, k) = 0$. We first observe that $r(i\beta) \neq 0$ since otherwise $i\beta$ would be a root for every $k \in \mathbb{R}$. This on the other hand, implies that $\mathcal{H}$ is empty. Hence $p(i\beta, k) = 0$ implies that,

$$k = \frac{(i\beta)^n + a_{n-1}(i\beta)^{n-1} + \cdots + a_0}{c_{n-1}(i\beta)^{n-1} + \cdots + c_0}. \tag{2.10}$$

Since $k \in \mathbb{R}$, β must be a root of,

$$\phi(\beta) = \mathbf{Im}\Big(\big(\lambda^n + s(\lambda)\big)\, r(\bar{\lambda})|_{\lambda = i\beta}\Big) = 0.$$

We note that $\phi(\beta) \in \mathbb{R}[\beta]$ has degree at most $2n-1$. Thus it can be written as,

$$\phi(\beta) = i^{2n-2}\beta^{2n-1} + i^{2n-4}d_{2n-3}\beta^{2n-3} + \cdots + i^2 d_3\beta^3 + d_1\beta,$$

for some set of real coefficients $\{d_{2n-3}, d_{2n-5}, \dots, d_1\} \subseteq \mathbb{R}$, noting that all exponentials of i are either 1 or -1. Now let us set $\tau(\beta) = i^{2n-2}\beta^{2n-2} + i^{2n-4}d_{2n-3}\beta^{2n-4} + \cdots + i^2 d_3\beta^2 + d_1 \in \mathbb{R}[\beta]$; hence, $\phi(\beta) = \beta\tau(\beta)$. Now we note that $\tau(\beta)$ is an even polynomial in β (having only even degrees). Letting

$$\upsilon(\beta) = i^{2n-2}\beta^{n-1} + i^{2n-4}d_{2n-3}\beta^{n-2} + \cdots + i^2 d_3\beta + d_1, \tag{2.11}$$

we have $\tau(\beta) = \upsilon(\beta^2)$. Thereby, the roots of τ are the square roots of those of υ. This implies that τ has two real roots if υ has a positive real root. We further observe that if β_0 is a positive real root of $\upsilon(\beta)$, then $\sqrt{\beta_0}$ and $-\sqrt{\beta_0}$ are the real roots of τ.

Let us now consider the scenario that leads to greatest upper bound on the number of connected components of $\mathcal{H}$. This scenario corresponds to the situation where $\phi(\beta)$ has $2n-1$ real roots; let these roots be $\{0, \beta_1, -\beta_1, \dots, \beta_{n-1}, -\beta_{n-1}\}$, where $\beta_j \in \mathbb{R}_+$ for each j. These roots can be mapped to feedback gains via relation (2.10); in fact, β_j and $-\beta_j$ are mapped to the same k. Thus adding the k that corresponds to the 0 root via relation (2.10), we will have at most n values for k. These values on the other hand, divide the real line into $n+1$ intervals. But only one of the two unbounded intervals can be stabilizing by Observation 2.3.4; the upper bound of n on the number of connected components of $\mathcal{H}$ now follows. $\square$

Remark. *Note that having n connected components for $\mathcal{H}$ implies that we can have a situation where $\{k_1, \dots, k_n\}$ are marginally stabilizing gains and the open intervals (k_j, k_{j+1}) are stabilizing. Figure 2.1 demonstrates this phenomena for two adjacent intervals: there is a gain k_0 such that the closed-loop system is marginally stable but (k', k_0) and (k_0, k'') are both* *stabilizing for some* $k', k'' \in \mathbb{R}$.[14] *The system parameters*[15] *are*

$$A = \begin{pmatrix} -0.825 & -1.21 & -\frac{625919}{4800000} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \; b = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \; c = \begin{pmatrix} 1 \\ 7.5 \\ 12.5 \end{pmatrix}.$$

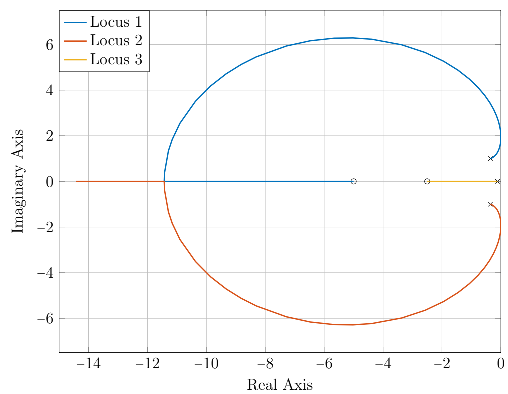

Figure 2.1: An example where the feedback gain k_0 leads to a marginally stable closed-loop system yet both intervals (k', k_0) and (k_0, k'') are stabilizing for some $k', k'' \in \mathbb{R}$.

We now proceed to show that the bound on the number of connected components of $\mathcal{H}$ can be tightened to $\lceil \frac{n}{2} \rceil$. The proof of this tighter bound follows a distinct line of reasoning, as necessitated by examples such as that shown in Figure 2.1. The result contains several technical details. Let us outline the idea behind the proof first: we denote by $\{0, \beta_1, -\beta_1, \dots, \beta_n, -\beta_n\}$ as the real roots of $\phi(\beta)$ and $\{k_1, \dots, k_n\}$ in an increasing order as the feedback gains acquired via relation (2.10) in Lemma 2.3.9:

a. We will first show that when two adjacent intervals (k_{j-1}, k_j) and (k_j, k_{j+1}) are stabilizing, then $p(\lambda, k_j) = \lambda^n + (a_{n-1} - k_j c_{n-1})\lambda^{n-1} + \dots + (a_0 - k_j c_0)$ would have the non-stable

[14]Note that Theorem 1 in [27] uses the same line of reasoning as the proof of Lemma 2.3.9 to arrive at the improved bound of $\lceil \frac{n}{2} \rceil$ for the number of connected components in $\mathcal{H}$, assuming that the intervals constructed above are stabilizing/non-stabilizing interlacing intervals. However, as Figure 2.1 depicts, this assumption is not valid in general. The $\lceil \frac{n}{2} \rceil$ bound is still valid, however, and can be obtained by utilizing the structure of the polynomial $\phi(\beta)$ and the relation between the feedback gain k and parameter β.

[15]These parameters are actually chosen carefully according to the analysis of Lemma 2.3.12.

$$\begin{tikzcd}
 & & & & \phi(\beta) \arrow[dllll] \arrow[dlll] \arrow[dl] \arrow[d] \arrow[drr] \arrow[drrr] \arrow[drrrr] & & & & \\
\beta_1 \arrow[d] & -\beta_1 \arrow[dl] & \cdots & \beta_j \arrow[d] & -\beta_j \arrow[dl] & \cdots & \beta_{n-1} \arrow[d] & -\beta_{n-1} \arrow[dl] & 0 \arrow[d] \\
k_1 \arrow[d] & & \cdots & k_j \arrow[d] & & \cdots & k_{n-1} \arrow[d] & & k_n \arrow[d] \\
p(\lambda, k_1) & & \cdots & p(\lambda, k_j) \arrow[dl] \arrow[dr] & & \cdots & p(\lambda, k_{n-1}) & & p(\lambda, k_n) \\
 & & \lambda_0 & \cdots & \lambda_{n-1} & & & &
\end{tikzcd}$$

Figure 2.2: The relation between roots of $\phi(\beta)$ and $p(\lambda, k)$, as the root locus intersects the imaginary axis for continuous time systems.

mode $\lambda_0 = i\beta$, i.e., the mode with zero real part, as a simple root of $p(\lambda, k_j) \in \mathbb{R}[\lambda]$. By Lemma 2.3.8, this would imply that we can find some C^∞ function $\eta : \mathcal{I} \to \mathbb{C}$ (with $\mathcal{I} \subset \mathbb{R}$ and $k_i \in \mathcal{I}$) such that $\eta(t)$ tracks the zero of $p(\lambda, k)$ locally with $\eta(k_j) = r$. Indeed, what we really need is that the curve of the root $i\beta$ is differentiable at $i\beta$.

b. If two adjacent intervals are both stabilizing, the curve $\eta(t)$ is tangent to the imaginary axis at t_0. We show that this observation leads to having $-\beta_j$ and β_j as multiple zeros of $\phi(\beta)$. Mind the subtlety here: λ_0 is the simple root of the polynomial $p(\lambda, k_j) \in \mathbb{R}[\lambda]$ in λ whereas $\pm\beta_j$ are multiple zeros of the polynomial $\phi(\beta) \in \mathbb{R}[\beta]$ in β (recall the expression for $\phi(\beta)$ in the proof of Lemma 2.3.9.). See Figure 2.2 for a demonstration of the relations.

c. Using the multiplicities of $\lambda_j{}'s$ as roots of $\phi(\beta)$ and a careful counting of the stabilizing/non-stabilizing intervals lead to the final result.

We shall state several propositions before we prove the main result. First we provide an asymptotic expansion of the zeros of $p(\lambda, k)$ with respect to k. Note that this is not simply a Taylor expansion around k, as a multiple root is not differentiable with respect to k. Recall the definitions of the polynomials $r(\lambda)$ and $s(\lambda)$ (2.9) used in the proof of Lemma 2.3.9.

Proposition 2.3.10. *Suppose for $k_0 \in \mathbb{R}$, λ_0 is a root of $p(\lambda, k_0) \in \mathbb{R}[\lambda]$ with multiplicity $m \in \mathbb{N}$ and $r(\lambda_0) \neq 0$. Then for k that is sufficiently close to k_0, $p(\lambda, k)$ will have m roots given by,*

$$\lambda_j = \lambda_0 + \left((k - k_j) \frac{r(\lambda_0)}{h(\lambda_0)} \right)^{1/m} \omega_j + o(|k - k_j|^{1/m}),$$

for $j \in \{1, \dots, m\}$, where ω_j's are the m^{th} roots of unity,[16] $h(\lambda) \in \mathbb{C}[\lambda]$ is such that $p(\lambda, k_0) = (\lambda - \lambda_0)^m h(\lambda)$, and $o(|k - k_j|^{1/m})$ signifies a function $f(k)$ for which $\lim_{k \to k_j} |f(k)|/|k - k_j|^{1/m} = 0$.[17]

Remark. *Before we prove the proposition, let us remark on how to interpret the exponent,*

$$\left((k - k_j) \frac{r(\lambda_0)}{h(\lambda_0)} \right)^{1/m}.$$

One may observe that $(k - k_j) r(\lambda_0)/h(\lambda_0)$ could be real or complex. In either case, we shall treat this exponent as a complex number with a positive magnitude and an angle in $[0, 2\pi)$. That is, we write,

$$(k - k_j) \frac{r(\lambda_0)}{h(\lambda_0)} = \left| (k - k_j) \frac{r(\lambda_0)}{h(\lambda_0)} \right| e^{i\theta},$$

where θ is the phase term. But consequently there would be m choices for $((k - k_j) r(\lambda_0)/h(\lambda_0))^{1/m}$, namely $|(k - k_j) r(\lambda_0)/h(\lambda_0)|^{1/m} e^{i(\theta + 2\pi l)/m}$ for $l = \{0, 1, \dots, m-1\}$. Note that these numbers differ multiplicatively from each other by $e^{i2\pi\nu/m}$, with $\nu \in \{0, \dots, m-1\}$; the expression above would not be affected by selecting any of these numbers. Moreover, this statement should be interpreted separately for $k \uparrow k_j$ and $k \downarrow k_j$ as the phase term would be different. The difference amounts to a rotation of λ_j's. This difference determines how the number of unstable roots changes when k increases from $k - \epsilon$ to $k + \epsilon$.

---

[16] That is, there are the zeros of $z^m - 1$.

[17] Recall root locus rules!

Proof. Since λ_0 is a root of multiplicity m, $p(\lambda, k_0) = (\lambda - \lambda_0)^m h(\lambda)$, where $h(\lambda) \in \mathbb{C}[\lambda]$ with $h(\lambda_0) \neq 0$. Hence, for $k \in \mathbb{R}$, we can write $p(\lambda, k)$ as

$$\begin{aligned} p(\lambda, k) &= \lambda^n + (a_{n-1} - kc_{n-1})\lambda^{n-1} + \cdots + a_0 - kc_0 \\ &= \lambda^n + (a_{n-1}\lambda^{n-1} - k_0 c_{n-1})\lambda^{n-1} + \cdots + a_0 - k_0 c_0 - (k - k_0)(c_{n-1}\lambda^{n-1} + \cdots + c_0) \\ &= p(\lambda, k_0) - (k - k_0) r(\lambda) \\ &= (\lambda - \lambda_0)^m h(\lambda) - (k - k_0) r(\lambda). \end{aligned}$$

Now suppose that $(\lambda - \lambda_0)^m h(\lambda) - (k - k_j) r(\lambda) = 0$. Putting $\hat{\lambda} = \lambda - \lambda_0$, $\hat{h}(\hat{\lambda}) = h(\hat{\lambda} + \lambda_0)/h(\lambda_0)$, $\hat{r}(\hat{\lambda}) = r(\hat{\lambda} + \lambda_0)/h(\lambda_0)$ and $t = k - k_j$, it suffices to show that the zeros of $\hat{\lambda}^m \hat{h}(\hat{\lambda}) - t\,\hat{r}(\hat{\lambda})$ are exactly,

$$\hat{\lambda}_j = \left(\frac{r(\lambda_0)}{h(\lambda_0)}\right)^{1/m} \omega_j + o(|t|^{1/m}), \text{ for } j \in \{1, \ldots, m\},$$

as $t \to 0$. Note that $\hat{h}(0) = 1$ and $\hat{h}(\hat{\lambda}) \in \mathbb{C}[\hat{\lambda}]$. Let $t > 0$ and observe that if z is a zero of

$$\psi_t(\lambda) \coloneqq \lambda^m \hat{h}(t^{1/m}\lambda) - \hat{r}(t^{1/m}\lambda),$$

then $t^{1/m} z$ would be a zero of $\lambda^m h(\lambda) - t\, r(\lambda)$. On a compact set $[0, T]$ ($T \in \mathbb{R}_+$), $\psi_t(\lambda) \to \psi_0(\lambda) = \lambda^M - r(0)$ uniformly. Let us denote the zeros of $\psi_0(\lambda) = \lambda^m - r(0)$ by

$$z_j = \big(r(0)\big)^{1/m} \omega_j, \text{ for } j \in \{1, 2, \ldots, m\},$$

where ω_j's are the mth roots of unity. Choose a sufficiently small $\varepsilon > 0$ such that the disks $B_{z_j}(\varepsilon)$ are disjoint. Since $\partial B_{z_j}(\varepsilon)$ is compact and $\psi_0(\lambda)$ does not vanish on $\partial B_{z_j}(\varepsilon)$, there is some $l > 0$ such that $|\psi_0(\lambda)| > l$. Since $\psi_t(\lambda) \to \psi_0(\lambda)$ uniformly on any compact subset of $\mathbb{C}$, there is some $t^* > 0$, such that $|\psi_t(\lambda) - \psi_0(\lambda)| < l$ for $t \in (-t^*, t^*)$ and $\lambda \in \bar{B}_{z_j}(\varepsilon)$. By Rouché's Theorem [67], there is exactly one zero for $\psi_t(\lambda)$ in each $B_{z_j}(\varepsilon)$ for $t \in (-t^*, t^*)$. As such, the zeros of $\psi_t(\lambda)$ are given by,

$$r(0)^{1/m}\omega_j + o(1).$$

It then follows that,

$$\lambda_j' = \left(t \frac{r(\lambda_0)}{h(\lambda_0)} \right)^{1/m} \omega_j + o(|t|^{1/m}), \text{ for } j = 1, \ldots, m;$$

the case where $t < 0$ can be treated similarly by considering $t\,h(\lambda) = (-t)(-h(\lambda))$. Hence,

$$\lambda_j = \lambda_0 + \left((k - k_j) \frac{r(\lambda_0)}{h(\lambda_0)} \right)^{1/m} \omega_j + o(|k - k_j|^{1/m}),$$

for $j = 1, \ldots, m$. $\square$

Before we proceed to the proof of the main result of this section, let us make an observation pertaining to the derivative of $f(\beta) \in \mathbb{R}[\beta]$ evaluated on the imaginary axis.

Proposition 2.3.11. *Consider* $f(\lambda) = \alpha_m \lambda^m + \cdots + \alpha_1 \lambda + \alpha_0 \in \mathbb{R}[\lambda]$ *and let* $\varphi(\beta) = \mathbf{Im}\left(f(\lambda)\big|_{\lambda = i\beta} \right) \in \mathbb{R}[\beta]$. *Then we have,*

$$\frac{d\varphi(\beta)}{d\beta} = \mathbf{Re}\left(\frac{df}{d\lambda}\bigg|_{\lambda = i\beta} \right).$$

Proof. When m is odd we have,

$$\varphi(\beta) = i^{m-1}\alpha_m \beta^m + i^{m-3}\alpha_{m-2}\beta^{m-2} + \cdots + i^2 \alpha_3 \beta^3 + \alpha_1 \beta.$$

Hence,

$$\varphi'(\beta) = m i^{m-1}\alpha_m \beta^{m-1} + (m-2) i^{m-3}\alpha_{m-2}\beta^{m-3} + \cdots + 3 i^2 \alpha_3 \beta^2 + \alpha_1.$$

But we also note that,

$$\begin{aligned}
\mathbf{Re}\left(\frac{df}{d\lambda}\bigg|_{\lambda = i\beta} \right) &= \mathbf{Re}\left(\left(m\alpha_m x^{m-1} + \cdots + \alpha_1 \right)|_{x = i\beta} \right) \\
&= \mathbf{Re}\left(m\alpha_m (i\beta)^{m-1} + \cdots + \alpha_1 \right) \\
&= m i^{m-1}\alpha_m \beta^{m-1} + (m-2) i^{m-3}\alpha_{m-2}\beta^{m-3} + \cdots + 3 i^2 \alpha_3 \beta^2 + \alpha_1 \\
&= \varphi'(\beta)
\end{aligned}$$

The case when m is even follows analogously. $\square$

We are now ready to prove the bound $\lceil \frac{n}{2} \rceil$ on the number of connected components of $\mathcal{H}$.

Lemma 2.3.12. *Let $\mathcal{H} \neq \varnothing$; then it has at most $\lceil \frac{n}{2} \rceil$ connected components.*

Proof. Consider two adjacent intervals (k_{j-1}, k_j) and (k_j, k_{j+1}) that are both stabilizing. Note that by assumption $p(\lambda_0, k_j) = 0$ for some pure imaginary λ_0. First, let us examine whether λ_0 can have multiplicity $m \geq 2$. By Proposition 2.3.10 for k sufficiently close to k_0, $p(\lambda, k)$ will have m roots given by,

$$\lambda_j = \lambda_0 + \left((k - k_j)\frac{r(\lambda_0)}{h(\lambda_0)}\right)^{1/m} \omega_j + o(|k - k_j|^{1/m}),$$

for $j = 1, \ldots, m$, where ω_j's are the mth roots of unity. When $m \geq 3$, as $k \to k_j$, m roots, from m equally spaced directions in the complex plane (see Figure 2.3), would tend to λ_0. In this

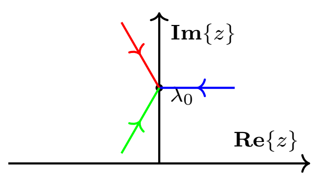

Figure 2.3: Roots coming from $m = 3$ equally spaced directions in the complex plane tend to λ_0

case, at least one direction would be in the right half plane, contradicting the assumption that both (k_{j-1}, k_j) and (k_j, k_{j+1}) are stabilizing. If $m = 2$ and $r(\lambda_0)/h(\lambda_0)$ is a complex number with a nontrivial imaginary part, i.e., the phase term is *not* a multiple of π, then the roots would tend to λ_0 from a nonvertical line; thereby one of intervals must be nonstabilizing. If $m = 2$ and $r(\lambda_0)/h(\lambda_0)$ is real, the roots would traverse a horizontal line passing through λ_0,[18] with one of the intervals as nonstabilizing. This is again a contradiction. Hence λ_0 must be simple. This on the other hand would imply the differentiability of the root with respect to k.

[18]Note that in this case either $k \uparrow k_j$ or $k \downarrow k_j$ would yield the quantity $(k - k_j)r(\lambda_0)/h(\lambda_0)$ in Proposition 2.3.10 positive, and the resulting expression would be a scalar multiple of the primitive 2^{nd} roots of unity, i.e., $\pm\alpha$ for some $\alpha \in \mathbb{R}$.

By the asymptotic formula above (with $m = 1$), the derivative of $\eta'(k_j) = r(\lambda_0)/h(\lambda_0)$; note that in this scenario $h(\lambda_0) = p'(\lambda_0, k_j)$. But the condition that both intervals are stabilizing implies that $\eta'(\lambda_0)$ is pure imaginary (the tangent of the curve should be the imaginary axis, see Figure 2.1); that is

$$\left.\frac{d\eta}{dk}\right|_{k_j} = i\gamma,$$

for some $\gamma \in \mathbb{R}$. This implies that if $\lambda_0 = i\beta$ ($\beta \in \mathbb{R}$) is a zero of $\phi(\beta)$; since $\eta'(\lambda_0)$ is pure imaginary, we would also have,

$$\mathbf{Re}\Big(\big(\lambda^n + s(\lambda) - k_j r(\lambda)\big)' r(\bar{\lambda})|_{\lambda=i\beta}\Big) = 0, \tag{2.12}$$

where $s(\lambda) = a_{n-1}\lambda^{n-1} + \cdots + a_0$. Putting $r(\lambda) = r_o(\lambda) + r_e(\lambda)$, where $r_o(\lambda)$ consists of the terms with odd degrees and $r_e(\lambda)$ consists terms of even degrees, we observe that $r(\bar{\lambda})|_{\lambda=i\beta} = r(-i\beta) = r_e(i\beta) - r_o(i\beta) = \big(r_e(\lambda) - r_o(\lambda)\big)|_{z=i\beta}$. Now by Proposition 2.3.11, we have

$$\begin{aligned}
\phi'(\beta) &= \mathbf{Re}\Big[\Big(\big(\lambda^n + s(\lambda)\big(r_e(\lambda) - r_o(\lambda)\big)\big)\Big)'|_{\lambda=i\beta}\Big] \\
&= \mathbf{Re}\Big\{\Big[\big(\lambda^n + s(\lambda)\big)'\big(r_e(\lambda) - r_o(\lambda)\big) \;+\; \big(\lambda^n + s(\lambda)\big)\big(r_e'(\lambda) - r_o'(\lambda)\big)\Big]|_{\lambda=i\beta}\Big\}.
\end{aligned}$$

Noting that $\big(\lambda^n + s(\lambda) - k\,r(\lambda)\big)|_{\lambda=\lambda_0} = 0$, we have

$$\phi'(\beta) = \mathbf{Re}\Big\{\Big[\big(\lambda^n + s(\lambda)\big)'\big(r_e(\lambda) - r_o(\lambda)\big) \;+\; \big(-k\,r(\lambda)\big)\big(r_e'(\lambda) - r_o'(\lambda)\big)\Big]|_{\lambda=i\beta}\Big\}.$$

Now $r_e'(\lambda)$ consists of terms with odd degrees and $r_o'(\lambda)$ of those with even degrees; this implies that,

$$\big(r_e'(\lambda) - r_o'(\lambda)\big)|_{\lambda=i\beta} = \big(-r_e'(\lambda) - r_o(\lambda)\big)|_{\lambda=-i\beta} = -r'(\bar{\lambda})|_{\lambda=i\beta},$$

minding that $r'(\bar{\lambda})$ refers to the derivative of $r'(\lambda)$ evaluated at $\bar{\lambda}$. Furthermore,

$$\mathbf{Re}\Big[\big(-kr'(\lambda)r(\bar{\lambda})\big)|_{\lambda=i\beta}\Big] = \mathbf{Re}\Big[\big(-kr'(\bar{\lambda})r(\lambda)\big)|_{\lambda=i\beta}\Big].$$

Combining all these observations, we conclude that $\phi'(\beta) = 0$.

Now let $\{0, u_1, -u_1, \ldots, u_\mu, -u_\mu, v_1, -v_1, \ldots, v_\nu, -v_\nu\}$ denote the roots of the polynomials $\phi(\beta)$, where u_i's and v_j's are nonnegative real numbers, and $v_1, -v_1, \ldots, v_\nu, -v_\nu$ are roots with multiplicity greater than 1. We must then have $2\mu + 2(2\nu) \leq 2n - 2$. These roots will be mapped to the gains k via the relation,

$$k = \left.\frac{q(\lambda)}{\lambda^n + r(\lambda)}\right|_{\lambda = i\beta}.$$

Now let $\Lambda = \{k_1, \ldots, k_{\mu'}\}$ be the set of distinct real gains corresponding to $\{u_1, -u_1, \ldots, u_\mu, -u_\mu\}$ and $\Pi = \{g_1, \ldots, g_{\nu'}\}$ be the set of distinct real gains corresponding to $\{v_1, -v_1, \ldots, v_\nu, -v_\nu\}$. Note that $\mu' \leq \mu$ and $\nu' \leq \nu$ since it is possible that multiple roots are mapped to the same gain; append the gain $k(0)$ corresponding to the zero root via (2.10) to Λ or Π if necessary. We must then have $\mu' + 2\nu' \leq n$. Now if $k_j \in \Lambda$, then one of the intervals (k_{j-1}, k_j) and (k_j, k_{j+1}) is not stabilizing. Let Υ be the collection of intervals that are not stabilizing. It follows then that if $k \in \Lambda$, k must be the end point of an interval in Υ. Note that only one of the unbounded intervals could be stabilizing; thereby,

$$\mu' \leq 1 + 2(|\Upsilon| - 1) = 2|\Upsilon| - 1.$$

Now, Lemma 2.3.9 implies that the set of feedback (real) gains k is divided into $\nu' + \mu' + 1$ intervals. As such, $\nu' + \mu' + 1 - |\Upsilon|$ is the number of stabilizing intervals, obtained by subtracting the number of non-stabilizing intervals from the total number of intervals. Hence,

$$\mu' + \nu' + 1 - |\Upsilon| \leq \mu' + \nu' + 1 - \frac{\mu' + 1}{2} = \nu' + \frac{\mu' + 1}{2} \leq \frac{2\nu' + \mu' + 1}{2} \leq \frac{n+1}{2} = \lceil \frac{n}{2} \rceil.$$

□

Remark. *We note that* $\lceil \frac{n}{2} \rceil$ *is a tight upper bound. Indeed, Figure 2.1 already indicates that there are two disjoint stabilizing intervals for a SISO system with* $n = 3$*. Figure 2.4 provides a more transparent view of this; the system parameters are,*

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -0.133 & -1.125 & -0.625 \end{pmatrix}, \ b = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \ c = \begin{pmatrix} 12.5 \\ 7.5 \\ 1 \end{pmatrix}.$$

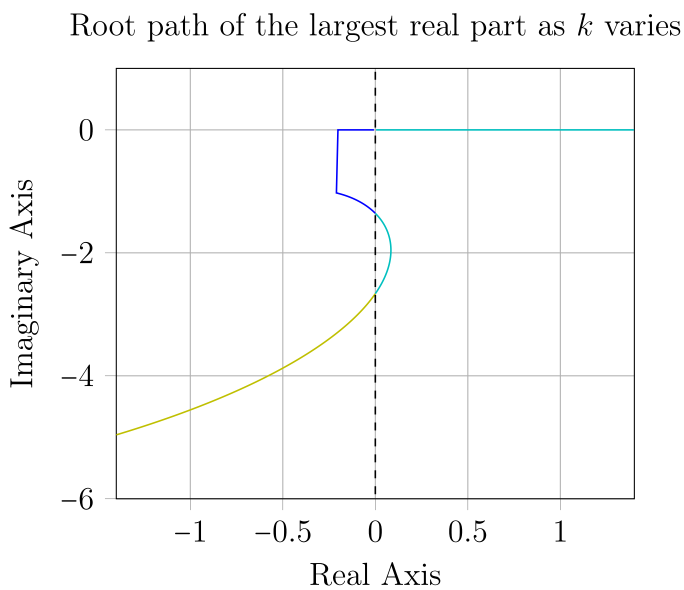

Figure 2.4: The figure depicts how the root with the largest real part varies with respect to the feedback gain k. The blue and yellow segments correspond to two stabilizing intervals.

We mention that although stabilizing and non-stabilizing intervals do not necessary interlace for all controllable and observable triplets $(A, b, c^\top)$, this property is indeed generic.[19] To this end, we first observe that the set $\mathcal{CO}$ of all controllable and observable triplets $(A, b, c^\top)$ is open in $\mathbb{M}_{n\times n}(\mathbb{R}) \times \mathbb{R}^n \times \mathbb{R}^n$.

Proposition 2.3.13. *The set $\mathcal{CO}$ is (Zariski) open.*

Proof. Putting $\mathbf{C}$ to denote the controllability matrix and $\mathbf{O}$ to denote the observability matrix. We observe $(\mathcal{CO})^c = \mathcal{C}^c \cup \mathcal{O}^c \cup \mathcal{C}^c\mathcal{O}^c$, where $\mathcal{C}^c$ denotes the collection of non-controllable but observable systems, $\mathcal{O}^c$ denotes the collection of non-observable but controllable systems, and $\mathcal{C}^c\mathcal{O}^c$ denotes the collection of non-controllable and non-observable systems. We note that $\mathcal{C}^c$ is the set where all the $n \times n$ minors of $\mathbf{C}$ vanish. Similarly for $\mathcal{O}^c$ and $\mathcal{C}^c\mathcal{O}^c$. Hence, $(\mathcal{CO})^c$ is an algebraic set. Consequently, $\mathcal{CO}$ is open. □

[19] A property is called generic when it holds on a nonempty Zariski open set.

We now observe the interlacing property is generic.

Lemma 2.3.14. *For a controllable and observable system $(A, b, c^\top)$, the property that the stabilizing and non-stabilizing intervals in $\mathcal{H}$ interlace is generic.*

Proof. First note that the only scenario for non-interlacing is to have two (or more) adjacent stabilizing intervals since if two intervals (k', k) and (k, k'') are both non-stabilizing, then in the counting process, we would only be considering (k', k'') as *one* non-stabilizing interval.[20] Further note that if $\mathcal{H} = \varnothing$, i.e., the whole real line is nonstabilizing, then the interlacing property holds trivially. Denote by $\mathcal{U} \subseteq \mathcal{CO}$ the subset for which the corresponding Hurwitz stabilizing set $\mathcal{H}$ does not have the interlacing property. As we have shown, for any $(A, b, c^\top) \in \mathcal{CO}$, if we have two adjacent stabilizing intervals, then the corresponding polynomial $\phi(\beta)$ must have a root with multiplicity greater than one. This means that the discriminant of $\phi(\beta)$, denoted by $\Delta(\phi(\beta))$, must vanish. This in turn is a polynomial in the entries of A, b, and $c^\top$. That is,

$$\mathcal{U} \subseteq \{(A, b, c^\top) \in \mathcal{CO} : \Delta(\phi(\beta)) = 0\} =: \mathcal{D}.$$

Note that $\mathcal{D}$ is an algebraic set. This suggests that the interlacing property holds on the nonempty Zariski open set $\mathcal{CO} \setminus \mathcal{D}$ and as such is generic. $\square$

An algorithm for characterizing the connected components of $\mathcal{H}$

Our analysis for deriving the bound $\lceil \frac{n}{2} \rceil$ for the number of connected components of $\mathcal{H}$ has direct algorithmic implications. We summarize the corresponding algorithm as follows:

---

[20] Here k, k', k'' are all critical points.

Algorithm 1: Identifying stabilizing intervals of $\mathcal{H}$

Find the real roots $\{\lambda_1, \dots, \lambda_l\}$ of (2.11).

Append $\{0\}$ to $\{\lambda_1, \dots, \lambda_l\}$ if necessary, we get $L = \{0, \lambda_1, \dots, \lambda_l\}$.

Map L to $\{k_1, \dots, k_{l'}\}$ (order this list in an increasing manner) by (2.10).

$k_{l'+1} \leftarrow +\infty$

Calculate the number of unstable roots for the interval $(-\infty, k_1)$ by computing roots of $p(\lambda, k_a)$, for an arbitrary $k_a < k_1$.

for $i = 1$ to l' **do**

  Check the multiplicity of λ_i that maps to k_i.

  Calculate the change in the number of unstable roots from $k_i - \epsilon$ to $k_i + \epsilon$ (Remark 2.3.1).

  Compute the number of unstable roots for the interval (k_i, k_{i+1}).

  if number of unstable roots for (k_i, k_{i+1}) is zero **then**

    The interval (k_i, k_{i+1}) is stabilizing.

  else

    The interval (k_i, k_{i+1}) is not stabilizing.

  end if

end for

The main computational cost of Algorithm 1 is finding the roots of a real polynomial to any desired accuracy. The specifics are beyond the scope of this chapter; see [43, 61] and references therein for the recent algorithmic developments in this direction. Algorithm 1 divides the set of feedback gains into separated intervals where the number of unstable roots of $p(\lambda, k)$ is invariant. At the boundaries separating these intervals, $p(\lambda, k)$ has at least an imaginary root for the corresponding feedback gain. The main idea behind Algorithm 1 is to analyze the root paths of $p(\lambda, k)$ at the boundaries of the separated intervals, using Remark 2.3.1. Consequently, Algorithm 1 keeps track of the number of unstable roots when k increases from $-\infty$ to $+\infty$. In this way, Algorithm 1 not only determines all stabilizing intervals, but also computes the number of unstable roots of $p(\lambda,k)$ for unstable intervals.

Let us demonstrate the progression of Algorithm 1 for the example in Remark 2.3.1. The characteristic polynomial of the closed-loop system for this example is then

$$p(\lambda,k) = \lambda^3 + 0.825\lambda^2 + 1.21\lambda + 0.3401666667 + k(\lambda^2 + 7.5\lambda + 12.5).$$

Furthermore,

$$\phi(\beta) = \lambda^5 - 7.5225\lambda^3 + 1.367899792\lambda,$$

with nonnegative roots $0, \frac{\sqrt{6018}}{40}$, where $\frac{\sqrt{6018}}{40}$ has multiplicity 2. The root $\beta_1 = 0$ is mapped to $k_1 = -\frac{625919}{6\times 10^7}$, and $\beta_2 = \frac{\sqrt{6018}}{40}$ to $k_2 = \frac{2041}{6000}$. We start from the unbounded interval (k_2, ∞): pick k and test the stability of the closed loop system; in this case, we conclude that (k_2,∞) is stabilizing. Since k_2 is acquired from a multiple root of $\phi(\beta)$, we examine the derivative $p'(\lambda,k_2)$; since it is pure imaginary, we conclude that the interval (k_1,k_2) is stabilizing. As such $(-\infty,k_1)$ is not stabilizing by Observation 2.3.4.

2.4 Properties of Schur stabilizing feedback gains

In this section, we study the properties of the set of static feedback gains (2.4) for discrete-time linear systems (2.2). Here is our first observation.

Lemma 2.4.1. *There is a homeomorphism $h : \mathcal{H}_{\mathbf{x}} \to \mathcal{S}_{\mathbf{x}}$.*

Proof. Without loss of generality, as we have done throughout this chapter, we assume that the pair (A,b) is in the controllable canonical form. Recall that the bilinear transform,

$$g : \lambda \mapsto \frac{\lambda+1}{\lambda-1}$$

is a diffeomorphism between the unit disk $\mathbb{D}$ and the open left-half plane $\mathbb{H}_-$ in $\mathbb{C}$. Clearly, $G \coloneqq (g,\dots,g) : \mathbb{D}^n \to \mathbb{H}_-^n$ defines a diffeomorphism between $\mathbb{D}^n \to \mathbb{H}_-^n$. Passing to the quotient space (modulo the action of the symmetric group), we have a diffeomorphism $\tilde{G} : \mathbb{D}^n/S_n \to \mathbb{H}_-^n/S_n$ given by $\tilde{G} \circ \pi = G$, where π is the canonical projection. Let ζ be the bijection between $\mathbb{R}^n$ and the set of monic nth degree polynomials, i.e., if $\alpha = (\alpha_0, \ldots, \alpha_{n-1}) \in \mathbb{R}^n$, then $\zeta(\alpha) = \lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_0$. Denote the sets,

$$\mathcal{E} = \{\alpha \in \mathbb{R}^n : \zeta(\alpha) \text{ has all roots in } \mathbb{D}\},$$
$$\mathcal{F} = \{\alpha \in \mathbb{R}^n : \zeta(\alpha) \text{ has all roots in } \mathbb{H}_-\}.$$

By Corollary 2.2.1.1, we have following commuting diagram:

$$\begin{tikzcd} \mathcal{E} \arrow[r, dashed, "\hat{\sigma}^{-1}\circ\tilde{G}\circ\sigma"] \arrow[d, "\hat{\sigma}"] & \mathcal{F} \arrow[d, "\hat{\sigma}"] \\ \{\mathbb{D}^n\}_*/S_n \arrow[r, "\tilde{G}"] & \{\mathbb{H}_-^n\}_*/S_n, \end{tikzcd}$$

where $\{\mathbb{D}^n\}_*/S_n \subseteq \mathbb{C}^n/S_n$ denotes the n-dimensional vector invariant under conjugation with entries inside the unit disk of $\mathbb{C}$; similarly for $\{\mathbb{H}_-^n\}_*/S_n$ (recall the notation in § 5.2).

Now, the map

$$\hat{\sigma}^{-1} \circ \tilde{G} \circ \hat{\sigma} : \mathcal{E} \to \mathcal{F},$$

defines a homeomorphism between $\mathcal{E}$ and $\mathcal{F}$. However, note that $\mathcal{S}_{\mathbf{x}} = a - \mathcal{E}$ and $\mathcal{H}_{\mathbf{x}} = a - \mathcal{F}$, where a is the last row of A. This completes the proof since a translation in $\mathbb{R}^n$ is a diffeomorphism. □

Remark. *The above result immediately implies that $\mathcal{S}_{\mathbf{x}}$ is open, connected and contractible since $\mathcal{H}_{\mathbf{x}}$ is. In our subsequent discussion, we will also outline more direct proofs for the above facts since they provide additional insights into the structure of the set of stabilizing feedback gains for discrete-time linear systems.*

In view of Lemma 2.4.1, one might be inclined to construct a similar homeomorphism between the sets $\mathcal{S}$ and $\mathcal{H}$. However, as it turns out, the technique adopted in Lemma 2.4.1 can not be generalized for this purpose. For example, when $k_0 \in \mathcal{S}$, it does follow that $k_0 c \in \mathcal{S}_{\mathbf{x}}$. However, under the homeomorphism constructed in Lemma 2.4.1, the image of $k_0 c$

is not necessarily a scalar multiple of c. For a concrete example, one may consider the triplet $(A, b, c^\top)$ given by

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ b = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \ c = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

We note $0 \in \mathcal{S}$ and $\mathcal{Z}_{p(\lambda, A-0bc^\top)} = \{0, \dots, 0\}$. Under the bilinear transform, the zeros will be mapped to $\{-1, \dots, -1\}$, which corresponds to characteristic polynomial $p(\lambda) = \lambda^4 - 4\lambda^3 + 6\lambda^2 - 4\lambda + 1$. However, no $k \in \mathcal{H}$ can yield a closed-loop system $A - kbc^\top$ with this characteristic polynomial.[21]

We now demonstrate that:

a. $\mathcal{S}$ and $\mathcal{S}_{\mathbf{x}}$ are both open in the Euclidean topology.

b. $\mathcal{S}$ and $\mathcal{S}_{\mathbf{x}}$ are both bounded.

c. $\mathcal{S}$ and $\mathcal{S}_{\mathbf{x}}$ are convex if the system has two states.

d. $\mathcal{S}_{\mathbf{x}}$ is connected and $\mathcal{S}$ has at most $\lceil \frac{n}{2} \rceil$ connected components.

