Analysis of Policy Gradient Descent for Control: Global Optimality via Convex Parameterization

Yue Sun and Maryam Fazel [*]

January 18, 2022

Abstract

Policy gradient descent is a popular approach in reinforcement learning due to its simplicity. Recent work has investigated the optimality and convergence properties of this method when applied in certain control problems. In this paper, we connect policy gradient descent (applied to a nonconvex problem formulation) with classical convex parameterizations in control theory, to show the *gradient dominance* property for the nonconvex cost function. Such a connection between nonconvex and convex landscapes holds for continuous/discrete time LQR, distributed optimal control, minimizing the $\mathcal{L}_2$ gain, and $\mathcal{H}_2/\mathcal{H}_\infty$ mixed/robust control, among others. To the best of our knowledge, this paper offers the first result unifying the landscape analysis of a broad class of control problems.

1 Introduction

This paper proposes a framework that builds the mapping between a few control problems with their associated convex parameterized form. With the mapping, we show that all stationary points of the cost functions, as functions of the policy, are global minima despite their nonconvexity. The fact allows first order optimization methods (i.e., policy gradient method) to converge the globally optimal controller. We give a comprehensive theory covering many control problems, including continuous/discrete time LQR, distributed optimal control, minimizing the $\mathcal{L}_2$ gain, and $\mathcal{H}_2/\mathcal{H}_\infty$ mixed/robust control that unifies the conclusion of each specific work.

We start from introducing linear quadratic regulator (LQR), which is one of the most well studied optimal control problems for decades [1]. Consider the continuous time linear time-invariant dynamical system,

$$\dot{x} = Ax + Bu, \;\; x(0) = x_0, \tag{1}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^p$ is the input, and A, B are constant matrices describing the dynamics. The goal of optimal control is to determine the input series $u(t)$ that minimizes some cost function (that typically depends on the state and input). In the infinite horizon LQR problem, we define constant matrices $Q \in \mathbf{S}_{++}^n, R \in \mathbf{S}_{++}^p$, and minimize the cost as a function of input

$$\mathrm{cost}(u(t)) := \mathbf{E}_{x_0} \int_0^\infty (x(t)^\top Q x(t) + u(t)^\top R u(t))\mathrm{d}t. \tag{2}$$

[*] Y. Sun (yuesun@uw.edu) and M. Fazel (mfazel@uw.edu) are with Department of Electrical and Computer Engineering, University of Washington, Seattle, USA.

The optimal controller is linear in the state, called static state feedback controller, and can be described as $u(t) = Kx(t)$ for a constant $K \in \mathbb{R}^{p\times n}$ [1].

We can define this cost function with variable K,

$$\mathcal{L}(K) := \mathbf{E}_{x_0} \int_0^\infty (x(t)^\top Qx(t) + u(t)^\top Ru(t))\mathrm{d}t, \text{ s.t. } u(t) = Kx(t). \tag{3}$$

Policy gradient method. The policy gradient descent method is to minimize $\mathcal{L}(K)$ by running first order method with respect to K. We initialize K_0 as a stabilizing controller, so that $\mathcal{L}(K_0)$ is well defined. For LQR as an example, the paper [2] demonstrates an algorithm to find a stabilizing controller by policy gradient method. They begin with an arbitrary controller, and define an alternative cost with a discount factor to make the cost finite, and run policy gradient method on that cost and later anneal the discount factor. After obtaining a proper initialization K_0, we run

$$K_{t+1} \leftarrow K_t - \eta_t \nabla\mathcal{L}(K_t).$$

It is shown that, the loss function is typically nonconvex in K, e.g., continuous/discrete time LQR [3,4]. However, gradient descent for nonconvex optimization is widely used in machine learning, or control tasks with the context of reinforcement learning.

Convex parameterization. In classical control theory literature, due to the nonconvex nature, policy gradient method is not commonly used. Instead, one can introduce another parameterization of the cost to make it convex, and apply convex optimization method with global convergence guarantee. This approach is in sharp contrast to how one would typically minimize a cost function through gradient descent on K. In the following discussion, we will first review the convex parameterization methods.

Solving LQR – classical way. LQR is usually solved by the algebraic Riccati equation (ARE) [5,6]. A large number of works have studied the solution of ARE, including approaches based on iterative algorithms [7], algebraic solution methods [8], and semidefinite programming [9]. Besides LQR, convex parameterizations such as Youla parametrization, Q-parameterization, or the more recent System Level Synthesis (SLS) have allowed the reformulation of certain control design problems as semidefinite programs. They are convex optimization methods, which natually can be globally minimized by first order algorithms. On the other hand, one has to know the system parameters to run them, and the papراmeterizations are case-by-case.

LQR with unknown system parameters: model-based and model-free. There are two major types of algorithms when system parameters are not known. The first type is model-based methods, when we first estimate the system parameters and then a controller is constructed based on the identified system. System identification has a long history, as reviewed in [10]. Recently the paper [11] gave sample complexity bounds for state-observed system. The papers [12–15] describe the joint system identification and optimal control approaches, where the algorithms estimate the system parameter from the intput and output, and apply controllers based on the estimated system. [16] compares the sample complexity of model-based and model-free algorithms for LQR, and shows that model-based method is more sample efficient.

The second type of method is model free method, when the controller is directly trained by observing the cost function or its gradient, without characterizing the dynamics. Here one does not necessarily estimate the system parameters A, B. The paper [17] is a review of reinforcement learning area and optimal control, which studies a few fixed point type dynamic programming methods .

Q-learning is a typical model free method for reinforcement learning, and it is applied to LQR as in [12,18,19].

Recent works on policy gradient descent. Policy gradient descent calls for an estimate of the cost and its gradient with respect to controller K. The goal is to show that gradient descent with respect to K converges to the optimal controller (we can call it K^*). The policy gradient descent is recently reviewed by [20,21]. The paper [4] provides a counterexample showing that minimizing the quadratic LQ cost as a function of K is not convex, quasi-convex or star-convex.

There has been recent evidence of the *empirical* success of first order methods in solving nonconvex reinforcement learning problems. The paper [22, Ch. 3] proposes the gradient based method for optimal control and extends to decentralized control. The paper [23] studies feedback control with dynamical controllers, and observes that gradient descent with Youla parameterization is robust within the set of stabilizing controllers while other parameterizations are not. On the *theoretical* side, despite the nonconvexity of $\mathcal{L}(K)$, for certain types of control problems, there are works showing the *gradient dominance* property, which enables first order methods to converge to the global optimum. The paper [4] gives the first result by proving the coercivity and the gradient dominance properties of $\mathcal{L}(K)$ for the discrete time LQR. Based on this, the paper [4] shows the linear convergence of gradient based method. The paper [24] is a survey of the zeroth order realization of policy gradient method on discrete time LQR with sample complexity analysis. Later the paper [3] shows a similar result for the continuous time case, papers [25,26] give a more detailed analysis for both discrete and continuous time LQR. The papers [27,28] show the convergence for two types of zero-sum LQ games. The paper [29] studies the convergence of gradient descent on $\mathcal{H}_2$ control with $\mathcal{H}_\infty$ constraint, and shows that gradient descent implicitly makes the controller robust. The paper [30] shows the convergence for finite-horizon distributed control under the quadratic invariance assumption.

Negative results – linear control problems where policy gradient descent does not converge to global minima. Not all control problems are easy to solve by policy gradient methods. The distributed control problem is an example. The controller K has a sparsity pattern, i.e., K is in a subspace. [31,32] show that, generally the set of stabilizing controllers is highly disconnected and the problem is NP-hard without the extra assumption of quadratic invariance in [30]. Similarly, the static output feedback controller design is NP-hard. The goal is to minimize the LQ cost, but we can only observe an output $y = Cx$ but cannot observe the full state, and we are only allowed to use a static output feedback controller $u = Ky$. If C is not full row rank, the set of stabilizing controllers is also highly disconnected [32]. If C is full row rank, the problem almost reduces to state feedback control since one can recover the state x from y, and the paper [33] shows that policy gradient finds the optimal controller. The paper [34] shows that, the cost of the LQR problem with output dynamical controller, i.e., the LQG cost, as a function of the fixed degree controller with state-space representation, has saddle points . Although a parameterization can construct an equivalent convex optimization problem, the map for such parameterization is generally not everywhere smooth, and the nonsmoothness breaks the gradient dominance and generates saddle points. The negative examples above will not be covered by our theorem.

Control for nonlinear systems is far more difficult, typically via dynamic programming and the Bellman equations [35], or recent deep RL that led to empirical success for controlling complex systems . Yet it is still mysterious how deep learning models work in this context, and recent theoretical studies have focused on linear systems in the hope of providing insights into more complex cases.

Motivation: Although we reviewed many papers that show convergence of policy gradient descent, they investigate different control problems and the proofs are given case by case. However, we observe that all the results are proven by the gradient dominance property, and all of them are solved by convex parameterization methods in classical control literature. Thus, we ask whether there is a proof that unifies the proofs of the gradient dominance property for different control problems, and bridges nonconvex methods with convex methods in classical control literature.

Contributions: In this paper, we make a connection between nonconvex policy gradient descent and known convex parameterization methods with a map between the two parameters. This map maintains the Lojasiewicz inequality when going from the convex landscape to the nonconvex landscape.

Our result is quite general—we show that continuous-time LQR is a special case that the main theorems apply to, and we generalize the guarantees provided by this method to a range of other control problems. The instances cover optimal control, robust control, mixed design and system level synthesis. To judge whether a nonconvex landscape can be optimized globally using policy gradient method, one can directly check if it is covered by the theorems, avoiding a case-by-case analysis. Also, as discussed in [4], theoretical guarantees for first-order methods naturally lead to guarantees for the more practical zeroth-order optimization or sampling-based methods, which do not need access to the gradient of the cost with respect to K.

Outline: The rest of this paper is structured as follows. Sec. 2 reviews the continuous-time LQR problem. Sec. 3 presents our main result on the the nonconvex cost, showing all stationary points are global minima. Sec. 4 lists more examples of control problems covered by the main theorem. Although Sec. 4 covered many problems, Sec. 5 further generalizes our main result using different parameterizations and Sec. 6 covers examples under the more generic result. Sec. 7 gives a proof sketch with intuitive connections between the nonconvex and convex formulations.

2 Review of convex parameterization for continuous-time LQR

Convex parameterization (e.g., solving optimal control by linear matrix inequalities (LMI) in [36]) is widely used in optimal control problems, and here we discuss its application for continuous time LQR [3]. We will introduce new variables, construct an equivalent convex optimization problem with new variables, and the pair of variables are proven to be linked by a bijection. In the next section we use the critical properties of the nonconvex and convex problems as an intuition to generalize to a more general form.

Define a continuous time linear time invariant system (1) where x is state and u is input signal, and x_0 is the initial state. We assume that $\Sigma := \mathbf{E}(x_0 x_0^\top) \succ 0$. This is a commonly used setup such as in the theoretical study [25, §3.3], and the practical work [22, Ch. 3]. With $\Sigma \succ 0$, the optimal controller is not state-dependent; when Σ is low rank, then a controller K that gives finite LQR cost does not stabilize the system for all initial state $x_0 \in \text{null}(\Sigma)$.

One can then consider minimizing the linear quadratic (LQ) cost (2) as a function of $u(t)$ where Q, R are positive definite matrices. The paper [1] proves that, the input signal that minimizes the cost function $\text{cost}(u)$ is given by a static state feedback controller, denoted by $u(t) = K^* x(t)$. K^* can be obtained by solving linear equations, called Riccati equations. Once we know that the optimal state feedback controller is static, we can write cost as $\mathcal{L}(K)$ as (3). It is a function of K, and we search only static state feedback controllers.

An alternative approach is reparameterization, to obtain a convex optimization problem, as used in [3]. We will review it here, starting from the Lyapunov equation. Suppose the initial state satisfies

$\mathbf{E}(x_0 x_0^\top) = \Sigma \succ 0$, and $\dot{x}(t) = Ax(t)$. Then with a matrix $P \in \mathbf{S}_{++}^{n \times n}$ (P is a positive definite matrix) as the variable, the Lyapunov equation is written as

$$AP + PA^\top + \Sigma = 0.$$

In our setup (1), we use a state feedback controller $u = Kx$, thus we have $\dot{x} = (A + BK)x$. We denote the set of stabilizing controllers as $\mathcal{S}_{K,\text{sta}}$, which is defined as

$$\mathcal{S}_{K,\text{sta}} = \{K : \ \text{Re}(\lambda_i(A + BK)) < 0, \ i = 1, ..., n\}.$$

If a state feedback controller is applied, the cost is only bounded when $K \in \mathcal{S}_{K,\text{sta}}$ and is coercive in $\mathcal{S}_{K,\text{sta}}$ [26]. Replace A by the closed loop system matrix $A + BK$ in the Lyapunov equation, and let $L = KP \in \mathbb{R}^{p \times n}$, we get

$$AP + PA^\top + BL + L^\top B^\top + \Sigma = 0.$$

Let $\mathcal{A}(P) = AP + PA^\top, \ \mathcal{B}(L) = BL + L^\top B^\top$, which are referred to as Lyapunov maps. Assume $\mathcal{A}$ is invertible, then we have the relation

$$\mathcal{A}(P) + \mathcal{B}(L) + \Sigma = 0. \tag{4}$$

Indeed, once we fix the system and any stabilizing controller A, B, K, the matrices P as well as $L = KP$ are uniquely determined. P is the Grammian matrix

$$P = \int_0^\infty e^{t(A+BK)} \ \Sigma \ e^{t(A+BK)^\top} \ \mathrm{d}t. \tag{5}$$

The matrix P is positive definite if $\Sigma \succ 0$. We are interested in the cost function $\mathcal{L}(K)$ when $K \in \mathcal{S}_{K,\text{sta}}$, which corresponds to (2) by inserting $u(t) = Kx(t)$; If K is not a stabilizing controller, we define $\mathcal{L}(K) = +\infty$.

$$\mathcal{L}(K) = \begin{cases} \mathbf{Tr}((Q + K^\top RK)P), & K \in \mathcal{S}_{K,\text{sta}}; \\ +\infty, & K \notin \mathcal{S}_{K,\text{sta}}. \end{cases} \tag{6}$$

One can construct a bijection from P, L to K, and prove that, if we minimize $f(L, P)$ subject to (4), the optimizer P^*, L^* will map to the optimal K^*, and this minimization problem is convex.

Convex parameterization for continuous time LQR: Suppose the dynamics and costs are (1) and (2), and let $\mathbf{E}(x_0 x_0^\top) = \Sigma \succ 0$. Denote the (static) state feedback controller by K, so that $u(t) = Kx(t)$. The optimal control problem is

$$\min_K \ \mathcal{L}(K), \quad \text{s.t.} \quad K \in \mathcal{S}_{K,\text{sta}} \tag{7}$$

where $\mathcal{L}(K)$ is the cost in (2) with $u = Kx$. This problem can be expressed as the following equivalent convex problem,

$$\min_{L,P,Z} \ f(L, P, Z) := \mathbf{Tr}(QP) + \mathbf{Tr}(ZR) \tag{8a}$$

$$\text{s.t. } \mathcal{A}(P) + \mathcal{B}(L) + \Sigma = 0, \ P \succ 0, \tag{8b}$$

$$\begin{bmatrix} Z & L \\ L^\top & P \end{bmatrix} \succeq 0. \tag{8c}$$

The connection between the two problems is distilled in Sec. 3. For all feasible (L, P, Z) triplets in (8), we can take the first two elements (L, P), and they form a bijection with all stabilizing controllers K in (7). The cost function values are equal under the bijection. So we can solve for L^*, P^*, and $K^* = L^*(P^*)^{-1}$.