Most of the proofs for the discrete time case have a similar flavor as their continuous counterparts. However, the proof for the upper bound on the number of connected components of $\mathcal{S}$ has a few distinct steps.

Lemma 2.4.2. *The set $\mathcal{S}$ is open in $\mathbb{R}$ and $\mathcal{S}_{\mathbf{x}}$ is open in $\mathbb{R}^n$.*

Proof. The proof proceeds similar to the proof of Lemma 2.3.1. We only need to observe that the composition map,

$$\upsilon := \max \circ |\cdot| : \mathbb{C}^n_* \to [0, \infty),$$

[21] Note that we are not claiming that such homeomorphism does not exist. The non-existence of such a homeomorphism requires a deeper understanding of these two topological subspaces. To the best of our knowledge such a homeomorphism has not been reported in the literature.

is continuous, even when adopted on the quotient space. That is, there is a unique continuous map $\tilde{\upsilon} : \mathbb{C}^n_*/S_n \to [0,\infty)$ such that $\upsilon = \tilde{\upsilon} \circ \pi$. Hence the map,

$$k \mapsto \mathcal{Z}_{p(\lambda, A-kbc^\top)} \mapsto \max\left(|\mathcal{Z}_{p(\lambda, A-kbc^\top)}|\right)$$

is continuous. The interval $[0,1)$ is open in the subspace topology of $[0,\infty)$ and $\mathcal{S}$ is the preimage of $[0,1)$ under the above map; thereby, $\mathcal{S}$ is open.

In order to show that $\mathcal{S}_\mathbf{x}$ is open in $\mathbb{R}^n$, we only need to observe that the map $F : \mathbb{R}^n \to [0,\infty)$, given by

$$k \mapsto \mathcal{Z}_{p(\lambda, A-bk^\top)} \mapsto \max\left(|\mathcal{Z}_{p(\lambda, A-bk^\top)}|\right),$$

is continuous and $\mathcal{S}_\mathbf{x} = F^{-1}([0,1))$. □

We now note that contrary to the continuous time case, the sets $\mathcal{S}$ and $\mathcal{S}_x$ are bounded.

Proposition 2.4.3. *The set $\mathcal{S}_\mathbf{x}$ is bounded in $\mathbb{R}^n$.*

Proof. It suffices to assume that (A,b) is in the controllable canonical form. For any $k \in \mathbb{R}^n$, the characteristic polynomial of the corresponding closed-loop system assumes the form,[22]

$$p(\lambda,k) = \lambda^n + (a_{n-1} - k_{n-1})\lambda^{n-1} + \cdots + (a_0 - k_0).$$

Let $\lambda_1, \ldots, \lambda_n$ denote the zeros of $p(\lambda,k)$. By Vieta's formula, the coefficients of $p(\lambda,k)$ are elementary symmetric functions of its roots $\lambda_1, \ldots, \lambda_n$. For every $k \in \mathcal{S}_\mathbf{x}$, $|\lambda_j| < 1$ for all j; as such, the coefficients of $p(\lambda,k)$, and by extension, $k_0, \ldots, k_{n-1}$, are bounded. □

Corollary 2.4.3.1. *For the controllable and observable output feedback system $(A,b,c^\top)$, the set $\mathcal{S}$ is bounded.*

Proof. We observe that $\mathcal{S} = \{k \in \mathbb{R} : kc \cap \mathcal{S}_\mathbf{x} \neq \varnothing\}$. □

In regards to convexity properties, it is known that the set $\mathcal{S}_\mathbf{x}$ is not convex in general [9, 25].

[22] The entries of k are consistent with the way the are indexed in the characteristic polynomial.

Observation 2.4.4. *The set $\mathcal{S}_{\mathbf{x}}$ is convex when $n = 2$.*

Proof. Without loss of generality, we assume that the pair (A, b) is in the controllable canonical form. Suppose that $k = (k_0, k_1)^\top$ and $e = (k_0', k_1')^\top$ are two stabilizing controllers. The characteristic polynomials of the corresponding closed-loop systems are then,

$$p_k(\lambda) = \lambda^2 + k_1\lambda + k_0, \text{ and } p_{k'}(\lambda) = \lambda^2 + k_1'\lambda + k_0'.$$

Note that by Vieta's formula, $k_0 < 1$ and $k_0' < 1$ since the zeros are inside the unit disk. For $\hat{k} = (1-\delta)k + \delta k'$ with $\delta \in [0,1]$, consider the corresponding characteristic equation $p_{\hat{k}}(\lambda) = (1-\delta)p_k(\lambda) + \delta p_{k'}(\lambda)$. If $p_{\hat{k}}$ has two conjugate zeros $z_1, \bar{z}_1$, then $|z_1| = |\bar{z}_1| < 1$ by Vieta's formula since $|z_1|^2 = (1-\delta)k_0 + \delta k_0' < 1$. On the other hand, if $p_{\hat{k}}$ has two real zeros, suppose that one of them is 1 or -1 (note that by Vieta's formula, the product of two zeros is strictly less than 1; hence the other zero is inside the open unit disk), i.e., $p_{\hat{k}}(1) = 0$ or $p_{\hat{k}}(-1) = 0$. Note that $p_k(1), p_k(-1), p_{\hat{k}}(1), p_{\hat{k}}(-1)$ are all positive since if p_k has two conjugate zeros, then $p_k(\lambda)$ is positive on the real line; if p_k has two real zeros, by the assumption that the zeros are in $(-1,1)$, $p_k(1), p_k(-1)$ are positive. This is a contradiction to the assumption that $p_{\hat{k}}(1) = 0$ or $p_{\hat{k}}(-1) = 0$ (as $p_{\hat{k}}(\lambda)$ is the convex combination of $p_k(\lambda)$ and $p_{k'}(\lambda)$). Hence the zeros of $p_{\hat{k}}$ must be in the open unit disk for every $\delta \in [0,1]$. □

2.4.1 Connectedness properties of $\mathcal{S}_{\mathbf{x}}$ and $\mathcal{S}$

The following topological property of the set $\mathcal{S}_{\mathbf{x}}$ has immediate algorithmic implications.

Lemma 2.4.5. *For the state feedback system, the set $\mathcal{S}_{\mathbf{x}}$ is connected and contractible in $\mathbb{R}^n$.*

Proof. The proof proceeds similar to the proof of Lemma 2.3.7. Putting $\Gamma = \{\lambda \in \mathbb{C} : |\lambda| < 1\}^n$, it suffices to show that Γ_*/S_n is connected and contractible.[23] But this is immediate since any $v \in \Gamma_*/S_n$ is connected to $(0,\ldots,0)$ by the convex line segment $(1-\delta)v + \delta\, 0$ $(\delta \in (0,1))$ in Γ_*. □

[23] As such, Γ_* is the subset of n-dimensional complex-valued vectors with entries having modulus less than one and closed under conjugation.

For the output feedback case, the set $\mathcal{S}$ is not connected in general. Following is an example of a SISO system with more than one path-connected component.[24]

Example 3. *Consider the LTI system* $(A, b, c^\intercal)$ *with,*

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \; b = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \; c = \begin{pmatrix} 0.5184 \\ -2.448 \\ 4.33 \\ -3.4 \end{pmatrix}.$$

The feedback controllers are then parametrized by intervals in $\mathbb{R}$. *Figure 2.5 depicts that the roots of the closed loop system are inside the unit disk for some interval, then become unstable, and subsequently reenter the unit disk as* k *varies; as such,* $\mathcal{S}$ *has two connected components.*

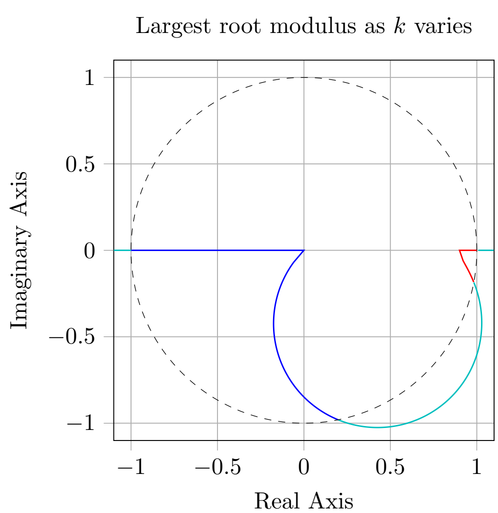

Figure 2.5: The figure depicts how the root with largest modulus varies with respect to the feedback gain k. The blue and red segments correspond to two stabilizing intervals.

[24]This example actually shows that the bound in Lemma 2.4.8 is tight.

Connected Components of $\mathcal{S}$ for SISO Systems

We now show that there is at most $\lceil \frac{n}{2} \rceil$ connected components in $\mathcal{S}$. We will follow a similar line of reasoning as for the continuous systems presented in § 2.3.1. We shall demonstrate the upper bound n first, followed by the upper bound of $\lceil \frac{n}{2} \rceil$. The essential ideas for proving the two results are similar to the strategy we followed in Lemmas 2.3.9 and 2.3.12, with some subtle differences.

Lemma 2.4.6. *The set $\mathcal{S}$ has at most n connected components.*

Proof. Note that in this case, $\mathcal{S}$ is a subset of $\mathbb{R}$; it also suffices to assume that the system is in the controllable canonical form. Consider the characteristic polynomial of a closed-loop system,

$$p(\lambda, k) = \lambda^n + (a_{n-1} - kc_{n-1})\lambda^{n-1} + \cdots + a_0 - kc_0,$$

where $a = (a_{n-1,}, \ldots, a_0)$ is the last row of A and c_j's are components of c. Let ζ be the bijection between $\mathbb{R}^n$ and the set of monic nth degree polynomials, and

$$\Gamma = \{a \in \mathbb{R}^n : \zeta(a) \text{ has roots on the unit disk in } \mathbb{C}\},$$

and parameterize the line $\ell(k) \coloneqq a - kc$. Suppose $\ell(k) \cap \Gamma$ for finitely many k's, listed in increasing order $\{k_1, \ldots, k_l\}$ (this will be proven subsequently). Let $n_{p(x,k)}(\mathbb{D})$ denote the number of roots of the closed loop characteristic polynomial on $\mathbb{D}$. Moreover, let γ be a counterclockwise oriented unit circle in $\mathbb{C}$, tracing the boundary of $\mathbb{D}$. For each $k \in (k_j, k_l)$, we define,

$$m(k) = \frac{1}{2\pi i} \int_\gamma \frac{p'(\lambda, k)}{p(\lambda, k)} \, d\lambda.$$

Note that $p(\lambda, k)$ does not vanish on γ, and by Cauchy's Argument Principle [54], $m(k)$ is the number of zeros of $p(\lambda, k)$ inside γ, that is, $m(k) = n_{p(\lambda,k)}(\mathbb{D})$. We further note that $m(k)$ is continuous in k, and as such, $n_{p(\lambda,k)}(\mathbb{D})$ is constant on each interval (k_j, k_{j+1}). Hence, either

$n_{p(\lambda,k)}(\mathbb{D}) = n$ or $n_{p(\lambda,k)}(\mathbb{D}) < n$, respectively, corresponding to stabilizing and non-stabilizing gains k.

Now by inspecting the number of intersections between $\ell(k)$ and Γ, we can derive an upper bound on the number of connected components of $\mathcal{S}$. In this direction, when $\ell(k)$ intersects Γ there is $\lambda_0 = e^{i\theta} \in \mathbb{C}$ such that,

$$\lambda_0^n + (a_{n-1} - kc_{n-1})\lambda_0^{n-1} + \cdots + (a_0 - kc_0) = 0,$$

and therefore,

$$k = \frac{\lambda_0^n + a_{n-1}\lambda_0^{n-1} + \cdots + a_0}{c_{n-1}\lambda_0^{n-1} + c_{n-2}\lambda_0^{n-2} + \cdots + c_0}. \tag{2.13}$$

Note that $r(\lambda_0) = c_{n-1}\lambda_0^{n-1} + \cdots + c_0 \neq 0$ (see details in the proof of Lemma 2.3.9). This implies that $h(\lambda) := \mathbf{Im}((\lambda_0^n + s(\lambda_0))\overline{r(\lambda_0)}) = 0$, where $\overline{r(\lambda_0)}$ is the complex conjugate of $r(\lambda_0)$. Substituting $\lambda_0 = e^{i\theta}$ into $(\lambda_0^n + s(\lambda_0))\overline{r(\lambda_0)}$, we have

$$\alpha_n e^{in\theta} + \alpha_{n-1}e^{i(n-1)\theta} + \cdots + \alpha_0 + \alpha_1 e^{-i\theta} + \cdots + \alpha_{-(n-1)}e^{-i(n-1)\theta} = 0,$$

where $(\alpha_n, \ldots, \alpha_1, \alpha_0, \alpha_{-1}, \ldots, \alpha_{-(n-1)})$ are the corresponding coefficients when we expand the product. This implies that

$$\beta_n \sin(n\theta) + \cdots + \beta_1 \sin(\theta) = 0,$$

where $(\beta_n, \ldots, \beta_1) \in \mathbb{R}^n$ are the corresponding real coefficients. We now note Chebyshev polynomials of second kind satisfy,

$$U_{n-1}(\cos(\theta))\sin(\theta) = \sin(n\theta),$$

where $U_{n-1}(\cos(\theta))$ is the Chebyshev polynomial of degree $n-1$ in $\cos(\theta)$. It thus follows that,

$$\sin(\theta)\left(\beta_1 + \beta_2 U_1(\cos(\theta)) + \cdots + \beta_n U_{n-1}(\cos(\theta))\right) =: \sin\theta\, g(\cos(\theta)), \tag{2.14}$$

where $g(\cos(\theta)) \in \mathbb{R}[\cos(\theta)]$ has degree $n-1$. By Fundamental Theorem of Algebra, there will be at most $n-1$ possible values for $\cos(\theta)$. Noting that $\theta = 0$ or $\theta = \pi$ also satisfy the above relation. Mind that for each value of $\cos(\theta)$, there are two possible θ's in $[0, 2\pi)$, yielding a conjugate pair $e^{i\theta}$ and $e^{-i\theta}$. But this conjugate pair will be mapped to the same gain k via (2.13). Thereby, we will have at most $n+1$ possible values for such k's. Since the set of stabilizing controllers is bounded, we have at most n connected components. $\square$

Remark. *Figure 2.6 depicts a similar situation as observed perviously for the continuous systems: two adjacent intervals* (k'', k_0) *and* (k_0, k') *could be both stabilizing but* k_0 *is only marginally stabilizing. The system parameters are,*

$$A = \begin{pmatrix} \frac{2909}{1000} & -\frac{283}{100} & \frac{26129688223-120\sqrt{6273911930105230}}{18036490000} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

$$b = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad c = \begin{pmatrix} 0.1343 \\ -0.1846 \\ 0.0623 \end{pmatrix}.$$

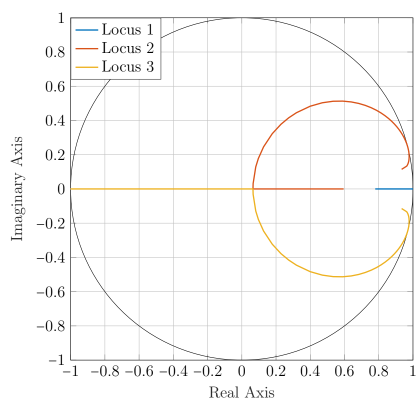

Figure 2.6: Two adjacent stabilizing intervals (k_{j-1}, k_j) and (k_j, k_{j+1}) where k_j is not stabilizing.

Now we prove that the bound on the number of connected components of $\mathcal{S}$ can be tightened to $\lceil \frac{n}{2} \rceil$. The strategy for proving this bound is similar to Lemma 2.3.12. That is, we need to examine the implications of having two adjacent stabilizing intervals (k_{j-1}, k_j) and (k_j, k_{j+1}); see Figure 2.6.

We first make an observation.

Proposition 2.4.7. *Let*

$$\varphi(\lambda) = a_{-n}\lambda^{-n} + a_{n-1}\lambda^{-(n-1)} + \cdots + a_0 + a_1\lambda + \cdots + a_n\lambda^n,$$

and

$$h(\theta) = \mathbf{Im}(\varphi(\lambda)|_{\lambda=e^{i\theta}}),$$

where $\varphi(\lambda) \in \mathbb{R}[\lambda]$*. Then* $h'(\theta) = \mathbf{Re}\left((\lambda\varphi'(\lambda))\big|_{\lambda=e^{i\theta}}\right)$*; note that* $h'(\theta)$ *denotes differentiation with respect to* θ *and* φ' *refers to differentiation with respect to* λ*.*

Proof. By linearity of the operations involved, it suffices to show this relation holds for λ^{-l} and λ^l, when $l \in \mathbb{N}$. For λ^j, $v(\theta) \coloneqq \mathbf{Im}(\lambda^l|_{\lambda=e^{i\theta}}) = \sin(l\theta)$ and $v'(\theta) = l\cos(l\theta)$; however,

$$\mathbf{Re}\left((\lambda(\lambda^l)')|_{\lambda=e^{i\theta}}\right) = \mathbf{Re}\left(l\lambda^l|_{\lambda=e^{i\theta}}\right) = l\cos(l\theta).$$

For λ^{-l}, $w(\theta) \coloneqq \mathbf{Im}(\lambda^{-l}|_{e^{i\theta}}) = -\sin(l\theta)$ and $w'(\theta) = -l\cos(l\theta)$. The proof now follows by observing that,

$$\mathbf{Re}\left((\lambda(\lambda^{-l})')|_{\lambda=e^{i\theta}}\right) = \mathbf{Re}\left(-l\lambda^{-l}|_{\lambda=e^{i\theta}}\right) = -l\cos(l\theta).$$

□

We are now in the position to prove the tight bound of $\lceil n/2 \rceil$ on the number of connected components of $\mathcal{S}$.

Lemma 2.4.8. *When* $\mathcal{S} \neq \varnothing$*, it has at most* $\lceil \frac{n}{2} \rceil$ *connected components.*

Proof. Similar to the continuous case, we need to explore the implications of having two adjacent intervals be stabilizing. Suppose (k_{j-1}, k_j) and (k_j, k_{j+1}) are two such intervals, where k's in each interval are stabilizing. By assumption, $p(\lambda_0, k_j) = 0$ for some $|\lambda_0| = 1$.

By Proposition 2.3.10, $\lambda_0 = e^{i\theta_0}$ is a simple root for $p(\lambda, k_j) \in \mathbb{R}[\lambda]$ since the unit disk has positive Gaussian curvature (see argument in the proof to Lemma 2.3.12). Let us denote by $s(\lambda) = a_{n-1}\lambda^{n-1} + \cdots + a_0$ and $r(\lambda) = c_{n-1}\lambda^{n-1} + \cdots + c_0$. Now the curve $\eta(t)$ is differentiable at λ_0 with

$$\left.\frac{d\eta}{dt}\right|_{t=k_j} = \frac{r(\lambda_0)}{(\lambda^n + s(\lambda))'|_{\lambda=\lambda_0}},$$

by appealing to the asymptotic formula or just formally differentiating. Geometrically, at k_j, the derivative should be orthogonal to λ_0 if the curve $\eta(k)$ is tangent to the unit sphere at λ_0. This implies that,

$$\left.\frac{r(\lambda)}{\left(\lambda^n + s(\lambda)\right)'}\right|_{\lambda=\lambda_0} = i\gamma\lambda_0,$$

where $\gamma \in \mathbb{R}$. Hence,

$$\mathbf{Re}\left(r(\bar{\lambda})\left(\lambda^n + s(\lambda)\right)' \lambda|_{\lambda=e^{i\theta_0}}\right) = 0. \tag{2.15}$$

Recall that k_j's correspond to the roots of $\sin(\theta)g(\cos(\theta))$ via (2.13). So if (k_{j-1}, k_j) and (k_j, k_{j+1}) are both stabilizing, then the θ_0 that maps to k_j must satisfy,

$$\mathbf{Im}\left(\left(\lambda^n + s(\lambda)\right) r(\bar{\lambda})|_{\lambda=e^{i\theta_0}}\right) = 0, \tag{2.16}$$

$$\mathbf{Re}\left(r(\bar{\lambda})p'_{k_j}(\lambda)\lambda|_{\lambda=e^{i\theta_0}}\right) = 0. \tag{2.17}$$

Let us now show that these two relations imply if $\cos(\theta_0)$ is a solution to $g(\cos(\theta_0))$ (recall that (2.16) is equivalent to $\sin(\theta)g(\cos(\theta)) = 0$), then $g'(\cos(\theta_0)) = 0$ due to (2.17). Note that $\theta_0 \neq 0$ or π by (2.15).

Let us consider the function $G(\theta)$ defined by

$$G(\theta) = \mathbf{Im}\left(\left(\lambda^n + s(\lambda)\right) r(\bar{\lambda})|_{\lambda=e^{i\theta}}\right).$$

Note that $G(\theta) = \sin(\theta)g(\cos(\theta))$. We adopt the notation $G'(\theta)$ to refer to differentiation with respect to θ. Let us define $q(\lambda) = r(1/\lambda)$ and observe that $q'(\lambda) = -r'(1/\lambda)/\lambda^2$. Then

$$G(\theta_0) = \mathbf{Im}\left(\left(\lambda^n + s(\lambda)\right) q(\lambda)|_{\lambda=e^{i\theta_0}}\right).$$

By Proposition 2.4.7, we have

$$
\begin{aligned}
G'(\theta_0) &= \mathbf{Re}\left(\lambda\left((\lambda^n + s(\lambda))q(\lambda)\right)'|_{\lambda=e^{i\theta_0}}\right) \\
&= \mathbf{Re}\left(\lambda\left(\lambda^n + s(\lambda)\right)' q(\lambda)|_{\lambda=e^{i\theta_0}}\right. \\
&\quad \left. +\lambda\left(\lambda^n + s(\lambda)\right) q'(\lambda)|_{\lambda=e^{i\theta_0}}\right).
\end{aligned}
$$

Noting that $\lambda_0^n + s(\lambda_0) - k_j r(\lambda_0) = 0$, i.e., $\lambda_0^n + s(\lambda_0) = k_j r(\lambda_0)$, it follows that,

$$
\begin{aligned}
G'(\theta_0) &= \mathbf{Re}\left(\left(\lambda(\lambda^n + s(\lambda))' r(\frac{1}{\lambda}) + \lambda k_j r(\lambda) q'(\lambda)\right)|_{\lambda=e^{i\theta_0}}\right) \\
&= \mathbf{Re}\left(\left(\lambda(\lambda^n + s(\lambda))' r(\frac{1}{\lambda}) - \frac{1}{\lambda} k_j r(\lambda) r'(\frac{1}{\lambda})\right)|_{\lambda=e^{i\theta_0}}\right) \\
&= \mathbf{Re}\left(\left(\lambda\left(\lambda^n + s(\lambda)\right)' r(\bar{\lambda}) - k_j r(\lambda)\bar{\lambda} r'(\bar{\lambda})\right)|_{\lambda=e^{i\theta_0}}\right) \\
&= \mathbf{Re}\left(\left(\lambda\left(\lambda^n + s(\lambda)\right)' r(\bar{\lambda}) - k_j r(\bar{\lambda})\lambda r'(\lambda)\right)|_{\lambda=e^{i\theta_0}}\right) \\
&= \mathbf{Re}\left(r(\bar{\lambda}) p'_{k_j}(\lambda)\lambda|_{\lambda=e^{i\theta_0}}\right).
\end{aligned}
$$

The above identity is precisely (2.17), i.e., $G'(\theta_0) = 0$. But by the chain rule,

$$
G'(\theta_0) = \cos(\theta_0) g(\cos(\theta_0)) - \sin^2(\theta_0) g'(\cos(\theta_0)) = 0.
$$

Since $\sin(\theta_0) \neq 0$, it follows that $g'(\cos(\theta_0)) = 0$. Putting $v = \cos(\theta_0)$, it now follows that v must be a multiple root if the two adjacent intervals are both stabilizing.

Now let $\{u_1, \dots, u_\mu, v_1, \dots, v_\nu\}$ denote the roots of the polynomial $g(v) = g(\cos(\theta)) = 0$, where v_j's are roots of multiplicity greater than 1. We have $\mu + 2\nu \leq n - 1$. These roots will be mapped to k_j's via (2.13). Let $\Lambda = \{k_j\}_{j=1}^{\mu'}$ and $\Pi = \{k_j\}_{j=1}^{\nu'}$ be the corresponding sets to $\{u_j\}$ and $\{v_j\}$. Since $\theta = 0$ or $\theta = \pi$ also corresponds to real k's, we add these two to the set Λ. Now $\mu' + 2\nu' \leq n + 1$. Let Υ be the collection of nonstabilizing intervals. Note that if $k_i \in \Lambda$, then one of the intervals (k_{j-1}, k_j) and (k_j, k_{j+1}) must be not stabilizing. Hence, k_j must be the endpoint of some interval in Υ. It now follows that the cardinality of Λ is given by

$$
\mu' = |\Lambda| \leq 2 + 2(|\Upsilon| - 2).
$$

On the other hand, the total number of stabilizing intervals would be $\mu' + \nu' + 1 - |\Upsilon|$ and thereby,

$$\mu' + \nu' + 1 - |\Upsilon| \leq \mu' + \nu' + 1 - \frac{\mu' + 2}{2} \leq \frac{n+1}{2}.$$

Consequently, the total number of stabilizing intervals is at most $\lceil \frac{n}{2} \rceil$.[25] $\square$

Remark. *As shown in Example 3, this bound is tight.*

We now observe the property that the stabilizing and non-stabilizing intervals in $\mathcal{S}$ interlace is also generic. The proof is almost verbatim to the one to Lemma 2.3.14 except that we need the nonemptyness of $\mathcal{U}$ for discrete systems; this is immediate from Example 3.

Lemma 2.4.9. *For a controllable and observable system $(A, b, c^\top)$, the property that stabilizing and non-stabilizing intervals in $\mathcal{S}$ interlace is generic.*

An algorithm for characterizing the connected components of $\mathcal{S}$

Similar to the continuous case, our analysis leads to an algorithm for identifying the stabilizing intervals for discrete-time linear SISO systems. We summarize this algorithm below.

[25] In particular, when $n = 2$, the set $\mathcal{S}$ corresponding to the triplet $(A, b, c^\top)$, if nonempty, is connected.

Algorithm 2: Identifying stabilizing intervals of $\mathcal{S}$

Find the real zeros $\{\lambda_1, \ldots, \lambda_l\}$ of (2.14).
Construct the list $\{\lambda_1, \bar{\lambda}_1, \ldots, \lambda_l, \bar{\lambda}_l\}$.
Append $\{-1, 1\}$ to this list if necessary, we get $L = \{1, -1, \lambda_1, \bar{\lambda}_1, \ldots, \lambda_l, \bar{\lambda}_l\}$.
Map L to $\{k_1, \ldots, k_{l'}\}$ (order this list in an increasing manner) by (2.13).
$k_{l'+1} \leftarrow +\infty$
The interval $(-\infty, k_1)$ is not stabilizing (Proposition 2.4.3).
Calculate the number of unstable roots for the interval $(-\infty, k_1)$ by computing roots of $p(\lambda, k_a)$, for an arbitrary $k_a < k_1$.
for $i = 1$ to l' **do**
    Check the multiplicity of λ_i that maps to k_i.
    Calculate the change in the number of unstable roots from $k_i - \epsilon$ to $k_i + \epsilon$ (Remark 2.3.1).
    Compute the number of unstable roots for the interval (k_i, k_{i+1}).
    if number of unstable roots for (k_i, k_{i+1}) is zero **then**
        The interval (k_i, k_{i+1}) is stabilizing.
    else
        The interval (k_i, k_{i+1}) is not stabilizing.
    end if
end for

Algorithm 2 divides the space of feedback gains into separated intervals where the number of unstable roots of $p(\lambda, k)$ is fixed. At the boundaries separating these intervals, $p(\lambda, k)$ has at least a root on the unit circle for the corresponding feedback gain. The main characteristic of Algorithm 2 is invoking Remark 2.3.1 for analyzing the root paths of $p(\lambda, k)$ at the boundaries of the separated intervals. Consequently, Algorithm 2 enables keeping track of the number of unstable roots when k increases from $-\infty$ to $+\infty$. As such, Algorithm 2 determines all stabilizing intervals. In addition, this algorithm computes the number of unstable roots of $p(\lambda, k)$ for unstable intervals. We demonstrate the progression of Algorithm 2 on the example in Remark 2.4.1. The characteristic polynomial of the closed-loop system is given by

$$p(\lambda, k) = \lambda^3 + 2.909\lambda^2 + 2.83\lambda - 0.9217272705 + k(0.1343\lambda^2 - 0.1846\lambda + 0.06229).$$

Furthermore,

$$g(\cos(\theta)) = 0.24916 \left(\cos(\theta)\right)^2 - 0.4840272752 \cos(\theta) + 0.2350722459,$$

has a root $\frac{\sqrt{6018}}{40}$ with multiplicity 2. Then $\sin(\theta_1) = 0$ is mapped to $k_1 = 0.06065638$ and $k_3 = 20.09687366$; $\cos(\theta_2) = \frac{\sqrt{6018}}{40}$ on the other hand maps to $k_2 = 0.6198635016$. By Corollary 2.4.3.1, the two unbounded intervals $(-\infty, k_1)$ and (k_3, ∞) are both non-stabilizing. We note that k_1 corresponds to $\sin(\theta) = 0$ and thereby (k_1, k_2) is stabilizing. Since k_2 is obtained from a multiple root of $g(\cos(\theta))$, we need to check condition (2.15). In this case, we conclude that the interval (k_2, k_3) is stabilizing.[26]

2.5 Conclusion

The motivation for this work stems from recent interest in devising learning type algorithms for control synthesis, that evolve over the set of stabilizing feedback gains. This in turn, has inspired the need to further examine the topological properties of these sets. We envisage that some of these properties might have been observed in the earlier literature in system theory and known to experts;[27] however, this work is an attempt to gather and prove these properties in a concise and rigorous manner using basic topology and the theory of polynomials. In this work, we have focused on topological and metrical properties of stabilizing state feedback gains and SISO output feedback gains for continuous and discrete time linear systems; extensions to MIMO case are elaborated upon in [12]. In this latter case, topolog-

[26] Alternatively, we note that k_3 is obtained from $\sin(\theta) = 0$, and as such, (k_2, k_3) must be stabilizing since (k_3, ∞) is not.

[27] Particularly those with deep affinity for root locus and geometry of polynomials.

ical arguments turn out to be even more dominant for characterizing the set of stabilizing feedback gains, with less reliance on the geometry of polynomials.

Chapter 3

LQR THROUGH THE LENS OF FIRST ORDER METHODS: DISCRETE-TIME CASE

3.1 Introduction

Linear-quadratic-regulator (LQR) has been one of the cornerstones of modern control theory [78] since Kalman's seminal work in 1960s [39]. LQR is formulated around an optimization problem for determining a sequence of (control) inputs to a linear system in order to minimize a given (integral) quadratic cost over an infinite horizon.[1] From the theoretical point of view, a fundamental property of LQR synthesis is that the resulting optimal input is in the form of a state feedback; as such, it can be represented as a constant feedback gain on the state of the system [1,39]. The state feedback gain that "solves" the infinite-horizon LQR problem, in turn, can be obtained by solving the algebraic Riccati equation (ARE). That is, in the traditional approach to LQR design, the state feedback gain is revealed after obtaining the "certificate" or "cost-to-go" for the underlying optimal control problem.[2] Historically, a large number of works have studied the solution of ARE, including approaches based on iterative algorithms [31], algebraic solution methods [45], and semidefinite programming [2].

Recently, there has been a surge of interest in constructing optimal control strategies directly, viewing control synthesis by policy gradient based algorithms.[3] Adopting such a point of view has been partially inspired by the success of learning algorithms, such as Reinforcement Learning (RL), where using principles of Dynamic Programming (DP), one can devise real-time model-free methods for both continuous-time and discrete-time LQR [10,

---

[1]We shall not delve into the finite-horizon LQR in this chapter.

[2]The analogy here would be solving the dual, followed by the recovery of the primal solution.

[3]One might as well extrapolate that these methods provide a streamline recipe for learning optimal feedback gains in real-time.

19, 36, 46, 48, 51, 52] (noting that LQR problem can be formulated as an RL problem [8, 52]). However, policy iterations are inherently prohibitive since the cost function has an infinite horizon, is undiscounted and unbounded per stage [8]. The RL perspective for LQR, particularly direct policy updates, has the merits in terms of extension to model-free setting by means of stochastic (zeroth-order) optimization [20, 57], computational scalability and rich parametrization of the feedback policy. In the meantime, LQR has proved to play an important role in theoretical RL by serving as a benchmark for demonstrating global convergence of policy gradient based algorithms.

In this chapter, inspired by the work [24], we examine first order methods for solving the LQR problem. In this direction, we first give a fairly complete account of the analytical properties of LQR cost function. More specifically, we show the cost function is smooth, coercive, gradient dominated over its effective domain,[4] and has quadratic growth under standard LQR assumptions.[5] We next examine three types of well-posed flows over the set of stabilizing controllers induced by this cost function, namely, gradient flow, natural gradient flow and quasi-Newton flow. These flows are closely related to the policy iteration schemes. Indeed, the conciseness and well-established theory of ordinary differential equations (ODEs) not only provide an elegant analysis of the convergence of numerical methods but also provide deeper insights into the corresponding optimization algorithms [29]. In this direction, we prove that the Lyapunov functionals for these flows decay at an exponential rate and the corresponding trajectories are exponentially stable in the sense of Lyapunov.

We then proceed to discuss the forward Euler discretization of these flows, realized as gradient descent, natural gradient descent and quasi-Newton iteration (hence, the state feedback gain can be updated iteratively). We show that the stepsizes in the gradient descent and natural gradient descent can be obtained via the Lyapunov equations in two consecutive updates; the coerciveness of the cost function on the other hand, ensures that the updated

[4]Gradient dominated property was first observed in [24]; in this chapter, we provide an alternate, control-theoretic proof of this fact.

[5]In [24] it was assumed that $Q \succ 0$; here we only require that $Q \succeq 0$ and the eigenvalues of A on the unit disk are (Q, A)-observable.

feedback gains remain stabilizing. As such, both the function values and feedback gains converge linearly to the corresponding global minimum. In the case of natural gradient descent in particular, we obtain a sequence of value matrices that monotonically decrease in the positive semidefinite cone.[6] Convergence rate of the quasi-Newton iteration is also analyzed,[7] proving that the corresponding iterates and function values converge Q-quadratically [8] to the global optima.[9]

A summary of our contributions compared with the existing works on direct policy updates is as follows:

a. We motivate the policy gradient algorithms from the perspective of *minimizing flows*. The flow perspective yields an elegant Lyapunov-type argument for the stability analysis and provide further insights into the discretization process and deriving iterative algorithms.

b. We relax the assumption that state penalization matrix Q is positive definite as adopted in [24] needed for the convergence proof. Instead, we assume that Q is positive semidefinite and that the eigenvalues of A on the unit disk in $\mathbb{C}$ are (Q, A)-observable; this is consistent with the standard LQR theory.

c. We improve the selection of the stepsize for the proposed direct policy updates. Our analysis utilizes the Lyapunov matrix equation, that is not only conceptually appealing but also simplify the subsequent convergence analysis. Moreover, we establish a connection between quasi-Newton and Hewer's algorithms; we show that quasi-Newton

[6] The terminology "natural gradient descent" (flow) is reserved for a particular choice of Riemannian metric; see §3.5 for details.

[7] The quasi-Newton iteration is consistent with Hewer's algorithm [31] if stepsize is set to be $1/2$, essentially a Newton's iteration. We provide a transparent motivation as to why the proposed algorithm is a "quasi-Newton" iteration over direct policy space and an alternative proof for the quadratic convergence.

[8] We have adopted the terminology of [58].

[9] The algorithm is referred to as "Gauss-Newton" in [24] and the convergence was shown to be linear rather than Q-quadratic.

algorithm enjoys a Q-quadratic convergence rate.

d. We provide a careful analysis of the analytical properties of LQR cost function when viewed as a matrix function over its domain. In this direction, we provide a theoretical treatment on how standard assumptions in LQR theory translate into the corresponding analytical properties of the matrix-valued optimization problem. Indeed, these properties provide a transparent and elegant point of view on the stabilizing property of the proposed iterative process.

The remainder of this chapter is as follows. The LQR problem statement and related definitions are provided in §3.3. §3.3.2 introduces the analytical properties of the LQR cost function. Subsequently, gradient flow, natural (Riemannian) gradient flow and quasi-Newton flow are introduced in §3.4, §3.5 and §3.6, respectively. Discrete realizations of these flows, namely, gradient descent, natural gradient descent and quasi-Newton iterations are addressed in §3.4.1, §5.5 and §3.6.1. §3.7 presents simulation results to illustrate the theoretical contributions of the chapter; in §3.8, we provide a few concluding remarks.

3.2 Notation and Preliminaries

We denote by $\mathbb{M}_{n\times m}(\mathbb{R})$ the set of $n \times m$ real matrices and $\mathbb{GL}_n(\mathbb{R})$ as the set of invertible square matrices; $\mathbb{R}^n$ denotes the n-dimensional real Euclidean space with the $n = 1$ case identified with real number. The set of non-negative numbers is denoted by $\mathbb{R}_+$ and natural numbers as $\mathbb{N}$; $\mathbb{S}_n$ denotes the set of $n \times n$ real symmetric matrices. Other notation includes $A^\intercal$, $\rho(A)$, $\text{rank}(A)$, $\mathbf{Tr}(A)$, $vec(A)$ representing the transpose, spectral radius, rank, trace, and vectorization of the matrix A, respectively; $A \otimes B$ is the Kronecker product of matrices A and B, and $\mathbf{rbd}\ \mathcal{K}$ designates the relative boundary of the set $\mathcal{K}$. The real inner product between a pair of vectors x and y is denoted by $\langle x, y\rangle$. $\|A\|_2$ denotes the spectral (operator) norm of a square matrix A and $\|A\|_F$ denotes its Frobenius norm.[10] Lastly, the notation

[10] 2-norm is assumed when we use $\|.\|$.

$A \geq B$ for two symmetric matrices refers to the positive semi-definiteness of their difference $A - B$; analogously for positive definiteness of this difference using $A > B$. We let $\lambda_i(A)$ denote the eigenvalues of a square matrix A. These eigenvalues are indexed in an increasing order with respect to their real parts, i.e.,

$$\mathbf{Re}(\lambda_1(A)) \leq \cdots \leq \mathbf{Re}(\lambda_n(A)).$$

If A is symmetic, the ordering becomes $\lambda_1(A) \leq \cdots \leq \lambda_n(A)$. When $A \geq 0$, $\|A\| = \lambda_n(A)$ and we shall use these interchangeably. We use $C^{\omega}(U)$ to denote the set of real analytic functions over an open set $U \subseteq \mathbb{R}^n$. A function $f : U \to \mathbb{R}$ is C^{∞}-smooth if it is infinitely differentiable. A function f is L-smooth when f is *continuously differentiable* and its gradient is L-Lipschitz, i.e., $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function with $\arg\min_{x \in \mathbb{R}^n} f(x) \neq \varnothing$. For $x_* \in \arg\min f(x)$, f is *gradient dominated* if $f(x) - f(x_*) \leq \tau \langle \nabla f(x), \nabla f(x) \rangle$ for all $x \in \mathbb{R}^n$ and a constant $\tau > 0$ [62]; f has *quadratic growth* about x_* if $\|x - x_*\|_2^2 \leq \kappa(f(x) - f(x_*))$ for all x and a constant $\kappa > 0$.
A pair (A, B) is stabilizable if there exists some K such that $A - BK$ is Schur. An eigenvalue λ of A is called (C, A)-observable if

$$\operatorname{rank}\left(\begin{pmatrix} A - \lambda I \\ C \end{pmatrix}\right) = n.$$

Given a pair of system matrices (A, B), $\mathcal{S}$ denotes the set of Schur stabilizing feedback gains,

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

We will frequently use a few linear algebraic facts on matrix equations and inequalities; some of these are collected in the following proposition.

Proposition 3.2.1. *The following relations hold:*

a. *For matrices A, B, C of appropriate dimensions, $vec(ABC) = (C^{\top} \otimes A)\, vec(B)$.*

b. For an $n \times n$ *positive definite* X*,*

$$M^\top XN + N^\top XM \geq -(aM^\top XM + \frac{1}{a}N^\top XN), \tag{3.1}$$

$$M^\top XN + N^\top XM \leq aM^\top XM + \frac{1}{a}N^\top XN, \tag{3.2}$$

where $M, N \in \mathbb{M}_{n\times m}(\mathbb{R})$ *with* $m \leq n$ *and* $a \in \mathbb{R}_+$*.*

c. Suppose that $A \in \mathbb{M}_{n\times n}(\mathbb{R})$ *has spectral radius bounded by* 1*, i.e.,* $\rho(A) < 1$*. Then*

$$A^\top XA + Q - X = 0,$$

has a unique solution,

$$X = \sum_{j=0}^{\infty} (A^\top)^j Q A^j,$$

and if $Q \succ 0$ *then* $X \succ 0$*. Moreover, if* $\widetilde{X}$ *satisfies*

$$A^\top \widetilde{X} A + \widetilde{Q} - \widetilde{X} = 0,$$

with $\widetilde{Q} \preceq Q$*, then* $\widetilde{X} \preceq X$*.*

d. Given arbitrary positive definite matrices X, Y *of the same size, one has,*

$$\lambda_1(Y)\,\mathbf{Tr}(X) \leq \mathbf{Tr}(XY) \leq \lambda_n(Y)\,\mathbf{Tr}(X). \tag{3.3}$$

The proofs of these observations can be found in [33].

3.3 Problem Setup and its Analytic Properties

In this section, we provide an overview of LQR, and in particular its modified initial state independent version, as well as a few analytic observations that are of independent interest. Although the reader might know of the extensive LQR literature, we note that some of these observations have only become necessary when the LQR optimization is viewed *directly* on the set of stabilizing feedback gains.

3.3.1 Discrete-time infinite horizon LQR

In the standard setup of LQR, we consider a (discrete-time) linear time invariant model of the form,

$$x_{k+1} = Ax_k + Bu_k, \tag{3.4}$$

where $A \in \mathbb{M}_{n\times n}(\mathbb{R})$, $B \in \mathbb{M}_{n\times m}(\mathbb{R})$ and (A, B) is stabilizable. The LQR problem is the optimization problem of devising a linear feedback gain $K \in \mathbb{M}_{m\times n}(\mathbb{R})$ for which $u_k = -Kx_k$, minimizing,[11]

$$J(x_0) = \sum_{k=0}^{\infty} \left[\langle x_k, Qx_k\rangle + \langle u_k, Ru_k\rangle\right],$$

where x_0 is the initial condition, and the quadratic cost is parameterized by $0 \le Q \in \mathbb{S}_n$ and $0 < R \in \mathbb{S}_m$; Q satisfies the condition: the eigenvalues of A lying on the unit disk is (Q, A)-observable. LQR is traditionally solved via dynamic programming or calculus of variations, leading to the celebrated Algebraic Riccati Equation (ARE) [1].[12]

In order to update the feedback gain (policy) directly, it will be conceptually appealing to consider the cost as a matrix function over the set of feedback gains. With this aim in mind, we may define $J_{x_0}: \mathbb{M}_{m\times n}(\mathbb{R}) \to \mathbb{R}$ as,

$$\begin{aligned} J_{x_0}(K) = \sum_{k=0}^{\infty} \big[&\langle (A - BK)^k x_0, Q(A - BK)^k x_0\rangle \\ &+ \langle K(A - BK)^k x_0, RK(A - BK)^k x_0\rangle\big], \end{aligned} \tag{3.5}$$

for some fixed initial condition $x_0 \in \mathbb{R}^n$. Our first task in this direct optimization setup is to determine the domain over which the function is well-defined. In other words, we are interested in the effective domain $\mathbf{dom}(J_{x_0}) = \{K \in \mathbb{M}_{m\times n}(\mathbb{R}) : J_{x_0}(K) < +\infty\}$. Addressing this seemingly natural analytical question turns out to be subtle. If K is stabilizing, i.e.,

[11] The condition that u_k has the form $-Kx_k$ is not set a priori in the LQR formulation; this feedback form is typically shown via the adoption of a dynamic programming step.

[12] For the dynamic programming case, one starts with the finite horizon case, apply the optimality principle, and then identify a solution concept for the infinite horizon case using a limit argument; calculus of variations provide another approach for deriving necessary conditions for LQ-type problems.

$\rho(A - BK) < 1$, then $K \in \mathbf{dom}(J_{x_0})$. In the meantime, for a non-stabilizing K, i.e., $\rho(A - BK) \geq 1$, when the system matrix $A - BK$ has both stable and unstable modes, if x_0 is chosen to be in the span of eigenspace corresponding to stable modes, $J_{x_0}(K) < \infty$. That is, $\{K : \rho(A - BK) < 1\}$ is a proper subset of $\mathbf{dom}(J_{x_0})$. Indeed, $\{K : \rho(A - BK) < 1\}$ is the interior of $\mathbf{dom}(J_{x_0})$ (see Proposition 3.1 and Lemma 3.2 in [11] for details). This implies that $J_{x_0}(K)$ is not differentiable everywhere on its domain. More precisely, $J_{x_0}(K)$ is differentiable on $\mathcal{S}$ but non-differentiable on $\mathbf{dom}(J_{x_0}) \setminus \mathcal{S}$. This complication is rather unnecessary as we are primarily interested in stabilizing controllers. This motivates us to examine initial condition independent formulation of LQR.[13]

Ideally, the objective function $f : \mathbb{M}_{m\times n}(\mathbb{R}) \to \mathbb{R}$ for our LQR calculus has an effective domain that coincides with the set of stabilizing feedback gains $\{K \in \mathbb{M}_{m\times n}(\mathbb{R}) : \rho(A-BK) < 1\}$. This can be achieved by choosing a set of linearly independent vectors $\{x_0^1, \dots, x_0^n\} \subseteq \mathbb{R}^n$ and defining,[14]

$$f(K) = \sum_{j=1}^{n} J_{x_0^j}(K). \tag{3.6}$$

As such, the function f would be infinite if K is not stabilizing (see Lemma 3.3.3 for details).

Remark. *The initial independent formulation is rather natural for general optimal control problems. In such problems, it is often desired to constrain the control synthesis to stabilizing feedback gains. For a learning algorithm that is built around a descent direction, such a formulation allows for an automatic enforcement of this stabilizing feature.*

Alternatively, we could let $x_0 \sim \mathcal{D}$, where $\mathcal{D}$ denotes some probability distribution, and let

$$f(K) = \mathbb{E}_{x_0\sim\mathcal{D}}(J_{x_0}).$$

As long as the samples span the whole space with probability 1, the function enjoys same properties as we have defined above.

[13] This is indeed necessary if we want to formulate an unconstrained optimization problem over the set of stabilizing feedback gains.

[14] Of course, one may choose the standard basis $\{e_1, \dots, e_n\}$, where e_i is the vector with zero entries except a "1" at the ith entry; the choice of an arbitrary basis simply retains flexibility.

In the pioneering works [41, 42, 49, 50] aiming to develop necessary optimality condition for continuous LTI systems, such initial condition independent formulation had been introduced. However, in these works the implications of such a formulation on differentiability and coerciveness of (4.3) *were not discussed.*

We shall now see that f (4.3) enjoys several favorable properties, e.g., f is differentiable over its effective domain and f diverges to infinity when K tends to the boundary of this domain, i.e., f is coercive. More importantly, for every $K \in \mathbf{dom}(f)$, the function $f(K)$ can be written as,

$$f(K) = \sum_{j=1}^{n} \mathbf{Tr}(X\mathbf{\Sigma}^j),$$

where $\mathbf{\Sigma}^j = x_0^j(x_0^j)^\intercal$ and X satisfies the Lyapunov equation,

$$(A - BK)^\intercal X(A - BK) + K^\intercal RK + Q = X. \tag{3.7}$$

Note that $J_{x_0^j}(K)$ does not necessarily admit the compact form $J_{x_0^j}(K) = \mathbf{Tr}(X\mathbf{\Sigma}^j)$ for every $K \in \mathbf{dom}(J_{x_0^j}(K))$. This is due to the fact that matrix X only makes (mathematical) sense if K is stabilizing, but $\mathbf{dom}(J_{x_0^j}(K))$ contains non-stabilizing feedback gains.

3.3.2 Analytical Properties of the LQR cost function

In this section, we investigate the properties of the LQR cost (4.3). We will observe that,

- f is a real analytic function over its domain.
- f is coercive and has compact sublevel sets.
- f is (natural) gradient dominated and has quadratic growth.

To simplify the notation, in the rest of this chapter, we shall denote

$$\mathbf{\Sigma} \coloneqq \sum_{j=1}^{n} \mathbf{\Sigma}^j, \quad A_K \coloneqq A - BK,$$

$$\mathbf{N}_K \coloneqq RK - B^\intercal X(A - BK);$$

when the context is clear, we will write $\mathbf{N}$ instead of $\mathbf{N}_K$ and in describing the iterative process, we shall denote $\mathbf{N}_{K_j}$ as $\mathbf{N}_j$. Let us recall some of the topological properties of the set of Schur stabilizing feedback gains $\mathcal{S}$; the proofs can be found in [13].[15]

Lemma 3.3.1. *The set $\mathcal{S}$ is regular open, contractible, and unbounded when $m \geq 2$ and the boundary $\partial\mathcal{S}$ is precisely the set $\mathcal{B} = \{K \in \mathbb{M}_{m\times n}(\mathbb{R}) : \rho(A - BK) = 1\}$.*

We now observe that $f(K)$ is real analytic over $\mathcal{S}$.

Lemma 3.3.2. *For the LQR cost* (4.3)*, we have $f \in C^\omega(\mathcal{S})$.*

Proof. For every $K \in \mathcal{S}$, let X be the solution to the Lyapunov equation (3.7). Then

$$vec(X) = \left(A_K^\top \otimes A_K^\top\right) vec(X) + vec(K^\top RK + Q). \tag{3.8}$$

Since the eigenvalues of $I \otimes I - A_K \otimes A_K$ are precisely $\{1 - \lambda_i(A_K)\lambda_j(A_K) : i, j = 1, \ldots, n\}$, $I \otimes I - A_K^\top \otimes A_K^\top$ is invertible. Hence,

$$vec(X) = (I \otimes I - A_K \otimes A_K)^{-1} vec\left(K^\top RK + Q\right).$$

By Cramer's rule, $X(K)$ is a rational function of polynomials in the entries of K and thus the map $K \mapsto X(K)$ is C^ω.[16] Hence, f can be viewed in terms of the composition,[17]

$$K \mapsto X(K) \mapsto \mathbf{Tr}(X\mathbf{\Sigma}).$$

As a composition of C^ω maps, f is thus real analytic.

□

With the initial condition independent formulation, the function f (4.3) diverges to infinity smoothly as K approaches the boundary $\partial\mathcal{S}$ or when K diverges to infinity.

---

[15] Strictly speaking, it is assumed (A, B) is controllable in [13]; however, by Kalman decomposition, one may easily deduce the results for a stabilizable pair (A, B).

[16] For a given K, $X(K)$ will be referred to as the the "cost matrix" as it characterizes the infinite horizon closed loop cost from the current state when this cost is finite. We shall use $X(K)$, X_K and X interchangebly if the context is clear.

[17] Mind that this perspective is only valid in $\mathcal{S}$.

Lemma 3.3.3. *The LQR cost (4.3) is coercive in the sense that,*

$$\lim_{K_j \to K \in \partial S} f(K_j) = +\infty,$$

or

$$f(K) \to \infty \text{ if } K \in \mathcal{S} \text{ and } \|K\| \to \infty.$$

Proof. Suppose that the sequence $\{K_j\} \subseteq \mathcal{S}$ and $K_j \to K \in \partial\mathcal{S}$. Denote the sequence $\{X_j\} \subseteq \mathbb{S}_n^{++}$ to be the corresponding sequence of value matrices. We claim that the sequence diverges to infinity in 2-norm. Namely, $\|X_j\|_2 \to +\infty$ as $j \to \infty$. To show this, it suffices to show the sequence contains no bounded subsequence. We prove by contradiction. Suppose not, i.e., there exists some bounded subsequent $\{X_{n_k}\}$; then by Weirestrass-Balzano [68], there exists some subsubsequence n_{k_j} such that $X_{n_{k_j}} \to X$ for some $X \geq 0$. By continuity, $X \geq 0$ solves the Lyapunov equation

$$A_K^\top X A_K + Q + K^\top R K = X.$$

But this is a contradiction: if (λ, v) is an eigen pair of A_K with $|\lambda| = 1$, then we have

$$v^\top \left(A_K^\top X A_K - X\right) v + v^\top (Q + K^\top R K) v = 0,$$

which implies that $Qv = 0$, $Kv = 0$ and $Av = \lambda v$. This is a contradiction to the (Q, A) observability of λ. Hence, $\{X_j\}$ must be unbounded. It thus follows that $\mathbf{Tr}(X_j \mathbf{\Sigma}) \to +\infty$ as $j \to +\infty$.

On the other hand,

$$\begin{aligned} f(K) &\geq \lambda_1 \left(\sum_{i=0}^{\infty} (A_K)^i \, \mathbf{\Sigma} \, (A_K^\top)^i \right) \mathbf{Tr}(Q + K^\top R K) \\ &\geq \lambda_1(\mathbf{\Sigma}) \, \lambda_1(R) \|K\|_F^2. \end{aligned}$$

Thereby, for any $M > 0$, $f(K) \geq M$ for $\|K\|_F$ sufficiently large. $\square$

With the coercive property in place, i.e., growth to infinity smoothly, we can continuously extend the function to $\mathbb{M}_{n\times n}(\mathbb{R})$ as an extended real-valued function which allows $+\infty$ as a function value. This in turn will imply that all sublevel sets of $f(K)$ are compact.[18]

Corollary 3.3.3.1. *The sublevel set $\mathcal{S}_\alpha = \{K \in \mathcal{S} : f(K) \le \alpha\}$ is compact for every $\alpha > 0$.*

Proof. By Lemma 3.3.3, we can continuously extend f to $\tilde{f} : \mathbb{M}_{m\times n}(\mathbb{R}) \to \mathbb{R} \cup \{+\infty\}$, where,

$$\tilde{f} = \begin{cases} f(K), & \text{if } K \in \mathcal{S}, \\ +\infty, & \text{if } K \in \mathcal{S}^c. \end{cases}$$

The sublevel sets $\tilde{\mathcal{S}}_\alpha = \{K \in \mathbb{M}_{n\times n}(\mathbb{R}) : \tilde{f}(K) \le \alpha\}$ of $\tilde{f}$, in the meantime, are compact by Proposition 11.12 in [5]. The proof is completed by observing that $\mathcal{S}_\alpha = \tilde{\mathcal{S}}_\alpha$ when α is finite. □

As $f \in C^\omega(\mathcal{S})$, the gradient of f can be characterized explicitly.

Proposition 3.3.4. *(Proposition 1 in [55]) For $K \in \mathcal{S}$, $\nabla f(K) = 2\left(RK - B^\intercal X(A - BK)\right)Y$, where Y solves the Lyapunov matrix equation,*

$$A_K Y A_K^\intercal - Y + \mathbf{\Sigma} = 0. \tag{3.9}$$

We emphasize that Proposition 3.3.4 only makes sense when $K \in \mathcal{S}$. One is then tempted to set $\nabla f(K) = 0$ to obtain a stationary point. However, since X is a function of K, whether or not $\nabla f(K) = 0$ is solvable in $\mathcal{S}$ needs clarification. Indeed, the analytical properties, in particular, the coerciveness, would guarantee the existence of global minimizer; the cost function structure on the other hand guarantees the uniqueness of such stationary point. We remark such existence of unique global minimizer can also be deduced by standard control theoretic argument [73]; what we are demonstrating is an alternative way of arriving such

[18] This can also be proved directly. The condition $f(K) \to +\infty$ as $\|K\| \to \infty$ implies that the sublevel sets of f are bounded; condition $f(K_j) \to +\infty$ as $K_j \to K \in \partial\mathcal{S}$ implies that every sublevel set is bounded away from the boundary and hence closed in the Euclidean topology. Note the continuity of f only guarantees that the sublevel sets are closed in $\mathcal{S}$.

a conclusion. To begin, we first observe a comparison equality concerning the difference of value matrices for two different stabilizing controllers.

Lemma 3.3.5 (Comparison Lemma)**.** *Suppose $K, \tilde{K} \in \mathcal{S}$ and denote X and $\tilde{X}$ be the corresponding value matices. Then we have*

$$\begin{aligned} A_{\tilde{K}}^{\top}(X - \tilde{X})A_{\tilde{K}} + (K - \tilde{K})^{\top}\mathbf{N}_K + \mathbf{N}_K^{\top}(K - \tilde{K}) \\ + (K - \tilde{K})^{\top}(R + B^{\top}XB)(K - \tilde{K}) = X - \tilde{X}. \end{aligned}$$

Proof. It suffices to take the difference of the following Lyapunov matrix equations

$$\begin{aligned} A_K^{\top} X A_K + Q + K^{\top} R K &= X, \\ A_{\tilde{K}}^{\top} \tilde{X} A_{\tilde{K}} + Q + \tilde{K}^{\top} R \tilde{K} &= \tilde{X}. \end{aligned}$$

Indeed, a few algebraic operations reveal that,

$$\begin{aligned} &A_{\tilde{K}}^{\top}(X - \tilde{X})A_{\tilde{K}} + (K - \tilde{K})^{\top}\mathbf{N}_K + \mathbf{N}_K^{\top}(K - \tilde{K}) \\ &\quad - (K - \tilde{K})^{\top}R(K - \tilde{K}) - (A_K - A_{\tilde{K}})^{\top}X(A_K - A_{\tilde{K}}) \\ &\quad = X - \tilde{X}. \end{aligned}$$

□

We now prove the existence and uniqueness of the stationary point.

Lemma 3.3.6. *The matrix $K_* = (B^{\top}X_*B + R)^{-1}B^{\top}X_*A$ is the unique global minimizer of $f(K)$,*[19] *where X_* is the corresponding solution of the Lyapunov equation* (5.2)*.*

Proof. Since (A, B) is stabilizable, $\mathcal{S}$ is nonempty. As such, for every finite $c > 0$, the set $\mathcal{S}_c = \{K \in \mathcal{S} : f(K) \le c\}$ is a nonempty compact set. Therefore, $f(K)$ achieves its minimum on $\mathcal{S}_c$ at K^* with $\nabla f(K_*) = 0$. Thereby, $K_* = (B^{\top}X_*B + R)^{-1}B^{\top}X_*A$ must be in $\mathcal{S}$ (this expression is now more precise!). To show uniqueness, suppose there exists $K_{*,1}$ and $K_{*,2}$ such

[19] As we are establishing the global minimizer is unique, throughout the chapter we shall use K_* to denote the global minimizer.

that the gradient vanishes at both points, namely $\mathbf{N}_{K_{*,1}} = \mathbf{N}_{K_{*,2}} = 0$.[20] By the Comparison Lemma, it follows that $X_{*,1} \geq X_{*,2}$ and $X_{*,2} \geq X_{*,1}$. This is a contradiction and thus the global minimizer is unique. □

Remark. *Lemma 3.3.6 can be regarded as an alternative means of deriving the expression for the optimal state feedback control for standard infinite horizon LQR problem. It is instructive to emphasis on the control theoretic assumptions in LQR guarantees the existence and uniqueness of the stationary point in the matrix-valued optimization problem:*

a. *The stabilizability guarantees that the domain* $\mathbf{dom}(f)$ *is nonempty.*

b. $Q \geq 0$ *and* (Q, A)*-observability of eigenvalues on the unit disk together guarantees the coerciveness and thus compactness of the corresponding sublevel sets. This in turn, implies the existence of stationary point (solvability of* $\nabla f(K) = 0$ *in* $\mathcal{S}$*).*

c. *The quadratic cost on the other hand implies the uniqueness of the optimal solution.*

We now observe that the LQR cost (4.3) is a *gradient dominated* function [62].[21] The proof of this property in [24] (Corollary 5) is based on a careful comparison of the cost difference in each time step between the optimal policy and a specified policy. Here, we provide an alternate proof of this important property. This alternate approach is more control-theoretic in the sense that it is mainly concerned with the properties of the Lyapunov equation. Moreover, this approach reveals that $f(K)$ has quadratic growth at K_*.

Lemma 3.3.7 (Natural Gradient Dominated[22])**.** *Let* K_* *be the optimal feedback gain. For* $K \in \mathcal{S}$*,*

$$\tau_1 \|K - K_*\|_F^2 \leq f(K) - f(K_*) \leq \tau_2 \langle \mathbf{N}_K, \mathbf{N}_K \rangle,$$

[20] This follows from $Y_K > 0$ for every $K\mathcal{S}$.

[21] This property is also referred as Polyak-Łojasiewicz condition, as a special case of what had been proposed in [53].

[22] The terminology "natural gradient" will be explained in §3.5.

where

$$\tau_1 = \lambda_1(Y)\lambda_1(R + B^\top XB), \tau_2 = \frac{\lambda_n(Y_*)}{\lambda_1(R + B^\top XB)},$$

$Y_* = \sum_{j=0}^{\infty}(A_{K_*})^j\,\mathbf{\Sigma}\,(A_{K_*}^\top)^j$, *and* X_* *solves the Lyapunov matrix equation*

$$A_{K_*}^\top X_* A_{K_*} - X_* + Q + K_*^\top R K_* = 0. \tag{3.10}$$

Proof. Recall that $\mathbf{N}_K \coloneqq RK - B^\top X A_K$ and X is the solution of (3.7). Now by Comparison Lemma 3.3.5, we have

$$\begin{aligned}&A_{K_*}^\top (X - X_*)A_{K_*} + (K - K_*)^\top \mathbf{N}_K + \mathbf{N}_K^\top (K - K_*)\\&- (K - K_*)^\top (R + B^\top XB)(K - K_*) = X - X_*.\end{aligned}$$

By Proposition 3.2.1 (part (3.2)), we note that for every $\alpha > 0$,

$$\begin{aligned}&(K - K_*)^\top \mathbf{N}_K + \mathbf{N}_K^\top (K - K_*)\\&\preceq \frac{1}{\alpha}(K - K_*)^\top (K - K_*) + \alpha \mathbf{N}_K^\top \mathbf{N}_K.\end{aligned}$$

Picking $\alpha = 1/(\lambda_1(R + B^\top XB)$, we then have,

$$\begin{aligned}&(K - K_*)^\top \mathbf{N}_K + \mathbf{N}_K^\top (K - K_*)\\&- (K - K_*)^\top R(K - K_*) - (A - A_{K_*})^\top X(A_K - A_{K_*})\\&\preceq \frac{1}{\lambda_1(R + B^\top XB)}\mathbf{N}_K^\top \mathbf{N}_K.\end{aligned}$$

By Proposition 3.2.1 (part (c)) , it thus follows that,

$$\begin{aligned}&f(K) - f(K_*)\\&\le \mathbf{Tr}\left((\sum_{j=0}^{\infty}(A_{K_*}^\top)^j\,(\frac{1}{\lambda_1(R + B^\top XB)}\mathbf{N}_K^\top \mathbf{N}_K)\,A_{K_*}^j)\mathbf{\Sigma}\right)\\&= \frac{1}{\lambda_1(R + B^\top XB)}\ \mathbf{Tr}(\mathbf{N}_K^\top \mathbf{N}_K(\sum_{j=0}^{\infty} A_*^j \mathbf{\Sigma}(A_*^\top)^j)),\end{aligned}$$

where in the last equality we have used the cyclic property of the matrix trace. We note that $Y_* = \sum_{j=0}^{\infty}(A_{K_*})^j \mathbf{\Sigma}(A_{K_*}^\top)^j$ is uniquely determined by the system parameters $A, B, Q, R, \mathbf{\Sigma}$. Now

$$\mathbf{Tr}((X - X_*)\mathbf{\Sigma}) \leq \frac{\lambda_n(Y_*)}{\lambda_1(R + B^\top X B)} \mathbf{Tr}(\mathbf{N}_K^\top \mathbf{N}_K).$$

where the last inequality follows from Proposition 3.2.1 (part (d)). To show quadratic growth, by Comparison Lemma 3.3.5,

$$\begin{aligned} &A_K^\top(X_* - X)A_K + (K_* - K)^\top \mathbf{N}_{K_*} + \mathbf{N}_{K_*}^\top(K_* - K) \\ &- (K - K_*)^\top(R + B^\top X_* B)(K - K_*) = (X_* - X). \end{aligned} \tag{3.11}$$

Since $\mathbf{N}_{K_*} = 0$, it thus follows

$$\begin{aligned} &f(K) - f(K_*) \\ &= \mathbf{Tr}\Bigg(\sum_{j=0}^{\infty}(A_K^\top)^j(K - K_*)^\top \\ &\qquad (R + B^\top X_* B)(K - K_*)A_K^j\Bigg) \\ &= \mathbf{Tr}\left((K - K_*)^\top(R + B^\top X_* B)(K - K_*)Y\right) \\ &\geq \lambda_1(Y)\lambda_1(R + B^\top X_* B)\|K - K_*\|_F^2. \end{aligned}$$

□

Now it is straightforward to observe the cost function $f(K)$ is globally *gradient dominated*.

Corollary 3.3.7.1 (Gradient Dominated)**.** *For every $K \in \mathcal{S}$, we have*

$$f(K) - f(K_*) \leq \kappa \langle \nabla f(K), \nabla f(K) \rangle,$$

where

$$\kappa = \frac{\lambda_n(Y_*)}{4[\lambda_1(\mathbf{\Sigma})]^2 \lambda_1(R + B^\top X_* B)}.$$

Proof. It suffices to observe

$$4\,\mathbf{Tr}(Y^\intercal \mathbf{N}_K^\intercal \mathbf{N}_K Y) \geq 4[\lambda_1(Y)]^2\,\mathbf{Tr}(\mathbf{N}_K^\intercal \mathbf{N}_K),$$

and $\lambda_1(Y) \geq \lambda_1(\mathbf{\Sigma})$ since $Y = \sum_{j=0}^{\infty}(A_K^\intercal)^j \mathbf{\Sigma} A_K^j$. The proof is completed by noting $R + B^\intercal X B \geq R + B^\intercal X_* B$. ☐

3.4 Gradient Flow on $\mathcal{S}$

In this section, we show that the LQR cost function (4.3) gives rise to a well-posed gradient flow,

$$\dot{K}_t = -\nabla f(K_t). \tag{3.12}$$

Let us first observe that (3.12) admits a unique solution for all time t.

Lemma 3.4.1. *For every $K_0 \in \mathcal{S}$ and $t_0 \in \mathbb{R}$, there exists a unique solution $K_t \in C^\infty(\mathbb{R}, \mathcal{S})$ for the initial value problem,*

$$\begin{cases} \dot{K}(t) = -\nabla f(K), \\ K(t_0) = K_0. \end{cases} \tag{3.13}$$

Proof. Note that $K \mapsto 2\mathbf{N}_K Y$ is C^∞ smooth. The statement now follows from Corollary 3.3.3.1 and Proposition 3.7 in [29]. ☐

We next show that the unique trajectory of (3.12) is in fact exponentially stable; without loss of generality, we assume that $t_0 = 0$.

Theorem 3.4.2. *For $K_0 \in \mathcal{S}$, denote by K_t as the solution of (3.13). Then the energy functional $V(K_t) \coloneqq f(K_t) - f(K_*)$ decays exponentially fast to the origin and the trajectory K_t is globally exponentially stable in the sense of Lyapunov, i.e.,*

$$\|K_t - K_*\|_F^2 \leq c e^{-\alpha t} \|K_0 - K_*\|_F^2,$$

where $\alpha, c \in \mathbb{R}_+$ are constants determined by the LQR parameters A, B, Q, R and initial condition K_0.

Proof. To prove the exponential decay of the energy functional, it suffices to observe

$$\dot{V}(K_t) = -\langle \nabla f(K_t), \nabla f(K_t) \rangle \leq -\alpha V(K),$$

with $\alpha = 1/\kappa$.[23] Hence, $f(K_t) - f(K_*) \leq e^{-\alpha t}(f(K_0) - f(K_*))$.

Now first observe that the Lyapunov functional is smooth, positive definite and radially unbounded;[24] thus K_t is globally asymptotic stable, i.e.,

$$\lim_{t \to \infty} K_t = K_*.$$

By Lemma 3.3.7,

$$\begin{aligned} \|K_t - K_*\|_F^2 &\leq \frac{1}{\tau_1}\big(f(K_t) - f(K_*)\big) \\ &\leq \frac{1}{\lambda_1(\mathbf{\Sigma})\lambda_1(R + B^\top X_* B)} e^{-\alpha t} V(K_0), \end{aligned}$$

where in the last inequality we have used $\lambda_1(Y) \geq \lambda_1(\mathbf{\Sigma})$ and $\lambda_1(R + B^\top X B) \geq \lambda_1(R + B^\top X_* B)$. The proof is completed by setting,

$$c = \frac{V(K_0)}{\lambda_1(\mathbf{\Sigma})\lambda_1(R + B^\top X_* B)\|K_0 - K_*\|_F^2}.$$

□

3.4.1 Discretization of Gradient Flow

In this section, we examine the discretization of the gradient flow (3.12). As we have observed in Theorem 3.4.2, both the energy functional and the trajectory of this flow converge exponentially to their respective global minimum. Ideally, a gradient descent algorithm converges linearly for the function values as well as the iterates. In this direction, the forward Euler discretization of the gradient flow yields,

$$K_{j+1} = K_j - \eta_j \nabla f(K_j), \tag{3.14}$$

---

[23]Recall κ is the constant in Corollary 3.3.7.1.

[24]In control literature, this is sometimes referred to as weakly coercive; nevertheless, as shown here, this is equivalent to being coercive.

where η_j is a nonnegative stepsize to be determined. We emphasize in (and only in) this section we assume $Q \succ 0$. The case $Q \succeq 0$ is treated in our work [17]. $Q \succ 0$ deverses its standalone treatment as in this case we could choose larger stepsize based on the smallest eigenvalue of Q.

The stepsize (or learning rate) should reflect two principles during the iterative process: (1) stay stabilizing and (2) sufficiently decrease the function value. In following, we shall see that the gradient dominated property leads to a stepsize that results in a sufficient decrease in the function values while the coerciveness guarantees that the acquired feedback gain is stabilizing. To begin, we observe that if $K_{j+1} = K_j - \eta_j \nabla f(K_j)$, provided that K_j and K_{j+1} are both stabilizing, the difference of the value matrix $X_{j+1} - X_j$ can be characterized as follows.[25]

Proposition 3.4.3. *If $K_{j+1} = K_j - \eta_j \nabla f(K_j)$ and K_j, K_{j+1} are both stabilizing, then $Z \coloneqq X_j - X_{j+1}$ solves the Lyapunov matrix equation,*

$$\begin{aligned} &A_{K_{j+1}} Z A_{K_{j+1}}^\top - Z + 2\eta_j \left(Y_j^\top \mathbf{N}_j^\top \mathbf{N}_j + \mathbf{N}_j^\top \mathbf{N}_j Y_j \right) \\ &- Y_j^\top \mathbf{N}_j^\top (4\eta_j^2 R + 4\eta_j^2 B^\top X_j B) \mathbf{N}_j Y_j = 0. \end{aligned}$$

Proof. It suffices to substitute $K_{j+1} - K_j = -2\eta_j \mathbf{N}_j Y_j$ into Comparison Lemma 3.3.5. $\square$

We now observe that with appropriately chosen η_j, we can guarantee a sufficient decrease in the function value while ensuring stabilization (for the analogous result in [24], see the third part of Theorem 7 and Lemma 24).

Lemma 3.4.4. *Consider the sequence $\{K_j\}$ generated by (3.14) with stepsize η_j. Denote by $\{X_j\}$ the corresponding Lyapunov matrix solutions with respect to $\{K_j\}$. When*

$$\eta_j < \sqrt{\frac{1}{c_j} + \frac{b_j^2}{4c_j^2}} - \frac{b_j}{2c_j}, \tag{3.15}$$

[25] This relationship is used in [24].

where

$$b_j = \lambda_n(R + B^\top X_j B)\frac{f(K_j)}{\lambda_1(Q)} + \frac{4f(K_j)\|B\mathbf{N}_j Y_j\|_2 \lambda_n(Y)}{\lambda_1(Q)\lambda_1(\mathbf{\Sigma})},$$
$$c_j = \lambda_n(R + B^\top X_j B)\frac{4\|B\mathbf{N}_j Y_j\|_2 \lambda_n(Y_j) f(K_j)}{\lambda_1(Q)},$$

the sequence $\{K_j\}$ *is stabilizing for every* $j \geq 0$*. In particular,*

$$f(K_{j+1}) - f(K_j) \leq 4\,\mathbf{Tr}(Y_j \mathbf{N}_j^\top \mathbf{N}_j Y_j)(\eta_j - b_j\eta_j^2 - c_j\eta_j^3).$$

Before presenting the proof of this result, we outline the underlying idea. The crucial property we shall leverage is the compactness of the sublevel sets, analogous to devising the stepsize. If we start at a stabilizing control gain K where the gradient does not vanish, and consider the ray of $\{K - \eta\nabla f(K) : \eta \geq 0\}$, by compactness of the sublevel set, there is some ζ for which $f(K') = f(K)$, where $K' \coloneqq K - \zeta\nabla f(K)$ (See Figure 3.1). What we shall demonstrate is that with the stepsize η_j given in the Lemma, if K_{j+1} stays in the compact sublevel set, then K_{j+1} must stay in the interior of the sublevel set, namely, $f(K_{j+1}) < f(K_j)$. We then proceed to examine two alternatives: (1) K_{j+1} is not stabilizing, or (2) K_{j+1} is stabilizing but $f(K_{j+1}) > f(K_j)$; either alternative would lead to a contradiction.

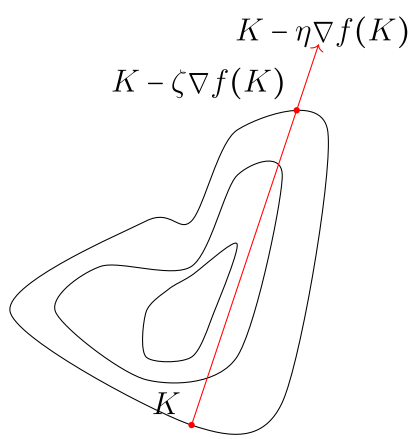

Figure 3.1: Gradient descent interacting with the level curves of f (4.3).

Proof. Suppose that the sequence generated by the choice of η_j is in fact stabilizing (to be proved subsequently!). This is crucial in our analysis as we use the Lyapunov matrix equation for the closed loop system, admitting a solution when K_j is stabilizing; without this assumption, the matrix X_j is not well-defined. By Proposition 3.4.3, we have,

$$\begin{aligned}
&f(K_{j+1}) - f(K_j) \\
&= \mathbf{Tr}\left(Y_{j+1}\left(-2\eta_j(Y_j^\top \mathbf{N}_j^\top \mathbf{N}_j + \mathbf{N}_j^\top \mathbf{N}_j Y_j)\right.\right. \\
&\quad \left.\left. + Y_j^\top \mathbf{N}_j^\top (4\eta_j^2 R + 4\eta_j^2 B^\top X_j B)\mathbf{N}_j Y_j\right)\right) \\
&\le 4\eta_j \, \mathbf{Tr}\left(\mathbf{N}_j^\top \mathbf{N}_j(-Y_j Y_{j+1}\right. \\
&\quad \left. + \eta_j \lambda_n(R + B^\top X_j B) Y_j Y_{j+1} Y_j)\right).
\end{aligned}$$

In order to determine a stepsize η_j such that $f(K_{j+1}) < f(K_j)$, we consider a univariate function,[26]

$$g(\eta) = \mathbf{Tr}\left(\mathbf{N}^\top \mathbf{N}(YY(\eta) - \eta a YY(\eta)Y)\right),$$

where $a := \lambda_n(R + B^\top X B)$, and $Y(\eta)$ is the solution of the matrix equation,[27]

$$\begin{aligned}
&(A - B(K - \eta 2\mathbf{N}Y))Y(\eta)(A - B(K - \eta 2\mathbf{N}Y)) \\
&+ \mathbf{\Sigma} - Y(\eta) = 0.
\end{aligned}$$

Note that in defining the function g we have dropped the indices as this function is used to determine stepsize for *every* iteration. Assuming that the choice of η ensures staying in the sublevel set of $f(K)$, i.e., $f(K - \eta 2\mathbf{N}Y) \le f(K)$, It can be shown with the proposed choices of stepsize, one has $g(\eta) > 0$ (see details in [11]). It remains to show that if η_j is chosen as above, our two opening assumptions are valid: (1) the sequence $\{K_j\}$ is stabilizing, and (2) K_{j+1} remains in the sublevel set of $f(K_j)$. We prove these by contradiction. First, note

[26] Note that since the products $Y_j Y_{j+1}$ and $\mathbf{N}_j^\top \mathbf{N}_j Y_j$ are not generally symmetric, the inequalities in Proposition 3.2.1 are not necessarily applicable.