3 Main result

Motivated by methods that use gradient descent in the policy space, we ask whether running a gradient-based algorithm and getting $\nabla_K \mathcal{L}(K) = 0$ for some K in fact gives the globally optimum K^*. The papers [3,4] show the coercivity and gradient dominance property of $\mathcal{L}(K)$ for discrete- and continuous-time LQR respectively. In this paper, we generalize these results from the special case of continuous-time LQR to a much broader set of control problems, showing the gradient dominance property of the nonconvex costs as functions of policy.

We present our main result in Theorem 1. We consider a pair of problems satisfying Assumptions 1, 2 or Assumptions 1, 3. In Sec. 4 we catalog a number of examples showing the generality of this result.

We begin by considering an abstract description of the pair of problems (7) and (8). These problem descriptions cover LQR as discussed in the last section, as well as more problems discussed in Sec. 4. Consider the problems

$$\min_{K} \quad \mathcal{L}(K), \quad \text{s.t. } K \in \mathcal{S}_K, \tag{9}$$

and

$$\min_{L,P,Z} \quad f(L,P,Z), \quad \text{s.t.} \quad (L,P,Z) \in \mathcal{S}, \tag{10}$$

where the sets $\mathcal{S}_K, \mathcal{S}$ capture the control constraints. They are defined differently for each specific example in Sec. 4. For example, for continuous time LQR, $\mathcal{S}_K$ is the set of all stabilizing controllers (7) and $\mathcal{S}$ is the intersection of (8b) & (8c). In infinite horizon problems, we need a stabilizing K so that $\mathcal{S}_K$ is equal to or a subset of the set of stabilizing controllers. We allow special cases when (10) depends only on L, P,

$$\min_{L,P} \quad f(L,P), \quad \text{s.t.} \quad (L,P) \in \mathcal{S}. \tag{11}$$

We distill the properties of the two problems (9) and (10) that will be critical for Theorem 1, and allow us to cover more problems as discussed in Sec. 4.

Assumption 1. *The feasible set $\mathcal{S}$ is convex in (L,P,Z). The cost function $f(L,P,Z)$ is convex, bounded, and differentiable over an open domain that contains the set $\mathcal{S}$.*

Assumption 1 indicates the second problem is convex. Next, we examine the connection between (7) and (8), formalized in the following assumption.

Assumption 2. *Let P be invertible*[1] *whenever $(L,P,Z) \in \mathcal{S}$. For all $(L,P,Z) \in \mathcal{S}$, $LP^{-1} \in \mathcal{S}_K$. Assume we can express $\mathcal{L}(K)$ as follows,*

$$\begin{aligned} \mathcal{L}(K) = \min_{L,P,Z} \; & f(L,P,Z) \\ \textit{s.t. } & (L,P,Z) \in \mathcal{S}, \; LP^{-1} = K. \end{aligned}$$

Denote $\nabla \mathcal{L}(K)[V] := \mathbf{Tr}(V^\top \nabla \mathcal{L}(K))$ as the directional derivative of $\mathcal{L}(K)$ is the direction V. With the assumptions above, we will present the main theorem.

[1] The invertibility of P holds for all instances in Sec. 4.

Theorem 1. *Suppose Assumptions 1,2 hold, and consider the two problems* (9) *and* (10). *Let* K^* *denote the global minimizer of* $\mathcal{L}(K)$ *in* S_K. *Then there exist constants* $C_1, C_2 > 0$ *and a direction* V *with* $\|V\|_F = 1$, *in the descent cone of* $\mathcal{S}_K$ *at* K *such that,*

1. *if* f *is convex, the gradient of* $\mathcal{L}$ *satisfies*[2]

$$\nabla\mathcal{L}(K)[V] \leq -C_1(\mathcal{L}(K) - \mathcal{L}(K^*)). \tag{12}$$

2. (a) *if* f *is* μ*-strongly convex, or*

   (b) *let* $\mathcal{P}_{\mathcal{S}}(-\nabla f(L,P,Z))$ *be the projection of* $-\nabla f(L,P,Z)$ *in the descent cone of* $\mathcal{S}$ *at* (L,P,Z), *if for any*

$$(L,P,Z) = \arg\min_{L',P',Z'} f(L',P',Z'),\ \textit{s.t.}\ (L',P',Z') \in \mathcal{S},\ L'(P')^{-1} = K,$$

   we have $\|\mathcal{P}_{\mathcal{S}}(-\nabla f(L,P,Z))\|_F^2 \geq \mu(f(L,P,Z) - f(L^*,P^*,Z^*))$,

   the gradient of $\mathcal{L}$ *satisfies*

$$\nabla\mathcal{L}(K)[V] \leq -C_2(\mu(\mathcal{L}(K) - \mathcal{L}(K^*)))^{1/2}. \tag{13}$$

Remark 1. The second part of the theorem contains two parts, leading to the same result. The first case works with the standard strongly convex objective function. The second case indicates the interaction between the function and the constraint. With the assumption, it is equivalent to a strongly convex objective function. An example is

$$\min_{x,y}\ y,\ \text{s.t.}\ y \geq x^2.$$

Remark 2. The constants in the above theorem can be computed or bounded in a case by case manner, as discussed further in Appendix B. They typically depend on the norm of system parameters and the radius of the feasible domain[3]. We study continuous time LQR as an example. Let the sublevel set be where $\mathcal{L}(K) \leq a$. Define

$$\nu = \frac{\lambda_{\min}^2(\Sigma)}{4}\left(\sigma_{\max}(A)\lambda_{\min}^{-1/2}(Q) + \sigma_{\max}(B)\lambda_{\min}^{-1/2}(R)\right)^{-2},$$

then

$$C_1 = \frac{\nu\lambda_{\min}^{1/2}(Q)\lambda_{\min}^{1/2}(R)}{4a^4} \cdot \min\left\{a^2,\ \nu\lambda_{\min}(Q)\right\}.$$

The paper [3] gives another convex formulation with strong convexity and we can get C_2 for that form,

$$C_2 = \frac{\nu}{2a^3}\min\left\{a^2,\ \nu\lambda_{\min}^{1/2}(Q)\lambda_{\min}^{1/2}(R)\right\}.$$

See Appendix B for more details.

[2] We always consider the directional derivative of a feasible direction within descent cone.

[3] Although the set can be unbounded, when we run gradient descent with respect to $\mathcal{L}(K)$, the cost is typically bounded by the initial value $\mathcal{L}(K_0)$ so the iterates are in a sublevel set, therefore boundedness of this sublevel set suffices for our purpose.

We know that $\|\nabla\mathcal{L}(K)\|_F \geq \left|\nabla\mathcal{L}(K)[\frac{V}{\|V\|_F}]\right|$ for any direction V. The lower bound $\|\nabla\mathcal{L}(K)\|_F \gtrsim (\mathcal{L}(K) - \mathcal{L}(K^*))^\alpha$ on the norm of the gradient is known as Lojasiewicz inequality [37]. The case when $\alpha = 1/2$ is also called the *gradient dominance* property. If this inequality holds for all K, all stationary points of the objective function are global minima, and an iterative method in which the norm of the gradient decreases to zero will have to converge to a global minimum. Nonconvex functions that satisfy Lojasiewicz inequality are easily optimized, compared to those with spurious local minima. In practice, Lojasiewicz inequality often holds in a neighborhood of a local minimum, and Lojasiewicz inequality is typically used as a tool for local convergence analysis (it is rare that Lojasiewicz inequality holds for $\mathcal{L}(K)$ globally, but it holds for example problems in this paper).

Next, we consider a stronger assumption covered by Assumption 2, where we assume that there is a bijection of a specific form between K and (L, P) (Assumption 3). This is true for many control problems including continuous time LQR. Theorem 1 also holds with Assumptions 1, 3. We emphasize the special case with the bijection for illustration. (In this section, the map between the variables is $K = LP^{-1}$, and in Sec. 5, we will present the result for a general $K = \Phi(P)$) In Sec. 7, we will illustrate the key proof steps: we use the fact that the convex function $f(L, P)$ is gradient dominant, and apply the bijection between K and (L, P) to calculate $\nabla\mathcal{L}(K)$.

Assumption 3. *1. (Bijection between the two feasible sets) Let P be invertible, and let $K = LP^{-1}$ define a bijection[4] $K \leftrightarrow (L, P)$, where there exists an auxiliary variable Z such that $(L, P, Z) \in \mathcal{S}$.*

2. (Equivalence of functions) Choose a controller $K \in \mathcal{S}_K$ with corresponding $(L, P) \in \mathcal{S}$. Then $\mathcal{L}(K) = \min_Z f(L, P, Z)$ subject to $(L, P, Z) \in \mathcal{S}$.

Theorem 1 suggests that when the original nonconvex optimization problem can be mapped to a convex optimization problem that satisfies Assumptions 1, 2 or 1, 3, all stationary points of the nonconvex objective are global minima. So if we can evaluate the gradient of nonconvex objective and run gradient descent algorithm, the iterates converge to the optimal controller.

4 Control problems covered by main theorem

Theorem 1 requires an optimal control problem (9), and its convex form (10) that satisfies a few assumptions. This is an abstract and general description that does not need the exact continuous time LQR formulation in Sec. 2. Sec. 2 implies that the continuous time LQR satisfies Assumptions 1,3, thus we can directly apply Theorem 1 to argue that the continuous time LQR cost $\mathcal{L}(K)$ satisfies (12).

In this section, we discuss more examples, showing that Theorem 1 covers a wide range of control design problems. This illustrates the **generality** of Theorem 1. If a new control problem is encountered, the assumptions for Theorem 1 can be checked, in order to directly conclude that the stationary points of the original cost function are all global minima, and further, the nonconvex function can be globally optimized by policy gradient descent.

[4] Note that generally $K = LP^{-1}$ cannot guarantee a bijection. However bijection is possible with the extra constraint $(L, P) \in \mathcal{S}$.

4.1 Discrete time infinite horizon LQR

We will show that minimizing the LQ cost as a function of the state feedback controller K, and the convex form, satisfy the assumptions for Theorem 1. So that all stationary points of the LQ cost as a function of K are global minima, same as the result in [4].

We consider a discrete time linear system

$$x(t+1) = Ax(t) + Bu(t),\ x(0) = x_0,$$

The goal is to find a state feedback controller K such that the cost function

$$\mathcal{L}(K) = \mathbf{E}_{x_0} \sum_{i=0}^{\infty} x(t)^\top Qx(t) + u(t)^\top Ru(t),\ u(t) = Kx(t)$$

is minimized. In other words, we will solve

$$\min_K\ \mathcal{L}(K),\ \text{ s.t. } K \text{ stabilizes.} \tag{14}$$

Here we assume that $\mathbf{E}(x_0 x_0^\top) = \Sigma$. Similar to the continuous time system, one can choose the same parameterization P, L, Z and another PSD matrix $G \in \mathbb{R}^{n\times n} \succeq 0$ and solve the following problem

$$\min_{L,P,Z,G}\ f(L,P,Z,G) := \mathbf{Tr}(QP) + \mathbf{Tr}(ZR), \tag{15a}$$

$$\text{s.t. } P \succ 0,\ G - P + \Sigma = 0, \tag{15b}$$

$$\begin{bmatrix} Z & L \\ L^\top & P \end{bmatrix} \succeq 0,\ \begin{bmatrix} G & AP+BL \\ (AP+BL)^\top & P \end{bmatrix} \succeq 0. \tag{15c}$$

The goal is to argue that $\mathcal{L}(K)$ and (15) has the connection such that Theorem 1 applies, so that the stationary point of $\mathcal{L}(K)$ has to be the global optimum.

Lemma 1. *The LQR problems* (14) *and* (15) *satisfy Assumptions 1, 2.*

Proof. (15) is a convex optimization problem. Now we prove Assumption 2, i.e., we prove that $L(K)$ equals the minimum of the problem (15) with an extra constraint $K = LP^{-1}$.

- We first minimize over Z, the minimizer is $Z = LP^{-1}L^\top$. Now we plug $Z = LP^{-1}L^\top$ into cost, replace L by KP and the cost becomes $\mathbf{Tr}((Q + K^\top RK)P)$.
- We will eliminate G by

  $$G - P + \Sigma = 0,\ \begin{bmatrix} G & AP+BL \\ (AP+BL)^\top & P \end{bmatrix} \succeq 0.$$

  Using Schur complement, it is equivalent to

  $$(AP+BL)P^{-1}(AP+BL)^\top - P + \Sigma \preceq 0.$$

  Plug in $L = KP$, we have

  $$(A+BK)P(A+BK)^\top - P + \Sigma \preceq 0.$$

  The cost does not involve G so it does not change.

- Now, we need to prove that $\mathcal{L}(K)$ is equal to

$$\begin{aligned}&\min_{P}\ \mathbf{Tr}((Q+K^\top RK)P),\\ &\text{s.t. } (A+BK)P(A+BK)^\top - P + \Sigma \preceq 0.\end{aligned} \tag{16}$$

  The constraint (16) can be written as

$$(A+BK)P(A+BK)^\top - P + \Theta = 0,\ \Theta \succeq \Sigma.$$

- Denote the solution to $(A+BK)P(A+BK)^\top - P + \Theta = 0$ as $P(\Theta)$. $P(\Theta)$ for all $\Theta \succeq \Sigma$ covers the feasible points of (16). $P(\Theta)$ is expressed as:

$$P(\Theta) = \sum_{t=0}^{\infty}(A+BK)^t\Theta((A+BK)^\top)^t.$$

  So $P(\Theta) \succeq P(\Sigma)$, for all $\Theta \succeq \Sigma$. Since Q and $K^\top RK$ are positive semidefinite, the cost $\mathbf{Tr}((Q+K^\top RK)P)$ achieves the minimum at $P = P(\Sigma)$.

- At the end, $P(\Sigma)$ is the Grammian $\mathbf{E}\sum_{t=0}^{\infty} x(t)x(t)^\top$ when $\mathbf{E}x(0)x(0)^\top = \Sigma$. We studied the connection between continous time Grammian (5) and the cost (6), and a similar result holds for discrete time LQR:

$$\mathbf{Tr}((Q+K^\top RK)P(\Sigma)) = \mathcal{L}(K).$$

□

We built the connection between minimizing $\mathcal{L}(K)$, and the convex optimization (15). We argued this pair of problems satisfies the assumptions of Theorem 1. Theorem 1 suggests that $\mathcal{L}(K)$ is gradient dominant, so we can approach K^* by gradient descent on K. This is essentially the conclusion of [4,25]. Note that the proof of discrete time LQR [4,25] and continuous time LQR [3,26] cannot trivially extend to each other, but our result can cover both continuous and discrete time cases.