[27] The function is not defined for every $\eta > 0$ but only for an interval for which $K - \eta 2\mathbf{N}Y$ is stabilizing.

that we can not have K_{j+1} be stabilizing while $K_{j+1} \notin S_{f(K_j)}$. Suppose that this is the case. The sublevel set $S_{K_j} \coloneqq \{K : f(K) \le f(K_j)\}$ is compact and the ray $\{K_j - \zeta \nabla f(K_j) : \zeta \ge 0\}$ intersects the boundary of S_{K_j} for some $\zeta > 0$; suppose that $K' = K_j - \zeta' \nabla f(K_j) \in \partial S_{K_j}$, where ζ' is the smallest positive real number for which this intersection occurs, i.e., the first time the ray intersects the boundary. It is clear ζ' must be greater than η_j as otherwise we would have $\zeta' < \eta_j$ and $f(K_j - \zeta' \nabla f(K_j)) < f(K_j)$, a contradiction[28]. Now we prove that K_j is stabilizing. If not, we must have

$$[0, \eta_j) \subseteq [0, \zeta'],$$

since otherwise, there exists $s' < \eta_j$ such that $s' = \zeta'$ and $f(K') = f(K_j)$, which would also contradict the inequality $f(K') < f(K_j)$. □

Theorem 3.4.5. *Putting* $d_j = \max(b_j, c_j)$ *where* b_j, c_j *are given in Lemma 3.4.4, if* $\eta_j = \sqrt{\frac{1}{3d_j} + \frac{1}{9}} - \frac{1}{3}$, *we have,*

$$f(K_j) - f(K_*) \le q^j (f(K_0) - f(K_*)),$$
$$\|K_j - K_*\|_F \le c_1 q^{j/2},$$

where $q \in (0,1)$ *and* $c_1 > 0$ *are constants.*

Remark. η_j *is acquired by noting that according to Lemma 3.4.4,*

$$f(K_j) - f(K_{j+1}) \ge 4\,\mathbf{Tr}(Y_j \mathbf{N}_j^\top \mathbf{N}_j Y_j)(\eta_j - d_j \eta_j^2 - d_j \eta_j^3).$$

Maximizing $\eta_j - d_j \eta_j^2 - d_j \eta_j^3$ *while esnuring* $1 - d_j \eta_j - d_j \eta_j^2 > 0$ *yields the desired quantity.*

Proof. Note the proposed stepsize rule satisfies $1 - 2d_j \eta_j - 3 d_j \eta_j^2 = 0$. Putting $r_j = f(K_j) - f(K_*)$, we observe that with the chosen stepsize η_j,

$$r_j - r_{j+1} \ge 4\,\mathbf{Tr}(Y_j \mathbf{N}_j^\top \mathbf{N}_j Y_j)(d_j \eta_j^2 + 2 d_j^2 \eta_j^3) \eqqcolon \nu_j r_j.$$

[28] Note what we proved above is: if a stepsize is strictly smaller than η_j, the function value is strictly decreasing if the gradient is not vanishing.

It follows that,

$$r_{j+1} \le (1-\nu_j) r_j =: q_j r_j.$$

It can be shown the proposed stepsize is bounded away from 0, i.e., $\eta_j \ge \epsilon$ for some constant $\epsilon > 0$ (see the detailed computation in Proposition $B2$ in [11].). Hence, the sequence $\{q_j\}$ is upper bounded away from 1[29], namely, for every j

$$q_j \le q < 1.$$

Thereby,

$$f(K_j) - f(K_*) \le q^j \left(f(K_0) - f(K_*) \right).$$

To see the convergence of iterates, it suffices to note by Lemma 3.3.7,

$$\|K_j - K_*\|_F^2 \le \frac{1}{\tau_1} \left(f(K_j) - f(K_*) \right).$$

□

3.5 Natural Gradient Flow on $\mathcal{S}$

If we inspect the proof of the Lyapunov stability of the gradient system (Theorem 3.4.2), the positive definite matrix Y does not affect the qualitative nature of these properties. Nevertheless, the matrix Y introduces a constant factor in the corresponding upper bounds. In this section, we consider a family of gradient systems of the form,

$$\dot{K}_t = -\nabla f(K) Y^{-\gamma} = -2\mathbf{N}_K Y^{1-\gamma}, \tag{3.16}$$

where $\gamma > 0$ is (real) scalar.[30] As discussed subsequently, such parameterized gradient system can achieve better convergence rate for different values of γ. Viewing such a gradient flow

[29] It is rather clear d_j is lower bounded away from 0. So $d_j \eta_j^2 + 2 d_j^2 \eta_j^3 > 0$.

[30] When $\gamma = 1$, this flow can be viewed as the continuous limit of the natural gradient descent as discussed in [24].

in the context of a flow on a Riemannian manifold is particularly pertinent.[31] In fact, as $\mathcal{S}$ is open, it is a *submanifold* in $\mathbb{M}_{m\times n}(\mathbb{R})$. We first observe that the inner product induced by Y^γ, i.e., $\langle M, N\rangle_{Y^\gamma} = \mathbf{Tr}(M^\top N Y^\gamma)$ is a well-defined Riemannian metric over $\mathcal{S}$.

Proposition 3.5.1. *Over $\mathcal{S}$, the inner product $\langle\cdot,\cdot\rangle_{Y(K)^\gamma}$ induces a Riemannian metric.*

Proof. Note that $Y(K)$ is positive definite for every $K \in \mathcal{S}$. It suffices to show that $Y(K)$ varies smoothly with K. But this follows from,

$$vec(Y) = (I \otimes I - A_K \otimes A_K)^{-1}\, vec(\mathbf{\Sigma}).$$

□

We can thus view $\mathcal{S}$ as a Riemannian manifold with metric induced by $\langle\cdot,\cdot\rangle_{Y^\gamma}$; the function $f : \mathcal{S} \to \mathbb{R}$ is then a scalar-valued function defined on this manifold. Let us now consider the gradient of f, denoted by $\mathrm{grad} f$, with respect to the Riemannian metric induced by $\langle\cdot,\cdot\rangle_{Y^\gamma}$ on $\mathcal{S}$.[32]

Proposition 3.5.2. *Over the Riemannian manifold $\left(\mathcal{S}, \langle\cdot,\cdot\rangle_{Y^\gamma}\right)$, $\mathrm{grad} f = 2(RK - B^\top X A_K)Y^{1-\gamma}$.*

Proof. It suffices to note that,

$$df(K)[E] = 2\,\mathbf{Tr}(E^T(RK - 2B^\top X A_K)Y) = \langle E, 2(RK - 2B^\top X A_K)Y^{1-\gamma}\rangle_{Y^\gamma}.$$

□

Now the gradient flow of interest on this manifold is,

$$\dot{K}_t = -\mathrm{grad} f(K_t).$$

We now observe that with respect to the Riemannian metric, the potential function decays at an exponential rate and the trajectory is exponentially Lyapunov stable (compare the different rate with the gradient flow in Theorem 3.4.2).

[31] We will see that in our case, it is better to choose γ other than $\gamma = 1$.

[32] We will use standard notitions in Riemmaninan manifold theory [74]. For example, df will denote 1-form and $\mathrm{grad} f$ will denote the gradient with respect to a Riemannian metric. As we are working in Euclidean space, we implicitly identify all tangent vectors by standard isomorphism, i.e., $T_K\mathcal{S} \approx \mathbb{M}_{m\times n}(\mathbb{R})$.

Lemma 3.5.3. *For $K_0 \in \mathcal{S}$, denote $K(t)$ as the solution of* (3.16)*. Then*

$$f(K_t) - f(K_*) \le e^{-rt}(f(K_0) - f(K_*)),$$
$$\|K_t - K_*\|_F^2 \le ce^{-rt}(f(K_0) - f(K_*)),$$

where r is a constant determined by the system parameters A, B, Q, R and K_0.

Proof. The proof proceeds similarly to Theorem 3.4.2. We only need to note that with respect to the Riemannian metric,

$$\begin{aligned}\dot{V}(K_t) = df(K_t)\dot{K}_t &= \langle \mathrm{grad} f(K_t), \dot{K}_t \rangle_{Y^\gamma} \\ &= -4\,\mathbf{Tr}(\mathbf{N}_t^\top \mathbf{N}_t Y_t^{2-\gamma}).\end{aligned}$$

According to Lemma 3.3.7, we now have,

$$\begin{aligned}\dot{V}(K_t) &\le -\frac{4\lambda_1(R + B^\top X_* B)\lambda_1(Y_t^{2-\gamma})}{\lambda_n(Y_*)} V(K_t) \\ &\le -\frac{4\lambda_1(R + B^\top X_* B)}{\lambda_n(Y_*)} V(K_t) =: r.\end{aligned}$$

The conclusion then follows. □

Remark. *Lemma 3.5.3 shows that the gradient descent* (3.13) *converges to the equilibrium point at an exponential rate $4\lambda_1(R + B^\top X_* B)\lambda_1^2(Y)/\lambda_n(Y_*)$. Hence, the natural gradient flow* (3.16) *modifies the exponential convergence rate of the gradient descent algorithm to $4\lambda_1(R + B^\top X_* B)\lambda_1(Y^{2-\gamma})/\lambda_n(Y_*)$, by a constant factor of $\lambda_1(Y^{2-\gamma})/\lambda_1^2(Y)$. This factor depends on the largest and smallest eigenvalues of the matrix Y. For example, if $\gamma \ge 3$, then $\lambda_1(Y^{2-\gamma}) = 1/\lambda_n(Y^{\gamma-2})$.*

Remark. *We note that the convergence rate of trajectory K_t is dependent on $\lambda_1(Y)$ and $\lambda_n(Y)$. For example, when $\gamma = 1$ and $\mathbf{\Sigma} = 2I$, then the natural gradient flow converges faster than the gradient flow since $\lambda_1(Y) > 1$. On the other hand, if $\gamma = 1$ and $\lambda_1(Y) < 1$, then gradient flow converges faster than natural gradient flow.*[33] *Simulation results in §3.7 show that this parameterized gradient flow offers a significant computational advantage for LQR.*

[33] This can be done by an $\mathbf{\Sigma}$ that has a spectrum bounded by 1.

We remark that in the particular case of $\gamma = 1$, the natural gradient flow has a favorable property with respect to the induced flow on the value matrix X_t. Consider again the flow,

$$\dot{K}_t = -2(RK_t - B^\intercal X A_{K_t}), \tag{3.17}$$

inducing the flow over the "value" matrix $X_t \coloneqq X(K_t)$ given by,

$$\dot{X}_t = \frac{dX_t}{dK}\dot{K}_t. \tag{3.18}$$

Lemma 3.5.4. *For $K_0 \in \mathcal{S}$, the gradient flow* (3.17) *induces a well-posed flow over the positive semidefinite cone X_t* (3.18)*. Moreover, the trajectory $\{X_t\}$ is monotonically decreasing in Loewner ordering.*

Proof. The well-posedness follows from the well-posedness of $\{K_t\}$. To show that the trajectory is monotonically decreasing, it suffices to observe,

$$\begin{aligned} \dot{X}_t = \frac{dX_t}{dK}\dot{K}_t &= \sum_{j=0}^{\infty} \left(A_{K_t}^\intercal\right)^j \left(\dot{K}_t^\intercal \mathbf{N}_t + \mathbf{N}_t \dot{K}_t\right)\left(A_{K_t}^j\right) \\ &= -\sum_{j=0}^{\infty} \left(A_{K_t}^\intercal\right)^j \left(2\mathbf{N}_t^\intercal \mathbf{N}_t + 2\mathbf{N}_t \mathbf{N}_t^\intercal\right) A_{K_t}^j \preceq 0. \end{aligned}$$

The second equality uses the Fréchet derivative of the map $K \mapsto X_K$ and the computation can be found in [11]. ☐

Note that this monotonicity does not hold in general for gradient flow: in this case the flow is dictated by $\mathbf{\Sigma}$ and along the trajectory, one can only guarantee that the function value $\mathbf{Tr}(X_t\mathbf{\Sigma})$ decreases.

3.5.1 Discretization of Natural Gradient Flow

In this section, we delve into the discretization of natural gradient flow; we shall only consider the case when $\gamma = 1$.[34] Specifically, we consider the gradient flow,

$$\dot{K}_t = -2(RK_t - B^\intercal X A_K).$$

[34]Other choices can be analyzed in a similar manner.

The forward Euler discretization yields,

$$K_{j+1} = K_j - 2\eta_j(RK_j - B^\intercal X_j A_{K_j}), \tag{3.19}$$

where η_j is the stepsize to be determined. In discretizing gradient flow, our guideline is to choose a stepsize such that the function value is sufficiently decreased while keeping iterates stabilizing. However, in natural gradient flow with $\gamma = 1$, we observe that by Lemma 3.5.4: if we follow the natural gradient flow, the value matrix is monotonic with respect to the semidefinite cone. This essentially means that taking a sufficiently small stepsize in the direction of the natural gradient would guarantee a decrease in the value of the Lyapunov matrix solution $X_{t+\delta} \preceq X_\delta$. The reader is also referred to [24] (Lemma 15) where a similar stepsize for the natural gradient update has been derived).

Proposition 3.5.5. *Consider the sequence $\{K_j\}$ generated by* (4.8)*. Denote by $\{X_j\}$ the corresponding Lyapunov matrix solution with respect to K_j. If $\eta_j < 1/(\lambda_n(R) + B^\intercal X_j B)$, then K_j is stabilizing for every $j \geq 0$ and $X_{j+1} \preceq X_j$. In particular, $Z \coloneqq X_j - X_{j+1}$ solves the Lyapunov matrix equation,*

$$Z = A_{K_{j+1}} Z A_{K_{j+1}}^\intercal + \mathbf{N}_j^\intercal(-4\eta_j I + 4\eta_j^2(R + B^\intercal X_j B))\mathbf{N}_j. \tag{3.20}$$

Proof. First, we suppose that the sequence generated by the choice of η_j is in fact stabilizing (to be proved subsequently). Then Equation (3.20) follows from substituting $K_j - K_{j+1} = 2\eta_j \mathbf{N}_j$ into Comparison Lemma 3.3.5. Hence, if $-4\eta_j I + 4\eta_j^2 R + 4\eta_j^2 B^\intercal X_j B \preceq 0$, then $X_{j+1} \preceq X_j$. This can be guaranteed by choosing,

$$\eta_j < \frac{1}{\lambda_n(R + B^\intercal X_j B)}.$$

It now remains to show that if η_j is chosen as above, the sequence will be stabilizing. Suppose that K_j is stabilizing. Note that the sublevel set $\mathcal{S}_{K_j} \coloneqq \{K : f(K) \leq f(K_j)\}$ is compact and the ray $K_j - \zeta M_j$ intersects the boundary of $\mathcal{S}_{K_j}$ for some $\zeta = \zeta' > 0$; suppose that $K' = K_j - \zeta' M_j \in \partial\mathcal{S}_{K_j}$. But this implies that

$$[0, \frac{1}{\lambda_n(R + B^\intercal X_j B)}) \subseteq [0, \zeta'],$$

since otherwise, there would exist $s' < 1/\lambda_n(B^\top X_j B + R)$ such that $s' = \zeta'$ and $f(K_j - s'M_j) = f(K_j)$, contradicting $f(K_j - s'M) < f(K_j)$. □

The problem of determining the optimal stepsize can be done by minimizing the expression,

$$-4\eta_j I + 4\eta_j^2(R + B^\top X_j B) \leq 0,$$

over the positive semidefinite cone. This is equivalent to minimizing,

$$-4\eta_j + 4\eta_j^2\left(\lambda_n(R + B^\top X_j B)\right),$$

at $\eta_j \in [0, 1/\lambda_n(R+B^\top X_j B))$. Obviously, the optimal stepsize should be $\eta_j = 1/(2\lambda_n(R + B^\top X_j B))$. With this choice of stepsize, the function value converges linearly to the optimal value function.

Theorem 3.5.6. *If $\eta_j = 1/(2\lambda_n(R + B^\top X_j B))$, we have,*

$$\begin{aligned} f(K_j) - f(K_*) &\leq q_0^j(f(K_0) - f(K_*)), \\ \|K_j - K_*\|_F &\leq c_2 q_0^{j/2}. \end{aligned}$$

where $q_0 = (1 - 4\lambda_1(R))/(\lambda_n(Y_*)\lambda_n(R + B^\top X_0 B))$ and c_2 is some positive constant.

Proof. Putting $r_j = f(K_j) - f(K_*)$, we observe that with the chosen η_j,

$$\begin{aligned} r_j - r_{j+1} &= \mathbf{Tr}((X_j - X_{j+1})\mathbf{\Sigma}) \\ &\geq \mathbf{Tr}(\frac{1}{\lambda_n(R + B^\top X_j B)}\mathbf{N}_j^\top \mathbf{N}_j Y_{j+1}) \\ &\geq \frac{\|Y_{j+1}\|}{\lambda_n(R + B^\top X_j B)}\mathbf{Tr}(\mathbf{N}_j^\top \mathbf{N}_j) \\ &\geq \frac{\lambda_1(R)}{\lambda_n(Y_*)\lambda_n(R + B^\top X_j B)} r_j. \end{aligned} \tag{3.21}$$

It thus follows that,

$$r_{j+1} \leq (1 - \frac{\lambda_1(R)}{\lambda_n(Y_*)\lambda_n(R + B^\top X_j B)})r_j =: q_j r_j.$$

Note that by the choice of stepsize, $\{X_j\}$ monotonically decreases over the positive semidefinite cone and thus $q_j \le q_0$ for $j \ge 1$, where,

$$q_0 = 1 - \frac{\lambda_1(R)}{\lambda_n(Y_*)\lambda_n(R + B^\top X_0 B)} \le 1 - \frac{\lambda_1(Q)\lambda_1(R)}{f(K_0)\lambda_n(R + B^\top X_0 B)};$$

in the last inequality we have used the estimate $\|Y_*\| \le f(K_0)/\lambda_1(Q)$ (the proof can be found in [11]). Thereby,

$$f(K_j) - f(K_*) \le q_0^j \left(f(K_0) - f(K_*)\right).$$

The proof to the convergence of the iterates is almost identical to the one in Theorem 3.4.5 and thus omitted. $\square$

Remark. *We note that the discretization of natural gradient flow can perform better than gradient descent. One can monitor the one step progression $r_j - r_{j+1}$ to confirm such a behavior. This is distinct from the continuous flows as if $\lambda_1(Y) > 1$, then gradient flow performs better than natural gradient flow.*

3.6 Quasi-Newton Flow on $\mathcal{S}$

In this section, we motivate a quasi-Newton flow over the set of stabilizing feedback gains (policy) $\mathcal{S}$.[35] As observed previously, the Hessian of the LQR cost $f(K)$ is not positive definite everywhere. As such, there is no well-defined notion of (global) Newton iteration over policy space. However, examining the Comparison Lemma 3.3.5 allows us to derive a local second-order approximation of the LQR cost under the Riemannian metric Y. With is metric, recall that the gradient of f is,

$$\mathrm{grad} f(K) = 2(RK - B^\top X A_K).$$

[35] The justification for calling this evolution a quasi-Newton flow becomes apparent subseqeuntly.

We now observe the second-order approximation of the cost function and the proof can be found in [11].[36]

Proposition 3.6.1. *When K and $K+\Delta K$ are both stabilizing for sufficiently small ΔK,*[37] *then,*

$$\begin{aligned} f(K+\Delta K) = f(K) &+ \langle gradf(K), \Delta K\rangle_Y \\ &+ \langle \Delta K, (R+B^\intercal XB)(\Delta K)\rangle_Y + \mathcal{R}(\Delta K), \end{aligned}$$

where $\|\mathcal{R}(\Delta K)\|$, the remainder of the approximation, is $O(\|\Delta K\|^2)$.

Proposition 3.6.1 essentially states that we have a somewhat "good" local second-order approximation of $f(K)$ with respect to the Riemannian metric Y. We may now devise a flow to minimize $f(K)$ by minimizing this second-order approximation, namely,

$$\begin{aligned} \dot{K}_t &= -(R+B^\intercal X_t B)^{-1}\mathrm{grad}f(K_t) \\ &= -2(R+B^\intercal X_t B)^{-1}\mathbf{N}_t. \end{aligned}$$

The analysis presented in §3.4 and §3.5 allow us to obtain a streamlined proof of the convergence of this flow; as such, we omit the proof.

3.6.1 Discretization of Quasi-Newton Flow

The quasi-Newton flow over $\mathcal{S}$ has interesting consequences in terms of its discretization: the forward Euler leads to the iterative procedure

$$K_{j+1} = K_j - \eta_j(R+B^\intercal X_j B)^{-1}\mathrm{grad}f(K_j) \tag{3.22}$$

with stepsize η_j to be determined; we shall show that with constant stepsize $\eta = \frac{1}{2}$, both the function value and the iterates will converge quadratically to the optima.

---

[36]Proposition 3.6.1 can be considered as a slight extension of Lemma 6 in [24]. However, the emphasis in [24] was on the asymptotic behavior of the first-order approximation; this setup was subsequently utilized for a different purpose in [24]. For our purpose, it is important to prove that for the second-order approximation, the remainder of the approximation is $O(\|\Delta K\|^2)$.

[37]By openness of $\mathcal{S}$, if ΔK is sufficiently small, $K+\Delta K$ is stabilizing provided that K is.

Remark. *The update is consistent with the Gauss-Newton updates proposed in [24]. We have chosen to refer to this update as quasi-Newton in this chapter as it is obtained by minimizing a local second-order approximation of the LQR cost at each iteration.*

We first observe that if $\eta \le 1$, the corresponding sequence of value matrices $\{X_j\}$ is monotonically decreasing over the positive semidefinite cone.

Proposition 3.6.2. *Consider the sequence* $\{K_j\}$ *generated by* (3.22). *Denote by* $\{X_j\}$ *the corresponding Lyapunov matrix solution with respect to* K_j. *If* $\eta_j < 1$, *then* K_j *is stabilizing for every* $j \ge 0$ *and* $X_{j+1} \preceq X_j$. *In particular* $Z \coloneqq X_{j+1} - X_j \preceq 0$ *solves the Lyapunov matrix equation,*

$$Z = A_{K_{j+1}} Z A_{K_{j+1}}^\top + (-4\eta_j + 4\eta_j^2)\mathbf{N}_j^\top (R + B^\top X_j B)^{-1}\mathbf{N}_j.$$

Proof. Suppose that with $\eta_j < 1$, the sequence generated by (3.22) are all stabilizing.[38] Substituting the update rule (3.22) into Comparison Lemma 3.3.5 yields,

$$Z = A_{K_{j+1}}^\top Z A_{K_{j+1}} + (-4\eta_j + 4\eta_j^2)\mathbf{N}_j^\top (R + B^\top X_j B)^{-1}\mathbf{N}_j.$$

It is now clear if $\eta_j < 1$, then $X_{j+1} - X_j \prec 0$. To show the choice of η_j guaranteeing the stability of $A - BK_j$, we may follow almost the same argument as in the proof of Proposition 3.5.5. □

The optimal stepsize for the quasi-Newton iteration is obtained by minimizing the quantity $-4\eta + 4\eta^2$. As such, the optimal stepsize is $\eta_j = 1/2$ for every j. The corresponding update is then equivalent to,

$$\begin{aligned} K_{j+1} &= K_j - \frac{1}{2}(R + B^\top X_j B)^{-1} 2\mathbf{N}_j \\ &= K_j - K_j + (R + B^\top X_j B)^{-1} B^\top X_j A \\ &= (R + B^\top X_j B)^{-1} B^\top X_j A. \end{aligned} \tag{3.23}$$

[38] Similar to the proof to Proposition 3.5.5, we need this assumption to make sense of defining the corresponding value matrix sequence $\{X_j\}$.

Remark. *With the optimal choice of stepsize as $\eta = 1/2$, the quasi-Newton over K coincides with the Hewer' algorithm [31], obtained by considering the Newton iteration over the ARE. We have thus provided an alternative point view of this algorithm: the algorithm can be obtained directly over the policy space even without the ARE.*

Theorem 3.6.3. *With stepsize $\eta = 1/2$, the update* (3.23) *converges to the global minimum at a Q-quadratic rate. Namely, there exists constants $c > 0, c_3 > 0$, such that,*

$$\begin{aligned} f(K_j) - f(K_*) &\le c(f(K_{j-1}) - f(K_*))^2 \\ \|K_j - K_*\|_F &\le c_3 \|K_{j-1} - K_*\|_F^2. \end{aligned}$$

Remark. *As we have remarked, the quasi-Netwon iteration with stepsize* 1/2 *coincides with Hewer's algorithm. So basically the Theorem follows from the proof in [31]. But our viewpoint indeed allows us to deduce a rather different proof from the perspective of cost function properties. For example, we have already proved if stepsize is smaller than* 1*, then the feedback controller remains stabilizing, which is traditionally proved by a rather complicate comparison among iterates. To show there exist constants c, c_3 independent of iteration index, a continuity argument would suffice.*[39]

Proof. By Proposition 3.6.2, the sequence $\{K_j\}$ generated with stepsize 1/2 is stabilizing and the corresponding value matrices are monotonically decreasing. Thus, $X_j \to X_*$ and the set $\mathcal{E} := \{X_j\} \cup \{X_*\}$ is compact. Further, we have

$$\begin{aligned} X_{j+1} - X_* = \sum_{\nu=0}^{\infty} \Big((A_{j+1}^\top)^\nu (K_{j+1} - K_*)^\top \\ (R + B^\top X_* B)(K_{j+1} - K_*)(A_{j+1})^\nu \Big). \end{aligned} \tag{3.24}$$

Now observe that the map ϕ,

$$X \mapsto (R + B^\top X B)^{-1} B^\top X A,$$

[39]The constant can also be estimated in terms of system parameters explicitly (see [45] and [11] for details). However, they assume rather complicate algebraic form.

is smooth over the set of positive definite matrices. Thus ϕ is Lipschitz over the compact set $\mathcal{E}$. Denoting the corresponding Lipschitz constant as $\hat{c}$, we have,

$$\begin{aligned}
&\|K_{j+1} - K_*\|_F \\
&= \|(R + B^\top X_j B)^{-1} B^\top X_j A - (R + B^\top X_* B)^{-1} B^\top X_* A\|_F \\
&= \|\phi(X_{j+1}) - \phi(X_*)\|_F \le \hat{c}\|X_j - X_*\|_F.
\end{aligned}$$

Consequently,

$$\begin{aligned}
&f(K_{j+1}) - f(K_*) \\
&\le \mathbf{Tr}(Y_{j+1}(K_{j+1} - K_*)^\top (R + B^\top X_* B)(K_{j+1} - K_*)) \\
&\le \hat{c}^2 \|Y_{j+1}\|_2 \|R + B^\top X_* B\|_2 \|X_j - X_*\|_F^2 \\
&\le \hat{c}^2 \|Y_{j+1}\|_2 \|R + B^\top X_* B\|_2 \frac{1}{\lambda_1^2(\mathbf{\Sigma})} \mathbf{Tr}\left(((X_j - X_*)\mathbf{\Sigma})^2\right) \\
&\le \hat{c}^2 \|Y_*\|_2 \|R + B^\top X_* B\|_2 \frac{1}{\lambda_1^2(\mathbf{\Sigma})} \left(\mathbf{Tr}\left((X_j - X_*)\mathbf{\Sigma}\right)\right)^2 \\
&=: c\left(f(K_j) - f(K_*)\right)^2.
\end{aligned}$$

□

3.7 Simulation Results

In this section, we provide a representative set of examples to demonstrate the results reported in this chapter.

We first demonstrate the exponential stability of the proposed continuous flows. The system is of form (4.1) with parameters (A, B), $A \in \mathbb{R}^{100\times 100}$ and $B = I$, guaranteeing the controllability of the system. The entries of A are sampled from a standard normal distribution $\mathcal{N}(0, 1)$. We also scale A when necessary to make it stable such that the initial feedback gain can be set as $K_0 = 0$. The cost matrices Q, R are taken to be identity with appropriate dimensions. For the natural gradient, we simulate the flow with two different Riemannian

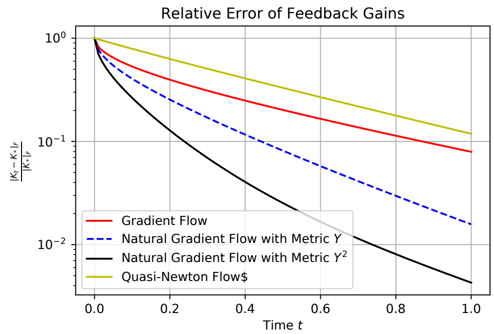

Figure 3.2: Exponential stability of trajectory K_t with the initial state matrix $\mathbf{\Sigma} = 0.5I$.

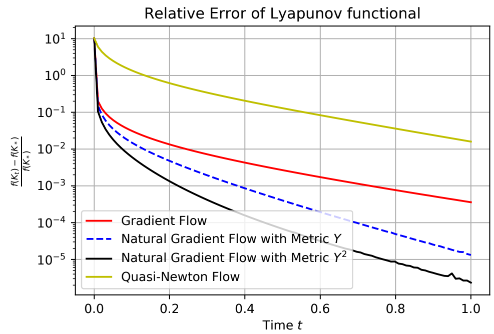

Figure 3.3: Exponential decay of the Lyapunov functional with the initial state matrix $\mathbf{\Sigma} = 0.5I$.

metrics, one induced by Y and the other by Y^2. Figure 3.2 demonstrates the exponential stability of the corresponding trajectories and Figure 3.3 depicts the exponential stability of the Lyapunov functionals for all flows when $\mathbf{\Sigma} = 0.5I$. The results are consistent with the observations discussed in §3.4, §3.5, and §3.6. In particular, since $\lambda_1(Y) < 1$, the natural gradient flow converges faster than the gradient flow, and amongst the natural gradient flows, the one with the metric induced by Y^2 outperforms the one with metric Y. Figures 3.4 and 3.5 show the convergence results with the same LQR parameters (A, B, Q, R), but the initial state matrix has chosen to be $\mathbf{\Sigma} = 2I$. These two figures underscore the observations in Remark 3.5: gradient flow outperforms natural gradient flows when $\lambda_1(Y) > 1$.

Next we examine the discrete realizations of these flows, namely, gradient descent, natural gradient descent and the quasi-Newton iteration (with the same setup for system parameters). With the adaptive stepsize proposed in Theorem 3.4.5, Figure 5.1 demonstrates that the sequence of feedback gains generated by gradient descent is stabilizing and converges to the global optimal feedback gain. Moreover, Figure 5.2 shows that the cost function $f(K)$ converges to $f(K_*)$ at a linear rate. In the meantime, Figures 4.3 and 4.4 demonstrate the linear convergence of the natural gradient descent algorithm. The stepsize is chosen adp-

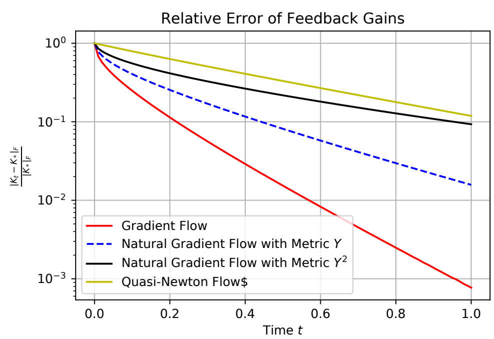

Figure 3.4: Exponential stability of trajectory K_t with the initial state matrix $\mathbf{\Sigma} = 2I$.

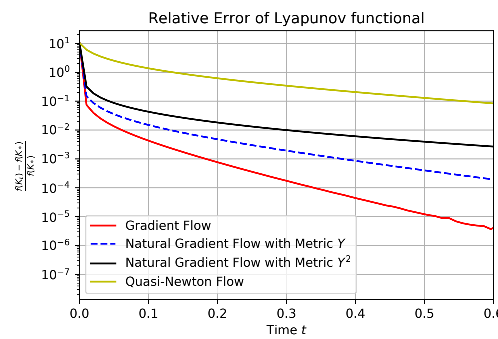

Figure 3.5: Exponential decay of the Lyapunov functional with the initial state matrix $\mathbf{\Sigma} = 2I$.

tively according to Theorem 3.5.6; we note the faster convergence of natural gradient descent compared with gradient descent. Figures 4.5 and 4.6 demonstrate the quadratic convergence for the quasi-Newton iteration. The stepsize is chosen to be 1/2; in this case, we recover the Hewer's algorithm, enjoying the fastest convergence rate.

3.8 Concluding Remarks

The chapter considers LQR through the lens of first order methods–an LQR calculus–where control synthesis is viewed directly in terms of optimizing an objective function over the set of stabilizing feedback gains. Using this narrative, we proceed to examine gradient descent and its various extensions for solving the LQR problem. The LQR objective is constructed over a set of linearly independent initial states to eliminate the dependency of the optimal policy on the initial state and encode closed loop stability. It is shown that the corresponding cost function is smooth, coercive and gradient dominated (this latter fact was previously reported in the literature; we provide an alternate approach for its proof). We next discussed three types of well-posed flows over the set of stabilizing controllers: gradient flow, natural gradient flow and the quasi-Newton flow. We subsequently examine the discretization of these flows,

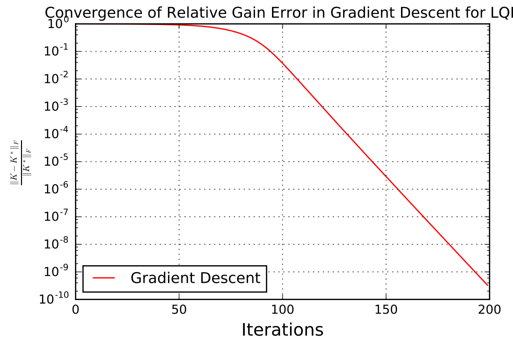

Figure 3.6: Convergence of the relative error for the feedback gain under gradient descent with adaptive stepsize given by 3.4.4

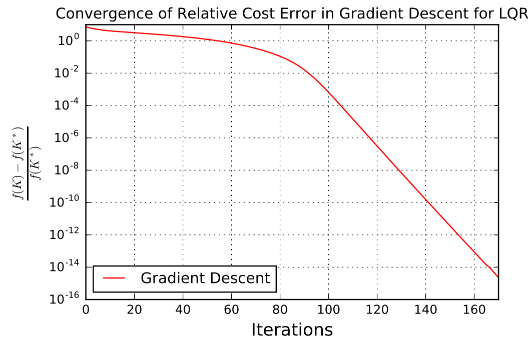

Figure 3.7: Convergence of the relative error for the LQR cost under gradient descent

and show that their realizations using the forward Euler method, i.e., gradient descent, natural gradient flow and quasi-Newton iterations, lead to algorithms with linear convergence rate and quadratic convergence rate, respectively. We postulate that adopting a direct policy update for control synthesis paves the way for more streamline approach for examining the intricate interplay between data, learning, and control. In the meantime, the approach hints at the use of direct policy updates for structured synthesis and online decision-making, not necessarily relying on the convex representation of the underlying control parameters.

In parallel, we can developed results for continuous LQR following the approaches adopted here. We refer the interested readers to our paper [15].

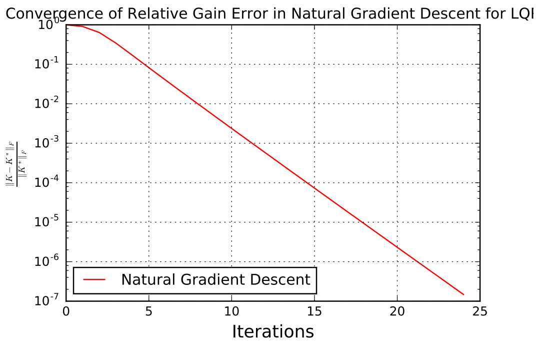

Figure 3.8: Convergence of the relative error for the feedback gain under natural gradient descent with adaptive stepsize given by 3.4.4

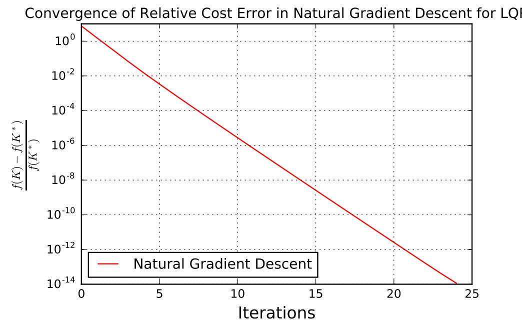

Figure 3.9: Convergence of the relative error for the LQR cost under natural gradient descent

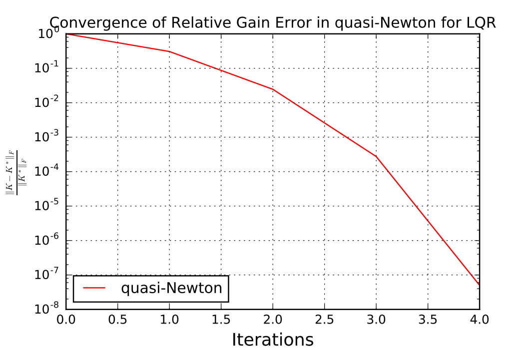

Figure 3.10: Convergence of the relative error for the feedback gain under quasi-Newton iteration with stepsize 1/2

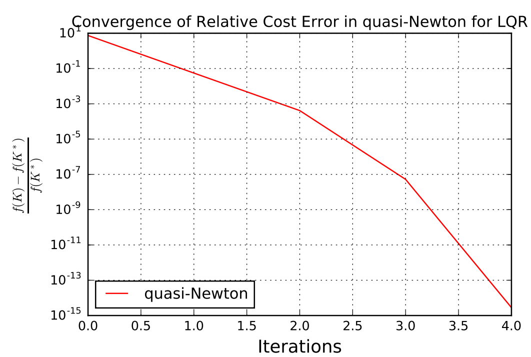

Figure 3.11: Convergence of the relative error for the LQR cost under quasi-Newton iteration with stepsize 1/2

Chapter 4

GLOBAL CONVERGENCE OF POLICY GRADIENT ALGORITHMS FOR INDEFINITE LEAST SQUARES STATIONARY OPTIMAL CONTROL

4.1 Introduction

[1]Least squares stationary optimal control provides an effective synthesis procedure for linear control systems since Kalman's original work in the 1960s [40]. This setting was later extended beyond positive semidefinite cost structure by Willems [78]. It is known that similar to standard LQR, this setup can be examined using the Algebraic Riccati Equation (ARE); DARE refers to the discrete analogue of this matrix equation. Historically, a large number of works have studied the solution of ARE and DARE, including approaches based on iterative algorithms [31],[2] algebraic solution methods [45], and semidefinite programming [2].

Although the cost function plays a fundamental role in the least squares optimal control, it is generally not "recommended" to *directly* compute the optimal gain (policy) using this cost without solving the associated Riccati equation. This approach, in the meantime, is in sharp contrast to how one would typically go about minimizing a cost function over the variable of interest in introductory optimization, say, through gradient descent. Recently, there has been a surge of interest in constructing optimal control strategies directly, viewing control synthesis through the lens of first order methods. Adopting such a point of view has been partially inspired by the application of learning algorithms in control, such as Reinforcement Learning (RL), where using principles of (approximate) dynamic programming, one can devise real-time model-free methods for both continuous-time and discrete-time op-

[1]The content of this chapter is published in [16]

[2]In Hewer's original work, the Q and R matrices are assumed to be positive definite. However, the algorithm still converges even for the indefinite cost structure [45].

timal control problems [8, 10, 19, 36, 46, 48, 51, 52]. The RL perspective not only provides more insights into the synthesis problem, but also can be extended to model-free settings by means of stochastic (zeroth-order) optimization [20, 57]. However, policy iteration is inherently prohibitive for an infinite horizon cost structure that is undiscounted and unbounded per stage [8].

The main contribution of this note is to extend policy based algorithms beyond positive (semi)definite cost structures considered in [11, 24]. More specifically, we show that under mild assumptions, even when the state and cost penalization matrices are indefinite in the least squares optimal control, gradient policy (respectively, natural gradient and quasi-Newton policies) converges to the global minimizer at a linear (respectively, linear and Q-quadratic) rate. Along the way, we devise a distinct approach for arguing the stability of the iterates as compared with those adopted in previous works.[3]

The note is organized as follows. In §4.2, we introduce the notation and preliminaries. §4.3 is devoted to the LQR setup, analytical properties of the cost function, a "mild" assumption, and its implications. In §4.4, we derive the corresponding stepsizes for gradient descent (GD), natural gradient descent (NGD), and quasi-Newton (QN) iterations; we then show the global linear (respectively, linear and Q-quadratic) convergence of gradient policy (respectively, natural gradient policy and quasi-Newton policy) under the proposed stepsizes. A numerical example is provided in §4.5 followed by concluding remarks in §5.9.

4.2 Notation and Preliminaries

We denote by $\mathbb{M}_{n\times m}(\mathbb{R})$ the set of $n \times m$ real matrices. $\mathbb{R}^n$ denotes the n-dimensional real Euclidean space; when $n = 1$, this set is identified with the set of real numbers. Other notation includes $A^\intercal$, $\rho(A)$, $\mathbf{Tr}(A)$, representing the transpose, spectral radius, and trace of the matrix A, respectively. The real inner product between a pair of vectors x and y is denoted by $\langle x, y\rangle$. $\|A\|_2$ denotes the spectral (operator) norm of a square matrix A and $\|A\|_F$

[3]The proposed technique also provides an alternative way to argue stability properties of the iterative process under standard LQR assumptions.

denotes its Frobenius norm.[4] Lastly, the notation $A \succeq B$ for two symmetric matrices refers to the positive semi-definiteness of their difference $A - B$; analogously for positive definiteness of this difference we use $A \succ B$. We let $\lambda_i(A)$ denote the eigenvalues of a square matrix A. These eigenvalues are indexed in an increasing order with respect to their real parts, i.e.,

$$\mathbf{Re}(\lambda_1(A)) \leq \cdots \leq \mathbf{Re}(\lambda_n(A)).$$

If A is symmetric, the ordering becomes $\lambda_1(A) \leq \cdots \leq \lambda_n(A)$. When $A \succeq 0$, $\|A\| = \lambda_n(A)$ and we shall use these interchangeably. We use $C^\omega(U)$ to denote the set of real analytic functions over an open set $U \subseteq \mathbb{R}^n$. A square matrix $A \in \mathbb{M}_{n\times n}(\mathbb{R})$ is *Schur* if $\rho(A) < 1$. A pair (A, B) is stabilizable if there exists some K for which $A - BK$ is Schur. Given a pair of system matrices (A, B), $\mathcal{S}$ denotes the set of Schur stabilizing feedback gains,

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

For the pair (A, B), we say that K is stabilizing if $A - BK$ is Schur; it is *marginally stabilizing* or *almost stabilizing* if $\rho(A - BK) = 1$. An eigenvalue λ of $A \in \mathbb{M}_{n\times n}(\mathbb{R})$ is called (C, A)-observable if

$$\operatorname{rank}\begin{pmatrix} A - \lambda I \\ C \end{pmatrix} = n,$$

for a given $C \in \mathbb{M}_{p\times n}(\mathbb{R})$; p is the dimension of the output of a linear system.

4.3 Problem Setup

In the standard least squares (stationary) optimal control, we consider a (discrete-time) linear time-invariant model of the form,

$$x_{k+1} = Ax_k + Bu_k, \tag{4.1}$$

where $A \in \mathbb{M}_{n\times n}(\mathbb{R})$, $B \in \mathbb{M}_{n\times m}(\mathbb{R})$ and (A, B) is stabilizable. The corresponding LQR problem is the optimization problem of devising a linear feedback gain $K \in \mathbb{M}_{m\times n}(\mathbb{R})$ for

[4] 2-norm is assumed when we use $\|.\|$.

which $u_k = -Kx_k$, minimizing,[5]

$$J(x_0) = \sum_{k=0}^{\infty} \left[\langle x_k, Qx_k \rangle + \langle u_k, Ru_k \rangle \right],$$

where x_0 is the initial condition, and the quadratic cost is parameterized by $Q = Q^\top$ and $R = R^\top$; note that we *do not* require positive (semi-)definiteness of Q and R. Such a generalization is not only of theoretical interest but also has important applications in network synthesis and stability theory [78]. In order to update the feedback gain (policy) directly, it is conceptually appealing to consider the cost as a matrix function over the set of feedback gains. With this aim in mind, we may define $J_{x_0}: \mathbb{M}_{m\times n}(\mathbb{R}) \to \mathbb{R}$ as,

$$J_{x_0}(K) = \sum_{j=0}^{\infty} \left[\langle (A - BK)^j x_0, (Q + K^\top RK)(A - BK)^j x_0 \rangle \right], \tag{4.2}$$

for some fixed initial condition $x_0 \in \mathbb{R}^n$. Note that the cost function J is a function of the policy K and initial condition x_0. Since we are interested in *optimal policy* independent of initial conditions, naturally, we should reformulate the cost function to reflect this independence. Indeed, this point has been discussed in [11] where it is argued that such a formulation is necessary for the cost function to be well defined (see details in §III [11]). The independence with respect to the initial condition can be achieved by either sampling x_0 from a distribution with a full-rank covariance [24], or choosing a spanning set $\{z_1, \ldots, z_n\} \subseteq \mathbb{R}^n$ [11], and defining the value function over $\mathcal{S}$ as,

$$f(K) = \sum_{i=1}^{n} J_{z_i}(K), \tag{4.3}$$

where $J_{z_i}(K)$ is the cost by choosing the initial state x_0 as z_i and letting $u_k = Kx_k$. Note that over the set $\mathcal{S}$, f admits a compact form $f(K) = \mathbf{Tr}(X\mathbf{\Sigma})$, where $\mathbf{\Sigma} = \sum_{i=1}^{n} z_i z_i^\top$ and X is the solution to the Lyapunov equation,

$$(A - BK)^\top X(A - BK) + Q + K^\top RK = X. \tag{4.4}$$

[5]The condition that u_k has the form $-Kx_k$ is not set a priori in the LQR formulation; this feedback form is typically shown via the adoption of a dynamic programming step.

The optimization problem of interest in this paper can now be stated as,

$$\text{minimize} \quad f(K) \quad \text{s.t.} \quad K \in \mathcal{S}.$$

How the cost function f behaves near the boundary $\partial\mathcal{S}$ is of paramount importance in the design of iterative algorithms for least squares optimal control problems. In the standard setting (with a definiteness assumption on the state and control costs), the cost function diverges to $+\infty$ when the feedback gain approaches the boundary of this set (see [11] for details). In fact, this property guarantees stability of the obtained solution via first order iterative algorithms for the suitable choice of stepsize. However, the behavior of f on the boundary $\partial\mathcal{S}$ could be more intricate for the indefinite cost structure. For example, if $K \in \partial\mathcal{S}$, i.e., $\rho(A - BK) = 1$, then it is possible that the cost is still finite. This happens when an eigenvalue of $A - BK$ on the unit disk in the complex plane is not $(Q + K^\intercal RK, A - BK)$-observable. To see this, we note that for every ω_i, the series

$$J_{\omega_i}(K) = \omega_i^\intercal \left(\sum_{j=0}^{\infty} ((A - BK)^\intercal)^j (Q + K^\intercal RK)(A - BK)^j \right) \omega_i,$$

is convergent to a finite (real) number if the marginally stable modes are not detectable. Even on $\bar{\mathcal{S}}^c$ (complement of closure of $\mathcal{S}$), f could be finite if all non-stable modes of $A - BK$ are not $(Q + K^\intercal RK, A - BK)$-observable. The complication suggests that the function value is no longer a valid proxy for stability. We remark that such a situation does not occur in the LQ setting examined in [11, 24], where it has been assumed that Q is positive definite.

4.3.1 Analytical properties of the indefinite cost function

In this section, we collect some useful analytic characterizations of $f(K)$. To simplify the notation, in the rest of this paper, we set,

$$A_K := A - BK, \quad \text{and} \quad \mathbf{N}_K := RK - B^\intercal X(A - BK);$$

when the context is clear, we will write $\mathbf{N}$ instead of $\mathbf{N}_K$; in describing the iterative process on the gain matrix (when K is updated), we shall denote $\mathbf{N}_{K_j}$ as $\mathbf{N}_j$.

Proposition 4.3.1. *The indefinite least squares optimal control problem (4.3) on the set of stabilizing feedback gains has the following properties:*

a. *The set $\mathcal{S}$ is regular open, contractible, and unbounded when $m \geq 2$ and the boundary $\partial\mathcal{S}$ [6] is precisely the set $\mathcal{B} = \{K \in \mathbb{M}_{m\times n}(\mathbb{R}) : \rho(A - BK) = 1\}$.*

b. *For the cost* (4.3)*, one has $f \in C^{\omega}(\mathcal{S})$.*

c. *The gradient of f (4.3) is given by*

$$\nabla f(K) = 2(RK - B^{\top} X A_K) Y_K,$$

where Y_K solves the Lyapunov matrix equation,

$$A_K Y A_K^{\top} + \mathbf{\Sigma} = Y. \tag{4.5}$$

d. *Let $K, \tilde{K} \in \bar{\mathcal{S}}$; suppose that the corresponding Lyapunov matrix equations* (5.2) *have symmetric solutions X and $\tilde{X}$, respectively.*[7] *Namely,*

$$A_K^{\top} X A_K + Q + K^{\top} R K = X,$$
$$A_{\tilde{K}}^{\top} \tilde{X} A_{\tilde{K}} + Q + \tilde{K}^{\top} R \tilde{K} = \tilde{X}.$$

Then we have

$$A_{\tilde{K}}^{\top}(X - \tilde{X}) A_{\tilde{K}} + (K - \tilde{K})^{\top} \mathbf{N}_K + \mathbf{N}_K^{\top}(K - \tilde{K})$$
$$- (K - \tilde{K})^{\top}(R + B^{\top} X B)(K - \tilde{K}) = X - \tilde{X}.$$

e. *Suppose that $K_* \in \arg\min_{K \in \mathcal{S}} f(K)$. Then*

$$\tau_1 \|K - K_*\|_F^2 \leq f(K) - f(K_*) \leq \tau_2 \langle \mathbf{N}_K, \mathbf{N}_K \rangle,$$

---

[6]The boundary of a set $\mathcal{E}$ is given by $\bar{\mathcal{E}} \setminus \mathcal{E}^{\circ}$, i.e., points belonging to the closure that are not in the interior. $\bar{\mathcal{E}}$ will denote the closure of a set $\mathcal{E}$.

[7]Note that the assumption clearly holds if $K, \tilde{K} \in \mathcal{S}$. It will also holds if $K \in \partial\mathcal{S}$ and the eigenvalues of $A - BK$ on the unit disk are not $(Q + K^{\top} R K, A - BK)$-observable.

where

$$\tau_1 = \lambda_1(Y)\lambda_1(R + B^\top XB),\ \tau_2 = \frac{\lambda_n(Y_*)}{\lambda_1(R + B^\top XB)},$$

and Y_* *solves the Lyapunov equation* (4.5) *with* K_*.

The proofs of these results can be found in [11]. We emphasize that (e) holds only if $\arg\min_{K\in\mathcal{S}} f(K) \neq \varnothing$, namely, there exists $K_* \in \mathcal{S}$ such that $f(K) \geq f(K_*)$ for every $K \in \mathcal{S}$. In the next subsection, we elaborate on a "mild" assumption to ensure that this condition holds.

4.3.2 A key assumption and its consequences

Throughout the manuscript, we have the following standing assumption.

Assumption 1. *There exists a* strict local minimizer *of* $f(K)$ *over* $\mathcal{S}$. *In other words, there exists some* $K_* \in \mathcal{S}$ *and an open neighborhood* $B_\delta(K_*) = \{K : \|K - K_*\|_F < \delta\}$, *such that* $f(K_*) < f(K)$ *for every* $K \in (B_\delta(K_*) \cap \mathcal{S}) \setminus \{K_*\}$.

Remark. *The seminal work of Willems [78] explores many facets of the least squares optimal control with indefinite* Q *and* R;[8] *in particular, this work examines conditions for which the above assumption holds. We will not discuss these conditions and instead refer the reader to [78] and references therein.*

We observe several implications of this assumption.

Proposition 4.3.2. *Suppose that* K_* *is the strict local minimizer of* $f(K)$ *over* $\mathcal{S}$ *and* X_* *is the corresponding value matrix. Then,*

a. $X_* = X_*^\top$,

b. $R + B^\top X_* B \succ 0$,

[8] An our adopted terminology is in his honor.

c. X_* *solves the DARE* (5.1),

$$X = A^\top XA + Q - A^\top XB(R + B^\top XB)^{-1}B^\top XA, \tag{4.6}$$

d. *The minimizer* K_* *is the unique global minimizer,*

e. X_* *is the maximal solution to DARE* (5.1) *and is unique among all* almost stabilizing solutions *of* (5.1).

Proof. Part (a) follows from having X_* solve the Lyapunov matrix equation (5.2) with $K = K_*$ and the fact that $Q + K^\top RK$ is symmetric. For parts (b) and (c), we first note that if K_* is a strict local minimizer in $\mathcal{S}$, since $f \in C^\omega(\mathcal{S})$, first-order and second-order optimality conditions imply $\nabla f(K_*) = 0$ and $\nabla^2 f(K_*) \succ 0$. By the Hessian formula in [11], we have $R + B^\top X_* B \succ 0$, i.e., (b) holds. Further, since $\nabla f(K_*) = \mathbf{N}_{K_*} Y_{K_*}$ and $Y_{K_*} \succ 0$, it follows that $N_{K_*} = 0$. Namely, $RK_* - B^\top X_* A_{K_*} = 0$. Substituting $K_* = (R + B^\top X_* B)^{-1} B^\top X_* A$ into the Lyapunov equation (5.2), we have X_* solves the DARE (5.1). For part (d), it suffices to observe that K_* is the unique stationary point. To this end, suppose that there exist $K_{*,1}$ and $K_{*,2}$ such that the gradient vanishes at both points, namely $\mathbf{N}_{K_{*,1}} = \mathbf{N}_{K_{*,2}} = 0$[9]. By part (d) in Proposition 5.3.1, we have

$$X_{*,1} - X_{*,2} = A_{K_{*,2}}^\top (X_{*,1} - X_{*,2}) A_{K_{*,2}} - (K_{*,1} - K_{*,2})^\top (R + B^\top X_{*,1} B)(K_{*,1} - K_{*,2}).$$

As $A_{K_{*,2}}$ is Schur, it follows that $X_{*,1} \succeq X_{*,2}$ and similarly $X_{*,2} \succeq X_{*,1}$. Hence, the stationary point is unique. Part (e) follows from standard DARE theory (see Chapters 12 and 13 in [45] for details.) □

4.4 Global Convergence of Policy Gradient Algorithms

In this section, we show the global convergence of gradient descent (GD), natural gradient descent (NGD), and quasi-Newton (QN) iterations for indefinite least squares optimal

[9] This follows from $Y_K \succ 0$ for every $K \in \mathcal{S}$.

control. In particular, under Assumption 1, it is shown that gradient descent (respectively, natural gradient descent and quasi-Newton) converges to the maximal solution of the DARE at a linear (respectively, linear and quadratic) rate. In this direction, first recall that the gradient, natural gradient and quasi-Newton directions [11] are given by,

$$\begin{aligned}\mathbf{g}(K) &\coloneqq 2(RK - B^\top X A_K)Y,\\ \mathbf{n}(K) &\coloneqq 2(RK - B^\top X A_K),\\ \mathbf{qn}(K) &\coloneqq 2(R + B^\top X B)^{-1}(RK - B^\top X A_K);\end{aligned}$$

GD, NGD and QN now refer to following update rules:

$$\text{GD}: \quad K_{j+1} = K_j - \eta_j \mathbf{g}(K_j), \tag{4.7}$$

$$\text{NGD}: \quad K_{j+1} = K_j - \eta_j \mathbf{n}(K_j), \tag{4.8}$$

$$\text{QN}: \quad K_{j+1} = K_j - \eta_j \mathbf{qn}(K_j), \tag{4.9}$$

where η_j's are stepsizes to be determined. We provide the convergence analysis for the case of natural gradient descent.

Theorem 4.4.1 (Natural Gradient Analysis)**.** *Consider the iterates* $\{K_j\}$ *generated by NGD* (4.8)*, with stepsize* $\eta_j = 1/(2\lambda_n(R + B^\top X_j B))$*, where* $\{X_j\}$ *solve the corresponding Lyapunov equations* (5.2)*. Then both the function values and gain iterates converge to their corresponding global minima at a linear rate. That is,*

$$\begin{aligned}f(K_j) - f(K_*) &\le q_1^j (f(K_0) - f(K_*)),\\ \|K_j - K_*\|_F^2 &\le c_1 q_1^j \|K_0 - K_*\|_F^2,\end{aligned}$$

for some $q_1 \in (0,1)$ *and* $c_1 > 0$*.*

Proof. The analysis provided in [11] for the one-step progression of NGD holds here; thus the convergence rate would remain the same if we can prove that the iterates remain stabilizing.

By induction, it suffices to argue that with the chosen stepsize, K_j is stabilizing provided that K_{j-1} is. Consider the ray $\{K_t = K_{j-1} - t\mathbf{n}(K_{j-1}) : t \ge 0\}$. Note that by openness of $\mathcal{S}$

and continuity of eigenvalues, there is a maximal interval $[0,\zeta)$[10] such that $K_{j-1}+t\mathbf{n}(K_{j-1})$ is stabilizing for $t\in[0,\zeta)$ and $K_{j-1}+\zeta\mathbf{n}(K_{j-1})$ is marginally stabilizing. Now suppose that $\zeta \le 1/(2\lambda_n(R_1+B_1^\top X_{i-1}B_1))$; take a sequence $t_l\in[0,\zeta)$ such that $t_l\to\zeta$. Consider the sequence of value matrices $\{X_{t_l}\}$ and denote by $\mathcal{L}$ as the set of all limit points of $\{X_{t_j}\}$. Observe that $X_*\le X_{t_l}\le X_{j-1}$. By Bolzno-Weierstrass [66], $\mathcal{L}$ is nonempty.[11] By continuity, any $Z\in\mathcal{L}$ solves,

$$Z=(A-BK_\zeta)^\top Z(A-BK_\zeta)+Q+K_\zeta^\top RK_\zeta.$$

Now by part (d) in Proposition 5.3.1, we have

$$Z-X_*=(A-BK_\zeta)^\top(Z-X_*)(A-BK_\zeta)+(K_\zeta-K_*)^\top(R+B^\top X_*B)(K_\zeta-K_*).$$

Suppose that (λ,v) is the eigenvalue-eigenvector pair of $A-BK_\zeta$ such that $(A-BK_\zeta)v=\lambda v$ and $|\lambda|=1$. Then it follows that,

$$v^\top(Z-X_*)v=v^\top(A-BK_\zeta)^\top(Z-X_*)(A-BK_\zeta)v+v^\top(K_\zeta-K_*)^\top(R+B^\top X_*B)(K_\zeta-K_*)v.$$

Thereby $(K_\zeta-K_*)v=0$ and $K_\zeta v=K_*v$. Consequently, $(A-BK_*)v=(A-BK_\zeta)v$. But this is a contradiction to the assumption that K_* is a stabilizing solution.

Hence $\{X_j\}$ is a monotonically non-increasing sequence bounded below by X_*. As such, the sequence of iterates $\{K_j\}$ and the sequence of function values $\{f(K_j)\}$ converge linearly to K_* and $f(K_*)$ following the arguments in [11]. □

We mention that the above stability argument can be applied for the sequence generated by the quasi-Newton iteration as well. The quadratic convergence rate for such a sequence would then follow from the proof in [11].

---

[10] We suppose that ζ is finite; if ζ is infinite, there is nothing left to prove.

[11] Note that it is not guaranteed that X_{t_j} is convergent. The limit points are also not necessarily well-ordered in the ordering induced by the p.s.d. cone.

Theorem 4.4.2 (Quasi-Newton Analysis)**.** *Suppose Assumption* 1 *holds. Consider the iterates* $\{K_j\}$ *generated by QN* (4.9) *with stepsize* $\eta_j = 1/2$*. Then both the function values and iterates converge to their respective global minima at a Q-quadratic rate. That is,*

$$f(K_j) - f(K_*) \leq q_2(f(K_{j-1}) - f(K_*))^2,$$
$$\|K_j - K_*\|_F^2 \leq c_2 q_2 \|K_{j-1} - K_*\|_F^4,$$

for some $q_2 > 0$ *and* $c_2 > 0$*.*

4.4.1 Convergence of Gradient Policy

The gradient policy analysis requires more work since the stepsize developed in [11] involves the smallest eigenvalue $\lambda_1(Q)$. However by carefully replacing "$\lambda_1(Q)$-related quantities" in [11], one can still prove the global linear convergence rate as follows.

Theorem 4.4.3 (Gradient Analysis)**.** *Suppose Assumption* 1 *holds. Consider the iterate* $\{K_j\}$ *generated by GD* (4.7) *with stepsize* η_j *specified in Theorem 4.4.6. Then both the function values and iterates converge to their respective global minima at a linear rate. That is,*

$$f(K_j) - f(K_*) \leq q_3^j(f(K_0) - f(K_*)),$$
$$\|K_j - K_*\|_F^2 \leq c_3 q_3^j \|K_0 - K_*\|_F^2,$$

for some $q_3 \in (0,1)$ *and* $c_3 > 0$*.*

This subsection is devoted to the proof of Theorem 4.4.3. As it was pointed out previously, the strategy adopted in [11,24] are no longer viable for an indefinite cost structure. However, as we will show, a perturbation bound would circumvent this issue and allows deriving the required stepsize, guaranteeing a decrease in function values while ensuring stabilization.

In the following, we shall drop all the subscripts as the stepsize will be valid for every iterate. Suppose now that we have a stabilizing policy K and the gradient direction is given by $\mathbf{g}(K) = 2\mathbf{N}Y$.[12] The main object that we work with in this section is the ray starting at K along the gradient direction,

$$\{K_\eta : K - \eta\mathbf{g}(K), \eta \geq 0\}.$$

We shall further denote $A_\eta = A - BK_\eta = A - B(K - \eta 2\mathbf{N}Y)$.

Here is an outline of our proof strategy:

a. By openness of $\mathcal{S}$ and continuity of eigenvalues, there exists a maximal interval $[0, c)$ such that K_η is stabilizing for every $\eta < c$ and K_c is marginally stabilizing; such a c could be either finite or infinite.

b. Now suppose that c above is known. Then for every $\eta < c$, $f(K_\eta)$ is well-defined and we can compute the difference,

$$f(K) - f(K_\eta) = 4\eta\,\mathbf{Tr}\left(\mathbf{N}^\top\mathbf{N}(YY_\eta - \eta aYY_\eta Y)\right),$$

where $a = \lambda_n(R + B^\top XB)$, and Y_η solves the Lyapunov matrix equation (4.5) with $A_K = A - BK_\eta$.

c. Next we define a univariate function $\phi : [0, c) \to \mathbb{R}$ by,

$$\phi(\eta) = \mathbf{Tr}\left(\mathbf{N}^\top\mathbf{N}(YY_\eta - \eta aYY_\eta Y)\right).$$

Note that $\phi(0) > 0$ if the gradient does not vanish at K. Now our goal is to characterize a step size $0 < \eta' < c$ such that $\phi(\eta') > 0$.

It is clear that the knowledge of c and characterizing η' above are crucial for stepsize analysis. We shall demonstrate that characterizing η' will suffice to provide a stepsize; the quadratic cost structure will implicitly enforce stabilization.

To begin, we observe a perturbation bound on Y_η, assuming that K_η is stabilizing.

---

[12] Note that the subscripts have been dropped; $\mathbf{N}$ and Y are both dependent on K.

Proposition 4.4.4. *Put* $\mu_1 = \|Y\|_2\|B\mathbf{N}Y\|_2^2/\lambda_1(\boldsymbol{\Sigma})$ *and* $\mu_2 = \|Y\|_2\|B\mathbf{N}Y\|_2\|A_K\|_2/\lambda_1(\boldsymbol{\Sigma})$, *and let*

$$\eta_0 = \frac{\sqrt{\mu_1 + \mu_2^2}}{4\mu_1} - \frac{\mu_2}{4\mu_1};$$

suppose that A_η *is Schur stable for every* $\eta \le \eta_0$. *Then for all* $\eta \le \eta_0$, *we have* $\|Y_\eta\|_2 \le \beta_0\|Y\|_2$ *with* $\beta_0 = 1/(1 - 4\mu_1\eta_0^2 - 4\mu_2\eta_0) > 0$.

Proof. Putting $\boldsymbol{\Upsilon} = B\mathbf{N}Y$ and taking the difference of the corresponding Lyapunov equations, we have

$$\begin{aligned} Y_\eta - Y - A_K(Y_\eta - Y)A_K^\top &= 2\eta\left(A_K Y_\eta \boldsymbol{\Upsilon}^\top + \boldsymbol{\Upsilon} Y_\eta A_K^\top\right) + 4\eta^2 \boldsymbol{\Upsilon} Y_\eta \boldsymbol{\Upsilon}^\top \\ &\le \|Y_\eta\|_2\left(4\eta\|\boldsymbol{\Upsilon}\|_2\|A_K\|_2 + 4\eta^2\|\boldsymbol{\Upsilon}\|_2^2\right) I \\ &\le \|Y_\eta\|_2\left(4\eta\|\boldsymbol{\Upsilon}\|_2\|A_K\|_2 + 4\eta^2\|\boldsymbol{\Upsilon}\|_2^2\right) \boldsymbol{\Sigma}/\lambda_1(\boldsymbol{\Sigma}). \end{aligned}$$

It thus follows that,

$$Y_\eta - Y \le \frac{\|Y_\eta\|_2\left(4\eta\|\boldsymbol{\Upsilon}\|_2\|A_K\|_2 + 4\eta^2\|\boldsymbol{\Upsilon}\|_2^2\right)}{\lambda_1(\boldsymbol{\Sigma})} Y.$$

Hence,

$$\|Y_\eta\|_2\left(1 - \frac{\|Y\|_2\left(4\eta\|B\mathbf{N}Y\|_2\|A_K\|_2 + 4\eta^2\|B\mathbf{N}Y\|_2^2\right)}{\lambda_1(\boldsymbol{\Sigma})}\right) \le \|Y\|_2.$$

The proof is completed by a direct computation showing that $1/\beta_0 = 1 - \mu_1\eta_0^2 - 4\mu_2\eta_0 > 0$ with the choice of η_0 and noting that for every $\eta \le \eta_0$,

$$1 - 4\mu_1\eta^2 - 4\mu_2\eta \ge 1 - 4\mu_1\eta_0^2 - 4\mu_2\eta_0.$$

□

The next lemma shows that if c is known, a positive stepsize can be chosen.

Lemma 4.4.5. *Let* c *be the largest real positive number such that* A_t *is Schur stable for every* $t \in [0, c)$ *and* A_c *is marginally Schur stable.*[13] *Let*

$$a_1 = a\beta_0\|Y\|_2 + 4\|\mathbf{N}\|_2\beta_0\|Y\|_2^2,\ a_2 = a4\|\mathbf{N}\|_2\beta_0\|Y\|_2^2;$$

[13]Here we have assumed that c is not $+\infty$. Of course, if $c = +\infty$, then any stepsize would lead to a stabilizing update.

then with $\eta_1 \le \min(c-\varepsilon, \eta_0, c_0)$*, where* $\varepsilon > 0$ *is an arbitrary positive real number and*

$$c_0 < \sqrt{\frac{1}{a_2} + \frac{a_1^2}{4a_2^2}} - \frac{a_1}{2a_2},$$

one has $\phi(\eta_1) \ge 0$*.*

Proof. The computation follows a similar method used in [11] by replacing the estimate of $Y(\theta)$ by the bound in the above proposition (see details in Lemma 5.5 in [11]). □

Finally, we show that $c > \min(\eta_0, c_0)$. This would then imply that one can choose the stepsize as $\eta = \min(\eta_0, c_0)$.

Theorem 4.4.6. *With the stepsize* $\eta = \min(\eta_0, c_0)$*,* M_η *remains stabilizing and* $\phi(\eta) \ge 0$*.*

Proof. Let $\eta = \min(\eta_0, c_0)$. It suffices to prove that for every $t \in [0, \eta]$, A_t is Schur stabilizing and $\phi(t) \ge 0$. We prove this by contradiction. Suppose that this is not the case. Then by continuity of eigenvalues, there exists a number $\eta' \le \eta$ such that A_s is stabilizing for every $s \in [0, \eta')$ and $K_{\eta'}$ is marginally stabilizing. If this is the case, the choice of η_0, c_0 guarantees that for every $s \in [0, \eta')$, $\phi(s)$ is well-defined and $\phi(s) \ge 0$. Now take a sequence $t_i \to \eta'$ and consider the corresponding sequence of value matrices $\{X_{t_i}\}$. Note that the sequence of function values $\mathbf{Tr}(X_{t_i}\mathbf{\Sigma})$ satisfies

$$\mathbf{Tr}(X_*\mathbf{\Sigma}) \le \mathbf{Tr}(X_{t_i}\mathbf{\Sigma}) \le \mathbf{Tr}(X\mathbf{\Sigma})$$

since $\phi(t) \ge 0$. But this implies that $\{X_{t_i}\}$ is a bounded sequence (note that the above inequality on function values does not guarantee the boundedness of the sequence; it is crucial that $X_{t_i} \ge X_*$). Hence by a similar argument adopted in the proof of Theorem 4.4.1, these observations establish a contradiction; as such, the proposed stepsize guarantees stabilization.

□

It is now straightforward to conclude the convergence rate of Theorem 4.4.3 by similar arguments as in [11].[14]

[14] Strictly speaking, we need to show our proposed stepsizes are bounded away from 0. Namely, that there is some constant $d > 0$ such that $\eta_j > d$ for every j. The computations are omitted here due to space limitation. In the meantime, one can be convinced of this fact by checking the asymptotics of η_0 and c_0.

Remark. *We briefly comment on the per iteration computational complexity of the above three algorithms. For model-based design, gradient and natural gradient descent involve solving Lyapunov matrix equations; quasi-Newton direction, on the other hand, requires a matrix inversion in addition to solving a Lyapunov equation. We refer the reader to [69] and references therein for detailed computational treatment for solving Lyapunov equations.*

4.5 A Numerical Example

In this section, we show the proposed convergence results by a numerical example. The system parameters are $A = 0.5I$, $B = I$, $R = I$ and

$$Q = \begin{pmatrix} 1.62370842 & 0.36712592 & -1.31209102 & 1.97803823 & -0.49297266 \\ 0.36712592 & 2.21878741 & 0.47525552 & -1.07142839 & 1.04343275 \\ -1.31209102 & 0.47525552 & 1.90887732 & -0.83057818 & 0.3818043 \\ 1.97803823 & -1.07142839 & -0.83057818 & 0.93847322 & -0.90779531 \\ -0.49297266 & 1.04343275 & 0.3818043 & -0.90779531 & -1.06295748 \end{pmatrix}.$$

Note that Q is indefinite and its (rounded) eigenvalues are $4.75, 2.55, 0.96, -1.1, -1.53$. Figures 5.1-5.2 show the global linear convergence of the gradient policy update. The global linear convergence of natural gradient policy are demonstrated in Figures 4.3-4.4. Figures 4.5-4.6 show the Q-quadratic convergence for the quasi-Newton policy update.

4.6 Concluding Remarks

This note considers policy gradient algorithms for the indefinite least squares stationary optimal control, e.g., indefinite LQR. We show the global linear (respectively, linear and Q-quadratic) convergence of gradient policy (respectively, natural gradient and quasi-Newton policies). Although these results are presented assuming the knowledge of the system matrices, gradient and natural gradient policies can be extended to model-free case by means of stochastic (zeroth order) optimization (see [24] for details). As such, this note extends the

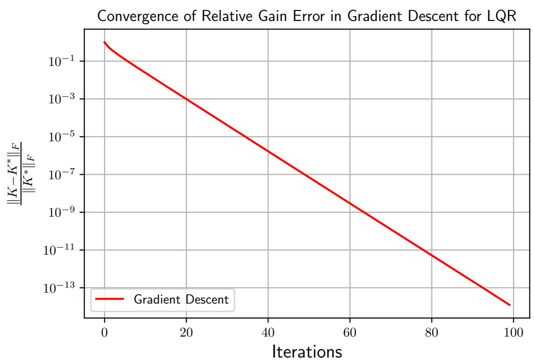

Figure 4.1: Convergence of the relative error for the feedback gain under gradient descent iteration

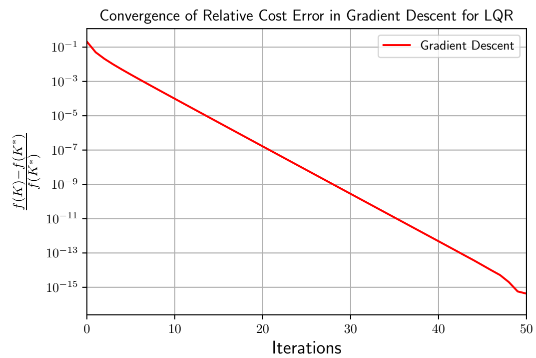

Figure 4.2: Convergence of the relative error for indefinite LQR cost under gradient descent iteration

results reported in [11, 24] for indefinite LQR. These extensions have important implications for optimal control, stability analysis and LQ games. Indeed, some of these observations have been utilized to show global convergence of sequential policy updates in LQ dynamic games [17].

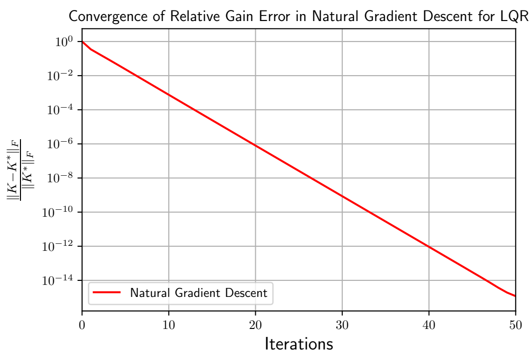

Figure 4.3: Convergence of the relative error for the feedback gain under natural gradient descent iteration

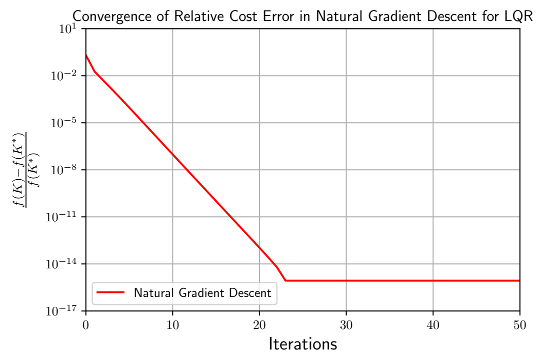

Figure 4.4: Convergence of the relative error for indefinite LQR cost under natural gradient descent iteration

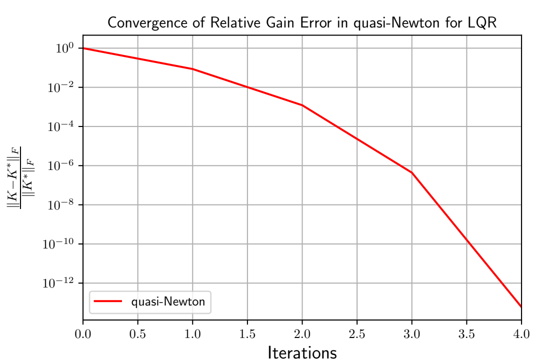

Figure 4.5: Convergence of the relative error for the feedback gain under quasi-Newton iteration with constant stepsize 1/2

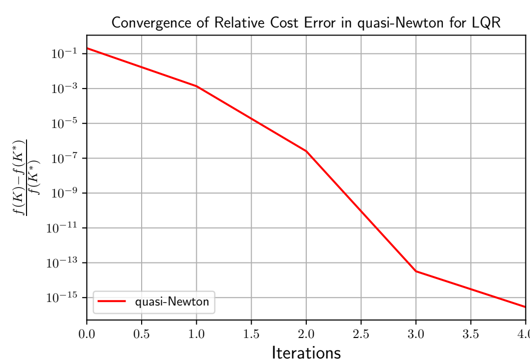

Figure 4.6: Convergence of the relative error for indefinite LQR cost under quasi-Newton iteration with constant stepsize 1/2

Chapter 5

GLOBAL CONVERGENCE OF POLICY GRADIENT FOR SEQUENTIAL ZERO-SUM LINEAR QUADRATIC DYNAMIC GAMES

5.1 Introduction

Linear-quadratic (LQ) dynamic and differential games exemplify situations where two players influence an underlying linear dynamics in order to respectively, minimize and maximize a given quadratic cost on the state and the control over an infinite time-horizon.[1] The LQ game setup has a rich history in system and control theory due to its wide range of applications [6, 22, 82]. In particular, $\mathcal{H}_\infty$ optimal control can be interpreted as a two player zero-sum LQ game [3]. As such, LQ games are generally approached via *generalized algebraic Riccati equation* (GARE) derived from optimal control theory [72]. Adopting a solution approach based on the Riccati equation, in the meantime, has broadly influenced the "data-driven" approaches for solving the generic LQ problem and its extensions. For instance, in the value-iteration for reinforcement learning (RL)—e.g., Q learning—one aims to first estimate the cost-to-go at a given time instance and through this estimate, update the state feedback gain.

Recently, there is a renewed interest in analyzing the classical LQ problem under the RL framework from the perspective of direct policy updates [11,21,24]. The conceptual simplicity of policy optimization offers advantages in terms of computational scalability, extensions to model-free settings and richer parametrization of feedback policies (e.g., structured controller design [11]). The conceptual simplicity of policy optimization offers advantages in terms

[1]We will adopt the convention of referring to the continuous time scenario as differential games. Moreover, in this paper, we focus on infinite horizon LQ games without a discount factor.

of computational scalability, extensions to model-free settings and rich parametrization of feedback policies (e.g., structured controller design [11]). For LQ games, sequential policy updates are relevant to the realm of multi-agent RL [35, 70]. Hence, understanding the performance of policy gradient based algorithms for LQ games is not only beneficial for the optimal control problem itself, but also serves as a benchmark in providing deeper insights into theoretical guarantees of multi-agent RL in more general settings [75].

However, the application of policy optimization algorithms in the game setting requires more intricate analysis due to the fact that the infinite horizon cost is undiscounted and unbounded per stage. Indeed, it is well known that devising direct policy iterations for undiscounted and unbounded per stage cost functions in the RL setting is nontrivial [8]. The cost structure of standard LQR, however, streamlines the design of policy based iterations [11, 24, 32].[2] Nevertheless in policy iteration, special care has to be exercised to ensure that the iterative policy updates are in fact stabilizing. The stabilization issues are particularly relevant in the LQ dynamic games as the policy space for LQ games is an open set admitting a cartesian product structure. Hence, in the policy updates for LQ dynamic games (say, in the RL setting), we must guarantee that the iterates jointly stay in the open set as otherwise the "simulator" would diverge. Recently, [81], using certain assumptions and relying on a projection step, have proposed a sequential direct policy updates with a sublinear convergence for LQ dynamic games.