4.2 LQR with Markov jump linear system

We generalize the discrete time linear system to multiple linear systems with transitions, called Markov jump linear system in this part. We show that, the LQR with Markov jump linear system can be covered by the conclusion of Theorem 1. It means all stationary points of the linear quadratic cost as a function of policy/controllers are global minima.

Markov jump linear system. Suppose there are N linear systems, the i-th one being

$$x(t+1) = A_i x(t) + B_i u(t).$$

Now we study the LQR of Markov jump linear system [38]. At each time t, the dynamics linking $x(t+1)$ and the past state and input $x(t), u(t)$ is given by

$$x(t+1) = A_{w(t)}x(t) + B_{w(t)}u(t),\ w(t) \in [N] := \{1, ..., N\}.$$

At time t, a system $w(t)$ from number 1 to N is randomly chosen by some probabilistic model. The transition of the linear systems, or the transition of $w(t)$, follows the following probabilistic model

$$\mathbf{Pr}(w(t+1)=j|w(t)=i)=\rho_{ij}\in[0,1],\ \forall t\geq 0.$$

Suppose $\mathbf{Pr}(w(0)=i)=p_i$. For the i-th system, we will use a state feedback controller K_i. Let $K=[K_1,...,K_N]$. Define the cost as

$$\mathcal{L}(K)=\mathbf{E}_{w,x_0}\sum_{t=0}^{\infty}x(t)^\top Qx(t)+u(t)^\top Ru(t),\ \text{s.t. }u(t)=K_{w(t)}x(t),\ \mathbf{Pr}(w(0)=i)=p_i.$$

The nonconvex problem we target to solve is

$$\min_K\ \mathcal{L}(K),\quad \text{s.t. }\mathcal{L}(K)\text{ is finite.}\tag{17}$$

Convex formulation. We propose the following convex formulation. Denote $X_0,X_1,...,X_N\in\mathbb{R}^{n\times n}$, $L_1,...,L_N\in\mathbb{R}^{p\times n}$, $Z_0,Z_1,...,Z_N\in\mathbb{R}^{p\times p}$, $U_{ji}\in\mathbb{R}^{n\times n}$ for $i,j\in[N]$. The following problem is convex:

$$\begin{aligned}&\min\ \mathbf{Tr}(QX_0)+\mathbf{Tr}(Z_0R),\\&\text{s.t. }X_0=\sum_{i=1}^N X_i,\ Z_0=\sum_{i=1}^N Z_i,\ \begin{bmatrix}Z_i & L_i\\ L_i^\top & X_i\end{bmatrix}\succeq 0,\\&X_i-p_i\Sigma=\sum_{j=1}^N U_{ji},\ \begin{bmatrix}\rho_{ji}^{-1}U_{ji} & A_jX_j+B_jL_j\\ (A_jX_j+B_jL_j)^\top & X_j\end{bmatrix}\succeq 0,\ \forall i,j\in[N].\end{aligned}$$

The mapping between the controller K_i and the new variables are $K_i=L_i(X_i)^{-1}$. When the convex problem is minimized, X_i^* represents the Grammian matrix $X_i^*=\sum_{t=0}^{\infty}\mathbf{E}(x(t)x(t)^\top\mathbf{1}_{w(t)=i})$.

We prove that (17) and the convex formulation satisfy Assumptions 1, 2 in Appendix C.1, so that we apply Theorem 1 to claim that all stationary points of $\mathcal{L}(K)$ are global minima.

4.3 Minimizing $\mathcal{L}_2$ gain

We quote from [36] the problem of minimizing the $\mathcal{L}_2$ gain with static state feedback controller K and the convex formulation. We can apply Theorem 1 to argue that all stationary points of $\mathcal{L}_2$ gain as a function of K are global minima. The $\mathcal{L}_2$ gain is also the $\mathcal{H}_\infty$ norm of transfer function [36, §6.3.2]. This problem has an associated convex optimization problem and we can show that they satisfy Assumptions 1,2.

We consider minimizing the $\mathcal{L}_2$ gain of a closed loop system. The continuous time linear dynamical system is

$$\dot{x}=Ax+Bu+B_ww,\ y=Cx+Du.$$

For any signal z, denote

$$\|z\|_2:=\left(\int_0^\infty\|z(t)\|_2^2\mathrm{d}t\right)^{1/2}$$

Suppose we use a state feedback controller $u = Kx$, and aim to find the optimal controller K^* that minimizes the $\mathcal{L}_2$ gain. We minimize the squared $\mathcal{L}_2$ gain as

$$\min_K \ \mathcal{L}(K) := (\sup_{\|w\|_2=1} \|y\|_2)^2, \text{ s.t. } u = Kx.$$

This problem can be further reformulated as the formulation in [36, Sec 7.5.1]

$$\begin{aligned} &\min_{L,P,\gamma} \ f(L,P,\gamma) := \gamma, \text{ s.t.} \\ &\begin{bmatrix} AP + PA^\top + BL + L^\top B^\top + B_w B_w^\top & (CP+DL)^\top \\ CP + DL & -\gamma I \end{bmatrix} \preceq 0. \end{aligned} \tag{19}$$

The minimum $\mathcal{L}_2$ gain is $\sqrt{\gamma^*}$ and $K^* = L^* P^{*-1}$. We will show in the Appendix C.2 that the above nonconvex and convex problems satisfy Assumptions 1,2. Thus we can claim that all stationary points of $\mathcal{L}(K)$ are global minima.

4.4 Dissipativity

We quote from [36] the problem of maximizing the dissipativity with static state feedback controller K and the convex formulation, and apply Theorem 1 to show that all stationary points of the dissipativity as a function of K are global minima.

We study the dynamical system

$$\dot{x} = Ax + Bu + B_w w, \ y = Cx + Du + D_w w \tag{20}$$

The notion of dissipativity can be found in [36, §6.3.3, §7.5.2]. Our goal is to maximize the dissipativity, which is defined and formulated as with a convex parameterization [36, §7.5.2].

The dissipativity is defined as all $\eta > 0$ (if it exists, we usually take the maximum one) that satisfy the following inequality for all w and all $T > 0$,

$$\int_0^T w^\top y - \eta w^\top w \mathrm{d}t \geq 0.$$

We use a state feedback controller K, and the goal is to find K^* that maximizes the dissipativity η. Same as before, let K be factorized as LP^{-1}. We can maximize the dissipativity η as a function of K. From the formulation in [36, §7.5.2], we maximize η subject to the dissipativity constraint (21),

$$\begin{aligned} &\max_{\eta,L,P} \ \eta, \\ &\text{s.t.} \begin{bmatrix} AP + PA^\top + BL + L^\top B^\top & B_w - PC^\top - (DL)^\top \\ B_w^\top - CP - DL & 2\eta I - (D + D^\top) \end{bmatrix} \preceq 0. \end{aligned} \tag{21}$$

We can claim that all stationary points of $\mathcal{L}(K)$ are global minima.

4.5 System level synthesis (SLS) for finite horizon time varying discrete time LQR

In this part, we switch to the discrete time system in finite horizon. We study the finite horizon time varying LQR problem, and its solution using SLS, and show that it satisfies Assumptions 1,3.

Hence we can apply Theorem 1 to conclude that all stationary points of the nonconvex objective functions are global minima.

This problem and its convex form are introduced in [39]. We consider the following linear dynamical system

$$x(t+1) = A(t)x(t) + B(t)u(t) + w(t) \tag{22}$$

over a finite horizon $0, \dots T$. Note the problem is different than the previous ones in (1) the system is time varying, (2) we only consider a finite horizon until T. Let the state be x and the input be u. Define

$$\begin{aligned} X &= \begin{bmatrix} x(0) \\ ... \\ x(T) \end{bmatrix}, \ U = \begin{bmatrix} u(0) \\ ... \\ u(T) \end{bmatrix}, \\ W &= \begin{bmatrix} x(0) \\ w(0) \\ ... \\ w(T-1) \end{bmatrix}, Z = \begin{bmatrix} 0 & 0 & ... & 0 & 0 \\ I & 0 & ... & 0 & 0 \\ 0 & I & ... & 0 & 0 \\ ... \\ 0 & 0 & ... & I & 0 \end{bmatrix}, \\ \mathcal{A} &= \operatorname{diag}(A(0), ..., A(T-1), 0), \\ \mathcal{B} &= \operatorname{diag}(B(0), ..., B(T-1), 0). \end{aligned}$$

Now we consider the time varying controller K that links state and input as

$$u(t) = \sum_{i=0}^{t} K(t, t-i)x(i), \tag{23}$$

and let

$$\mathcal{K} = \begin{bmatrix} K(0,0) & 0 & ... & 0 \\ K(1,1) & K(1,0) & ... & 0 \\ ... \\ K(T,T) & K(T,T-1) & ... & K(T,0) \end{bmatrix}.$$

We will minimize some cost function with the constraint. For example, in the discrete time LQR regime (more examples of nonquadratic cost in [39, Sec 2.2]), let the input be (23) and define

$$\mathcal{L}(\mathcal{K}) = \sum_{t=0}^{T} x(t)^\top Q(t)x(t) + u(t)^\top R(t)u(t), \tag{24}$$

here $Q(t), R(t) \succeq 0$. We will minimize $\mathcal{L}(\mathcal{K})$ where $\mathcal{K}$ is the variable.

Parameterization: The dynamics (22) can be written as

$$X = Z\mathcal{A}X + Z\mathcal{B}U + W = Z(\mathcal{A} + \mathcal{B}\mathcal{K})X + W$$

We define the mapping from W to X, U by

$$\begin{bmatrix} X \\ U \end{bmatrix} = \begin{bmatrix} \Phi_X \\ \Phi_U \end{bmatrix} W.$$

where Φ_X, Φ_U are block lower triangular. There is a constraint on Φ_X, Φ_U:

$$\begin{bmatrix} I - Z\mathcal{A} & -Z\mathcal{B} \end{bmatrix} \begin{bmatrix} \Phi_X \\ \Phi_U \end{bmatrix} = I. \tag{25}$$

It is proven in [39, Thm. 2.1] that $\mathcal{K} = \Phi_U \Phi_X^{-1}$, $\mathcal{K}$ and Φ_X, Φ_U is a bijection given Φ_X, Φ_U satisfying (25).

Let $\mathcal{Q} = \mathrm{diag}(Q(0), ..., Q(T))$, $\mathcal{R} = \mathrm{diag}(R(0), ..., R(T))$, the LQR cost with $x(0) \sim \mathcal{N}(0, \Sigma)$ and no noise is

$$f(\Phi_X, \Phi_U) = \left\| \mathrm{diag}(\mathcal{Q}^{1/2}, \mathcal{R}^{1/2}) \begin{bmatrix} \Phi_X(:,0) \\ \Phi_U(:,0) \end{bmatrix} \Sigma^{1/2} \right\|_F^2,$$

$\Phi_X(:,0), \Phi_U(:,0)$ are the first n columns of Φ_X, Φ_U. The LQR cost with $x(0), w(t)$ being i.i.d from $\mathcal{N}(0, \Sigma)$ is

$$f(\Phi_X, \Phi_U) = \left\| \mathrm{diag}(\mathcal{Q}^{1/2}, \mathcal{R}^{1/2}) \begin{bmatrix} \Phi_X \\ \Phi_U \end{bmatrix} (I_{T+1} \otimes \Sigma^{1/2}) \right\|_F^2.$$

The symbol $\otimes$ means Kronecker product. If we solve

$$\min_{\mathcal{K}} \; \mathcal{L}(\mathcal{K}), \; \mathcal{K} \text{ is block lower left triangular}$$

with the above two costs of $w(t)$, both can be minimized with constraint (25):

$$\begin{gathered} \min_{\Phi_X, \Phi_U} \; f(\Phi_X, \Phi_U), \; \text{s.t.} \; \begin{bmatrix} I - Z\mathcal{A} & -Z\mathcal{B} \end{bmatrix} \begin{bmatrix} \Phi_X \\ \Phi_U \end{bmatrix} = I, \\ \Phi_X, \Phi_U \text{ are block lower left triangular.} \end{gathered}$$

This problem is convex. The theorem [39, Thm. 2.1] suggests the relation between $\mathcal{L}$ and f satisfying Assumption 3 for Theorem 1. With Theorem 1, we can argue that all stationary points of $\mathcal{L}(\mathcal{K})$ are global minimum.

The paper [40] proposes some generalization of SLS. It introduces a localization constraint, where the state is constrained in a convex set. For example, the constraint is [40, Eq. (9)]

$$\Phi_X(:,0)x_0 \in \mathcal{P}$$

for a convex set $\mathcal{P}$. We can add it to the problem in convex parameterized problem and map it as a constraint in the controller K space. The nonconvex problem is still gradient dominant.

5 A more general description of Assumption 2

In this section, we will give a more general theorem, based on replacing the map $K = LP^{-1}$ by arbitrary function Φ defined below. This allows the theorem to cover more examples in Sec. 6.

We chose $K = LP^{-1}$ because this is frequently used for the convex parameterization of the optimal control problem. For example, with the continuous time LQR problem motivated in Sec. 2, the mapping between K and L, P is almost the only widely used convex parameterization method.

If we choose another change of variable, the resulting objective function is usually not convex in the new variables.

On the other hand, although the mapping $K = LP^{-1}$ is studied, we can generalize Theorem 1 with arbitrary mappings if the reformulated problem is convex – the new mappings still have to satisfy a few assumptions to preserve the Lojasiewicz inequality.

Here we will propose the following assumptions which replace the mapping $K = LP^{-1}$ by an abstract mapping Φ.

Suppose we consider the problems

$$\min_{K} \quad \mathcal{L}(K), \quad \text{s.t. } K \in \mathcal{S}_K, \tag{26}$$

and

$$\min_{P} \quad f(P), \quad \text{s.t.} \quad P \in \mathcal{S}. \tag{27}$$

The matrix P can be a concatenation of many variables, just as a shortlisted expression. For example, P represents (P, L, Z) of continuous LQR. We will study the original optimization problem (26), and map it to a convex optimization problem (27) where the mapping between K and the variable of the other problem P is abstractly denoted by $K = \Phi(P)$ in (28).

Assumption 4. *The feasible set $\mathcal{S}$ is convex in P. The cost function $f(P)$ is convex, finite and differentiable in $P \in \mathcal{S}$. $\mathcal{L}(K)$ is Lipschitz in K.*

Assumption 5. *For all $P \in \mathcal{S}$, $\Phi(P) \in \mathcal{S}_K$. Assume we can express $\mathcal{L}(K)$ as:*

$$\mathcal{L}(K) = \min_{P} \ f(P), \ \textit{s.t. } P \in \mathcal{S}, \ K = \Phi(P). \tag{28}$$

And we assume the first order Taylor expansion of the mapping Φ is well defined as

$$\Phi(P + \mathrm{d}P) = \Phi(P) + \Psi(P)[\mathrm{d}P] + o(\mathrm{d}P).$$

for any $P \in \mathcal{S}$ and any perturbation $\mathrm{d}P$ such that $\mathrm{d}P$ is in the descent cone of $\mathcal{S}$ at P.

We mentioned that, P represents (P, L, Z) in continuous LQR. And we can see that Assumption 2 is very similar to Assumption 5. We just apply $\Phi(P, L, Z) = LP^{-1}$ and get Assumption 2.