Contributions. In this paper, we first clarify the setting for discussing sequential LQ dynamic games, particularly addressing issues pertaining to stabilization. In the process, we identify assumptions pertaining to $\mathcal{H}_\infty$ optimal control (e.g., definiteness assumptions of Q, R_1, R_2) are indeed unnecessary from a purely dynamic game perspective. We thus consider LQ game with indefinite cost structure Q, R_1, R_2. We then propose leader-follower type algorithms that resemble the Stackelberg leadership model [4]. Specifically, in the proposed

[2]In [11, 24] it has been assumed that the quadratic state cost is via a positive definite Q; this is not "standard" as only the detectability of the pair (Q, A) and $Q \geq 0$ is required for LQR synthesis.

iterative algorithms for LQ games, one player is designated as a leader and the other as a follower. We require that the leader plays natural gradient or quasi-Newton policies while the follower can play any first-order based policies. In particular, we do not require a specific player to be the leader as long as the algorithm can be initialized appropriately. We prove that if the leader performs a natural gradient policy update, then the proposed leader-follower algorithm has a global sublinear and asymptotic linear convergence rate. In the meantime, if the leader adopts a quasi-Newton policy update, the algorithm converges at a Q-quadratic rate [3]. Compared with the results presented in [81], the contributions of this work include the following:

(1) We remove the "nonstandard assumption" that the NE point (K_*, L_*) satisfies $Q - L_*^\intercal R_2 L_* \succ 0$.[4] We note that such an assumption is not standard in the LQ literature and as such, needs further justification beyond its algorithmic implications. In fact, in the analysis presented in [81], it is crucial for the convergence of the algorithm to a priori know the positive number $\varepsilon > 0$ for which $\{L : Q - L_*^\intercal R_2 L_* \succeq \varepsilon I\}$; moreover, one has to be able to project onto this set.

(2) Our setting allows for larger stepsizes for the policy iteration in LQ dynamic games, greatly improving its practical performance. This is facilitated by providing insights into the stabilizing policy updates through a careful analysis of the corresponding indefinite GARE.

(3) We clarify the interplay between key concepts in control and optimization in the convergence analysis of the proposed iterative algorithms for LQ dynamic games. This is inline with our belief that identifying the role of concepts such stabilizability and detectability in the convergence analysis of "data-guided" algorithms for decision making problems with an embedded dynamic system is of paramount importance.

[3] We adopt the definition for Q-quadratic used in [58]

[4] It is noted that in [81] that the projection step is not generally required in implementations; however the convergence analysis presented in [81] is based on such a projection.

(4) We show that the quasi-Newton policy has a Q-quadratic convergence rate for LQ dynamic games. This result might be of independent interest for discrete-time GARE, data-driven or not. To the best of our knowledge, the proposed algorithm is the first iterative approach for discrete-time GARE with a Q-quadratic convergence.

(5) Finally, we show that in the proposed iterative algorithms for LQ dynamic games, any player can assume the role of the "leader" whereas in [81], it is required that the player maximizing the cost be designated as the leader. As such, we clarify the algorithmic source of asymmetry in the leader-follower setup for solving this class of dynamic game problems.

The paper is structured as follows. in § 5.2 we introduce notations and preliminaries. § 5.3 devotes to the basic LQ game setup, the analytical properties of the cost function and issues pertainining to stability when sequential algorithms are of consideration. § 5.4 introduces the oracle model adopted in this paper and its motivation. § 5.5 proposes natural gradient *ascent* for leader L and proves its global convergence. § 5.6 discusses quasi Newton iteration and its global Q-quadratic convergence. § 5.7 we discuss natural gradient *descent* and quasi Newton iteration for leader L. A set of representative numerical examples are supplied in § 5.8. We conclude the paper in § 5.9.

5.2 Notation and Preliminaries

We use the symbols $\prec, \preceq, \succ, \succeq$ to denote the ordering induced by the positive semidefinite (p.s.d.) cone. Namely, $A \succeq B$ means that $A - B$ is positive semidefinite. For a symmetric matrix $M \in \mathbb{R}^{n\times n}$, we denote the eigenvalues in non-increasing order, i.e., $\lambda_1(M) \leq \ldots \leq \lambda_n(M)$.

Let us recall relevant definitions and results from control theory. A matrix $A \in \mathbb{R}^{n\times n}$ is Schur if all the eigenvalues of A are *inside* the *open unit disk* of $\mathbb{C}$, i.e., $\rho(A) < 1$ where $\rho(\cdot)$ denotes the spectral radius. A pair (A, B) with $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ is stabilizable if there exists some $K \in \mathbb{R}^{m\times n}$ such that $A - BK$ is Schur. A pair (C, A) is detectable if $(A^\top, C^\top)$ is stabilizable. An eigenvalue λ of A is (C, A)-observable if

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

A matrix $K \in \mathbb{R}^{m\times n}$ is stabilizing for system pair (A, B) if $A - BK$ is Schur; it is *marginally stabilizing* if $\rho(A - BK) = 1$. For fixed $A \in \mathbb{R}^{n\times n}$ and $Q \in \mathbb{R}^{n\times n}$, the Lyapunov matrix is of the form

$$A^\top XA + Q - X = 0.$$

For a system pair (A, B), $Q \in \mathbb{R}^n$ and $R \in GL_n(\mathbb{R})$, the discrete algebraic Riccati equation (DARE) is of the form

$$A^\top XA - X - A^\top XB(R + B^\top XB)^{-1}B^\top XA + Q = 0. \tag{5.1}$$

Next, we recall a result on standard linear-quadratic-regulator (LQR) control.

Theorem 5.2.1. *If $Q \geq 0$, $R > 0$, (A, B) is stabilizable and the spectrum of A on the unit disk (centered at the origin) in $\mathbb{C}$ is (Q, A)-observable, then there exists a unique maximal solution X^+ to DARE* (5.1).[5] *Moreover, the infinite-horizon LQR cost is $x_0^\top X^+ x_0$ and the optimal feedback control K_* is stabilizing and characterized by $K_* = (R + B^\top X^+ B)^{-1}B^\top X^+ A$.*

In the presentation, we shall refer a solution X_0 to (5.1) as *stabilizing* if the corresponding feedback gain $K_0 = (R + B^\top X_0 B)^{-1}B^\top X_0 A$ is stabilizing; a solution X_0 is *almost stabilizing* if the corresponding gain $K_0 = (R + B^\top X_0 B)^{-1}B^\top X_0 A$ is marginally stabilizing.

In the sequential LQ game setup, one key difference from standard LQR is that Q may be indefinite in the corresponding DARE. As such, the following generalization of the above theorem becomes particularly relevant.

Theorem 5.2.2. *Suppose that $Q = Q^\top$, $R > 0$, (A, B) is stabilizable and there exists a solution X to DARE* (5.1)*. Then there exists a maximal solution X^+ to DARE such that*

[5] Where the notion of maximality is with respect to the p.s.d. ordering.

the LQR cost is given by $x_0^\top X^+ x_0$*. Moreover the optimal feedback control is given by* $K_* = (R + B^\top X B)^{-1} B^\top X^+ A$ *and the eigenvalues of* $A - BK_*$ *lie inside the closed unit disk of* $\mathbb{C}$*.*

For DAREs with an indefinite Q, we recall a theorem concerning the existence of solutions.

Theorem 5.2.3 (Theorem 13.1.1 in [45]). *Suppose that* (A, B) *is stabilizable,* $R = R^\top$ *is invertible,* $Q = Q^\top$ *(no definiteness assumption), and there exists a symmetric solution* $\hat{X}$ *to the matrix inequality,*

$$\mathcal{R}(X) := A^\top X A + Q - X - A^\top X B (R + B^\top X B)^{-1} B^\top X A \geq 0,$$

with $R + B^\top \tilde{X} B > 0$*. Then there exists a maximal solution* X^+ *to* (5.1) *such that* $R + B^\top X^+ B > 0$*. Moreover, all eigenvalues of* $A - B(R + B^\top X B)^{-1} B^\top X^+ A$ *are inside the closed unit disk.*

The map $\mathcal{R}_{A,B,Q,R} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ will be referred as *Riccati map* in our analysis; we emphasize its dependency on system parameters A, B, Q, R. In our subsequent analysis, these system parameters will not remain constant as the corresponding feedback gains are iteratively updated.

5.3 LQ Dynamic Games and some of its Analytic Properties

In this section, we review the setup of zero-sum LQ games. In particular we discuss the modified sequential formulation of LQ games and make a few analytical observations that are of independent interest. We note that some of these observations have only become necessary in the context of sequential policy updates for LQ dynamic games.

To simplify notations, throughout the paper we shall denote

$$\begin{aligned} A_{K,L} &:= A - B_1 K - B_2 L, \\ Q_{K,L} &:= Q + K^\top R_1 K - L^\top R_2 L, \end{aligned}$$

and note in this notation,

$$\begin{aligned} A_{K,0} &= A - B_1 K, & A_{0,L} &= A - B_2 L, \\ Q_{K,0} &= Q + K^\top R_1 K, & Q_{0,L} &= Q - L^\top R_2 L. \end{aligned}$$

5.3.1 Zero-sum LQ Dynamic Games

In the standard setup of LQ game, we consider a discrete-time linear time invariant model of the form,

$$x(k+1) = Ax(k) - B_1 u_1(k) - B_2 u_2(k), \qquad x(0) = x_0,$$

where $A \in \mathbb{R}^{n\times n}$, $B_1 \in \mathbb{R}^{n\times m_1}$, $B_2 \in \mathbb{R}^{n\times m_2}$, $u_1(k)$ and $u_2(k)$ are strategies played by two players. The cost incurred for both players is the quadratic cost

$$J(u_1, u_2, x_0) = \sum_{k=0}^{\infty} \left(\langle x(k), Qx(k) \rangle + \langle u_1(k), R_1 u_1(k) \rangle - \langle u_2(k), R_2 u_2(k) \rangle \right),$$

where $Q = Q^\top, R_1 = R_1^\top, R_2 = R_2^\top$ and x_0 is the initial condition; the underlying inner product is denoted by $\langle \cdot, \cdot \rangle$.

Remark. *We emphasize that we depart from traditional LQ game literature [3, 4, 72] by putting* no *assumptions on the definiteness of Q, R_1, R_2. In our exposition, we still put a negative sign in front of the term $\langle u_2, R_2 u_2 \rangle$. But this is just to be conformable with standard game literature.*

In this setting, player $u_1(k)$ chooses its policy to minimize J while player $u_2(k)$ aims to maximize it. The players' strategy space that we will be particularly interested in are closed-loop static linear policies, namely, policies of the form $u_1(k) = Kx(k)$ and $u_2(k) = Lx(k)$, where $K \in \mathbb{R}^{m_1\times n}$ and $L \in \mathbb{R}^{m_2\times n}$. Note that the cost function is guaranteed to be finite over the set of *Schur stabilizing feedback gains*,

$$\mathcal{S} = \{(K, L) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n} : \rho(A - B_1 K - B_2 L) < 1\}.$$

Indeed, if $(K, L) \in \mathcal{S}$, with initial condition x_0, the cost will be given by,

$$J(u_1, u_2, x_0) = x_0^\top \left(\sum_{j=0}^{\infty} [A_{K,L}^\top]^j Q_{K,L} A_{K,L}^j \right) x_0 = x_0^\top X x_0,$$

where X solves the *Lyapunov matrix equation*,

$$X = A_{K,L}^\top X A_{K,L} + Q + K^\top R_1 K - L^\top R_2 L. \tag{5.2}$$

Note that (5.2) has a unique solution if $(K, L) \in \mathcal{S}$. We say that a pair of strategies (K, L) is *admissible* if $(K, L) \in \mathcal{S}$.

Remark. *An elegant result in $\mathcal{H}_\infty$ control theory states that the minimax problem,*

$$\inf_{u_1 \in \ell_2(\mathbb{N})} \sup_{u_2 \in \ell_2(\mathbb{N})} \{J(u_1, u_2) \,|\, x_{u_1,u_2} \in \ell_2(\mathbb{N})\},$$

where $\ell_2(\mathbb{N})$ denotes the Banach space of all square summable sequences and x_{u_1,u_2} denotes the state trajectory after adopting control signals u_1, u_2, has a unique saddle point for all initial conditions if and only if there exists two static linear feedback gains K_* and L_* such $u_1(k) = K_* x(k)$ and $u_2(k) = L_* x(k)$ satisfying the saddle point condition [71]. Hence, the restriction of the optimization process to static linear policies is without loss of generality when we have standard assumptions, i.e., $Q \succeq 0, R_1 \succ 0, R_2 \succ 0$.

A stabilizing *Nash equilibrium* for the zero-sum game is the pair of actions $\{u_1^*(k), u_2^*(k)\} = \{K_* x(k), L_* x(k)\}$ such that,

$$J(u_1^*(k), u_2(k)) \le J(u_1^*(k), u_2^*(k)) \le J(u_1(k), u_2^*(k)), \tag{5.3}$$

for all initial states x_0 and all $u_1(k), u_2(k)$ for which $(u_1(k), u_2^*(k))$ and $(u_1^*(k), u_2(k))$ are both admissible pairs. We *emphasize* that it is important that $(u_1(k), u_2^*(k))$ and $(u_1^*(k), u_2(k))$ are stabilizing action pairs in the inequality (5.3). To demonstrate this delicate situation, we denote by $\mathcal{S}_{\pi_i}$ the projection of $\mathcal{S}$ onto the ith coordinate, i.e.,

$$\mathcal{S}_{\pi_1} = \{K : \exists L \text{ such that } A - B_1 K - B_2 L \text{ is Schur}\},$$
$$\mathcal{S}_{\pi_2} = \{L : \exists K \text{ such that } A - B_1 K - B_2 L \text{ is Schur}\},$$

and $\mathcal{S}_{\hat{K}}, \mathcal{S}_{\hat{L}}$ as sets defined by,

$$\mathcal{S}_{\hat{K}} = \{L : A - B_1 \hat{K} - B_2 L \text{ is Schur}\},$$
$$\mathcal{S}_{\hat{L}} = \{K : A - B_2 \hat{L} - B_1 K \text{ is Schur}\}.$$

Clearly, $\mathcal{S}_{K_*} \subset \mathcal{S}_{\pi_2}$ and $\mathcal{S}_{L_*} \subset \mathcal{S}_{\pi_1}$. This means that it is not the case that for all $K \in \mathcal{S}_{\pi_1}$ and all $L \in \mathcal{S}_{\pi_2}$, the corresponding actions $u_1(k) = Kx(k)$ and $u_2(k) = Lx(k)$ yield

$$J(u_1^*(k), u_2(k)) \leq J(u_1^*(k), u_2^*(k)) \leq J(u_1(k), u_2^*(k)).$$

This is simply due to the fact that $(\hat{K}, L_*)$ is not guaranteed to be stabilizing for all $\hat{K} \in \mathcal{S}_{\pi_1}$.

Note that the cost function J is a function of polices K, L and initial condition x_0. Since we are interested in the Nash equilibrium independent of the initial conditions, naturally, we should formulate cost functions for both players to reflect this independence. Indeed, this point has been discussed in [11] where it has been argued that such a formulation is in general necessary for the cost functions to be well defined (see details in §III [11]). The independence with respect to initial conditions can be achieved by either sampling x_0 from a distribution with full-rank covariance [24], or choosing a spanning set $\{w_1, \dots, w_n\} \subseteq \mathbb{R}^n$ [11], and defining the value function over $\mathcal{S}$ as,

$$f(K, L) = \sum_{i=1}^{n} J_{w_i}(K, L),$$

where $J_{w_i}(K, L)$ is the cost choosing initial state w_i, $u_1(k) = Kx(k)$ and $u_2(k) = Lx(k)$. Note that over the set $\mathcal{S}$ the value of function f admits a compact form,

$$f(K, L) = \mathbf{Tr}(X\mathbf{\Sigma}),$$

where $\mathbf{\Sigma} = \sum_{i=1}^{n} w_i w_i^\top$ and X is the solution to (5.2).

5.3.2 Stabilizing Policies in Sequential Zero-Sum LQ Games

Another subtle situation arising in sequential zero-sum LQ dynamic game is as follows: there is clearly no incentive for player 1 to destabilize the dynamics. However, from the perspective of player 2, making the states diverge to infinity could be desirable as the player aims to maximize the cost. For player 1, in the situation where $Q - L^\top R_2 L$ is not positive semidefinite, it is also possible that the best policy is not the one in $\mathcal{S}_{\pi_1}$. Hence, in round j, in order to guarantee that the game can be continued, it is important that both players choose their respective policies in $\mathcal{S}_{\pi_1}$ and $\mathcal{S}_{\pi_2}$. We may then stipulate that both players play *Schur stable* policies. We can justify this constraint by insisting that both players have an incentive to stabilize the system in the first place. This can also be encoded in the cost function for the players. That is, we may define the cost functions for player 1 and player 2 by,

$$f_1(K,L) = \delta_{\mathcal{S}_{\pi_1}}(K) + f(K,L),$$
$$f_2(K,L) = -\delta_{\mathcal{S}_{\pi_2}}(L) + f(K,L),$$

where $\delta_{\mathcal{S}_{\pi_i}}(x)$ is the indicator function of the set

$$\delta_{\mathcal{S}_{\pi_i}}(x) = \begin{cases} 0, & x \in \mathcal{S}_{\pi_i}, \\ +\infty, & x \notin \mathcal{S}_{\pi_i}. \end{cases}$$

Then we have two cost functions defined everywhere for both players and assume a finite value on $\mathcal{S}$ which agree with each other, i.e., $f(K,L)$. We still need to be careful in realizing that there are points for which the function value is indeterminate. For example, it is possible to find a point $(\hat{K},\hat{L})$ such that $f(\hat{K},\hat{L}) = -\infty$; then $f_1(\hat{K},\hat{L}) = +\infty - \infty$. To resolve this complication, we shall declare the function value to be the first summand; namely, if $f_1(\hat{K},\hat{L}) = +\infty - \infty$, then $f_1(\hat{K},\hat{L}) \equiv +\infty$.

From the perspective of sequential algorithm design, these newly introduced cost functions would constrain both players to play policies in $\mathcal{S}$. It might be tempting to design projection based algorithms. However, this can be difficult since describing the sets $\mathcal{S}_{\pi_1}$ and $\mathcal{S}_{\pi_2}$ for given system $(A,[B_1,B_2])$ is not straightforward and moreover the sets are generally not convex. We shall see later that by exploiting the problem structure, we can design sequential algorithms for both players to guarantee this condition without any projection step.

5.3.3 Analytic Properties of the Cost Function

In this section we shall clarify analytical properties of the cost functions in terms of the polices played by the two players; that is, we consider the cost function $f(K,L)$ over $\mathcal{S}$.[6] The observations are collected in following proposition:

Proposition 5.3.1. *a. The set $\mathcal{S}$ is open, contractible (i.e., path-connected and simply connected) and in general non-convex.*

b. One has $f \in C^{\omega}(\mathcal{S})$.

c. On the set $\mathcal{S}$, the gradients of f with respect to its arguments are given by,

$$\partial_K f(K,L) = (R_1 K - B_1^\top X A_{K,L})Y,$$
$$\partial_L f(K,L) = (-R_2 L - B_2^\top X A_{K,L})Y,$$

where X solves the Lyapunov equation (5.2) and Y solves the Lyapunov equation,

$$A_{K,L} Y A_{K,L}^\top + \mathbf{\Sigma} = 0. \tag{5.4}$$

Proof. The proof essentially follows from the ones presented in [11] by regarding the system pairs to be $(A, [B_1, B_2])$. □

As the set $\mathcal{S}$ is generally not convex, Item *a* of Proposition 5.3.1 assures us that algorithms based on local search (*e.g.* gradient descent) can potentially reach the Nash equilibrium. If $\mathcal{S}$ had more than one path-connected components, it will be impossible to guarantee the convergence to Nash equilibrium under random initialization. Moreover this observation implies that f is not convex-concave as the domain is not even convex.

We now observe that $Y(K,L)$ is a smooth function in (K,L) and is positive definite everywhere on $\mathcal{S}$. Hence $Y(K,L)$ is a well-defined Riemannian metric on $\mathcal{S}$. Under this

[6] In our formulation, the two players have different cost functions. But over the set $\mathcal{S}$, the cost functions coincide.

metric, we can thereby identify the gradient. In learning and statistics literature, such a gradient is referred as a "natural gradient." We shall use $\mathbf{N}_{f,K}$ and $\mathbf{N}_{f,L}$ to denote the natural gradient of f over K and L, respectively. Namely,

$$\mathbf{N}_{f,K}(K,L) = R_1 K - B_1^\top X A_{K,L},$$
$$\mathbf{N}_{f,L}(K,L) = -R_2 L - B_2^\top X A_{K,L}.$$

5.3.4 A Key Assumption and its Implications

Throughout the manuscript, we have the following standing assumption.

Assumption 1. *There exists a* stabilizing Nash Equilibrium $(K_*, L_*) \in \mathcal{S}$ *for the zero-sum game over the system dynamic* $(A, [B_1, B_2])$. *Moreover, the corresponding value matrix* $X_* = X_*(K_*, L_*)$ *satisfies at least one of the following conditions:*

(a1): $R_1 + B_1^\top X_* B_1 \succ 0$ *and* $R_2 - B_2^\top X_* B_2 + B_2^\top X_* (R_1 + B_1^\top X_* B_1)^{-1} B_1 B_2 \succ 0.$

(a2): $-R_2 + B_2^\top X_* B_1 \prec 0$ *and* $R_1 + B_1^\top X_* B_1 - B_1^\top X_* B_2 (-R_2 + B_2^\top X_* B_2)^{-1} B_2^\top X_* B_1 \succ 0.$

Remark. *The existence of a stabilizing Nash is a necessary assumption adopted in the LQ literature [3]. However, we do not constrain the value matrix* X_* *to be positive semidefinite, as assumed in [3, 72, 81]. The definiteness is useful when the LQ game formulation is tied to* $\mathcal{H}_\infty$ *control. However, from the LQ game perspective, this association seems unnecessary. Conditions* $(a1)$ *or* $(a2)$ *are necessary if it is desired to extract* unique *policies from the optimal value matrix* X_*. *Namely, if the* total derivative *of* f *vanishes, i.e.,*

$$\begin{pmatrix} R_1 & 0 \\ 0 & -R_2 \end{pmatrix}\begin{pmatrix} K \\ L \end{pmatrix} - \begin{pmatrix} B_1^\top \\ B_2^\top \end{pmatrix} XA + \begin{pmatrix} B_1^\top \\ B_2^\top \end{pmatrix} X \begin{pmatrix} B_1 & B_2 \end{pmatrix} = 0,$$

conditions $(a1)$ *or* $(a2)$ *are sufficient to guarantee the uniqueness of the solution in* $\mathcal{S}$. *Indeed, assumptions* $(a1)$ *or* $(a2)$ *are "almost necessary." If* (K_*, L_*) *is a NE, then* $f(\cdot, L_*)$ *achieves a local minimum at* K_*, *i.e.,* $\nabla_{KK} f(K_*, L_*)[E,E] = \langle E, (R_1 + B_1^\top X_* B_1) E Y_* \rangle \geq 0$

(note that by assumption, K_* is in the interior of $\mathcal{S}$ and thus the second-order partial derivative is well-defined). Similarly, $\nabla_{LL} f(K_*, L_*) = -R_2 + B_2^\top X_* B_2 \leq 0$. We relax these two necessary conditions to hold as strictly positive (respectively, negative) definite.[7] *In fact, in the sequential LQ formulation, the inequalities in these two conditions correspond to certain "quasi-Newton" directions and as such play a central role in our convergence analysis (see §5.5 and §5.6 for details.).*

Moreover, we shall subsequently see that assumptions $(a1)$ and $(a2)$ lead to distinct choices of leaders in the sequential algorithms. More specifically, if we assume condition $(a1)$, the leader of the sequential algorithm should be player L; for assumption $(a2)$, player K should be the designated leader.

We observe several implications of this assumption.

Proposition 5.3.2. *Under Assumption 1, we have following implications:*

a. *The pair $(A, [B_1, B_2])$ is stabilizable.*

b. *X_* is symmetric and solves the Generalized Algebraic Riccati Equation (GARE),*

$$\begin{pmatrix} B_1^\top X A \\ B_2^\top X A \end{pmatrix}^\top \begin{pmatrix} R_1 + B_1^\top X B_1 & B_1^\top X B_2 \\ B_2^\top X B_1 & -R_2 + B_2^\top X B_2 \end{pmatrix}^{-1} \begin{pmatrix} B_1^\top X A \\ B_2^\top X A \end{pmatrix} + A^\top X A + Q - X = 0. \tag{5.5}$$

c. *X_* is unique among all* almost stabilizing solutions *of* (5.5).

Proof. The statement in (a) is immediate since $A - B_1 K_* - B_2 L_*$ is Schur.
In order to show (b), we note that since (K_*, L_*) is a stabilizing Nash Equilibrium, then X_* is the solution of the Lyapunov equation (5.2); it thus follows that X_* is symmetric.

[7] If we do not relax the semidefiniteness conditions, the NE would be solutions to GARE involving Moore-Penrose inverse. This will introduce other complications than practically needed.

Further, note that the partial gradients of f vanish at (K_*, L_*), namely $(K_*, L_*) \in \mathcal{S}$ solves the equations

$$\nabla_K f(K,L) = (R_1 K - B_1^\intercal X A_{K,L})Y = 0,$$
$$\nabla_L f(K,L) = (-R_2 L - B_2^\intercal X A_{K,L})Y = 0.$$

Substituting this in the Lyapunov equation (5.2), it follows that (K_*, L_*) solves the GARE (5.5). Note that the inverse,

$$\begin{pmatrix} R_1 + B_1^\intercal X B_1 & B_1^\intercal X B_2 \\ B_2^\intercal X B_1 & -R_2 + B_2^\intercal X B_2 \end{pmatrix}^{-1}$$

is well-defined at X_* by the conditions $a1$ or $a2$ in the assumption.

For the statement in (c), by Lemma 3.1 [72], X_* is the unique stabilizing solution. It remains to show that there does not exist almost stabilizing solution to (5.5). Suppose there exists a pair $(K, L) \in \partial\mathcal{S}$, i.e., $\rho(A - B_1K - B_2L) = 1$, solving (5.5) with solution X. Then taking the difference between the identity (5.5) at (K_*, L_*) and (K, L), we have,

$$A_*^\intercal (X_* - X) A_{K,L} = X_* - X.$$

Since $A_{K,L}$ is marginally stable and A_* is stable, then $I \otimes I - A_{K,L}^\intercal \otimes A_*$ is invertible and thus $X_* - X = 0$. □

5.4 Oracle Models for Sequential LQ Games

In this work, we assume that both players have access to oracles that return either gradient, natural gradient or quasi-Newton directions. Suppose that $\mathcal{O}_K$ and $\mathcal{O}_L$ are the orcales for the two players respectively. The players will query their respective oracles in a sequential manner: if player 1 query the oracle, we assume the policy played by player 2 is fixed during the query and this policy is transparent to oracle $\mathcal{O}_K$ of player 1. As f is in general not convex-concave, if two players have the same oracles and play greedily using the information they have acquired, theoretically, there is no guarantee that they will eventually converge to the Nash equilibrium. In order to obtain theoretical guarantees, we assume that player 1 can access an oracle that computes the minimizer of $f(K, L)$ over K for a fixed L. This oracle can be constructed out of the simple first-order oracles by repeatedly performing gradient descent/natural gradient descent/quasi-Newton type steps [16]. More explicitly, we shall assume that for player 1, if player 2's policy is $\hat{L}$, the oracle can return $K \leftarrow \arg\min_K f(K, \hat{L})$.

5.4.1 Motivation

We shall present the motivation for equipping player 1 with a more powerful oracle model. As finding the Nash equilibrium is equivalent to solving the saddle point of $f(K, L)$, from the perspective of player 2, we may associate a value function independent of player 1. Namely, we may define a function of the form,

$$g(L) = \begin{cases} \inf_{K \in \mathcal{S}_L} f(K, L), & \text{if } L \in \mathcal{S}_{\pi_2}, \\ -\infty, & \text{otherwise.} \end{cases}$$

If $g(L)$ possesses a *smoothness* property, we may consider *projected gradient ascent* over the policy space. However, reflecting over $g(L)$ would reveal that $g(L)$ is not necessarily even continuous on $\mathcal{S}_{\pi_2}$. For example, if $Q - L^\top R_2 L \prec 0$, then $\inf_K f(K, L)$ could be $-\infty$. But on the other hand, by Danskin's Theorem, $g(L)$ is differentiable at $L \in \mathcal{S}_{\pi_2}$, where $f(K, L)$ admits a unique minimizer over K.

Lemma 5.4.1. *Suppose that $U \subseteq \mathbf{dom}(g)$ is an open subset such that for every $L \in U$,*

$$\arg\min_{K \in \mathcal{S}_L} f(K, L)$$

exits and is unique. Then $g(L)$ is differentiable and its gradient is,

$$\nabla g(L) = \nabla_L f(K_L, L), \text{ where } K_L = \arg\min_K f(K, L).$$

Traditionally Danskin's theorem would require that for every L, the minimization of K is over a common compact set. This is not the situation in our case as $\mathcal{S}_L$ is not compact nor common over L[8]. The statement of Lemma 5.4.1 instead follows from a variant of Danskin's Theorem in [7].

Proof. We only need to observe that $f(K,L)$ is C^∞ in both variables and thus Fréchet differentiable. Hence, the assumptions of Hypothesis $D2$ in [7] are satisfied. By Theorem $D2$ in [7], $g(L)$ is directionally differentiable in every direction. As the minimizer K_L is unique, the directional derivative is uniform in every direction and consequently, $g(L)$ is differentiable. □

The next issue that needs to be addressed is whether $\mathcal{U}$ is empty. It turns out that by standard LQR theory, $\{L \in \mathbf{dom}(g) : Q - L^\top R_2 L \succ 0\}$ is a subset in U. We can thus outline an update rule assuming that $g(L)$ is Lipschitz, namely, a *projected gradient ascent* over L as,

$$L_{j+1} = P_{\mathcal{S}_{\pi_2}}\left(L_j + \eta_j \nabla g(L_j)\right),$$

where $\nabla g(L_j)$ is given by

$$\nabla g(L_j) = (-R_2 L - B_2^T X_{K_{L_j},L_j} A_{K_{L_j},L_j}) Y_{K_{L_j},L_j},$$

where $K_{L_j} = \arg\min_K f(K, L_j)$. As already noted, we do not have a full description of the nonconvex set $\mathcal{S}_{\pi_2}$ and a projection would rather be prohibitive. What we shall propose instead, are update rules that guarantee all of its iterates stay in the set $\mathcal{S}_{\pi_2}$; this will be achieved without a projection step by exploiting the problem structure.

Another interesting interpretation of our setup is to consider this game from the perspective of player 2: we have a game played by player 2 with a greedy adversary. Each time player 2 chooses a policy L', the adversary (player 1) would try to act greedily, i.e., minimize the cost $f(K, L')$ over K. The goal for player 2 is to achieve the Nash equilibrium point for himself/herself and the greedy adversary. The information player 2 can acquire from the game (i.e., oracle) is the first-order information (function value and gradient). As such,

[8]Namely, it is not necessarily true $\mathcal{S}_{L_1} = \mathcal{S}_{L_2}$ if $L_1 \neq L_2$.

player 2 has an obligation to guarantee that along the iterates $\{L_j\}$, the oracle could return meaningful first-order information of $g(L)$, i.e., $g(L_j)$ is differentiable for every j.

5.5 Algorithm: Natural Gradient Policy on L

Throughout §5.5 and §5.6, we assume that condition $(a1)$ in our assumption holds and an oracle $\mathcal{O}_K$ that returns the stabilizing minimizer $f(K, L)$ for any fixed L, provided that such minimizer exists. Note that the unique minimizer corresponds to the maximal solution X^+ to the algebraic Riccati equation (with fixed L), namely,

$$X = A_{0,L}^\intercal X A_{0,L} + Q_{(0,L)} - A_{0,L}^\intercal X B_1 (R_1 + B_1^\intercal X B_1)^{-1} B_1^\intercal X A_{0,L}.$$

Note even in standard LQ game setting, $Q_{0,L}$ becomes indefinite in the iterative process. Fortunately, policy gradient based algorithms can still be utilized to find the minimizer [16].

5.5.1 *Algorithm*

The algorithm is given by:

Algorithm 3 Natural Gradient Policy for LQ Game

1: Initialize L_0 such that $(A - B_1 L_0, B_2)$ is stabliizable and the DARE $\mathcal{R}_{A_{(0,L_0)},B_1,Q_{(0,L_0)},R_1}(X) = 0$ has a stabilizing solution X^+ with $R_1 + B_1^\intercal X^+ B_1 > 0$.

2: **if** $j \geq 1$ **then**

3: Set: $K_{j-1} \leftarrow \arg\min_K f(K, L_{j-1})$

4: Set:

$$\begin{aligned} L_j &= L_{j-1} + \eta_j \mathbf{N}_g(L_j) \\ &\equiv L_{j-1} + \eta_j \mathbf{N}_{f,L}(K_{j-1}, L_{j-1}). \end{aligned}$$

5: **end if**

We note that the initialization step is generally nontrivial. However, if we further assume that (A, B_1) is stabilizable, $Q \geq 0$ and those eigenvalues of A lying on the unit disk are (Q, A)-detectable, then we can choose $L_0 = 0$[9]. For the general case, we may need to check invariant subspaces of system parameters (see [45] for details.)

5.5.2 Convergence Analysis

To simplify notations let,

$$\begin{aligned}
\mathbf{U}_{K,L} &= R_1 K - B^\top X_{K,L} A_{K,L}, \\
\mathbf{V}_{K,L} &= -R_2 L - B^\top X_{K,L} A_{K,L};
\end{aligned}$$

note $\mathbf{N}_{f,K}(K, L) = 2\mathbf{U}_{K,L}$ and $\mathbf{N}_{f,L}(K, L) = 2\mathbf{V}_{K,L}$. First a useful observation.

Lemma 5.5.1 (NG Comparison Lemma)**.** *Suppose that (K, L) and $(\hat{K}, \hat{L})$ are both stabilizing and let X and $\hat{X}$ be the corresponding value matrices. Then*

a.

$$\begin{aligned}
X - \hat{X} = A_{\hat{K},\hat{L}}^\top (X - \hat{X}) A_{\hat{K},\hat{L}} &+ (K - \hat{K})^\top \mathbf{U}_{K,L} \\
&+ \mathbf{U}_{K,L}^\top (K - \hat{K}) - (K - \hat{K})^\top R_1 (K - \hat{K}) \\
&+ (L - \hat{L})^\top \mathbf{V}_{K,L} + \mathbf{V}_{K,L}^\top (L - \hat{L}) \\
&+ (L - \hat{L})^\top R_2 (L - \hat{L}) \\
&- (A_{K,L} - A_{\hat{K},\hat{L}})^\top X (A_{K,L} - A_{\hat{K},\hat{L}}).
\end{aligned}$$

[9] This is indeed the standard assumption in the LQ literature.

b.

$$
\begin{aligned}
X - \hat{X} = A_{K,L}^{\top}(X - \hat{X})A_{K,L} &+ (K - \hat{K})^{\top}\mathbf{U}_{\hat{K},\hat{L}} \\
&+ \mathbf{U}_{\hat{K},\hat{L}}^{\top}(K - \hat{K}) + (K - \hat{K})^{\top}R_1(K - \hat{K}) \\
&+ (L - \hat{L})^{\top}\mathbf{V}_{\hat{K},\hat{L}} + \mathbf{V}_{\hat{K},\hat{L}}^{\top}(L - \hat{L}) \\
&- (L - \hat{L})^{\top}R_2(L - \hat{L}) \\
&+ (A_{K,L} - A_{\hat{K},\hat{L}})^{\top}\hat{X}(A_{K,L} - A_{\hat{K},\hat{L}}).
\end{aligned}
$$

Remark. *Item* (b) *of this lemma was observed in [81]. Our presentation offers a control theoretic perspective on its proof.*

Proof. We prove item (a); item (b) can be proved in a similar manner.
It suffices to take the difference of the Lyapunov equations:

$$
\begin{aligned}
A_{K,L}^{\top}XA_{K,L} + Q + K^{\top}R_1K - L^{\top}R_2L = X, \\
A_{\hat{K},\hat{L}}^{\top}\hat{X}A_{\hat{K},\hat{L}} + Q + \hat{K}^{\top}R_1\hat{K} - \hat{L}^{\top}R_2\hat{L} = \hat{X}.
\end{aligned}
$$

Indeed, a few algebraic operations reveal that

$$
\begin{aligned}
X - \hat{X} = A_{\hat{K},\hat{L}}^{\top}(X - \hat{X})A_{\hat{K},\hat{L}} &+ (K - \hat{K})^{\top}\mathbf{U}_{K,L} \\
&+ \mathbf{U}_{K,L}^{\top}(K - \hat{K}) - (K - \hat{K})^{\top}(R_1)(K - \hat{K}) \\
&+ (L - \hat{L})^{\top}\mathbf{V}_{K,L} + \mathbf{V}_{K,L}^{\top}(L - \hat{L}) + (L - \hat{L})^{\top}R_2(L - \hat{L}) \\
&- (A_{K,L} - A_{\hat{K},\hat{L}})^{\top}X(A_{K,L} - A_{\hat{K},\hat{L}}).
\end{aligned}
$$

□

We now observe another version of comparison lemma when $L, \tilde{L} \in \mathbf{dom}(g)$. Indeed, this lemma will play a more prominent role in our convergence analysis.