Remark 3. Note that, because of (28), the assumption does not trivially hold for any smooth mapping Φ in the very general context. For example, the paper [41] proposes the sum-of-squares method for solving polynomial optimizations, which has a convex parameterization of lifting the problem to a higher dimensional space. We explain the idea in a simple paradigm. let $x \in \mathbb{R}^2$ and the objective function is power 2. The objective function is

$$\mathcal{L}(x) = a_1 x_1^2 + a_2 x_1 x_2 + a_3 x_2^2.$$

One can define a matrix $X \in \mathbf{S}^{2\times 2} \succeq 0$ and a cost function that is linear in X,

$$f(X) = \begin{bmatrix} a_1 & a_2/2 \\ a_2/2 & a_3 \end{bmatrix} X.$$

It can be proven that X^* is rank-1, and it maps to $\begin{bmatrix} x_1^2 & x_1x_2 \\ x_1x_2 & x_2^2 \end{bmatrix}$. However, the map creates many meaningless points while lifting the dimension – extra points when X is rank-2 that are not mapped from the original problem x_1, x_2, and the extra points do not necessarily satisfy (28).

The following conclusion holds with the above Assumptions 4, 5. It generalizes beyond the specific mapping $\Phi(P, L, Z) = LP^{-1}$ to a more general definition, and we propose some instances of convex formulations with different Φ in the next section. We propose the following theorem and the proof is in Appendix A.

Theorem 2. *Denote $\Delta K = \Psi(P)[P^* - P]$. Let $\nabla\mathcal{L}(K)[\Delta K]$ be the directional derivative of $\mathcal{L}(K)$ in direction ΔK. Then with Assumptions 4, 5 we have*

$$\nabla\mathcal{L}(K)[\Delta K] \leq \mathcal{L}(K^*) - \mathcal{L}(K).$$

If K is not optimal, then the right hand side is strictly less than 0, which means the directional derivative of $\mathcal{L}$ is not 0. Therefore $\nabla\mathcal{L}(K) = 0$ holds only at the global minima.

Remark 4. Theorem 2 means that,

$$\|\nabla\mathcal{L}(K)\|_F \geq -\nabla\mathcal{L}(K)[\frac{\Delta K}{\|\Delta K\|_F}] \geq C(K)(\mathcal{L}(K) - \mathcal{L}(K^*))$$

where

$$\begin{aligned} C(K) &= \|\Psi(P)[P^* - P]\|_F^{-1} = \|\Psi(\Phi^{-1}(K))[\Phi^{-1}(K^*) - \Phi^{-1}(K)]\|_F^{-1} \\ &\geq \|\Psi(P)\|_{\text{op}}^{-1}\|P^* - P\|_F^{-1} \end{aligned}$$

For continuous time LQR, P represents the list of variables (P, L, Z) there. Remember

$$\nu = \frac{\lambda_{\min}^2(\Sigma)}{4}\left(\sigma_{\max}(A)\lambda_{\min}^{-1/2}(Q) + \sigma_{\max}(B)\lambda_{\min}^{-1/2}(R)\right)^{-2}.$$

In the sublevel set where $\mathcal{L}(K) \leq a$, we have that

$$\begin{aligned} \|\Psi(P)\|_{\text{op}} &\leq \frac{2a}{\nu}\max\left\{1, \frac{a^2}{\nu(\lambda_{\min}(Q)\lambda_{\min}(R))^{1/2}}\right\}, \\ \|P^* - P\|_F &\leq \frac{a}{\lambda_{\min}^{1/2}(Q)}\max\left\{\lambda_{\min}^{-1/2}(Q), \lambda_{\min}^{-1/2}(R)\right\}. \end{aligned}$$

6 Control problems with generalized map

This section will cover examples where the parameterization is based on the general map Φ, not necessarily $\Phi(P, L) = LP^{-1}$. We can apply Theorem 2 to these problems.

6.1 Distributed finite horizon LQR

The paper [22, Ch. 3] is an *empirical* study (i.e., proposing an algorithm without a proof of convergence) of the gradient descent method for distributed control synthesis. For such a problem, the controller is distributed with a graph structure, showing the accessibility of the distributed controllers to the states: if controller i has no access to state j, then $K_{ij} = 0$, otherwise $K_{ij} \in \mathbb{R}$. Thus there is an extra subspace constraint regarding the graph structure of K, and [22, Ch. 3] applies projected gradient descent on (2) with respect to K. It allows a fixed or random of initial state as in (2). Generally it is NP-hard to find a global optimum with the subspace constraint, so the paper only proposes an algorithm without a proof.

With an extra condition called quadratic invariance, the problem is not NP-hard. We review the solutions in [30] with the connection to our framework.

We consider the time varying linear system

$$\begin{aligned} x(t+1) &= A(t)x(t) + B(t)u(t) + w(t), \\ y(t) &= C(t)x(t). \end{aligned}$$

This is in finite time horizon $t = 0, ..., T$. The state evolution is same as the setup in our SLS example (Sec. 4.5), and we can use the same notations $X, U, W, Z, \mathcal{A}, \mathcal{B}$. We further define

$$Y = \begin{bmatrix} y(0) \\ ... \\ y(T) \end{bmatrix}, \ V = \begin{bmatrix} v(0) \\ ... \\ v(T) \end{bmatrix}, \ \mathcal{C} = \mathrm{diag}(C(0), ..., C(T)).$$

Now we will consider the control policy

$$u(t) = \sum_{i=0}^{t} K(t, t-i)y(i).$$

The search space of policy is same as SLS, and we define $\mathcal{K}$ matrix in the same way. The paper [30] studies the problem under the context of distributed control. One searches for the controller $K \in \mathcal{S}_K$ where $\mathcal{S}_K$ a subset of controllers. In distributed control, there is a graph model for controllers such that the i-th controller might not be able to access the state j for (i, j) in a set of indices $\mathcal{S}_{\text{idx}}$. In this case, $K_{i,j} = 0$ is an extra constraint for the control problem. Therefore, if one searches for the optimal controller in $\mathcal{S}_K$, we can define the subspace

$$\mathcal{S}_K := \{K \mid K_{i,j} = 0, \ \forall (i, j) \in \mathcal{S}_{\text{idx}}\}.$$

The extra constraint is not always easily handled, but [30, §3] proposes an extra assumption, called quadratic invariance (QI), and introduces the equivalent convex optimization.

Remember we defined

$$\mathcal{K} = \begin{bmatrix} K(0,0) & 0 & ... & 0 \\ K(1,1) & K(1,0) & ... & 0 \\ ... & & & \\ K(T,T) & K(T,T-1) & ... & K(T,0) \end{bmatrix}, \ \mathcal{C} = \mathrm{diag}(C(0), ..., C(T)).$$

And we define

$$P_{11} = (I - Z\mathcal{A})^{-1}, \ P_{12} = (I - Z\mathcal{A})^{-1}Z\mathcal{B}.$$

QI means that, for all $\mathcal{K} \in \mathcal{S}_K$, $\mathcal{K}\mathcal{C}P_{12}\mathcal{K} \in \mathcal{S}_K$.

The cost function is:

$$\mathcal{L}(\mathcal{K}) = \sum_{t=0}^{T} y(t)^\top Q(t)y(t) + u(t)^\top R(t)u(t).$$

Define

$$\Phi(\mathcal{G}) = (I + \mathcal{G}\mathcal{C}P_{12})^{-1}\mathcal{G}.$$

Then we can get a new variable $\mathcal{G}$ and a function Φ. With $\mathcal{K} = \Phi(\mathcal{G})$, the cost can be proven to be convex in $\mathcal{G}$. The variable $\mathcal{G}$ is in the same subspace as $\mathcal{K}$ determined by $\mathcal{S}_K$. Indeed, the mapping satisfies Assumptions 4, 5, and the exact formulation of the two optimization problems are described in [30, Append. A, Lem. 5]. Define $\mathcal{Q} = \mathrm{diag}(Q(0), ..., Q(T))$, $\mathcal{R} = \mathrm{diag}(R(0), ..., R(T))$. Let $w(t)$ be Gaussian random vectors with stationary covariance, $w(t_1)$ and $w(t_2)$ are independent $\forall t_1 \neq t_2$. $\Sigma_w = I_T \otimes \mathrm{Cov}(w)$ ($\otimes$ means Kronecker product), $\Sigma_x = \mathrm{diag}(\mathbf{E}(x_0 x_0^\top), 0, ..., 0)$. The convex cost function takes the form

$$f(\mathcal{G}) = \left\| \mathcal{Q}^{1/2}\mathcal{C}(I + P_{12}\mathcal{G}\mathcal{C})P_{11} \begin{bmatrix} \Sigma_w^{1/2} & \Sigma_x^{1/2} \end{bmatrix} \right\|_F^2 + \left\| \mathcal{R}^{1/2}\mathcal{G}\mathcal{C}P_{11} \begin{bmatrix} \Sigma_w^{1/2} & \Sigma_x^{1/2} \end{bmatrix} \right\|_F^2.$$

In summary, we have a pair of problems: 1) minimize $\mathcal{L}(\mathcal{K})$ over $\mathcal{K}$ and 2) minimize $f(\mathcal{G})$ over $\mathcal{G}$. They are related under the Assumptions 1, 3 of Theorem 1. Thus we can claim via Theorem 1 that, all stationary points of $\mathcal{L}(\mathcal{K})$ are global minima.

6.2 Multi-objective and mixed controller design

In this part, we study a few synthesis problems with dynamical controllers, where the objectives are about (e.g., norms of) transfer functions of the closed form system. We study the dynamical system with state, disturbance, input, output, and controller's input x, w, u, z, y with the following dynamics[5]

$$\begin{bmatrix} \dot{x} \\ z \\ y \end{bmatrix} = \begin{bmatrix} A & B_w & B \\ C_z & D & E \\ C & F & 0 \end{bmatrix} \begin{bmatrix} x \\ w \\ u \end{bmatrix}.$$

The controller follows

$$\begin{bmatrix} \dot{x}_c \\ u \end{bmatrix} = \begin{bmatrix} A_c & B_c \\ C_c & D_c \end{bmatrix} \begin{bmatrix} x_c \\ y \end{bmatrix}. \tag{29}$$

We will denote the transfer function of the closed loop system as $\mathscr{T}$, and the control problems below are typically related to $\mathscr{T}$.

In the next few subsections, we will present a few control problems:

1. The **variables** are the controller parameters $\begin{bmatrix} A_c & B_c \\ C_c & D_c \end{bmatrix}$.

2. The **objective functions** are $\mathcal{H}_2$ norm, $\mathcal{H}_\infty$ norm of $\mathscr{T}$ and the weighted sum of norms.

3. The book [42, eq(4.2.15)] defines the parameterization of the problem, by introducing the **variables that typically make the objective functions convex**:

$$v = [X, Y, K, L, M, N].$$

[5] In the mixed design problems (Sec. 6.2.3, 6.2.4), there are multiple dynamical systems.

4. **Mapping of the variables.** Define invertible matrices U, V such that $UV^\top = I - XY$. The matrices A_c, B_c, C_c, D_c are the unique solution of

$$\begin{bmatrix} K & L \\ M & N \end{bmatrix} = \begin{bmatrix} U & XB \\ 0 & I \end{bmatrix} \begin{bmatrix} A_c & B_c \\ C_c & D_c \end{bmatrix} \begin{bmatrix} V^\top & 0 \\ CY & I \end{bmatrix} + \begin{bmatrix} XAY & 0 \\ 0 & 0 \end{bmatrix}. \tag{30}$$

The change of variable enables us to make some control problems as convex optimization, listed below. For simplicity of notation, let

$$\mathscr{X} = \begin{bmatrix} Y & I \\ I & X \end{bmatrix}, \ \mathscr{A} = \begin{bmatrix} AY + BM & A + BNC \\ K & AX + LC \end{bmatrix}, \tag{31}$$

$$\mathscr{B} = \begin{bmatrix} B_w + BNF \\ XB_w + LF \end{bmatrix}, \ \mathscr{C} = \begin{bmatrix} C_z Y + EM & C_z + ENC \end{bmatrix}, \ \mathscr{D} = D + ENF. \tag{32}$$

Remark 5. The mapping in (30) can be written as

$$\begin{bmatrix} A_c & B_c \\ C_c & D_c \end{bmatrix} = \Phi(v) := \begin{bmatrix} U & XB \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} K - XAY & L \\ M & N \end{bmatrix} \begin{bmatrix} V^\top & 0 \\ CY & I \end{bmatrix}^{-1}$$

where Φ plays the role in (28). We propose a few control problems with convex forms in the next few subsections. The variables of nonconvex objective functions are A_c, B_c, C_c, D_c, the new objective functions with respect to $v = [X, Y, K, L, M, N]$ are convex, and the two forms satisfy Assumptions 4, 5 if the map is smooth (which is not always true, see the remark below). Thus the cost functions with respect to matrix $\begin{bmatrix} A_c & B_c \\ C_c & D_c \end{bmatrix}$ are gradient dominant.

Remark 6. In the following subsections, we refer to the result of [42] that, the optimal $\mathcal{H}_\infty$ design, $\mathcal{H}_2$ design and the multi-objective and robust designs, can be made convex optimization problems with the proposed way. However, this map is not guaranteed to be smooth. When matrices U, V are close to singular, the inverses of

$$\begin{bmatrix} U & XB \\ 0 & I \end{bmatrix}, \ \begin{bmatrix} V^\top & 0 \\ CY & I \end{bmatrix}$$

are not well-defined. This makes the nonconvex objective function not gradient dominant. For example, we review the LQG problem [34]. The dynamics takes the form

$$\dot{x}(t) = Ax(t) + Bu(t) + w(t), \ y(t) = Cx(t) + v(t)$$

where $w(t) \sim \mathcal{N}(0, W)$, $v(t) \sim \mathcal{N}(0, V)$. The controller takes the form (29). The cost function is

$$\lim_{T \to \infty} \frac{1}{T} \mathbf{E} \left(\int_{t=0}^{T} (x(t)^\top Q x(t) + u(t)^\top R u(t)) dt \right).$$

The set of stabilizing controllers of LQG problem can be non-connected, and the cost has saddle points [34]. Thus, to apply our theorem and claim that the objectives with respect to controller K are all gradient dominant, we have to restrict the problem in the set where the map is smooth, typically around the global minimum. We will review some controller design problems based on this map in the following subsections.

6.2.1 $\mathcal{H}_\infty$ design

([42, §4.2.3]) The goal in this part is to minimize the $\mathcal{H}_\infty$ norm of the closed loop system's transfer function by designing the optimal controller. Let the transfer function of the closed form system be $\mathscr{T}$. The problem with its raw, nonconvex form is to minimize the $\|\mathscr{T}\|_{\mathcal{H}_\infty}$ over A_c, B_c, C_c, D_c, and we will propose the convex formulation – the change of variable trick such that the argument becomes v. The problem takes the form:

$$\begin{aligned} &\min\ \gamma, \\ &\text{s.t.}\quad \mathscr{X} \succeq 0,\ \begin{bmatrix} \mathscr{A}^\top + \mathscr{A} & \mathscr{B} & \mathscr{C}^\top \\ \mathscr{B}^\top & -\gamma I & \mathscr{D}^\top \\ \mathscr{C} & \mathscr{D} & -\gamma I \end{bmatrix} \preceq 0. \end{aligned}$$

If we fix all other parameters listed in v and optimize over γ, then γ^* (that depends on v) is the $\mathcal{H}_\infty$ norm value of the closed loop system with the mapping from v to controller by (30). If we minimize over γ and v, then we can get optimal $\mathcal{H}_\infty$ design.

6.2.2 $\mathcal{H}_2$ design

([42, §4.2.5]) This part is similar to $\mathcal{H}_\infty$ design. Suppose the goal is to minimize $\|\mathscr{T}\|_{\mathcal{H}_2}$, one can alternatively solve

$$\begin{aligned} &\min\ \gamma, \\ &\text{s.t.}\quad \begin{bmatrix} \mathscr{A}^\top + \mathscr{A} & \mathscr{B} \\ \mathscr{B}^\top & -\gamma I \end{bmatrix} \preceq 0,\ \mathscr{D} = 0,\ \begin{bmatrix} \mathscr{X} & \mathscr{C}^\top \\ \mathscr{C} & Z \end{bmatrix} \succeq 0,\ \mathbf{Tr}(Z) \le \gamma. \end{aligned}$$

If we fix all other parameters and optimize over γ, Z, then γ^* (that depends on v) is the $\mathcal{H}_2$ norm value of the closed loop system with the mapping from v to controller by (30). If we minimize over γ, Z and v, then we can get optimal $\mathcal{H}_2$ design.

6.2.3 Multi-objective synthesis

([42, §4.3]) Let the system be

$$\begin{bmatrix} \dot{x} \\ z_1 \\ z_2 \\ y \end{bmatrix} = \begin{bmatrix} A & B_1 & B_2 & B \\ C_1 & D_1 & D_{12} & E_1 \\ C_2 & D_{21} & D_2 & E_2 \\ C & F_1 & F_2 & 0 \end{bmatrix} \begin{bmatrix} x \\ w_1 \\ w_2 \\ u \end{bmatrix} \tag{33}$$

Now we study the mixed design for $\mathcal{H}_\infty$ design from z_1 to w_1 and $\mathcal{H}_2$ design from z_2 to w_2. We keep the mapping (30) and the change of parameter (31), but replace (32) by

$$\mathscr{B}_i = \begin{bmatrix} B_i + BNF_i \\ XB_i + LF_i \end{bmatrix},\ \mathscr{C}_i = \begin{bmatrix} C_iY + E_iM & C_i + E_iNC \end{bmatrix},\ \mathscr{D}_i = D_i + E_iNF_i.$$

for $i = 1, 2$. Suppose we are given a positive number λ and hope to study $\|\mathscr{T}_1\|_{\mathcal{H}_\infty} + \lambda\|\mathscr{T}_2\|_{\mathcal{H}_2}$ where $\mathscr{T}_i$ is the transfer function of the i-th system (z_1 to w_1, z_2 to w_2), then we can write

$$\min\ \gamma_1 + \lambda\gamma_2, \tag{34}$$

$$\text{s.t.}\ \begin{bmatrix} \mathscr{A}^\top + \mathscr{A} & \mathscr{B}_1 & \mathscr{C}_1^\top \\ \mathscr{B}_1^\top & -\gamma_1 I & \mathscr{D}_1^\top \\ \mathscr{C}_1 & \mathscr{D}_1 & -\gamma_1 I \end{bmatrix} \preceq 0, \tag{35}$$

$$\begin{bmatrix} \mathscr{A}^\top + \mathscr{A} & \mathscr{B}_2 \\ \mathscr{B}_2^\top & -\gamma_2 I \end{bmatrix} \preceq 0,\ \mathscr{D}_2 = 0,\ \begin{bmatrix} \mathscr{X} & \mathscr{C}_2^\top \\ \mathscr{C}_2 & Z \end{bmatrix} \succeq 0,\ \mathbf{Tr}(Z) \leq \gamma_2. \tag{36}$$

If we fix all other parameters and optimize over γ_1, γ_2, Z, then the function value is the mixed $\mathcal{H}_\infty/\mathcal{H}_2$ norm value of the closed loop system with the mapping from v to controller by (30). If we minimize over γ_1, γ_2, Z and v, then we can get the optimal mixed $\mathcal{H}_\infty/\mathcal{H}_2$ design.

6.2.4 Robust state feedback control

([42, §8.1.2]) We study the robust state feedback control problem, where the robustness corresponds to a system with uncertain parameters, denoted by Δ below. We apply the system model (33). "State feedback" means that $C = I$ and $F_1, F_2 = 0$. Let N_Δ be a positive integer. The connection between w_1 and z_1 is an uncertain channel

$$w_1(t) = \Delta(t) z_1(t)$$

for any

$$\Delta(t) \in \Delta_c := \text{conv}\{0, \Delta_1, ..., \Delta_{N_\Delta}\}.$$

The goal is to minimize a certain norm of the transfer function from z_2 to w_2, which can be $\mathcal{H}_2$ norm, $\mathcal{H}_\infty$ norm studied in the previous part. We consider minimizing the norms under an extra constraint when the closed loop system achieves stability with z_1 to w_1 (z_1 with finite norm) and robust quadratic performance with z_2 to w_2 via a matrix P_p. The robust quadratic performance is defined as: there exists a matrix P_p,

$$P_p = \begin{bmatrix} \tilde{Q}_p & \tilde{S}_p \\ \tilde{S}_p^\top & \tilde{R}_p \end{bmatrix},\ P_p^{-1} = \begin{bmatrix} Q_p & S_p \\ S_p^\top & R_p \end{bmatrix}$$

such that $\tilde{R}_p \succ 0, Q_p \prec 0$, and

$$\int_0^\infty \begin{bmatrix} w_2(t) \\ z_2(t) \end{bmatrix}^\top P_p \begin{bmatrix} w_2(t) \\ z_2(t) \end{bmatrix} \mathrm{d}t \leq \epsilon \|w_2\|_{\mathcal{H}_2}^2$$

for some $\epsilon > 0$.

Define new variables Q, S, R in addition to $v = [X, Y, K, L, M, N]$, and let $\mathscr{M}$ replace

$$\mathscr{M} \leftarrow \begin{bmatrix} -(AY + BM)^\top & -(C_1 Y + E_1 M)^\top & -(C_2 Y + E_2 M)^\top \\ I & 0 & 0 \\ -B_1^\top & -D_1^\top & -D_{21}^\top \\ 0 & I & 0 \\ -B_2^\top & -D_{12}^\top & -D_2^\top \\ 0 & 0 & I \end{bmatrix}.$$

The constraints, which is proven to be convex in [42, §8.1.2] can be written as

$$R \succ 0,\ Q \prec 0,\ \begin{bmatrix} I \\ -\Delta_j \end{bmatrix}^\top \begin{bmatrix} Q & S \\ S^\top & R \end{bmatrix} \begin{bmatrix} I \\ -\Delta_j \end{bmatrix} \prec 0,\ \forall j \in [N_\Delta]$$

$$Y \succ 0,\ \mathscr{M}^\top \begin{bmatrix} 0 & I & 0 & 0 & 0 & 0 \\ I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & Q & S & 0 & 0 \\ 0 & 0 & S^\top & R & 0 & 0 \\ 0 & 0 & 0 & 0 & Q_p & S_p \\ 0 & 0 & 0 & 0 & S_p^\top & R_p \end{bmatrix} \mathscr{M} \succ 0.$$

For example, if we aim to minimize the $\mathcal{H}_2$ norm of the transfer function from z_2 to w_2, then we can minimize γ_2 subject to (36) and the constraints above. The main theorem of this paper suggests that, with the convex formulation, if we apply policy gradient descent with respect to $\mathcal{H}_2$ norm of the transfer function from z_2 to w_2 with robust stability of system 1 (z_1 with finite $\mathcal{H}_2$ norm) and robust quadratic performance constraints of system 2, then policy gradient descent converges to globally optimal controller.

6.2.5 Discrete time system

([42, §4.6]) Suppose we study the discrete time system, and we define the system in a similar way of defining the continuous time system:

$$\begin{bmatrix} x(t+1) \\ z_1(t) \\ z_2(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} A & B_1 & B_2 & B \\ C_1 & D_1 & D_{12} & E_1 \\ C_2 & D_{21} & D_2 & E_2 \\ C & F_1 & F_2 & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ w_1(t) \\ w_2(t) \\ u(t) \end{bmatrix},\ \begin{bmatrix} x_c(t+1) \\ u(t) \end{bmatrix} = \begin{bmatrix} A_c & B_c \\ C_c & D_c \end{bmatrix} \begin{bmatrix} x_c(t) \\ y(t) \end{bmatrix}.$$

Now we study the mixed design for $\mathcal{H}_\infty$ design from z_1 to w_1 and $\mathcal{H}_2$ design from z_2 to w_2. Suppose we are given a positive number λ and hope to study $\|\mathscr{T}_1\|_{\mathcal{H}_\infty} + \lambda \|\mathscr{T}_2\|_{\mathcal{H}_2}$ where $\mathscr{T}_i$ is the transfer function of the i-th system (z_1 to w_1, z_2 to w_2), then we can write

$$\begin{aligned} \min\ & \gamma_1 + \lambda \gamma_2, \\ \text{s.t.}\ & \begin{bmatrix} \mathscr{X} & 0 & \mathscr{A}^\top & \mathscr{C}_1^\top \\ 0 & \gamma_1 I & \mathscr{B}_1^\top & \mathscr{D}_1 \\ \mathscr{A} & \mathscr{B}_1 & \mathscr{X} & 0 \\ \mathscr{C}_1 & \mathscr{D}_1 & 0 & \gamma_1 I \end{bmatrix} \succ 0,\ \mathbf{Tr}(Z) \le \gamma_2, \\ & \begin{bmatrix} \mathscr{X} & \mathscr{A} & \mathscr{B}_2 \\ \mathscr{A}^\top & \mathscr{X} & 0 \\ \mathscr{B}_2^\top & 0 & \gamma_2 I \end{bmatrix} \succ 0,\ \begin{bmatrix} \mathscr{X} & 0 & \mathscr{C}_2 \\ 0 & \mathscr{X} & \mathscr{D}_2 \\ \mathscr{C}_2 & \mathscr{D}_2 & Z \end{bmatrix} \succ 0. \end{aligned}$$

If we fix all other parameters and optimize over γ_1, γ_2, Z, then the function value is the mixed $\mathcal{H}_\infty/\mathcal{H}_2$ value of the closed loop system with the mapping from v. If we minimize over γ_1, γ_2, Z and v, then we can get the optimal mixed $\mathcal{H}_\infty/\mathcal{H}_2$ design.

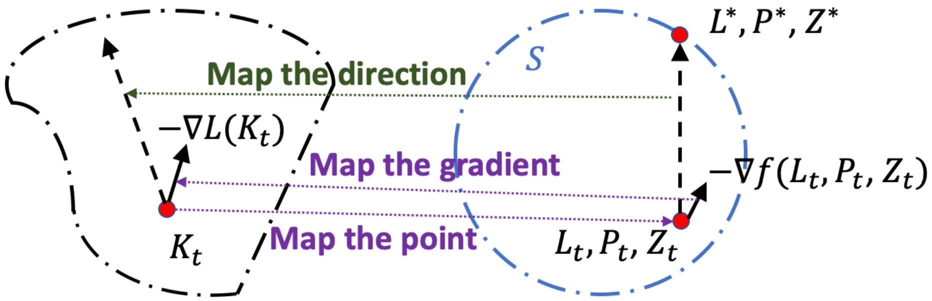

Figure 1: Mapping between nonconvex and convex landscapes. Suppose we run gradient descent at iteration t, for any controller K, we can map it to L, P, Z in the other parameterized space. and then we map the direction $(L^*, P^*, Z^*) - (L, P, Z)$ and the gradient $\nabla f(L, P, Z)$ back to the original K space. Since in (L, P, Z) space the objective function is convex, then $\langle \nabla f(L, P, Z), (L^*, P^*, Z^*) - (L, P, Z)\rangle < 0$. We prove that similar correlation holds for the nonconvex objective.

7 Proof sketch

We put the full proof of Theorem 1 in Appendix A, and give a sketch of the proof in this section. We illustrate the idea in Figure 1, which, on the high level, maps the original space of controller K where the cost is nonconvex, and the parameterized space with L, P, Z where the cost is convex.

For simplicity, we sketch the proof using Assumptions 1,3. For any point K, we can find a point (L, P, Z) in the parameterized space. If it is not the optimizer, we can find the line segment linking (L, P, Z) and the optimizer (L^*, P^*, Z^*). Note that the optimization problem is convex in this space so that $\langle \nabla f(L, P, Z), (L^*, P^*, Z^*) - (L, P, Z)\rangle$ is upper bounded by $f(L^*, P^*, Z^*) - f(L, P, Z)$. Then with the help of our assumptions, we can map the directional derivative back to the original K space, and show that the directional derivative in $\mathcal{L}(K)$ is not 0.

Before concluding, we remark that the assumptions in Theorem 1 come from an optimization theory perspective, and we do not dive into the control theoretic interpretations of the constants and assumptions. Our approach has the benefit that it unifies the analysis of many control problems in a single abstract result, showing that all stationary points of the objective functions are global minima, so that one can apply policy gradient method to find the globally optimal policy. The future work is to refine the analysis to obtain the best case-specific convergence rates, and to provide an interpretation of the associated constants in terms of control theoretic notions. We are also interested in understanding the cost landscape of LQG problem [34] and its connection with the convex parameterization, and investigating the second order landscape analysis via the mapping to the convex problem.

Acknowledgements

This work is supported in part by the NSF TRIPODS II grant DMS 2023166, NSF TRIPODS CCF 1740551, NSF CCF 2007036, and the Moorthy Professorship at UW ECE.

References

[1] R. E. Kalman *et al.*, "Contributions to the theory of optimal control," *Bol. soc. mat. mexicana*, vol. 5, no. 2, pp. 102–119, 1960.

[2] J. C. Perdomo, J. Umenberger, and M. Simchowitz, "Stabilizing dynamical systems via policy gradient methods," in *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021.

[3] H. Mohammadi, A. Zare, M. Soltanolkotabi, and M. R. Jovanović, "Global exponential convergence of gradient methods over the nonconvex landscape of the linear quadratic regulator," in *2019 IEEE 58th Conference on Decision and Control (CDC).* IEEE, 2019, pp. 7474–7479.

[4] M. Fazel, R. Ge, S. M. Kakade, and M. Mesbahi, "Global convergence of policy gradient methods for linearized control problems," *arXiv preprint arXiv:1801.05039*, 2018.

[5] R. F. Stengel, *Optimal control and estimation.* Courier Corporation, 1994.

[6] G. E. Dullerud and F. Paganini, *A course in robust control theory: a convex approach.* Springer Science & Business Media, 2013, vol. 36.

[7] G. Hewer, "An iterative technique for the computation of the steady state gains for the discrete optimal regulator," *IEEE Transactions on Automatic Control*, vol. 16, no. 4, pp. 382–384, 1971.

[8] P. Lancaster and L. Rodman, *Algebraic riccati equations.* Clarendon press, 1995.

[9] V. Balakrishnan and L. Vandenberghe, "Semidefinite programming duality and linear time-invariant systems," *IEEE Transactions on Automatic Control*, vol. 48, no. 1, pp. 30–41, 2003.