Lemma 5.5.2 (Comparison Lemma 2)**.** *Suppose that* $L, \tilde{L} \in \mathbf{dom}(g)$*, namely there exists* $K, \tilde{K}$ *such that*

$$
K = \underset{K' \in \mathcal{S}_L}{\arg\min} f(K', L), \qquad \tilde{K} = \underset{K' \in \mathcal{S}_{\tilde{L}}}{\arg\min} f(K', \tilde{L}).
$$

Further, suppose that the algebraic Riccati map $\mathcal{R}_{A_{0,L},B_1,Q_{(0,L)},R_1}(\tilde{X})$ is well-defined, i.e., $R_1 + B_1^\top \tilde{X} B_1$ is invertible. Recall that the Riccati map is given by,

$$\mathcal{R}_{A_{0,L},B_1,Q_{(0,L)},R_1}(\tilde{X}) = A_{0,L}^\top \tilde{X} A_{0,L} - \tilde{X} + Q_{(0,L)} - A_{0,L}^\top \tilde{X} B_1 (R_1 + B_1^\top \tilde{X} B_1)^{-1} B_1^\top \tilde{X} A_{0,L}.$$

Let X and $\tilde{X}$ be the corresponding value matrix. Putting

$$\mathbf{E} \coloneqq R_1 + B_1^\top \tilde{X} B_1, \qquad \mathbf{F} \coloneqq B_1^\top \tilde{X}(A - B_2 L),$$

then

$$\begin{aligned} X - \tilde{X} = A_{K,L}^\top (X - \tilde{X}) A_{K,L} + \mathcal{R}_{A_{0,L},B_1,Q_{(0,L)},R_1}(\tilde{X}) \\ + (\mathbf{E}K - \mathbf{F})^\top \mathbf{E}^{-1} (\mathbf{E}K - \mathbf{F}). \end{aligned}$$

Moreover,

$$\begin{aligned} \mathcal{R}_{A_{0,L},B_1,Q_{(0,L)},R_1}(\tilde{X}) = (L - \tilde{L})^\top \mathbf{V}_{\tilde{K},\tilde{L}} + \mathbf{V}_{\tilde{K},\tilde{L}}^\top (L - \tilde{L}) \\ - (L - \tilde{L})^\top \mathbf{O}_{\tilde{X}} (L - \tilde{L}), \end{aligned}$$

where

$$\mathbf{O}_{\tilde{X}} \coloneqq R_2 - B_2^\top \tilde{X} B_2 + B_2^\top \tilde{X} B_1 (R_1 + B_1^\top \tilde{X} B_1)^{-1} B_1^\top \tilde{X} B_2.$$

Proof. Note that X solves the Lyapunov matrix equation,

$$X = A_{K,L}^\top X A_{K,L} + Q - L^\top R_2 L + K^\top R_1 K,$$

with $K = (R_1 + B_1^\top X B_1)^{-1} B_1^\top X (A - B_2 L)$. Then

$$\begin{aligned} X - \tilde{X} - A_{K,L}^\top (X - \tilde{X}) A_{K,L} &= A_{K,L}^\top \tilde{X} A_{K,L} - \tilde{X} + Q_{(0,L)} + K^\top R_1 K \\ &= A_{0,L}^\top \tilde{X} A_{0,L} - \tilde{X} + Q_{(0,L)} + K^\top (R_1 + B_1^\top \tilde{X} B_1) K - K^\top B_1^\top \tilde{X} A_{0,L} - A_{0,L}^\top \tilde{X} B_1 K \\ &= \mathcal{R}_{A_{0,L},B_1,Q_{(0,L)},R_1}(\tilde{X}) + A_{0,L}^\top \tilde{X} B_1 (R_1 + B_1^\top \tilde{X} B_1)^{-1} B_1^\top \tilde{X} A_{0,L} \\ &\quad + K^\top (R_1 + B_1^\top \tilde{X} B_1) K - K^\top B_1^\top \tilde{X} A_{0,L} - A_{0,L}^\top \tilde{X} B_1 K \\ &= R_{A_{0,L},B_1,Q_{(0,L)},R_1}(\tilde{X}) + (\mathbf{E}K - \mathbf{F})^\top \mathbf{E}^{-1} (\mathbf{E}K - \mathbf{F}). \end{aligned}$$

Since $\tilde{X}$ satisfies the algebraic Riccati equation

$$\tilde{X} = A_{0,\tilde{L}}^\top \tilde{X} A_{0,\tilde{L}} + Q_{\tilde{L}} - A_{0,\tilde{L}}^\top \tilde{X} B_1 (R_1 + B_1^\top \tilde{X} B_1)^{-1} B_1^\top \tilde{X} A_{0,\tilde{L}},$$

it follows that,

$$\begin{aligned}\mathcal{R}_{A_{0,L},B_1,Q_{(0,L)},R_1}(\tilde{X}) &= A_{0,L}^\top \tilde{X} A_{0,L} - L^\top R_2 L - A_{0,L}^\top \tilde{X} B_1 (R_1 + B_1^\top \tilde{X} B_1)^{-1} B_1^\top \tilde{X} A_{0,L} \\ &\quad - A_{0,\tilde{L}}^\top \tilde{X} A_{0,\tilde{L}} + \tilde{L}^\top R_2 \tilde{L} + A_{0,\tilde{L}}^\top \tilde{X} B_1 (R_1 + B_1^\top \tilde{X} B_1)^{-1} B_1^\top \tilde{X} A_{0,\tilde{L}} \\ &= (L - \tilde{L})^\top \mathbf{V}_{\tilde{K},\tilde{L}} + \mathbf{V}_{\tilde{K},\tilde{L}}^\top (L - \tilde{L}) - (L - \tilde{L})^\top \mathbf{O}_{\tilde{X}} (L - \tilde{L}).\end{aligned}$$

□

Let us now prove the convergence of the proposed algorithm. In the following analysis, to simplify notations K_j's are exclusively used as the unique *stabilizing minimizers*,[10] for $f(K, L_j)$ over K.[11] and further we put

$$\begin{aligned} Q_j &\coloneqq Q_{L_j} = Q - L_j^\top R_2 L_j,\ \Delta \coloneqq X_{K_{j-1},L_{j-1}}, \\ \mathbf{O}_{j-1} &\coloneqq \mathbf{O}_{X_{K_{j-1},L_{j-1}}},\end{aligned}$$

where we recall the definition of $\mathbf{O}_{X_{K_{j-1},L_{j-1}}}$. We shall first show that if Algorithm 3 is initialized appropriately, then with stepsize

$$\eta_{j-1} \leq \frac{1}{\lambda_n(\mathbf{O}_{j-1})},$$

Algorithm 3 generates a sequence $\{L_j\}$ satisfying properties listed in the following lemma.

Lemma 5.5.3. *Suppose that Algorithm 3 is initialized appropriately. With stepsize $\eta_{j-1} \leq \frac{1}{\lambda_n(\mathbf{O}_{j-1})}$, we then have,*

a. $(A - B_2 L_j, B_1)$ is stabilizable for every $j \geq 1$.

---

[10] Depending on the structure of our problem, it is possible that there exists non-stabilizing minimizers. But here we are only concerned with minimizers in the set $\mathcal{S}$.

[11] Of course, we need to guarantee that for L_j, there exists a unique minimizer.

b. $\mathbf{O}_j \succ 0$ *for every* $j \geq 1$.

c. For every $j \geq 1$, $f(K, L_j)$ *is bounded below over* K *and there exists a unique minimizer* K_j, *which forms a stabilizing pair* (K_j, L_j). *Namely, the DARE*

$$\mathcal{R}_{A_{0,L_j},B_1,Q_j,R_1}(X) = 0, \tag{5.6}$$

admits a stabilizing maximal solution X^+ *satisfying* $R + B_1^\intercal X^+ B_1 \succ 0$. *In other words,* $g(L_j)$ *is finite and differentiable for every* $j \geq 0$

d. Putting $\Lambda = X_{K_j,L_j}$, $\mathbf{E}_j = R_1 + B_1^\intercal X_{K_j,L_j} B_1$ *and* $\mathbf{F}_j = B_1^\intercal X_{K_j,L_j}(A - B_2 L_j)$ *we have*

$$\begin{aligned}\Lambda - \Delta = {} & A_{K_j,L_j}^\intercal (\Lambda - \Delta) A_{K_j,L_j} \\ & + \mathbf{V}_{K_{j-1},L_{j-1}}^\intercal \left(4\eta_{j-1} I - 4\eta_j^2 \mathbf{O}_{j-1}\right) \mathbf{V}_{K_{j-1},L_{j-1}} \\ & + (\mathbf{E}_{j-1} K_j - \mathbf{F}_{j-1})^{-1} \mathbf{E}_{j-1}^{-1} (\mathbf{E}_{j-1} K_j - \mathbf{F}_{j-1}).\end{aligned}$$

Proof. It suffices to prove the lemma by induction since all items holds at $j = 0$ (by initialization of the algorithm). We shall first suppose that $(A - B_2 L_j, B_1)$ is stabilizable, i.e., (a) holds. Note that this property is not automatically guaranteed and we subsequently provide an analysis to carefully remove this assumption.[12]

First note that by our assumption, Δ is the maximal stabilizing solution of the DARE,[13]

$$\mathcal{R}_{A_{0,L_{j-1}},B_1,Q_{j-1},R_1}(X) = 0$$

and $K_{j-1} = (R_1 + B_1^\intercal \Delta B_1)^{-1} B_1^\intercal \Delta (A - B_2 L_{j-1})$. Now adopt the update rule,

$$L_j = L_{j-1} - \eta_{j-1} N_g(L_{j-1}) = L_{j-1} - 2\eta_{j-1} \mathbf{V}_{K_{j-1},L_{j-1}}.$$

[12] A side remark on our proof strategy: in linear system theory, a number of synthesis results are developed under the assumption of stabilizability of the system. We will utilize these observations here; however, to use those tools, we must assume that $(A - B_2 L_j, B_1)$ is stabilizable. But this is also one of our goals to show. The reader might recognize certain circular line of reasoning here. Indeed, one of our contributions is pseudo trick devised to circumvent this issue: we first assume stabilizability and then use the results developed to arrive at a contradiction if the system had not been stabilizable.

[13] This can be considered as an LQR problem for system $(A - B_2 L_{j-1}, B_1)$ with state cost matrix $Q - L_{j-1}^\intercal R_2 L_{j-1}$.

by Lemma 5.5.2 it follows

$$\begin{aligned}\mathcal{R}_{A_{0,L_j},B_1,Q_j,R_1}(X_{j-1}) \\ = \mathbf{V}_{K_{j-1},L_{j-1}}^\top \left(4\eta_{j-1} I - 4\eta_{j-1}^2 \mathbf{O}_{j-1})\right) \mathbf{V}_{K_{j-1},L_{j-1}}.\end{aligned} \tag{5.7}$$

Thereby with the stepsize $\eta_{j-1} \leq \frac{1}{\lambda_n(\mathbf{O})}$, we have

$$\begin{aligned}\mathcal{R}_{A_{0,L_j},B_1,Q_j,R_1}(X_{j-1}) \\ = \mathbf{V}_{K_{j-1},L_{j-1}} \left(4\eta_{j-1} I - 4\eta_{j-1}^2 \mathbf{O}_{j-1}\right) \mathbf{V}_{K_{j-1},L_{j-1}} \geq 0.\end{aligned} \tag{5.8}$$

By Theorem 5.2.3 (note that the inequality (5.8) is crucial for applying the theorem), there exists a maximal solution $X^+ \geq \Delta$ to the DARE (5.6) and moreover, with

$$K^+ = (R_1 + B_1^\top X^+ B_1)^{-1} B_1^\top X^+ (A - B_2 L_j),$$

the eigenvalues of $A - B_2 L_j - B_1 K^+$ are in the closed unit disk of $\mathbb{C}$. Equivalently, X^+ solves the following Lyapunov equation,

$$A_{K^+,L_j}^\top X^+ A_{K^+,L_j} + Q + (K^+)^\top R_1 K^+ - L_j^\top R_2 L_j = X^+.$$

Item (d) thereby follows from the first part of Lemma 5.5.2. We now observe that K^+ is indeed stabilizing. Suppose that this is not the case; then there exists $v \in \mathbb{C}^n$ such that $(A - B_2 L_j - B_1 K^+)v = \lambda v$ with $|\lambda| = 1$. Hence,

$$\begin{aligned}& v^\top \left(A_{K^+,L_j}^\top (X^+ - \Delta) A_{K^+,L_j}\right) v \\ & + v^\top \left(\mathbf{V}_{K_{j-1},L_{j-1}}^\top \left(4\eta_{j-1} I - 4\eta_j^2 \mathbf{O}_{j-1}\right) \mathbf{V}_{K_{j-1},L_{j-1}}\right) v \\ & \leq v^\top (X^+ - \Delta) v.\end{aligned}$$

This would imply that $\mathbf{V}_{K_{j-1},L_{j-1}} v = 0$. But this means that,

$$L_j v = L_{j-1} v. \tag{5.9}$$

By Lemma 5.5.1, we have

$$\begin{aligned} X^+ - \Delta &= A_{K^+,L_j}^\top (X^+ - \Delta) A_{K^+,L_j} \\ &+ \mathbf{V}_{K_{j-1},L_{j-1}}^\top \left(4\eta_{j-1} I - 4\eta_j^2 R_2\right) \mathbf{V}_{K_{j-1},L_{j-1}} \\ &+ (K^+ - K_{j-1})^\top R_1 (K^+ - K_{j-1}) \\ &+ (A_{K^+,L_j} - A_{K_{j-1},L_{j-1}})^\top \Delta (A_{K^+,L_j} - A_{K_{j-1},L_{j-1}}). \end{aligned}$$

Multiplying $v^\top$ and v on each side and combining the resulting expression with (5.9), we obtain,

$$K_{j-1} v = K^+ v.$$

But now we have

$$(A - B_1 K_{j-1} - B_2 L_{j-1}) v = Av - B_1 K^+ v - B_2 L_j v = \lambda v.$$

This is a contradiction to the Schur stability of (K_{j-1}, L_{j-1}). For item (b), note that $X_j \le X_*$, so $R_2 - B_2^\top X_j B_2 \succ 0$ and consequently $\mathbf{O}_j \succ 0$ as $R_1 + B_1^\top X_j B_1 \ge R_1 + B_1^\top X_{j-1} B_1 \succ 0$. Hence, we have completed the proof for items $(b), (c), (d)$ under the assumption of item (a).

We now argue that with the stepsize $\eta_{j-1} \le 1/\lambda_n(\mathbf{O}_{j-1})$, this assumption of stabilizability is indeed valid; namely, item (a) holds. Consider the ray $\{L_t : L_t = L_{j-1} + t2\mathbf{V}_{K_{j-1},L_{j-1}}\}$. We first note that there exists a maximal half-open interval $[0, \sigma)$ such that $(A - B_2 L_t, B_1)$ is stabilizable for every $t \in [0, \sigma)$ and $(A - B_2 L_\sigma, B_1)$ is not stabilizable (this is due to the fact that stabilizability is an open condition; see Proposition 5.11.1 for a proof). Now suppose that $\sigma < \frac{1}{\lambda_n(\mathbf{O}_{j-1})}$. We may take a sequence $t_l \uparrow \sigma$, and note that $(A - B_2 L_{t_l}, B_1)$ is stabilizable for every t_l. Let us denote the corresponding sequence of solutions to the DARE by $\{Z_{t_l}\}$. By our previous arguments, $\Delta \le Z_{t_l} \le X_*$, where X_* is the corresponding value matrix at Nash equilibrium point (K_*, L_*). Denote by $\mathcal{L}$ as the set of all limit points of the sequence $\{Z_{t_l}\}_{l=1}^\infty$. By Weirestrass-Balzano, the set $\mathcal{L}$ is nonempty as the sequence is bounded. Clearly, for every $Z \in \mathcal{L}$, $\Delta \le Z \le X_*$. By continuity, Z must solve the DARE,

$$\mathcal{R}_{A_{0,L_\sigma}, B_1, Q_{L_\sigma}, R_1}(X) = 0. \tag{5.10}$$

Putting $K' = (R_1 + B_1^\top Z B_1)^{-1} B_1^\top Z(A - B_2 L_j)$, we claim that $A - B_2 L_\sigma - B_1 K'$ is Schur stable. This is a consequence of (d) and the Comparison Lemma 5.5.1: it suffices to observe that A_{K',L_σ} is marginally stable satisfying (d) and,

$$\begin{aligned} Z - \Delta \geq\ & A_{K',L_\sigma}^\top (Z - \Delta) A_{K',L_\sigma} \\ & + \mathbf{V}_{K_{j-1},L_{j-1}}^\top \left(4\sigma I - 4\sigma^2 \mathbf{O}_{j-1}\right) \mathbf{V}_{K_{j-1},L_{j-1}}. \end{aligned}$$

Proceeding similar to the above line of reasoning, we can show that $A - B_2 L_\sigma - B_1 K'$ is Schur stable. But this contradicts our standing assumption that $(A - B_2 L_\sigma, B_1)$ is not stabilizable. Hence, for all $\eta_{j-1} \leq 1/\lambda_n(\mathbf{O}_{j-1})$, the pair $(A - B_2 L_j, B_1)$ is indeed stabilizable. □

We are now ready to state the convergence rate for the algorithm.

Theorem 5.5.4. *If the stepsize is taken as $\eta_j = 1/(2\lambda_n(\mathbf{O_{j-1}}))$, then,*

$$\sum_{j=0}^{\infty} \|\mathbf{N}_g(L_j)\|_F^2 \leq \frac{1}{\eta} \left(g(L_*) - g(L_0)\right),$$

where $\eta \in \mathbb{R}_+$ is some positive constant.

Remark. *This theorem suggests the gradient will vanish at a sublinear rate. As we know, there is a unique stationary point of g; this means the sublinear convergence to global maximum, i.e., Nash equilibrium point.*

Proof. Let $\eta = \inf_j 1/\lambda_n(\mathbf{O}_j)$ and note that $\mathbf{O}_j \geq R_2 - B_2^\top X_* B_2$, so $\eta > 0$. It suffices to note that by Lemma 5.5.3, we have

$$\begin{aligned} g(L_j) - g(L_{j-1}) &= \mathbf{Tr}((X_{K_j,L_j} - X_{K_{j-1},L_{j-1}})\mathbf{\Sigma}) \\ &\geq \frac{1}{\lambda_n(\mathbf{O}_{j-1})} \mathbf{Tr}\left[Y_{K_j,L_j}\left(N_g(L_{j-1})^\top N_g(L_{j-1})\right)\right] \\ &\geq \eta \|N_g(L_{j-1})\|_F^2. \end{aligned}$$

Telescoping the sum and noting that $g(L)$ is bounded above by $g(L_*)$, we have

$$\sum_{j=0}^{\infty} \|N_g(L_j)\|_F^2 \leq \frac{1}{\eta}\left(g(L_*) - g(L_0)\right) < \infty.$$

□

We observe that the convergence rate is asymptotically linear. This is a consequence of the local curvature of $g(L)$. Indeed, if we compute the Hessian at $g(L_*)$, the action of the Hessian (see Appendix 5.10 for details) is given by

$$\nabla^2 g(L_*)[E, E] = -2\langle \mathbf{O}_{X_*} E, EY_* \rangle.$$

As $\nabla^2 g(L_*)$ is negative definite, $-g(L)$ is locally strongly convex around a convex neighborhood of L_*. It thus follows that gradient descent enjoys a linear convergence rate around L_*. Moreover, the local curvature of $g(L)$ around L_* implies the iterates L_j converges to L_* asymptotically and thereby, the sequence K_j, L_j converges to the Nash Equilibrium (K_*, L_*) [14].

Corollary 5.5.4.1. *Suppose* (K_j, L_j) *is the sequence of policies generated by Algorithm 3 and* (K_*, L_*) *denotes the NE. Then we have* $(K_j, L_j) \to (K_*, L_*)$ *as* $j \to \infty$.

Proof. By item d in Lemma 5.5.3, the sequence of value matrices $\{X_j\}$ is monotonically increasing with respect to the p.s.d. ordering and bounded above. Then it follows $X_j \to X_*$. Consider the following map

$$\phi : X \mapsto \begin{pmatrix} R_1 + B_1^\top X B_1 & B_1^\top X B_2 \\ B_2^\top X B_1 & -R_2 + B_2^\top X B_2 \end{pmatrix}^{-1} (B_1, B_2)^\top X A.$$

The map is continuous on its domain. So $\phi(X_j) \to \phi(X_*)$. The proof is completed by noting that $\phi(X_j) = (K_j, L_j)^\top$. □

5.6 Algorithm: quasi-Newton Iterations of L

In this section, we shall assume that the oracle $\mathcal{O}_L$ returns the quasi-Newton direction. The motivation of quasi-Newton is to investigate the second-order local approximation of $g(L)$. Indeed, we may observe that,

$$g(L + \Delta L) \approx g(L) + \langle Y_{L+\Delta L} N_g(L), \Delta L \rangle - \langle \mathbf{O}_L \Delta L, \Delta L \rangle.$$

---

[14] In general, convergence of function value does not imply convergence of iterates.

The approximation can be acquired by a computation of Hessian matrix of $g(L)$. The details are deferred to Appendix.

Algorithm 4 quasi-Newton Policy for LQ Game

1: Initialize L_0 such that $(A - B_1 L_0, B_2)$ is stabilizable and the DARE $\mathcal{R}_{A_{0,L_0},B_1,Q_{L_0},R_1}(X) = 0$ has a stabilizing solution X^+ with $R_1 + B_1^\top X^+ B_1 > 0$.
2: **if** $j \geq 1$ **then**
3: Set: $K_j \leftarrow \arg\min_K f(K, L_{j-1})$.
4: Set: $L_j = L_{j-1} + \eta_{j-1} \mathbf{O}_{j-1}^{-1} 2(-R_2 L_{j-1} - B_2^\top X_{K_{j-1},L_{j-1}} A_{K_{j-1},L_{j-1}})$.
5: **end if**

5.6.1 Convergence Analysis

We first prove a result that can be considered as a counterpart to Lemma 5.5.3.

Lemma 5.6.1. *Suppose that Algorithm 4 is initialized appropriately. With stepsize* $\eta_{j-1} \leq \frac{1}{\lambda_n(\mathbf{O}_{j-1})}$, *we then have,*

a. $(A - B_2 L_j, B_1)$ *is stabilizable for every* $j \geq 1$.

b. $\mathbf{O}_j > 0$ *for every* $j \geq 1$.

c. For every $j \geq 1$, $f(K, L_j)$ *is bounded below over* K *and there exists a unique minimizer* K_j*, which forms a stabilizing pair* (K_j, L_j)*. Namely, the DARE*

$$\mathcal{R}_{A_{0,L_j},B_1,Q_j,R_1}(X) = 0,$$

admits a stabilizing maximal solution X^+ *satisfying* $R + B_1^\top X^+ B_1 > 0$*. In other words,* $g(L_j)$ *is finite and differentiable for every* $j \geq 0$.

d. Putting $\Lambda = X_{K_j,L_j}$, $\mathbf{E}_j = R_1 + B_1^\intercal X_{K_j,L_j} B_1$ *and* $\mathbf{F}_j = B_1^\intercal X_{K_j,L_j}(A - B_2 L_j)$ *we have*

$$\begin{aligned}&\Lambda - \Delta - A_{K_j,L_j}^\intercal (\Lambda - \Delta) A_{K_j,L_j} \\ &= \mathbf{V}_{K_{j-1},L_{j-1}}^\intercal \left(4\eta_{j-1}\mathbf{O}_{j-1}^{-1} - 4\eta_j^2 \mathbf{O}_{j-1}^{-1}\right) \mathbf{V}_{K_{j-1},L_{j-1}} \\ &\quad + (\mathbf{E}_{j-1}K_j - \mathbf{F}_{j-1})^\intercal \mathbf{E}_{j-1}^{-1} (\mathbf{E}_{j-1}K_j - \mathbf{F}_{j-1}).\end{aligned}$$

Proof. The proof proceeds similar to Lemma 5.5.3. The key difference is that the algebraic Riccati map assumes a new form with the quasi-Newton update. Namely, with quasi-Newton iteration,

$$\begin{aligned}&\mathcal{R}_{A_{0,L_j},B_1,Q_j,R_1}(X_{j-1}) \\ &= \left(4\eta_{j-1} I - 4\eta_{j-1}^2\right) \mathbf{V}_{K_{j-1},L_{j-1}}^\intercal \mathbf{O}_{j-1}^{-1} \mathbf{V}_{K_{j-1},L_{j-1}}.\end{aligned}$$

The statements then follows from almost same arguments as in Lemma 5.5.3. ☐

We are now ready to state the convergence rate for the algorithm.

Theorem 5.6.2. *If the stepsize is taken as* $\eta = 1/2$, *then*

$$g(L_*) - g(L_j) \le q(g(L_*) - g(L_{j-1}))^2,$$

for some $q > 0$.

Proof. By Lemma 5.6.1, the sequence of value matrices $\{X_j\}$ is monotonically nondecreasing and bounded above. Thereby $X_j \to X_*$ as $j \to \infty$. It follows the set $\mathcal{E} = \{X_j\} \cup \{X_*\}$ is compact. Substituting $K_{j-1} = (R_1 + B_1^\intercal X_{j-1} B_1)^{-1} B_1^\intercal X_{j-1}(A - B_2 L_{j-1})$ into the update rule, we get

$$L_j = -\mathbf{O}_{j-1}^{-1} B_2^{-1} X_{j-1}(A - B_1(R_1 + B_1^\intercal X_{j-1} B_1)^{-1} B_1^\intercal X_{j-1} A).$$

By Lemma 5.5.2 (take $\tilde{X} = X_*$ and $X = X_j$ and note $\mathbf{V}_{K_*,L_*} = 0$),

$$X_* - X_j \le \sum_{\nu=0}^{\infty} (A_j^\intercal)^\nu \left((L_* - L_j)^\intercal \mathbf{O}_* (L_* - L_j)\right) A_j^\nu,$$

where

$$\mathbf{O}_* = R_2 - B_2^\top X_* B_2 + B_2^\top X_* B_1 (R_1 + B_1^\top X_* B_1)^{-1} B_1^\top X_* B_2.$$

It follows that,

$$\begin{aligned} g(L_*) - g(L_j) &\leq \mathbf{Tr}(Y_*(L_* - L_j)^\top \mathbf{O}_*(L_* - L_j)) \\ &\leq \lambda_n(Y_*)\lambda_n(\mathbf{O}_*)\,\mathbf{Tr}((L_* - L_j)^\top (L_* - L_j)). \end{aligned}$$

We observe

$$\begin{aligned} L_* - L_j = &-\mathbf{O}_*^{-1} B_2^\top X_*(A - B_1(R_1 + B_1^\top X_* B_1)^{-1} B_1^\top X_* A) \\ &+ \mathbf{O}_{j-1}^{-1} B_2^\top X_{j-1}(A - B_1(R_1 + B_1^\top X_{j-1} B_1)^{-1} B_1^\top X_{j-1} A), \end{aligned}$$

and further, note that the map ϕ given by

$$X \mapsto -\mathbf{O}_X^{-1} B_2^\top X \left(A - B_1(R_1 + B_1^\top X B_1)^{-1} B_1^\top X A\right)$$

is smooth where,

$$\mathbf{O}_X = R_2 - B_2^\top X B_2 + B_2^\top X B_1 (R_1 + B_1^\top X B_1)^{-1} B_1^\top X B_2.$$

So over the compact set $\mathcal{E}$, we can find a Lipschitz constant β of ϕ, namely, for every $X, X' \in \mathcal{E}$, we have $\|\phi(X) - \phi(X')\|_F \leq \beta \|X - X'\|_F$. Then

$$\|L_* - L_j\|_F^2 = \|\phi(X_*) - \phi(X_j)\|_F^2 \leq \beta^2 \|X_* - X_{j-1}\|_F^2.$$

Hence

$$g(L_*) - g(L_j) \leq c\|X_* - X_{j-1}\|_F^2 \leq q\left(g(L_*) - g(L_{j-1})\right)^2,$$

where $c, q > 0$ are constants. □

By virtually the same argument of Corollary 5.5.4.1, we may conclude that the sequence of policies $\{K_j, L_j\}$ generated by quasi Newton iteration converges to NE at a Q-quadratic rate.

Corollary 5.6.2.1. *Suppose $\{K_j, L_j\}$ is the sequence of policies generated by Algorithm 4. Then we have $(K_j, L_j) \to (K_*, L_*)$.*

5.7 Switching the Leader in the Sequential Algorithms

We shall demonstrate in this section if the condition $a2$ in the assumption holds, it might not be guaranteed that $g(L)$ is differentiable in a neighborhood of L_*. In this case, however, choosing the leader to be player K would converge. The analysis proceeds in a similar manner. First, we observe that we can define a value function,

$$h(K) = \sup_{L \in \mathcal{S}_K} f(K, L).$$

Following virtually the same argument, we can the establish the following.

Proposition 5.7.1. *Suppose that $\mathcal{U} \subseteq \mathbf{dom}(h)$ is an open set such that for every $K \in \mathcal{U}$, there is a unique maximizer of $f(K, L)$ over L. Then $h(K)$ is differentiable and the gradient is given by*

$$\nabla h(K) = \nabla_K f(K, L_K), \text{ where } L_K = \arg\max_{L \in \mathcal{S}_K} f(K, L).$$

The algorithm for player K using natural gradient policy can be described similarly to Algorithm 3.

Algorithm 5 Natural Gradient Policy with Leader K

1: Initialize K_0 such that $(A - B_1 K_0, B_2)$ is stabilizable and the DARE

$$\mathcal{R}_{A_{(K_0,0)}, B_2, -Q_{(K_0,0)}, R_2}(X) = 0.$$

has a maximal symmetric solution Z^+ with $R_2 + B_2^\top Z^+ B_2 \succ 0$.

2: **if** $j \geq 1$ **then**

3: Set: $L_{j-1} \leftarrow \arg\max_L f(K_{j-1}, L)$.

4: Set: $K_j = K_{j-1} - \eta_j N_h(L_j) \equiv K_{j-1} - \eta_j N_{f,K}(K_{j-1}, L_{j-1})$.

5: **end if**

We observe that for fixed K', if L' is the unique stabilizing maximizer of $f(K', L)$ over L, then substituting $\nabla_L f(K', L') = 0$ into the Lyapunov matrix equation, L' solves the following

ARE,

$$Z = A_{(K',0)} Z A_{(K',0)} + Q + K'^{\top} R_1 K' - A^{\top}_{(K',0)} Z B_2(-R_2 + B_2^{\top} Z B_2)^{-1} B_2^{\top} Z A_{(K',0)}. \tag{5.11}$$

To utilize the theory developed in standard ARE, which concerns a minimization problem, we may consider following modification:

$$W = A^{\top}_{(K',0)} W A_{(K',0)} - Q - K'^{\top} R_1 K' - A^{\top}_{(K',0)} W B_2(R_2 + B_2^{\top} W B_2)^{-1} B_2^{\top} W A_{(K',0)}. \tag{5.12}$$

We observe that if W solves (5.12), then $-W$ solves (5.11). Now the analysis can be done almost in parallel.

Lemma 5.7.2. *Suppose that L_{j-1} is the unique stabilizing maximizer of $f(K_{j-1}, L)$, i.e., $L_{j-1} = \arg\max_L f(K_{j-1}, L)$. Putting $\Delta = X_{K_{j-1},L_{j-1}}$ and*

$$\mathbf{P}_{j-1} = R_1 + B_1^{\top}\Delta B_1 + B_1^{\top}\Delta B_2(R_2 - B_2^{\top}\Delta B_2)^{-1} B_2^{\top}\Delta B_1,$$

with stepsize $\eta_{j-1} \le \frac{1}{\lambda_n(\mathbf{P}_{j-1})}$, we then have,

a. $(A - B_1 K_j, B_2)$ *is stabilizable.*

b. $f(K_j, L)$ *is bounded above over L and there exists a unique stabilizing maximizer L_j, namely (K_j, L_j) is a stabilizing pair.*

c. Putting $\Lambda = X_{K_j,L_j}$, we have

$$\Delta - \Lambda \ge A^{\top}_{K_j,L_j}(\Delta - \Lambda) A_{K_j,L_j} + \mathbf{U}^{\top}_{K_{j-1},L_{j-1}}\left(4\eta_{j-1} I - 4\eta_j^2 \mathbf{P}_{j-1}\right)\mathbf{U}_{K_{j-1},L_{j-1}}.$$

Proof. The proof proceeds similarly to Lemma 5.5.3. Indeed, putting $\tilde{\Delta} = -\Delta$ and $\tilde{\Lambda} = -\Lambda$, we observe $\tilde{\Delta}$ and $\tilde{\Lambda}$ solves the DARE (5.12) and also note the DARE (5.12) has the same form with the DARE considered in Lemma 5.5.3. Further, note that the update rule is equivalent to

$$
\begin{aligned}
K_j &= K_{j-1} - \eta_{j-1} 2(R_1 K_{j-1} - B_1^\top \Delta A_{K_{j-1},L_{j-1}}) \\
&= K_{j-1} - \eta_{j-1} 2(R_1 K_{j-1} + B_1^\top \tilde{\Delta} A_{K_{j-1},L_{j-1}}) \\
&= K_{j-1} + 2\eta_{j-1}(-R_1 K_{j-1} - B_1^\top \tilde{\Delta} A_{K_{j-1},L_{j-1}}).
\end{aligned}
$$

In view of these observations, by the same machinery we employed in Lemma 5.5.3, we conclude

$$
\begin{aligned}
\tilde{\Lambda} - \tilde{\Delta} \geq\ & A_{K_j,L_j}^\top (\tilde{\Lambda} - \tilde{\Delta}) A_{K_j,L_j} \\
&+ \tilde{\mathbf{U}}_{K_{j-1},L_{j-1}}^\top \left(4\eta_{j-1} I - 4\eta_j^2 \tilde{\mathbf{P}}_{j-1}\right) \tilde{\mathbf{U}}_{K_{j-1},L_{j-1}},
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{\mathbf{U}}_{K_{j-1},L_{j-1}} &= -R_1 K_{j-1} - B^\top \tilde{\Delta} A_{K_{j-1},L_{j-1}}, \\
\tilde{\mathbf{P}}_{j-1} &= R_1 - B_1^\top \tilde{\Delta} B_1 \\
&\quad + B_1^\top \tilde{\Delta} B_2 (R_2 + B_2^\top \tilde{\Delta} B_2)^{-1} B_2^\top \tilde{\Delta} B_1.
\end{aligned}
$$

It thus follows that,

$$
\begin{aligned}
-\Lambda + \Delta \geq\ & A_{K_j,L_j}^\top (\Delta - \Lambda) A_{K_j,L_j} \\
&+ \mathbf{U}_{K_{j-1},L_{j-1}}^\top \left(4\eta_{j-1} I - 4\eta_j^2 \mathbf{P}_{j-1}\right) \mathbf{U}_{K_{j-1},L_{j-1}}.
\end{aligned}
$$

□

Now it is straightforward to conclude the sublinear convergence rate of Algorithm 5.

Lemma 5.7.3. *Suppose $\{K_j\}$ are the iterates generated by Algorithm 5. Then we have*

$$
\sum_{j=0}^{\infty} \mathbf{N}_h(K_j) \leq \eta \left(h(K_0) - h(K_*)\right),
$$

where $\eta > 0$ is some positive number.

The analysis of quasi-Newton method with K as the leader proceeds in a similar manner as Algorithm 4; as such we omit the details here.

5.8 Simulation Results

In this section, we provide a representative set of examples to demonstrate the resutls repoarted in this paper.

The linear dynamical system of form with parameters $(A, [B_1, B_2]))$ where $A = 0.5I \in \mathbb{R}^{5\times 5}$,

$$B_1 = \begin{pmatrix} I_{3\times 3} \\ 0_{2\times 3} \end{pmatrix}, \qquad B_2 = \begin{pmatrix} 0_{2\times 2} \\ I_{2\times 2} \end{pmatrix}.$$

The cost matrices Q, R_1, R_2 are given by

$$Q = \begin{pmatrix} -3.01760074 & -2.81426740e-03 & 4.32973286e-3 & -1.06481328e-2 & -6.85531223e-3 \\ -2.81426740e-3 & -3.00443882 & 8.68684314e-3 & 2.52575795e-4 & -3.03010149e-3 \\ 4.32973286e-3 & 8.68684314e-3 & 2.99678507 & -1.12890593e-2 & 3.76770625e-3 \\ -1.06481328e-2 & 2.52575795e-4 & -1.12890593e-2 & 2.99982394 & -4.41594781e-03 \\ -6.85531223e-3 & -3.03010149e-3 & 3.76770625e-3 & -4.41594781e-3 & 3.00709409 \end{pmatrix}$$

$$R_1 = \begin{pmatrix} -1 & & \\ & -1 & \\ & & -1 \end{pmatrix}, \qquad R_2 = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix},$$

where Q is indefinite with rounded eigenvalues $\{-3.01820999, -3.00387344, 2.98692324, 3.00247667, 3.014347$ Note the choices of R_1 and R_2 are also interesting as player 1 has negative definite cost matrix. The value matrix X_* at Nash Equilibrium is indefinite with rounded eigenvalues $\{-2.6132689, -2.59737491, 2.57855823, 2.59584142, 2.60897107\}$.

Figure 5.1 and Figure 5.2 demonstrate the convergence of natural gradient *ascent* with leader L. Figure 5.3 and Figure 5.4 shows the Q-quadratic convergence rate of quasi Newton iteration. The results are consistent with the theories developed in § 5.5 and § 5.6. Figures 5.5 5.6 5.7 5.8 shows the global convergence of natural gradient *descent* and quasi-Newton when player K assumes the leader.

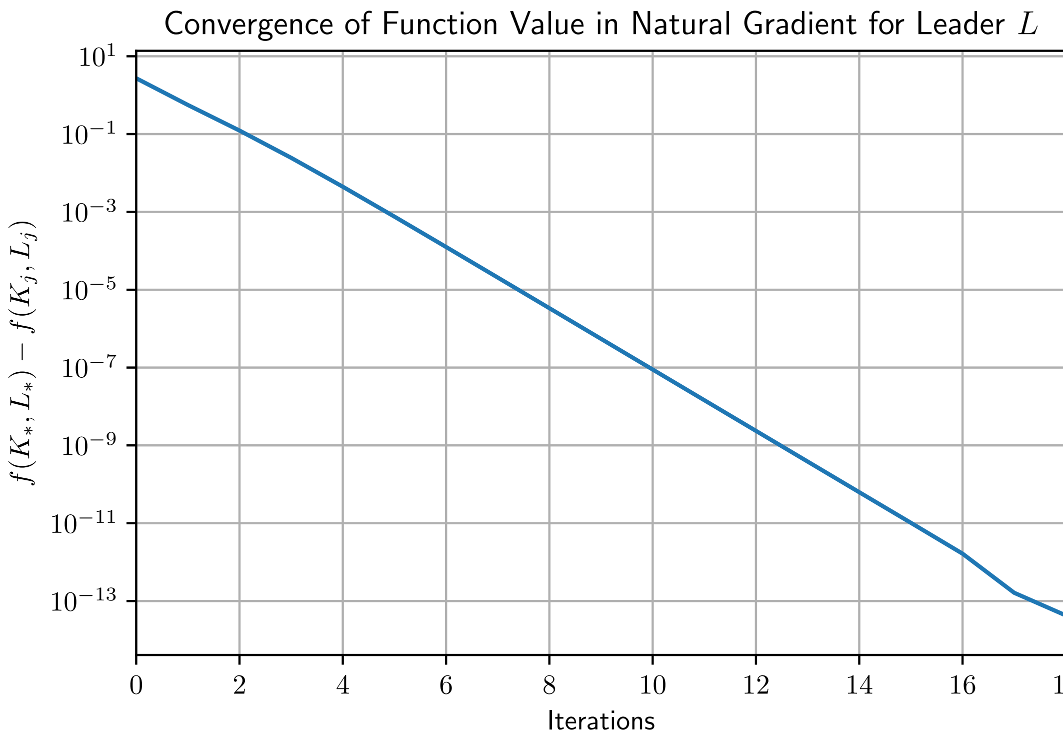

Figure 5.1: Convergence of function value under natural gradient sscent for player L with stepsize given in Theorem 5.5.4

5.9 Concluding Remarks

The papers considers sequential policy-based algorithms for LQ dynamic games. We prove global convergence of several policy gradient-based algorithms as well as identifying the role of control theoretic constructs in their analysis. Moreover, we have clarified a number of intricate issues pertaining to stabilization for LQ games and indefinite cost structure, while removing restrictive assumptions and circumventing the projection step.