[10] L. Ljung, "System identification: theory for the user," *PTR Prentice Hall, Upper Saddle River, NJ*, pp. 1–14, 1999.

[11] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht, "Learning without mixing: Towards a sharp analysis of linear system identification," in *Conference On Learning Theory*, 2018, pp. 439–473.

[12] S. J. Bradtke, B. E. Ydstie, and A. G. Barto, "Adaptive linear quadratic control using policy iteration," in *Proceedings of 1994 American Control Conference-ACC'94*, vol. 3. IEEE, 1994, pp. 3475–3479.

[13] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "On the sample complexity of the linear quadratic regulator," *arXiv preprint arXiv:1710.01688*, 2017.

[14] S. Dean, S. Tu, N. Matni, and B. Recht, "Safely learning to control the constrained linear quadratic regulator," in *2019 American Control Conference (ACC).* IEEE, 2019, pp. 5582–5588.

[15] H. Mania, S. Tu, and B. Recht, "Certainty equivalent control of LQR is efficient," *arXiv preprint arXiv:1902.07826*, 2019.

[16] S. Tu and B. Recht, "The gap between model-based and model-free methods on the linear quadratic regulator: An asymptotic viewpoint," in *Conference on Learning Theory.* PMLR, 2019, pp. 3036–3083.

[17] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE circuits and systems magazine*, vol. 9, no. 3, pp. 32–50, 2009.

[18] J. Y. Lee, J. B. Park, and Y. H. Choi, "Integral q-learning and explorized policy iteration for adaptive optimal control of continuous-time linear systems," *Automatica*, vol. 48, no. 11, pp. 2850–2859, 2012.

[19] D. Lee and J. Hu, "Primal-dual q-learning framework for LQR design," *IEEE Transactions on Automatic Control*, vol. 64, no. 9, pp. 3756–3763, 2018.

[20] S. M. Kakade, "A natural policy gradient," in *Advances in neural information processing systems*, 2002, pp. 1531–1538.

[21] A. Rajeswaran, K. Lowrey, E. V. Todorov, and S. M. Kakade, "Towards generalization and simplicity in continuous control," in *Advances in Neural Information Processing Systems*, 2017, pp. 6550–6561.

[22] K. Mårtensson, "Gradient methods for large-scale and distributed linear quadratic control," Ph.D. dissertation, Lund University, 2012.

[23] J. W. Roberts, I. R. Manchester, and R. Tedrake, "Feedback controller parameterizations for reinforcement learning," in *2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*. IEEE, 2011, pp. 310–317.

[24] D. Malik, A. Pananjady, K. Bhatia, K. Khamaru, P. Bartlett, and M. Wainwright, "Derivative-free methods for policy optimization: Guarantees for linear quadratic systems," in *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 2019, pp. 2916–2925.

[25] J. Bu, A. Mesbahi, M. Fazel, and M. Mesbahi, "LQR through the lens of first order methods: Discrete-time case," *arXiv preprint arXiv:1907.08921*, 2019.

[26] J. Bu, A. Mesbahi, and M. Mesbahi, "Policy gradient-based algorithms for continuous-time linear quadratic control," *arXiv preprint arXiv:2006.09178*, 2020.

[27] J. Bu, L. J. Ratliff, and M. Mesbahi, "Global convergence of policy gradient for sequential zero-sum linear quadratic dynamic games," *arXiv preprint arXiv:1911.04672*, 2019.

[28] K. Zhang, Z. Yang, and T. Basar, "Policy optimization provably converges to nash equilibria in zero-sum linear quadratic games," *Advances in Neural Information Processing Systems*, vol. 32, pp. 11 602–11 614, 2019.

[29] K. Zhang, B. Hu, and T. Basar, "Policy optimization for $\mathcal{H}_2$ linear control with $\mathcal{H}_\infty$ robustness guarantee: Implicit regularization and global convergence," in *Learning for Dynamics and Control*, 2020, pp. 179–190.

[30] L. Furieri, Y. Zheng, and M. Kamgarpour, "Learning the globally optimal distributed LQ regulator," in *Learning for Dynamics and Control*. PMLR, 2020, pp. 287–297.

[31] Y. Li, Y. Tang, R. Zhang, and N. Li, "Distributed reinforcement learning for decentralized linear quadratic control: A derivative-free policy optimization approach," *IEEE Transactions on Automatic Control*, 2021.

[32] H. Feng and J. Lavaei, "Connectivity properties of the set of stabilizing static decentralized controllers," *SIAM Journal on Control and Optimization*, vol. 58, no. 5, pp. 2790–2820, 2020.

[33] J. Duan, J. Li, and L. Zhao, "Optimization landscape of gradient descent for discrete-time static output feedback," *arXiv preprint arXiv:2109.13132*, 2021.

[34] Y. Tang, Y. Zheng, and N. Li, "Analysis of the optimization landscape of linear quadratic gaussian (lqg) control," in *Learning for Dynamics and Control.* PMLR, 2021, pp. 599–610.

[35] R. Bellman, "Dynamic programming," *Science*, vol. 153, no. 3731, pp. 34–37, 1966.

[36] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory.* SIAM, 1994.

[37] S. Lojasiewicz, "A topological property of real analytic subsets," *Coll. du CNRS, Les équations aux dérivées partielles*, vol. 117, pp. 87–89, 1963.

[38] J. P. Jansch-Porto, B. Hu, and G. E. Dullerud, "Convergence guarantees of policy optimization methods for markovian jump linear systems," in *2020 American Control Conference (ACC).* IEEE, 2020, pp. 2882–2887.

[39] J. Anderson, J. C. Doyle, S. H. Low, and N. Matni, "System level synthesis," *Annual Reviews in Control*, vol. 47, pp. 364–393, 2019.

[40] C. A. Alonso, F. Yang, and N. Matni, "Data-driven distributed and localized model predictive control," *arXiv preprint arXiv:2112.12229*, 2021.

[41] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM Journal on optimization*, vol. 11, no. 3, pp. 796–817, 2001.

[42] C. Scherer and S. Weiland, "Linear matrix inequalities in control," *Lecture Notes, Dutch Institute for Systems and Control, Delft, The Netherlands*, vol. 3, no. 2, 2000.

[43] A. Zare, H. Mohammadi, N. K. Dhingra, T. T. Georgiou, and M. R. Jovanović, "Proximal algorithms for large-scale statistical modeling and sensor/actuator selection," *IEEE Transactions on Automatic Control*, vol. 65, no. 8, pp. 3441–3456, 2019.

[44] O. L. d. V. Costa, *Discrete-time Markov jump linear systems*, ser. Probability and its applications (Springer-Verlag). London: Springer, 2005.

[45] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to derivative-free optimization.* SIAM, 2009.

[46] Y. Nesterov and V. Spokoiny, "Random gradient-free minimization of convex functions," *Foundations of Computational Mathematics*, vol. 17, no. 2, pp. 527–566, 2017.

A Proof of the main theorems

Theorem 1. *Suppose assumptions 1,2 hold, and consider the two problems* (9) *and* (10)*. Let K^* denote the global minimizer of $\mathcal{L}(K)$ in S_K. Then there exist constants $C_1, C_2 > 0$ independent of the suboptimality $\mathcal{L}(K) - \mathcal{L}(K^*)$, and a direction V, with $\|V\|_F = 1$, in the descent cone of $\mathcal{S}_K$ at K such that,*

1. *if f is convex, the gradient of $\mathcal{L}$ satisfies*[6]
$$\nabla \mathcal{L}(K)[V] \leq -C_1(\mathcal{L}(K) - \mathcal{L}(K^*)). \tag{37}$$

2. *if f is μ-strongly convex, the gradient satisfies*
$$\nabla \mathcal{L}(K)[V] \leq -C_2(\mu(\mathcal{L}(K) - \mathcal{L}(K^*)))^{1/2}. \tag{38}$$

where (the constants can be bounded with simple constraints bounding norms of L, P or K)

$$\begin{aligned} C_1 &= (2\max\{\|L - L^*\|_F \sigma_{\min}^{-1}(P), \|P - P^*\|_F \sigma_{\min}^{-2}(P)\sigma_{\max}(L)\})^{-1}, \\ C_2 &= (2\max\{\sigma_{\min}^{-1}(P), \sigma_{\min}^{-2}(P)\sigma_{\max}(L)\})^{-1}. \end{aligned}$$

Proof. Let $f(x)$ be any convex function. Denote $\mathcal{P}_{\mathcal{S}}(\nabla f(x))$ as the projection of $\nabla f(x)$ onto the descent cone of $\mathcal{S}$ at x, and we know $\|\mathcal{P}_{\mathcal{S}}(\nabla f(x))\| \geq \nabla f(x)[\frac{\Delta}{\|\Delta\|}]$ for any $-\Delta$ in the descent cone of $\mathcal{S}$ at x. We will find the direction Δ and bound the directional deravative. First, for any convex function $f(x)$, let the minimum be x^*, and $x - x^* = \Delta$. Let $\nabla f(x) = g$. For any non-stationary point, $f(x) \leq f(x^*) + g^\top \Delta$. Since $\mathcal{S}$ is a convex set, $-\Delta$ belongs to the descent cone of $\mathcal{S}$ at x, so the direction $-\frac{\Delta}{\|\Delta\|}$ is feasible, $f(x) - f(x - t\frac{\Delta}{\|\Delta\|}) \leq tg^\top \frac{\Delta}{\|\Delta\|}$ when $t \to_+ 0$, so that $f(x)[\frac{\Delta}{\|\Delta\|}] = g^\top \frac{\Delta}{\|\Delta\|} \geq \frac{f(x) - f(x^*)}{\|x - x^*\|}$. We will apply the inequality for $f(L, P, Z)$.

Let K^* be the optimal K and (L^*, P^*, Z^*) be the optimal point in the parameterized space. We have $\mathcal{L}(K^*) = f(L^*, P^*, Z^*)$.

We denote $\mathscr{Z}(L, P) \in \operatorname{argmin}_Z f(L, P, Z)$ subject to $(L, P, Z) \in \mathcal{S}$ (if there are multiple minimizers we pick any one). With either Assumption 3 or 2, we can define the mapping from K to (L, P, Z) respectively in one of the following ways:

1. (Assumption 3) let K map to (L, P) with $K = LP^{-1}$ and $Z = \mathscr{Z}(L, P)$.
2. (Assumption 2) let
$$\begin{aligned} &(L, P, Z) = \operatorname{argmin}_{L', P', Z'} \ f(L', P', Z') \\ &\text{s.t. } (L', P', Z') \in \mathcal{S}, \ P' \succ 0, \ L' P'^{-1} = K. \end{aligned}$$

Note f is convex, so

$$\begin{aligned} &\nabla f(L, P, Z)[(L, P, Z) - (L^*, P^*, Z^*)] \\ &\geq f(L, P, Z) - f(L^*, P^*, Z^*) \\ &= f(L, P, \mathscr{Z}(L, P)) - f(L^*, P^*, \mathscr{Z}(L^*, P^*)) \\ &= \mathcal{L}(K) - \mathcal{L}(K^*). \end{aligned} \tag{39}$$

[6] We always consider the directional derivative of a feasible direction within descent cone.

Now we consider the directional derivative in K space. By definition,

$$\nabla\mathcal{L}(K)[V] = \lim_{t\to 0^+}(\mathcal{L}(K+tV) - \mathcal{L}(K))/t.$$

Let $\Delta L = L^* - L$, $\Delta P = P^* - P$, and $V = \Delta LP^{-1} - LP^{-1}\Delta PP^{-1}$. Then

$$\begin{aligned}\nabla\mathcal{L}(K)[V] &= \lim_{t\to 0^+}(\mathcal{L}(K+tV) - \mathcal{L}(K))/t \\ &= \lim_{t\to 0^+}(\mathcal{L}(LP^{-1} + t(\Delta LP^{-1} - LP^{-1}\Delta PP^{-1})) - \mathcal{L}(LP^{-1}))/t \\ &= \lim_{t\to 0^+}(\mathcal{L}((L+t\Delta L)(P+t\Delta P)^{-1}) - \mathcal{L}(LP^{-1}))/t.\end{aligned}$$

The last line uses $(P+t\Delta P)^{-1} = P^{-1} - tP^{-1}\Delta PP^{-1} + o(t)$. Denote $\Delta(L,P,Z) = (L^*,P^*,Z^*) - (L,P,Z)$, $\Delta(L,P,Z)$ is in the descent cone of $\mathcal{S}$ at (L,P,Z) due to the convexity of $\mathcal{S}$. With Assumption 3, we continue with

$$\begin{aligned}\nabla\mathcal{L}(K)[V] &= \lim_{t\to 0^+}(f(L+t\Delta L, P+t\Delta P, \mathscr{Z}(L+t\Delta L, P+t\Delta P)) - f(L,P,\mathscr{Z}(L,P)))/t \\ &\le \lim_{t\to 0^+}(f(L+t\Delta L, P+t\Delta P, \mathscr{Z}(L,P)+t\Delta Z) - f(L,P,\mathscr{Z}(L,P)))/t \\ &= \nabla f(L,P,Z)[\Delta(L,P,Z)].\end{aligned}$$

With Assumption 2, we continue with

$$\begin{aligned}\nabla\mathcal{L}(K)[V] = \lim_{t\to 0^+}\min_{L',P',Z'}\ & f(L',P',Z') - f(L,P,Z) \\ \text{s.t. } & (L',P',Z')\in\mathcal{S},\ P'\succ 0, \\ & L'P'^{-1} = (L+t\Delta L)(P+t\Delta P)^{-1}.\end{aligned}$$

$(L+t\Delta L, P+t\Delta P, \mathscr{Z}(L,P)+t\Delta Z)$ is a feasible point of the optimization problem, thus is less than or equal to the minimum, and then

$$\begin{aligned}\nabla\mathcal{L}(K)[V] &\le \lim_{t\to 0^+}(f(L+t\Delta L, P+t\Delta P, \mathscr{Z}(L,P)+t\Delta Z) - f(L,P,\mathscr{Z}(L,P)))/t \\ &= \nabla f(L,P,Z)[\Delta(L,P,Z)].\end{aligned}$$

So the following inequality holds.

$$\nabla\mathcal{L}(K)[V] \le \nabla f(L,P,Z)[\Delta(L,P,Z)] \le -(f(L,P,Z) - f(L^*,P^*,Z^*)) = -(\mathcal{L}(K) - \mathcal{L}(K^*)) < 0.$$

After normalization, we have

$$\nabla\mathcal{L}(K)[\frac{V}{\|V\|_F}] \ge \frac{1}{\|V\|_F}(\mathcal{L}(K) - \mathcal{L}(K^*)). \tag{40}$$

With $V = \Delta LP^{-1} - LP^{-1}\Delta PP^{-1}$, we can get $\|V\|_F \le 1/C_1$.

If $f(L,P,Z)$ is μ strongly convex, then we can restrict f in the line segment $(L,P,Z) - (L^*,P^*,Z^*)$ and get

$$
\begin{aligned}
(\frac{\nabla\mathcal{L}(K)[V]}{\|V\|_F})^2 &\geq \frac{1}{\|V\|_F^2}(\nabla f(L,P,Z)[\Delta(L,P,Z)])^2 \\
&\geq \frac{\mu\|\Delta(L,P,Z)\|_F^2}{\|V\|_F^2}\cdot(f(L,P,Z)-f(L^*,P^*,Z^*)) \\
&= \frac{\mu(\|L^*-L\|_F^2+\|P^*-P\|_F^2+\|Z^*-Z\|_F^2)}{\|(L^*-L)P^{-1}-LP^{-1}(P^*-P)P^{-1}\|_F^2}\cdot(f(L,P,Z)-f(L^*,P^*,Z^*)) \\
&\geq \frac{\mu(\|L^*-L\|_F^2+\|P^*-P\|_F^2)}{\|(L^*-L)P^{-1}-LP^{-1}(P^*-P)P^{-1}\|_F^2}\cdot(f(L,P,Z)-f(L^*,P^*,Z^*)) \\
&\geq \frac{\mu(f(L,P,Z)-f(L^*,P^*,Z^*))}{(2\max\{\sigma_{\min}^{-1}(P),\sigma_{\min}^{-2}(P)\sigma_{\max}(L)\})^2}.
\end{aligned}
$$

Now we will prove with the following assumption: let $\mathcal{P}_{\mathcal{S}}(-\nabla f(L,P,Z))$ be the projection of $-\nabla f(L,P,Z)$ in the descent cone of $\mathcal{S}$ at (L,P,Z), if for any

$$
(L,P,Z) = \arg\min_{L',P',Z'} f(L',P',Z'), \text{ s.t. } (L',P',Z')\in\mathcal{S},\ L'(P')^{-1}=K,
$$

we have $\|\mathcal{P}_{\mathcal{S}}(-\nabla f(L,P,Z))\|_F^2 \geq \mu(f(L,P,Z)-f(L^*,P^*,Z^*))$.

Now we denote

$$
\Delta(L,P,Z) = (\Delta L,\Delta P,\Delta Z) = \frac{\mathcal{P}_{\mathcal{S}}(-\nabla f(L,P,Z))}{\|\mathcal{P}_{\mathcal{S}}(-\nabla f(L,P,Z))\|}
$$

and $V=\Delta LP^{-1}-LP^{-1}\Delta PP^{-1}$. The proof is similar to strongly convex case:

$$
\begin{aligned}
(\frac{\nabla\mathcal{L}(K)[V]}{\|V\|_F})^2 &\geq \frac{1}{\|V\|_F^2}(\nabla f(L,P,Z)[\Delta(L,P,Z)])^2 \\
&\geq \frac{\mu\|\Delta(L,P,Z)\|_F^2}{\|V\|_F^2}\cdot(f(L,P,Z)-f(L^*,P^*,Z^*)) \\
&= \frac{\mu(\|\Delta L\|_F^2+\|\Delta P\|_F^2+\|\Delta Z\|_F^2)}{\|(\Delta L)P^{-1}-LP^{-1}(\Delta P)P^{-1}\|_F^2}\cdot(f(L,P,Z)-f(L^*,P^*,Z^*)) \\
&\geq \frac{\mu(\|\Delta L\|_F^2+\|\Delta P\|_F^2)}{\|(\Delta L)P^{-1}-LP^{-1}(\Delta P)P^{-1}\|_F^2}\cdot(f(L,P,Z)-f(L^*,P^*,Z^*)) \\
&\geq \frac{\mu(f(L,P,Z)-f(L^*,P^*,Z^*))}{(2\max\{\sigma_{\min}^{-1}(P),\sigma_{\min}^{-2}(P)\sigma_{\max}(L)\})^2}.
\end{aligned}
$$

□

Theorem 2. *Denote $\Delta K=\Psi(P)[P^*-P]$. Let $\nabla\mathcal{L}(K)[\Delta K]$ be the directional derivative of $\mathcal{L}(K)$ in direction ΔK. Then with Assumptions 4, 5 we have*

$$
\nabla\mathcal{L}(K)[\Delta K] \leq \mathcal{L}(K^*)-\mathcal{L}(K).
$$

Proof. Suppose $f(P)$ is convex in P, and the optimizer of (27) is P^*. Denote

$$
P=\text{argmin}_{P'}\ f(P'),\ \text{s.t. } P'\in\mathcal{S},\ K=\Phi(P'),
$$

and

$$\Delta P = P^* - P, \ \Delta K = \Psi(P)[\Delta P].$$

We take the directional derivative and get (explanation of key steps below the last line)

$$\begin{aligned}
\nabla \mathcal{L}(K)[\Delta K] &= \lim_{t \to 0^+} \frac{\mathcal{L}(K + t\Delta K) - \mathcal{L}(K)}{t} \\
&= \lim_{t \to 0^+} \frac{\mathcal{L}(K + t\Psi(P)[\Delta P]) - f(P)}{t} \quad (41) \\
&= \lim_{t \to 0^+} \frac{\mathcal{L}(\Phi(P) + t\Psi(P)[\Delta P]) - f(P)}{t} \quad (42) \\
&= \lim_{t \to 0^+} \frac{\mathcal{L}(\Phi(P + t\Delta P) - o(t)) - f(P)}{t} \quad (43) \\
&= \lim_{t \to 0^+} \frac{\mathcal{L}(\Phi(P + t\Delta P)) - f(P)}{t} \\
&= \lim_{t \to 0^+} \frac{\min_{P' \in \mathcal{S}, \ \Phi(P + t\Delta P) = \Phi(P')} \ f(P') - f(P)}{t} \quad (44) \\
&= \lim_{t \to 0^+} \frac{\min_{P' \in \mathcal{S}, \ \Phi(P + t\Delta P) = \Phi(P')} \ f(P') - f(P + t\Delta P) + f(P + t\Delta P) - f(P)}{t} \\
&= \lim_{t \to 0^+} \frac{\min_{P' \in \mathcal{S}, \ \Phi(P + t\Delta P) = \Phi(P')} \ f(P') - f(P + t\Delta P)}{t} + \nabla f(P)[\Delta P]. \quad (45)
\end{aligned}$$

(41) and (42) replace ΔK and K with expressions in P and ΔP. (43) applies the Taylor expansion of Φ:

$$\Phi(P + t\Delta P) - (\Phi(P) + t\Psi(P)[\Delta P]) = o(t).$$

(44) applies Assumption 5, and we plug in $K = \Phi(P + t\Delta P)$. (45) applied the definition of directional derivative

$$\nabla f(P)[\Delta P] = \lim_{t \to 0^+} \frac{f(P + t\Delta P) - f(P)}{t}.$$

Now we bound the first term of (45). Note that, since $P + t\Delta P$ for $t > 0$ and $t \to 0^+$ belongs to the line segment from P to P^*. Since $\mathcal{S}$ is a convex set, we know that the line segment between to feasible points P^* and P is in $\mathcal{S}$. then

$$P + t\Delta P \in \left\{ P' \mid P' \in \mathcal{S}, \ \Phi(P + t\Delta P) = \Phi(P') \right\},$$

so that $f(P + t\Delta P)$ is no less than the minimum of the optimization problem (28),

$$\lim_{t \to 0^+} \frac{\min_{P' \in \mathcal{S}, \ \Phi(P + t\Delta P) = \Phi(P')} \ f(P') - f(P + t\Delta P)}{t} \leq 0.$$

$\nabla f(P)[\Delta P]$ is the directional derivative of $f(P)$ in the direction of $P^* - P$, for a convex function f, if P is not an optimizer, $\nabla f(P)[\Delta P]$ is upper bounded by $f(P^*) - f(P) = \mathcal{L}(K^*) - \mathcal{L}(K) < 0$. □

B Constants for continuous time LQR

Theorem 1 asks for two constants C_1, C_2. They are bounded differently for different examples. As an instance, we will calculate the constants for continuous time LQR, quoted from [3, Appendix B]. First $P \succ 0$, so we replace singular value by eigenvalue with P,

$$\begin{aligned}C_1 &= (2\max\{\|L - L^*\|_F \lambda_{\min}^{-1}(P), \|P - P^*\|_F \lambda_{\min}^{-2}(P)\sigma_{\max}(L)\})^{-1},\\ C_2 &= (2\max\{\lambda_{\min}^{-1}(P), \lambda_{\min}^{-2}(P)\sigma_{\max}(L)\})^{-1}.\end{aligned}$$

We need upper bounds for P, L and a lower bound for $\lambda_{\min}(P)$ to guarantee C_1, C_2 being finite. We will show the bounds within the sublevel set that $\{K : \mathcal{L}(K) \le a\}$. Since we can randomly initialize a feasible K_0 and run (projected) gradient descent method with respect to K, if $\mathcal{L}(K)$ is gradient dominant, it is reasonable to assume that during all iterations of the optimization algorithm, the function value is always upper bounded by $\mathcal{L}(K_0)$, or some values not too larger than $\mathcal{L}(K_0)$. So our derivation with a finite sublevel set is reasonable.

Suppose the matrices $Q, R \succ 0$, and we consider the sublevel set when $\mathcal{L}(K) \le a$. The sublevel set gives $\mathbf{Tr}(QP) + \mathbf{Tr}(LP^{-1}L^\top R) \le a$, so

$$\begin{aligned}\lambda_{\min}(R)\lambda_{\max}^{-1}(P)\|L\|_F^2 &\le \lambda_{\min}(R)\|LP^{-1/2}\|_F^2\\ &\le \mathbf{Tr}(LP^{-1}L^\top R)\\ &\le \mathbf{Tr}(QP) + \mathbf{Tr}(LP^{-1}L^\top R) \le a.\end{aligned}$$

So $\|L\|_F \le a(\lambda_{\max}(P)\lambda_{\min}^{-1}(R))^{1/2}$, and we know from [3, eq(34)] $\mathbf{Tr}(P) \le a\lambda_{\min}^{-1}(Q)$. So we can bound P, L

$$\begin{aligned}\mathbf{Tr}(P) &\le a\lambda_{\min}^{-1}(Q),\\ \|L\|_F &\le a(\lambda_{\min}(Q)\lambda_{\min}(R))^{-1/2}.\end{aligned}$$

Define

$$\nu = \frac{\lambda_{\min}^2(\Sigma)}{4}\left(\sigma_{\max}(A)\lambda_{\min}^{-1/2}(Q) + \sigma_{\max}(B)\lambda_{\min}^{-1/2}(R)\right)^{-2}$$

[43, eq(38,40)] suggests $\lambda_{\min}(P) \ge \nu/a$. In summary, we upper bounded L, and upper and lower bounded P in the sublevel set $\mathcal{L}(K) \le a$, and those bounds are also true for L^*, P^*. We can complete the calculation by inserting the bounds into C_1.

$$\begin{aligned}C_1 &= (2\max\{\|L - L^*\|_F \lambda_{\min}^{-1}(P), \|P - P^*\|_F \lambda_{\min}^{-2}(P)\sigma_{\max}(L)\})^{-1}\\ &\ge \frac{\nu\lambda_{\min}^{1/2}(Q)\lambda_{\min}^{1/2}(R)}{4a^4}\cdot\min\left\{a^2,\ \nu\lambda_{\min}(Q)\right\}.\end{aligned}$$

C_2 is calculated similarly with upper bound on P, L, P^{-1}.

$$\begin{aligned}C_2 &= (2\max\{\lambda_{\min}^{-1}(P), \lambda_{\min}^{-2}(P)\sigma_{\max}(L)\})^{-1}\\ &\ge \frac{\nu}{2a^3}\min\left\{a^2,\ \nu\lambda_{\min}^{1/2}(Q)\lambda_{\min}^{1/2}(R)\right\}.\end{aligned}$$

B.1 Strongly convex parameter of continuous time LQR

In our previous convex formulation of continuous time LQR (8), we translate the objective function as a linear function in the new variables L, P, Z. The problem (8) can be slightly reformulated as

$$\min_{L,P} \; f(L,P) := \mathbf{Tr}(QP) + \mathbf{Tr}(LP^{-1}L^\top R), \tag{46a}$$

$$\text{s.t. } \mathcal{A}(P) + \mathcal{B}(L) + \Sigma = 0, \; P \succ 0. \tag{46b}$$

Compared with (8), (46) does not contain the variable Z. Below, we will prove that the new objective function $f(L,P)$, restricted within the feasible set, is a strongly convex function, which is not the case for the linear objective (8). In Theorem 1, there is another result with strongly convex f and the gradient domminance parameter depends on the strongly convex parameter μ. We also calculate μ of $f(L,P)$ below.

Lemma 2. *Define a sublevel set of of f at level a, consisting of all L, P such that $f(L,P) \le a$. Define*

$$\nu = \frac{\lambda_{\min}^2(\Sigma)}{4}\left(\sigma_{\max}(A)\lambda_{\min}^{-1/2}(Q) + \sigma_{\max}(B)\lambda_{\min}^{-1/2}(R)\right)^{-2}, \; \eta = \|\mathcal{B}\|\left(\nu^{1/2}\lambda_{\min}(\Sigma)\lambda_{\min}(Q)\lambda_{\min}^{1/2}(R)\right)^{-1},$$

$$\mu_0 = \frac{2\lambda_{\min}(Q)\lambda_{\min}(R)}{a(1+a^2\eta)^2}, \; \mu \ge (\|\mathcal{A}^{-1}\circ\mathcal{B}\| + 1)^{-1}\mu_0.$$

The function $f(L,P)$ restricted within the feasible sublevel set (46) *is μ strongly convex.*

Proof. Denote $\mathcal{A}^{-1}$ as the inverse of $\mathcal{A}$, a linear operator such that $\mathcal{A}^{-1}(\mathcal{A}(P)) = P$. [3, Proposition 1] concludes that the following function $h(\cdot)$ is μ_0 strongly convex.

$$h(L) = f(L, -\mathcal{A}^{-1}(\mathcal{B}(L) + \Sigma)). \tag{47}$$

Define a perturbation direction $(\tilde{L}, \tilde{P})$ such that $(L + \tilde{L}, P + \tilde{P})$ is feasible. Any feasible perturbation at the point L, P will satisfy $\mathcal{A}(\tilde{P}) + \mathcal{B}(\tilde{L}) = 0$, so $\tilde{P} = -\mathcal{A}^{-1}(\mathcal{B}(\tilde{L}))$.

Let the strongly convex parameter of f in the feasible directions be μ, we will show its connection with μ_0.