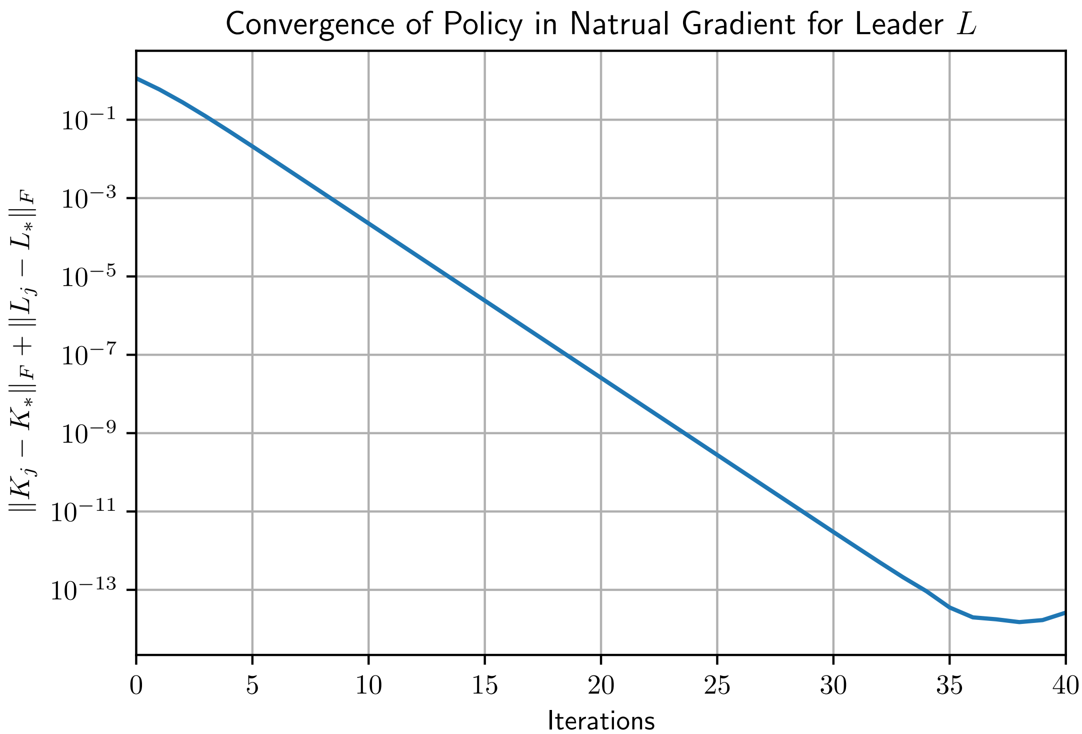

Figure 5.2: Convergence of polices under natural gradient sscent for player L with stepsize given in Theorem 5.5.4

5.10 *Hessian of* $g(L)$

In this section, we compute the Hessian of $g(L)$ at a point of differentiation of L_0. Indeed, we shall assume stronger assumptions of L_0: there is a unique stabilizing minimizer of $f(K, L_0)$ over L_0, denoted by K_0. Throughout the section, we denote $A_0 = A - B_1K_0 - B_2L_0$. Note, by assumption the DARE is solvable

$$\begin{aligned} X &= (A - B_2L_0)^\top X(A - B_2L_0) + Q - L_0^\top R_2 L_0 \\ &\quad + (A - B_2L_0)^\top X B_1(R_1 + B_1^\top X B_1)^{-1} B_1^\top X(A - B_2L_0), \end{aligned}$$

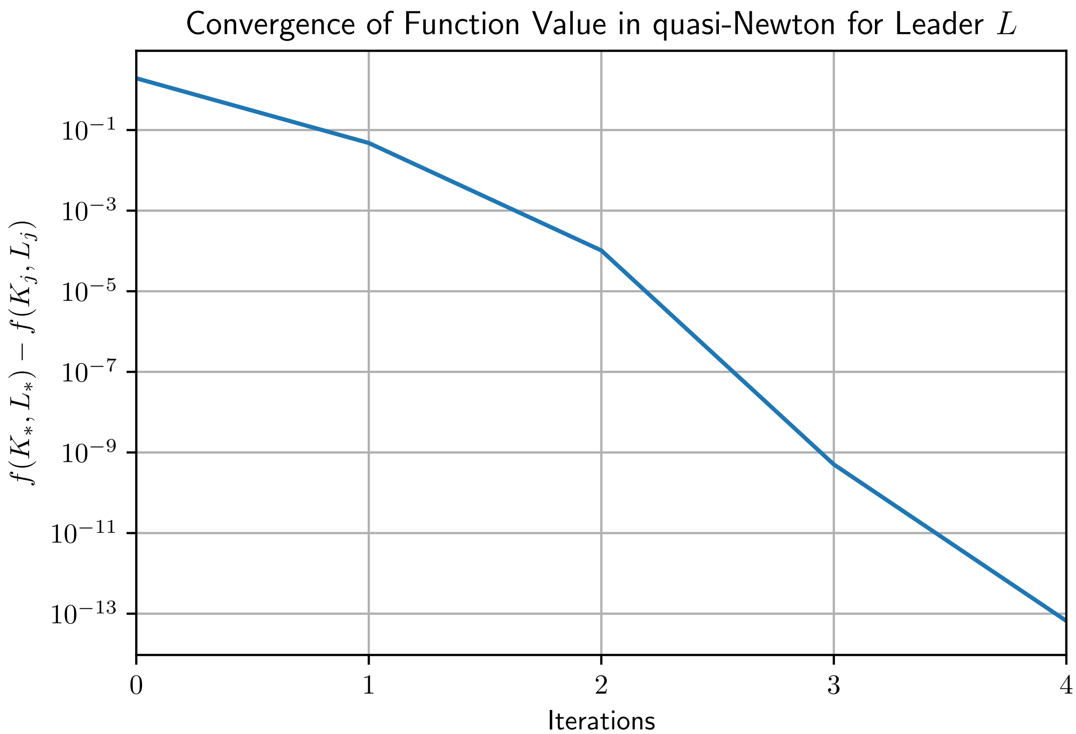

Figure 5.3: Convergence of function value under quasi Newton iteration for player L with stepsize 1/2

and the maximal solution X_0 is stabilizing, i.e., $K_0 = (R_1 + B_1^\top X_0 B_1)^{-1} B_1^\top X(A - B_2 L_0)$ is stabilizing the system $(A - B_2 L_0, B_1)$. As we have noted, the gradient of $g(L_0)$ is given by

$$\nabla g(L_0) = 2(-R_2 L_0 - B_2^\top X_0 (A - B_2 L_0 - B_1 K_0)) Y_0,$$

where Y_0 is the solution to the Lyapunov matrix equation

$$A_0 Y A_0^\top + \mathbf{\Sigma} = Y.$$

We now compute the Fréchet derivative of $\phi(L_0) = 2(-R_2 L_0 - B_2^\top X_0 A_0)$. Note $\phi : \mathbb{R}^{m_2 \times n} \to \mathbb{R}^{m_2 \times n}$, the Fréchet derivative $D\phi(L_0)$ is a bounded linear map in $\mathcal{L}(\mathbb{R}^{m_2 \times n}, \mathbb{R}^{m_2 \times n})$. So the

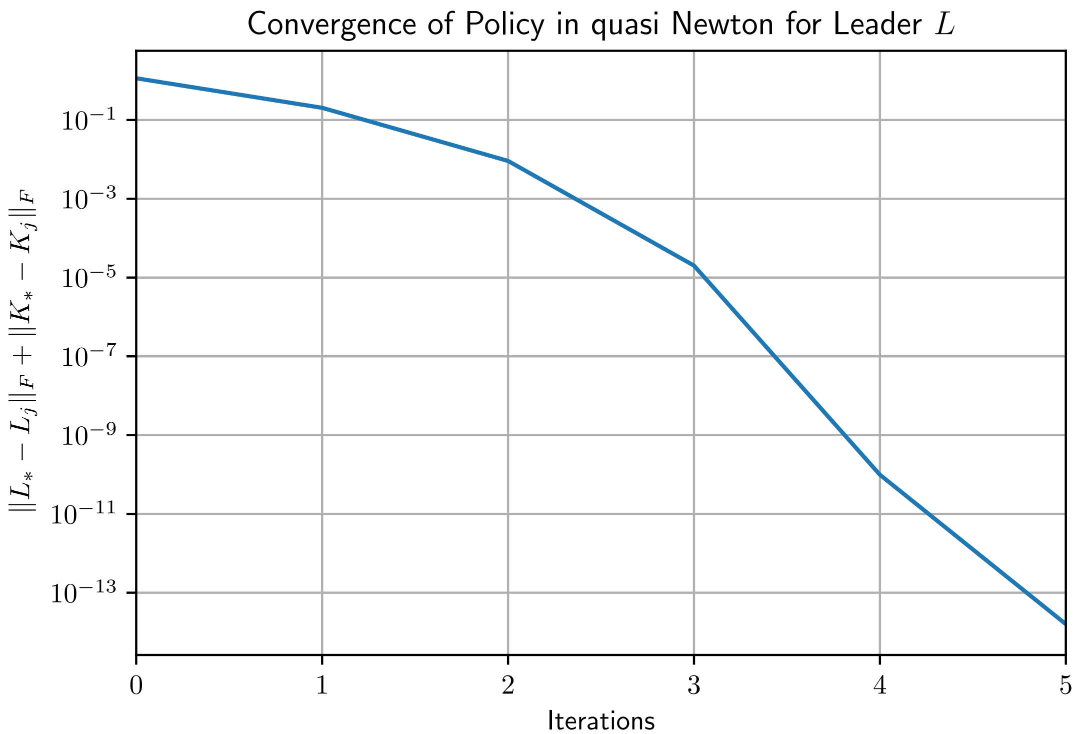

Figure 5.4: Convergence of policy error under quasi Newton iteration for player L with stepsize 1/2

action of $D\phi(L_0)$ at any $E \in \mathbb{R}^{m_2 \times n}$, denoted by $D\phi(L_0)[E] =: \mathbf{D} \in \mathbb{R}^{m_2 \times n}$ is given by

$$\begin{aligned}\mathbf{D} = 2(&-R_2 E - B_2^\top X_0'(E)(A - B_2 L_0 - B_1 K_0) \\ &- B_2^\top X_0(-B_2 E) - B_2^\top X_0(-B_2 K_0'(E))),\end{aligned}$$

where $X_0'(E) \in \mathbb{R}^{n \times n}$ (respectively, $K_0'(E)$) is the action of the Fréchet derivative of X_0 (respectively, K_0) with respect to L_0. Here we concern X_0, K_0 as maps of L. Now $X_0'(E)$

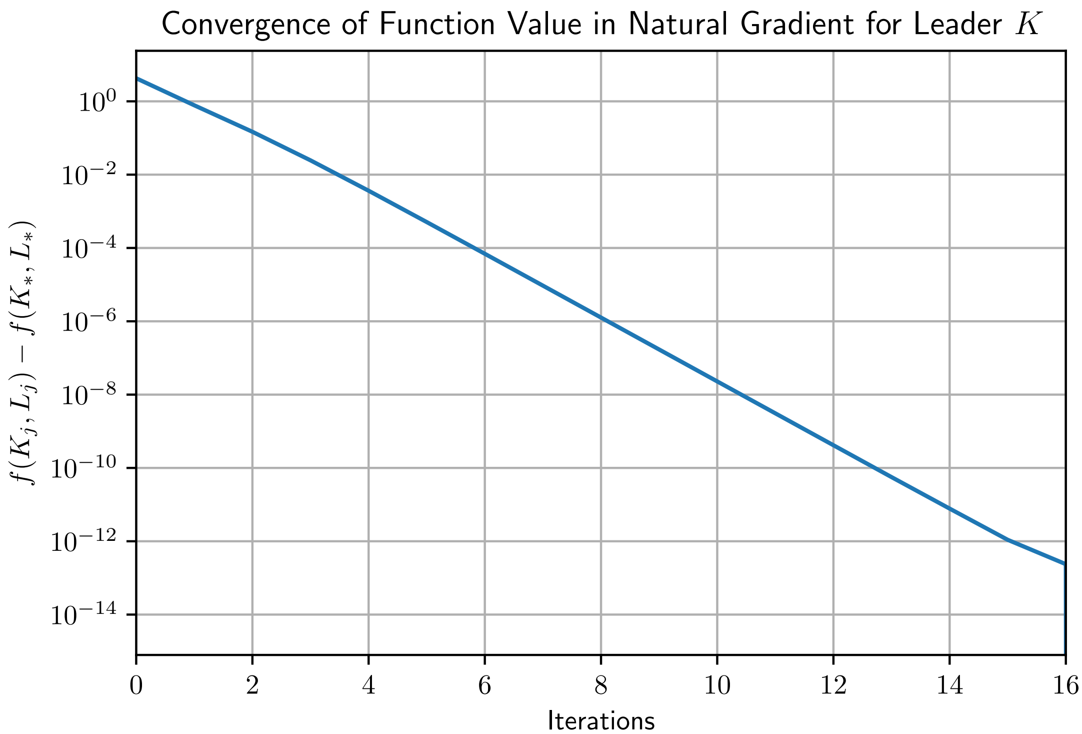

Figure 5.5: Convergence of function value under natural gradient descent for player K with stepsize in Lemma 5.7.3

satisfies

$$\begin{aligned}
X_0'(E) = A_0^\top X_0'(E) A_0 - E^\top R_2 L_0 - L_0^\top R_2 E + K_0'(E)^\top R_1 K_0 \\
+ K_0^\top R_1 K_0'(E) - (B_2 E + B_1 K_0'(E))^\top X_0 A_0 \\
- A_0^\top X_0 (B_1 K_0'(E) + B_2 E).
\end{aligned}$$

Noting $R_1 K_0 - B_1^\top X_0 A_0 = 0$, we have $X_0'(E)$ is the solution to the following Lyapunov equation

$$\begin{aligned}
X_0'(E) = A_0^\top X_0'(E) A_0 - E^\top (R_2 L_0 + B_2^\top X_0 A_0) \\
- L_0^\top (R_2 E + A_0^\top X_0 B_2 E).
\end{aligned}$$

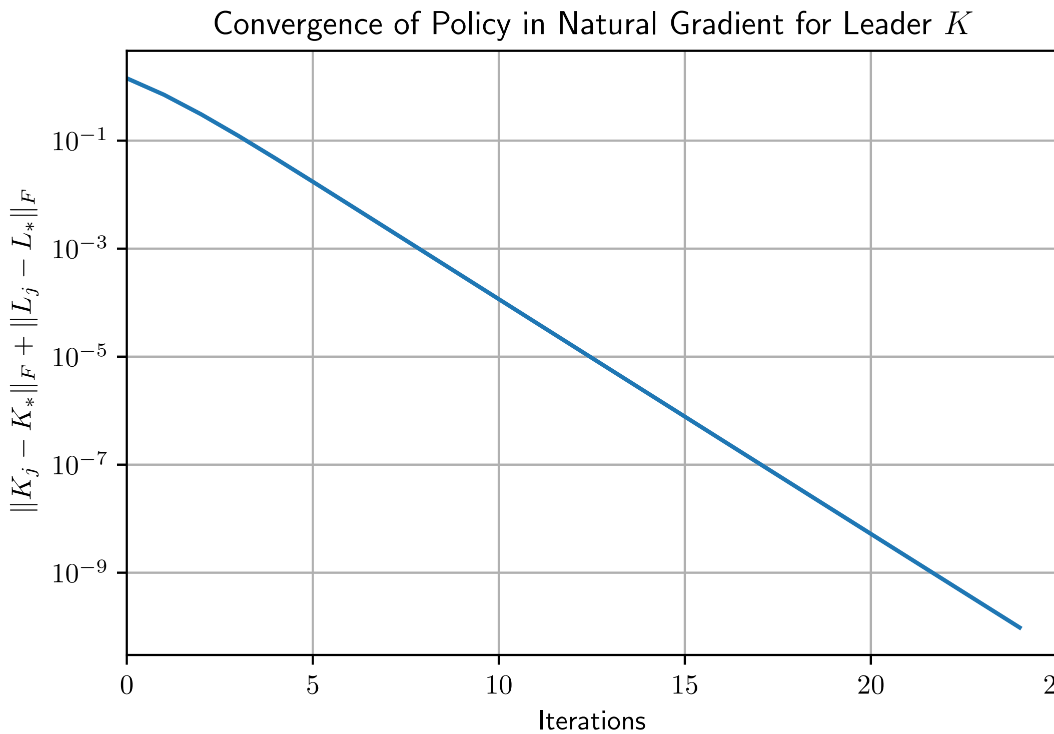

Figure 5.6: Convergence of policy under natural gradient descent for player K with stepsize in Lemma 5.7.3

As A_0 is Schur, the solution exists and is unique

$$X_0'(E) = \sum_{j=0}^{\infty} (A_0^\top)^j [-E^\top (R_2 L_0 + B_2^\top X_0 A_0) - L_0^\top (R_2 E + A_0^\top X_0 B_2 E)] A_0^j.$$

Similarly, we may compute

$$\begin{aligned} K_0'(E) &= (R_1 + B_1^\top X_0 B_1)^{-1} B_1^\top X_0 (-B_2 E) \\ &\quad + (R_1 + B_1 X_0 B_1)^{-1} B_1^\top X_0'(E)(A - B_2 L_0) \\ &\quad + (R_1 + B_1^\top X_0 B_1)^{-1} B_1^\top X_0'(E) B_1 (R_1 + B_1^\top X_0 B_1)^{-1} B_1^\top X_0 (A - B_2 L_0), \end{aligned}$$

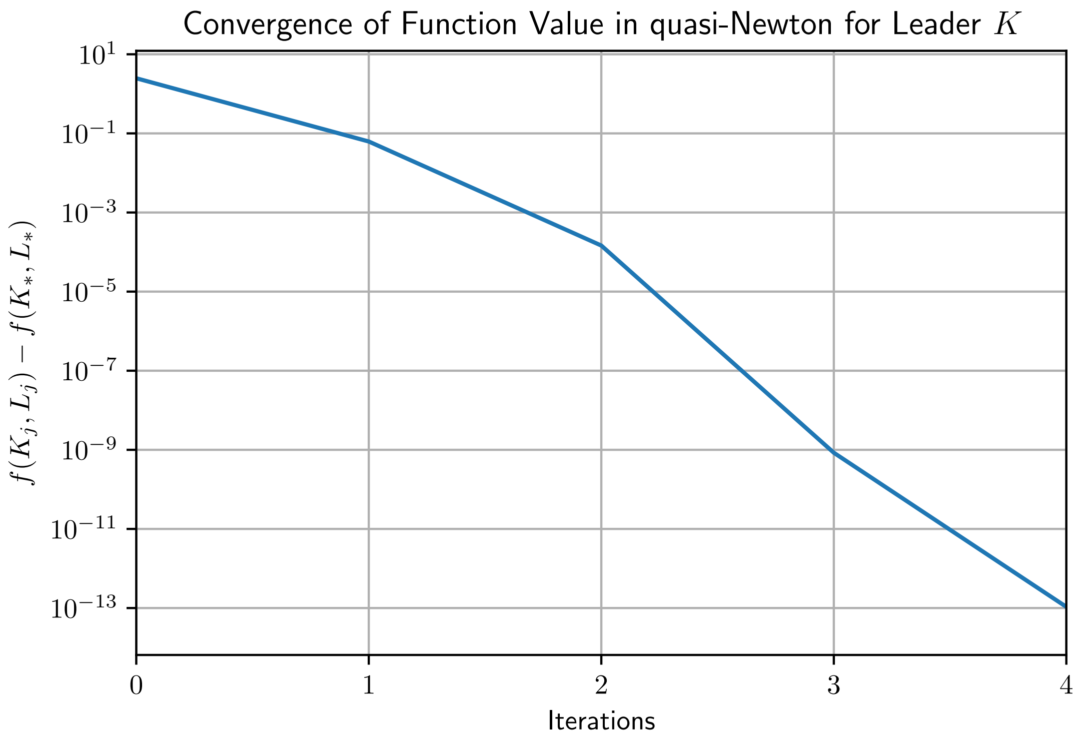

Figure 5.7: Convergence of function value under quasi Newton iteration for player K with stepsize 1/2

and

$$Y_0'(E) = \sum_{j=0}^{\infty} A_0^j \left[A_0 Y_0 (-B_2 E)^\top + (-B_2 E) Y_0 A_0^\top \right] (A_0^\top)^j.$$

Combining the computations, we have the action of the Hessian is given by

$$\begin{aligned}
&\langle \nabla^2 g(L) E, E \rangle \\
&= 2\langle -R_2 + B_2^\top X_0 B_2 - B_2^\top X_0 B_1 (R_1 + B_1^\top X_0 B_1)^{-1} B_1^\top X_0 B_2) E, E \rangle \\
&+ 2\langle -B_2^\top X_0'(E) A_0 - B_2^\top X_0 B_1 (R_1 + B_1 X_0 B_1)^{-1} B_1^\top X_0'(E)(A - B_2 L_0), E \rangle \\
&+ 2\langle B_2^\top X_0 B_1 (R_1 + B_1^\top X_0 B_1)^{-1} B_1^\top X_0'(E) B_1 (R_1 + B_1^\top X_0 B_1)^{-1} B_1^\top X_0 (A - B_2 L_0), E \rangle \\
&+ 2\langle (-R_2 L_0 - B_2^\top X_0 A_0) Y_0'(E), E \rangle
\end{aligned}$$

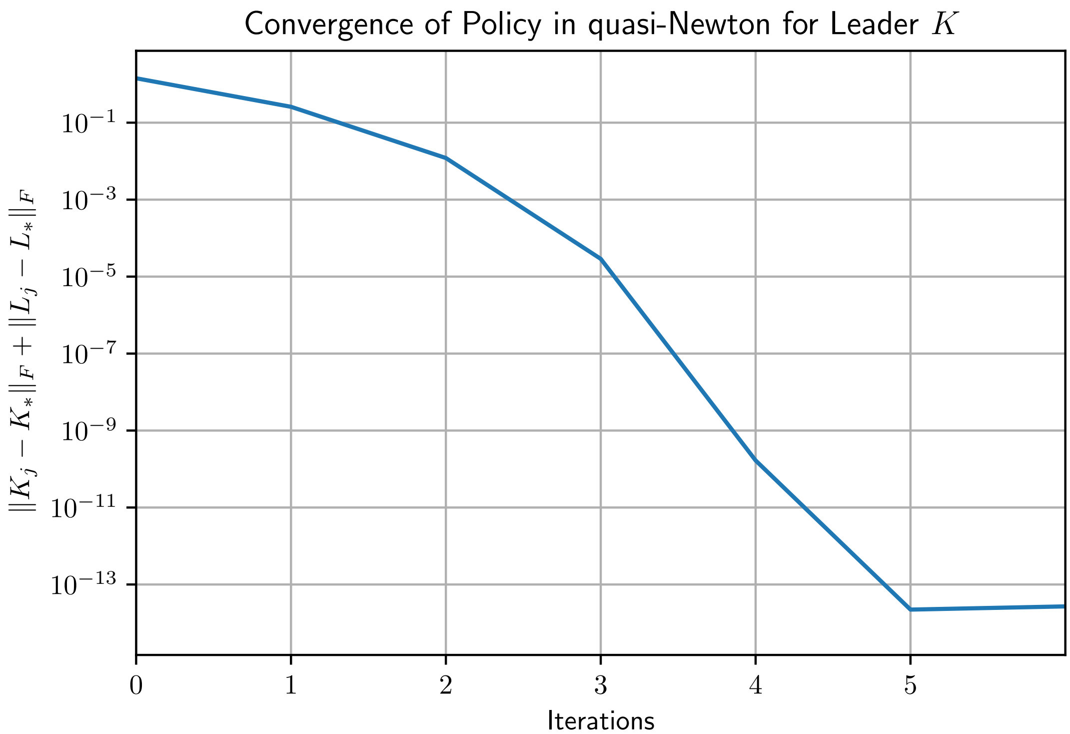

Figure 5.8: Convergence of poicy under quasi Newton iteration for player K with stepsize 1/2

It is instructive to note that at L_*, $X_*'(E) = 0$ since $-R_2 L_* - B_2^\intercal X_* A_* = 0$. So the action of Hessian at L_* is given by

$$\nabla^2 g(L_*)[E, E] = \langle \big(-R_2 + B_2^\intercal X_0 B_2 - B_2^\intercal X_0 B_1 (R_1 + B_1^\intercal X_0 B_1)^{-1} B_1^\intercal X_0 B_2)\big) E Y_0, E\rangle.$$

That is, $\nabla^2 g(L_*)$ is a positive definite operator by condition $(a1)$ in the assumption. Hence, $-g(L)$ is locally strongly convex in a convex neighborhood.

Remark. *If we ignore the formality, the above computation is nothing but a linear approximation.*

5.11 A Useful Control Theoretic Observation

In the proof to Lemma 5.5.3 and 5.6.1, we have used the fact that the set $\{L : (A - B_2L, B_1)$ is stabilizable$\}$ is open; here is the justification.

Proposition 5.11.1. *Suppose $A \in \mathbb{R}^{n\times n}$, $B_1 \in \mathbb{R}^{n\times m_1}$ and $B_2 \in \mathbb{R}^{n\times m_2}$ are fixed. Then the set*

$$\mathcal{L} = \{L \in \mathbb{R}^{m_2\times n} : (A - B_2L, B_1) \text{ is stabilizable}\}$$

is open in $\mathbb{R}^{m_2\times n}$.

Proof. Recall a pair $(A - B_2L, B_1)$ is stabilizable if and only if there exists some $F \in \mathbb{R}^{m_1\times n}$ such that $A - B_2L - B_1F$ is Schur. So $(A - B_2L, B_1)$ is stabilizable if and only if there exists $X \succ 0$ and $F \in \mathbb{R}^{m_1\times n}$ such that

$$(A - B_2L - B_1F)^\top X(A - B_2L - B_1F) - X \prec 0.$$

Now consider the map $\psi : \mathbb{S}_n^{++} \times \mathbb{R}^{m_2\times n} \times \mathbb{R}^{m_1\times n} \to \mathbb{R}$ by

$$\begin{aligned}(X, L, F) &\mapsto (A - B_2L - B_1F)^\top X(A - B_2L - B_1F) - X \\ &\mapsto \lambda_{\max}\left((A - B_2L - B_1F)^\top X(A - B_2L - B_1F) - X\right).\end{aligned}$$

The map ψ is continuous as it is a composition of continuous maps. It thus follows that $\psi^{-1}((-\infty, 0))$ is open. We now observe that $\mathcal{L} \equiv \pi_2(\psi^{-1}(-\infty, 0))$ where π_2 is the projection onto the second coordinate. Since projection map is open map[15], $\mathcal{L}$ is open. $\square$

[15] A map $f : X \to Y$ is open if $f(U)$ is open in Y whenever $U \subseteq X$ is an open set.

Chapter 6

CONCLUSION REMARKS

We have studied the geometry in feedback control and first-order algorithms for LQR, least squares stationary optimal control and LQ dynamic games. Our presentation focuses on the fundamental issues related to optimal control problems when direct policy updates are adopted. For examples, issues related to set of stabilizing feedback policies and stabilization during the iterative updating process. Overall, we present a complete account of topological and metrical properties of the set of stabilizing controls, propose stepsize criteria for LQR, least squares stationary optimal control and LQ dynamic games and prove the global convergence of the gradient-based algorithms.

BIBLIOGRAPHY

[1] Brian D. O. Anderson and John B. Moore. *Optimal Control: Linear Quadratic Methods.* Prentice-Hall, Inc., Upper Saddle River, NJ, 1990.

[2] V. Balakrishnan and L. Vandenberghe. Semidefinite programming duality and linear time-invariant systems. *IEEE Transactions on Automatic Control*, 48(1):30–41, 2003.

[3] Tamer Başar and Pierre Bernhard. *H_∞ Optimal Control and related Minimax Design Problems: a Dynamic Game Approach.* Springer Science & Business Media, 2008.

[4] Tamer Basar and Geert Jan Olsder. *Dynamic Noncooperative Game Theory.* SIAM, 1999.

[5] Heinz H Bauschke and Patrick L Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer Science & Business Media, 2nd edition, 2017.

[6] Pierre Bernhard. Linear quadratic zero-sum two-person differential games. *Encyclopedia of Systems and Control*, pages 649–653, 2015.

[7] Pierre Bernhard and Alain Rapaport. On a theorem of Danskin with an application to a theorem of von Neumann-Sion. *Nonlinear Analysis*, 24(8):1163–1182, 1995.

[8] D.P. Bertsekas. *Dynamic Programming and Optimal Control*, volume II. Athena Scientific, 4th edition, 2012.

[9] S Białas and J Garloff. Convex combinations of stable polynomials. *Journal of the Franklin Institute*, 319(3):373–377, 1985.

[10] S. J. Bradtke, B. E. Ydstie, and A. G. Barto. Adaptive linear quadratic control using policy iteration. In *Proceedings of 1994 American Control Conference*, volume 3, pages 3475–3479, 1994.

[11] 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*, 2019.

[12] Jingjing Bu, Afshin Mesbahi, and Mehran Mesbahi. On topological and metrical properties of stabilizing feedback gains: the MIMO case. *https://arxiv.org/abs/1904.02737*, 2019.

[13] Jingjing Bu, Afshin Mesbahi, and Mehran Mesbahi. On topological and metrical properties of stabilizing feedback gains: the MIMO case. *arXiv preprint arXiv:1904.02737*, 2019.

[14] Jingjing Bu, Afshin Mesbahi, and Mehran Mesbahi. On topological properties of the set of stabilizing feedback gain. *IEEE Transactions on Automatic Control*, 2020.

[15] Jingjing Bu, Afshin Mesbahi, and Mehran Mesbahi. Policy gradient-based algorithms for continuous-time linear quadratic control. *arXiv preprint arXiv:2006.09178*, 2020.

[16] Jingjing Bu and Mehran Mesbahi. Global convergence of policy gradient algorithms for indefinite least squares stationary optimal control. *IEEE Control Systems Letters*, 4(3):638–643, 2020.

[17] Jingjing Bu, Lillian J Ratliff, and Mehran Mesbahi. Global convergence of policy gradient for sequential zero-sum linear quadratic dynamic games. *arXiv preprint arXiv:1911.04672*, 2019.

[18] Christopher I. Byrnes. Algebraic and geometric aspects of the analysis of feedback systems. In Christopher I. Byrnes and Clyde F. Martin, editors, *Geometrical Methods for the Theory of Linear Systems*, pages 85–124. Springer Netherlands, 1980.

[19] Tae Yoon Chun, Jae Young Lee, Jin Bae Park, and Yoon Ho Choi. Stability and monotone convergence of generalised policy iteration for discrete-time linear quadratic regulations. *International Journal of Control*, 89(3):437–450, 2016.

[20] Andrew R Conn, Katya Scheinberg, and Luis N Vicente. *Introduction to Derivative-free Optimization*, volume 8. SIAM, 2009.

[21] Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. *arXiv:1710.01688*, 2017.

[22] Jacob Engwerda. *LQ Dynamic Optimization and Differential Games*. John Wiley & Sons, 2005.

[23] A. Fam and J. Meditch. A canonical parameter space for linear systems design. *IEEE Transactions on Automatic Control*, 23(3):454–458, 1978.

[24] Maryam Fazel, Rong Ge, Sham Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *Proceedings of the 35th International Conference on Machine Learning*, pages 1467–1476, 2018.

[25] Harriet Fell. On the zeros of convex combinations of polynomials. *Pacific Journal of Mathematics*, 89(1):43–50, 1980.

[26] Han Feng and Javad Lavaei. Connectivity properties of the set of stabilizing static decentralized controllers. *Preprint*, 2019.

[27] 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)*, 2019.

[28] Gary Harris and Clyde Martin. The roots of a polynomial vary continuously as a function of the coefficients. *Proceedings of the American Mathematical Society*, 100(2):390–392, 1987.

[29] Uwe Helmke and John B Moore. *Optimization and Dynamical Systems*. Springer Science & Business Media, London, 1994.

[30] R. Hermann and C. Martin. Applications of algebraic geometry to systems theory–part I. *IEEE Transactions on Automatic Control*, 22(1):19–25, 1977.

[31] G. 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.

[32] G. 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.

[33] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, NY, 2nd edition, 2012.

[34] Y.S. Hung and D.L. Chu. Relationships between discrete-time and continuous-time algebraic Riccati inequalities. *Linear Algebra and its Applications*, 270(1):287–313, 1998.

[35] Max Jaderberg, Wojciech M Czarnecki, Iain Dunning, Luke Marris, Guy Lever, Antonio Garcia Castaneda, Charles Beattie, Neil C Rabinowitz, Ari S Morcos, Avraham Ruderman, et al. Human-level performance in 3d multiplayer games with population-based reinforcement learning. *Science*, 364(6443):859–865, 2019.

[36] 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.

[37] E. Jonckheere. On the existence of a negative semidefinite, antistabilizing solution to the discrete-time algebraic Riccati equation. *IEEE Transactions on Automatic Control*, 26(3):707–712, 1981.

[38] Thomas Kailath. *Linear Systems*. Prentice–Hall, Englewood Cliffs, NJ, 1980.

[39] Rudolf Emil Kalman. Contributions to the theory of optimal control. *Boletinde la Sociedad Matematica Mexicana*, 5(1):102–119, 1960.

[40] Rudolf Emil Kalman et al. Contributions to the theory of optimal control. *Bol. Soc. Mat. Mexicana*, 5(2):102–119, 1960.

[41] D Kleinman and Michael Athans. The design of suboptimal linear time-varying systems. *IEEE Transactions on Automatic Control*, 13(2):150–159, 1968.

[42] C Knapp and S Basuthakur. On optimal output feedback. *IEEE Transactions on Automatic Control*, 17(6):823–825, 1972.

[43] Alexander Kobel, Fabrice Rouillier, and Michael Sagraloff. Computing real roots of real polynomials... and now for real! In *Proceedings of the ACM on International Symposium on Symbolic and Algebraic Computation*, pages 303–310. ACM, 2016.

[44] S.G. Krantz and H.R. Parks. *The Implicit Function Theorem: History, Theory, and Applications*. Modern Birkhäuser Classics. Springer New York, 2012.

[45] Peter Lancaster and Leiba Rodman. *Algebraic Riccati Equations*. Oxford University Press, New York, NY, 1995.

[46] D. Lee and J. Hu. Primal-dual Q-learning framework for LQR design. *IEEE Transactions on Automatic Control*, pages 1–1, 2018.

[47] J. M. Lee. *Introduction to Topological Manifolds*. Springer-Verlag, New York, NY, 2nd edition, 2011.

[48] Jae Young Lee, Jin Bae Park, and Yoon Ho Choi. Integral Q-learning and explorized policy iteration for adaptive optimal control of continuous-time linear systems. *Automatica*, 48(11):2850–2859, 2012.

[49] W Levine, T Johnson, and Michael Athans. Optimal limited state variable feedback controllers for linear systems. *IEEE Transactions on Automatic Control*, 16(6):785–793, 1971.

[50] 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.

[51] F. L. Lewis and D. Vrabie. Reinforcement learning and adaptive dynamic programming for feedback control. *IEEE Circuits and Systems Magazine*, 9(3):32–50, 2009.

[52] Frank L Lewis, Draguna Vrabie, and Kyriakos G Vamvoudakis. Reinforcement learning and feedback control: using natural decision methods to design optimal adaptive controllers. *IEEE Control Systems*, 32(6):76–105, 2012.

[53] Stanislaw Lojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux dérivées partielles*, 117:87–89, 1963.

[54] M. Marden. *Geometry of Polynomials*. American Mathematical Society, Providence, RI, 2nd edition, 1966.

[55] Karl Mårtensson. *Gradient Methods for Large-Scale and Distributed Linear Quadratic Control*. PhD thesis, Department of Automatic Control, Lund University, Sweden, 2012.

[56] Volker Mehrmann. A step toward a unified treatment of continuous and discrete time control problems. *Linear Algebra and its Applications*, 241-243:749–779, 1996.

[57] Arkadi Nemirovski, Anatoli Juditsky, Guanghui Lan, and Alexander Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.

[58] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[59] RAIMUND J Ober. Topology of the set of asymptotically stable minimal systems. *International Journal of Control*, 46(1):263–280, 1987.

[60] Atsumi Ohara and Shun-Ichi Amari. Differential geometric structures of stable state feedback systems with dual connections. In *2nd IFAC Workshop on System Structure and Control*, volume 25, pages 176–179, 1992.

[61] Victor Y Pan and Elias P Tsigaridas. Nearly optimal refinement of real roots of a univariate polynomial. *Journal of Symbolic Computation*, 74:181–204, 2016.

[62] Boris T Polyak. Gradient methods for the minimisation of functionals. *USSR Computational Mathematics and Mathematical Physics*, 3(4):864–878, 1963.

[63] M. N. Prakash and A. T. Fam. A geometric approach to stabilization by output feedback. *International Journal of Control*, 37(1):111–125, 1983.

[64] Qazi Ibadur Rahman and Gerhard Schmeisser. *Analytic Theory of Polynomials*. Oxford University Press, New York, NY, 2002.

[65] Jorma Rissanen. Control system synthesis by analogue computer based on the generalized linear feedback concept. *Proceedings of the Symposium on Analog Comutation Applied to Study of Chemical Processes*, pages 1–13, 1960.

[66] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 1976.

[67] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, NY, 3rd edition, 1987.

[68] Walter Rudin et al. *Principles of mathematical analysis*, volume 3. McGraw-hill New York, 1964.

[69] Valeria Simoncini. Computational methods for linear matrix equations. *SIAM Review*, 58(3):377–441, 2016.

[70] Sriram Srinivasan, Marc Lanctot, Vinicius Zambaldi, Julien Pérolat, Karl Tuyls, Rémi Munos, and Michael Bowling. Actor-critic policy optimization in partially observable multiagent environments. In *Advances in Neural Information Processing Systems*, pages 3422–3435, 2018.

[71] Anton Stoorvogel. *The H_∞ Control Problem: a State Space Approach*. Citeseer, 1990.

[72] Anton A Stoorvogel and Arie JTM Weeren. The discrete-time Riccati equation related to the H_∞ control problem. *IEEE Transactions on Automatic Control*, 39(3):686–691, 1994.

[73] Harry L Trentelman, Anton A Stoorvogel, and Malo Hautus. *Control Theory for Linear Systems*. Springer Science & Business Media, 2012.

[74] Loring W Tu. *Differential Geometry: Connections, Curvature, and Characteristic Classes*, volume 275. Springer, 2017.

[75] Kyriakos G Vamvoudakis, Hamidreza Modares, Bahare Kiumarsi, and Frank L Lewis. Game Theory-Based control system algorithms with Real-Time reinforcement learning: How to solve multiplayer games online. *IEEE Control Syst.*, 37(1):33–52, 2017.

[76] M. Vidyasagar, H. Schneider, and B. Francis. Algebraic and topological aspects of feedback stabilization. *IEEE Transactions on Automatic Control*, 27(4):880–894, 1982.

[77] Ė̇rnest Borisovich Vinberg. *A course in algebra.* Number 56. American Mathematical Soc., 2003.

[78] Jan Willems. Least squares stationary optimal control and the algebraic riccati equation. *IEEE Transactions on Automatic Control*, 16(6):621–634, 1971.

[79] Harald K. Wimmer. On the existence of a least and negative-semidefinite solution of the discrete-time algebraic Riccati equation. *Journal of Mathematical Systems, Estimation, and Control*, 5(4):445–457, 1995.

[80] Jerzy Zabczyk. *Mathematical Control Theory: An Introduction.* Birkhauser, Boston, MA, 1992.

[81] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Policy optimization provably converges to Nash equilibria in zero-sum linear quadratic games. *arXiv:1906.00729*, 2019.

[82] Pingjian Zhang. Some results on two-person zero-sum linear quadratic differential games. *SIAM J. Control Optim.*, 43(6):2157–2165, 2005.