Let L be perturbed by $\tilde{L}$. Apply chain rule to (47),

$$\nabla^2 h(L)[\tilde{L}, \tilde{L}] = \nabla^2 f(L,P)[(\tilde{L}, -\mathcal{A}^{-1}(\mathcal{B}(\tilde{L}))), (\tilde{L}, -\mathcal{A}^{-1}(\mathcal{B}(\tilde{L})))], \tag{48}$$

Here $\nabla^2 h(L)[\tilde{L}, \tilde{L}]$ is the Hessian operator of h at L acting on $\tilde{L}, \tilde{L}$, which equals $\langle \tilde{L}, \nabla^2 h(L)\tilde{L}\rangle$. The right hand side is defined similarly. Due to the strong convexity of h,

$$\nabla^2 h(L)[\tilde{L}, \tilde{L}] \ge \frac{\mu_0\|\tilde{L}\|_F^2}{2}. \tag{49}$$

We perturb f at (L,P) in direction $(\tilde{L}, \tilde{P}) = (\tilde{L}, -\mathcal{A}^{-1}(\mathcal{B}(\tilde{L})))$. The strongly convex parameter of f in feasible directions is defined as the positive number μ such that

$$\nabla^2 f(L,P)[(\tilde{L}, \tilde{P}), (\tilde{L}, \tilde{P})] \ge \frac{\mu(\|\tilde{P}\|_F^2 + \|\tilde{L}\|_F^2)}{2}$$

for all $(\tilde{L}, \tilde{P})$ such that $\tilde{P} = -\mathcal{A}^{-1}(\mathcal{B}(\tilde{L}))$. The directional Hessian is

$$\nabla^2 f(L,P)[(\tilde{L},\tilde{P}),(\tilde{L},\tilde{P})] = \nabla^2 f(L,P)[(\tilde{L}, -\mathcal{A}^{-1}(\mathcal{B}(\tilde{L}))),(\tilde{L}, -\mathcal{A}^{-1}(\mathcal{B}(\tilde{L})))]. \tag{50}$$

(50) equals (48). So that we apply (49),

$$\begin{aligned} \nabla^2 f(L,P)[(\tilde{L},\tilde{P}),(\tilde{L},\tilde{P})] &\geq \frac{\mu_0 \|\tilde{L}\|_F^2}{2} \\ = \frac{\|\tilde{P}\|_F^2 + \|\tilde{L}\|_F^2}{2} &\cdot \frac{\mu_0\|\tilde{L}\|_F^2}{\|\tilde{P}\|_F^2 + \|\tilde{L}\|_F^2} \\ = \frac{\|\tilde{P}\|_F^2 + \|\tilde{L}\|_F^2}{2} &\cdot \frac{\mu_0\|\tilde{L}\|_F^2}{\|\mathcal{A}^{-1}(\mathcal{B}(\tilde{L}))\|_F^2 + \|\tilde{L}\|_F^2}. \end{aligned}$$

So

$$\mu \geq (\|\mathcal{A}^{-1} \circ \mathcal{B}\| + 1)^{-1}\mu_0.$$

□

C Checking the assumptions for examples

C.1 Markov jump linear system

Example 1. *(Assumptions 1,2) We study the system*

$$x(t+1) = A_{w(t)}x(t) + B_{w(t)}u(t),\ w(t) \in [N].$$

The transition model is

$$\mathbf{Pr}(w(t+1) = j | w(t) = i) = \rho_{ij} \in [0,1],\ \forall t \geq 0.$$

Let $\mathbf{Pr}(w(0) = i) = p_i$, $K = [K_1, ..., K_N]$*. Define the cost as*

$$\mathcal{L}(K) = \mathbf{E}_{w,x_0} \sum_{t=0}^{\infty} x(t)^\top Q x(t) + u(t)^\top R u(t),\ u(t) = K_{w(t)}x(t),\ \mathbf{Pr}(w(0) = i) = p_i.$$

Let the convex formulation be

$$\min\ \mathbf{Tr}(QX_0) + \mathbf{Tr}(Z_0 R), \tag{51a}$$

$$s.t.\ X_0 = \sum_{i=1}^N X_i,\ Z_0 = \sum_{i=1}^N Z_i,\ \begin{bmatrix} Z_i & L_i \\ L_i^\top & X_i \end{bmatrix} \succeq 0, \tag{51b}$$

$$X_i - p_i\Sigma = \sum_{j=1}^N U_{ji},\ \begin{bmatrix} \rho_{ji}^{-1}U_{ji} & A_jX_j + B_jL_j \\ (A_jX_j + B_jL_j)^\top & X_j \end{bmatrix} \succeq 0,\ \forall i,j \in [N]. \tag{51c}$$

Then the pair of problems satisfy Assumptions 1,2.

Proof. We start from the Grammian matrices below. Let $Y_i(t) = \mathbf{E}(x(t)x(t)^\top \mathbf{1}_{w(t)=i})$, and $X_i = \sum_{t=0}^\infty Y_i(t)$. Then [38] suggests

$$Y_i(t+1) = \sum_{j=1}^N \rho_{ji}(A_j + B_jK_j)Y_j(t)(A_j + B_jK_j)^\top.$$

Then we can sum over the equation over time t,

$$\begin{aligned}
&\sum_{j=1}^N \rho_{ji}(A_j + B_jK_j)(\sum_{t=0}^\infty Y_j(t))(A_j + B_jK_j)^\top \\
=& \sum_{t=0}^\infty \sum_{j=1}^N \rho_{ji}(A_j + B_jK_j)Y_j(t)(A_j + B_jK_j)^\top \\
=& \sum_{t=0}^\infty Y_i(t+1) = \sum_{t=1}^\infty Y_i(t) \\
=& \sum_{t=0}^\infty Y_i(t) - Y_i(0)
\end{aligned}$$

So that

$$\sum_{j=1}^N \rho_{ji}(A_j + B_jK_j)X_j(A_j + B_jK_j)^\top = X_i - Y_i(0).$$

Let $L_i = K_iX_i$. We will relax the recursion as

$$\sum_{j=1}^N \rho_{ji}(A_jX_j + B_jL_j)X_j^{-1}(A_jX_j + B_jL_j)^\top \preceq X_i - Y_i(0). \tag{52}$$

In our setting $\mathbf{E}(x(0)x(0)^\top) = \Sigma$ so that $Y_i(0) = p_i\Sigma$.

Next, we will show that, if we solve the problem (51) with the extra constraints $K_i = L_iX_i^{-1}$, then the function value is equal to the LQ cost of the system with controllers K_i's.

First, if we minimize over Z_i's, then we have $Z_i = L_iX_iL_i^\top$. Moreover, the constraints (51c) are equivalent to the relaxation (52). Suppose the equal signs are achieved in (52), then X_i's will be the Grammian of the system $\sum_{t=0}^\infty \mathbf{E}(x(t)x(t)^\top \mathbf{1}_{w(t)=i})$ and hence the function value is equal to the LQ cost [44, §4.4.2, Prop. 4.8]. Now, it remains to show that, if not all (52) (with enumerating different j's) achieve equal signs, then the function value will only increase and not be optimal.

We define N matrices $W_1, ..., W_N$, such that $W_i \succeq Y_i(0) = p_i\Sigma$, and

$$\sum_{j=1}^N \rho_{ji}(A_j + B_jK_j)X_j(A_j + B_jK_j)^\top = X_i - W_i.$$

This corresponds to the Markov jump system

$$\tilde{x}(t+1) = A_{w(t)}\tilde{x}(t) + B_{w(t)}u(t),\ w(t) \in [N].$$

with the same parameters, transition probability matrix, controllers and a different initial condition

$$\mathbf{E}(\tilde{x}(t)\tilde{x}(t)^\top \mathbf{1}_{w(t)=i}) = W_i \succeq p_i\Sigma = \mathbf{E}(x(t)x(t)^\top \mathbf{1}_{w(t)=i}). \tag{53}$$

Let $\tilde{Y}_i(t) = \mathbf{E}(\tilde{x}(t)\tilde{x}(t)^\top \mathbf{1}_{w(t)=i})$ (so that $\tilde{Y}_i(0) = W_i$), and let $\tilde{X}_i = \sum_{t=0}^\infty \tilde{Y}_i(t)$. We will show that $\tilde{Y}_i(t) \succeq Y_i(t)$ for all $i = 1, ..., N$ and all $t \geq 0$.

We use induction over t. When $t = 0$, we assumed in (53) that $\tilde{Y}_i(0) \succeq Y_i(0)$ hold for all $i \in [N]$. And we have the recursions

$$\tilde{Y}_i(t+1) = \sum_{j=1}^N \rho_{ji}(A_j + B_jK_j)\tilde{Y}_j(t)(A_j + B_jK_j)^\top,$$

$$Y_i(t+1) = \sum_{j=1}^N \rho_{ji}(A_j + B_jK_j)Y_j(t)(A_j + B_jK_j)^\top.$$

If $\tilde{Y}_i(t) \succeq Y_i(t)$ for a certain $t \geq 0$ and for all $i \in [N]$, then the recursion implies that $\tilde{Y}_i(t+1) \succeq Y_i(t+1)$ for all $i \in [N]$. We sum over t and get $\tilde{X}_i \succeq X_i$, so that the objective function with $\tilde{X}_i$'s is larger than with X_i's unless $\tilde{X}_i = X_i$ for all $i \in [N]$.

As a result, the optimization problem (51) with the extra constraints $K_i = L_iX_i^{-1}$ achieves minimum when $Z_i = L_iX_iL_i^\top$ and (52) achieves equality for all $i \in [N]$. This means all X_i's are the Grammians $\sum_{t=0}^\infty \mathbf{E}(x(t)x(t)^\top \mathbf{1}_{w(t)=i})$ of the system, so that the objective function value is equal to LQ cost.

□

C.2 Minimizing $\mathcal{L}_2$ gain

Example 2. *(Assumptions 1,2) We consider minimizing the $\mathcal{L}_2$ gain of a closed loop system. The input output system is*

$$\dot{x} = Ax + Bu + B_ww, \; y = Cx + Du \tag{54}$$

and we use the state feedback controller $u = Kx$, and let

$$\mathcal{L}(K) := (\sup_{\|w\|_2=1} \|y\|_2)^2.$$

If we minimize the function $\mathcal{L}(K)$, this problem can be reformulated as

$$\begin{aligned} &\min_{L,P,\gamma} \; f(L,P,\gamma) := \gamma \\ &s.t. \begin{bmatrix} AP + PA^\top + BL + L^\top B^\top + B_wB_w^\top & (CP+DL)^\top \\ CP + DL & -\gamma I \end{bmatrix} \preceq 0. \end{aligned}$$

And $K^* = L^*P^{*-1}$. This pair of problems satisfy Assumptions 1,2.

Proof. We will check Assumption 2, which means checking

$$\mathcal{L}(K) \overset{?}{=} \min_{L,P,\gamma} \gamma \tag{55a}$$

$$\text{s.t.} \begin{bmatrix} AP + PA^\top + BL + L^\top B^\top + B_wB_w^\top & (CP+DL)^\top \\ CP + DL & -\gamma I \end{bmatrix} \preceq 0, \; LP^{-1} = K. \tag{55b}$$

Note that, the intermediate step in [36, Sec 7.5.1] is

$$\mathcal{L}(K) = \min_{P,\gamma} \gamma, \quad \text{s.t.} \tag{56a}$$

$$\begin{bmatrix} (A+BK)P + P(A+BK)^\top + B_w B_w^\top & P^\top (C+DK)^\top \\ (C+DK)P & -\gamma I \end{bmatrix} \preceq 0. \tag{56b}$$

Denote the optimizer of (55) by $\hat{L}, \hat{P}, \hat{\gamma}$, and the optimizer of (56) by $\check{P}, \check{\gamma}$.

Note $\hat{\gamma} \le \check{\gamma}$ because $(\gamma, L, P) = (\check{\gamma}, K\check{P}, \check{P})$ is feasible in (55). If (55) is not true (the equal sign cannot be satisfied), then $\hat{\gamma} < \check{\gamma}$, we can replace $\check{P}, \check{\gamma}$ with $\hat{P}, \hat{\gamma}$ in (56) and it's still feasible due to the feasibility in (55). Thus the optimality condition of $\check{P}, \check{\gamma}$ in (56) is violated, which contradicts the assumption that (55) is not true. Then we claim that (55) is true. The dissipativity uses the same change of variable and we omit the proof in [36]. □

D System level synthesis with infinite horizon

We studied the landscape of the optimal control problem where the variables are matrices (which are finite dimensional), and SLS for finite horizon problem was an example. Generally, SLS also works with the infinite horizon problem. In this regime, the variables are *transfer functions* and they are infinite dimensional. In practice, when the problem is made convex, one can parameterize the transfer function (say as finite impulse response) and minimize the cost with respect to the finite dimensional parameters. However, Theorem 1 does not apply to the infinite dimensional optimization problems, and it is not obvious that the finite dimensional parameterization satisfies the assumptions for our main theorem. We review the infinite horizon SLS here. A future direction is to judge whether the Lojasiewicz inequality holds in the space of transfer function or its parameterized form, and how to analyze it using SLS.

Example 3. *(System level synthesis with infinite horizon [39]) Suppose one has a discrete time dynamical system with*

$$x(t+1) = Ax(t) + Bu(t) + w(t).$$

One can apply a dynamic controller $K(z)$. The goal is to find the optimial controller which minimizes the LQR cost where $u(z) = K(z)x(z)$

$$\mathcal{L}(K) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} x(t)^\top Q x(t) + u(t)^\top R u(t).$$

Suppose x_0, w_t are i.i.d. from $\mathcal{N}(0, \Sigma)$. The SLS defines two transfer functions $\Phi_X(z), \Phi_U(z)$, and solve the following convex optimization problem

$$\begin{aligned} &\min_{\Phi_X(z), \Phi_U(z)} \left\| \begin{bmatrix} Q^{1/2}\Phi_X(z) \\ R^{1/2}\Phi_U(z) \end{bmatrix} \Sigma^{1/2} \right\|_{\mathcal{H}_2}, \\ &s.t. \ \begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \Phi_X(z) \\ \Phi_U(z) \end{bmatrix} = I, \\ &\Phi_X(z), \Phi_U(z) \in \frac{1}{z}\mathcal{RH}_\infty. \end{aligned}$$

Let the optimizer be $\Phi_U^*(z), \Phi_X^*(z)$. The optimal controller is $K^*(z) = \Phi_U^*(z)(\Phi_X^*(z))^{-1}$.

E Conditions of convexifiable nonconvex cost

We consider the pair of problems in Theorem 1, and ask the question: what property of the nonconvex cost function $\mathcal{L}(K)$ allows us to reformulate the problem (9) as a *convex* optimization problem (10)? In this section we propose the following lemma.

Lemma 3. *Suppose Assumptions 1, 3 hold, and $\mathcal{L}(LP^{-1})$ as a function of L, P is differentiable.We define the notation $\nabla^2_{L,P}\mathcal{L}(LP^{-1})[\Gamma_1, \Gamma_2]$ as in (57). If $\nabla^2_{L,P}\mathcal{L}(LP^{-1})[\Gamma_1, \Gamma_2] > 0$ for all $(L, P) \in \mathcal{S}$ and all (Γ_1, Γ_2) such that $\mathcal{A}(\Gamma_2) + \mathcal{B}(\Gamma_1) = 0$, then we can define a convex function $f(L, P)$ so that Assumption 1 holds.*

For the convex formulation with the above lemma, we can apply Theorem 1 so that all stationary points of $\mathcal{L}(K)$ are global minimum.

Proof. Suppose we observe the simple version (11). We know from Assumption 3 that, $f(L, P) = \mathcal{L}(K) = \mathcal{L}(LP^{-1})$ is convex in L, P. We take the Hessian and ask for

$$\nabla \begin{bmatrix} \nabla\mathcal{L}(LP^{-1})P^{-1} \\ -P^{-1}L^\top\nabla\mathcal{L}(LP^{-1})P^{-1} \end{bmatrix} \succ 0.$$

Note that this is a tensor and it is positive definite. For simplicity, we analyze the directional Hessian as the following. We expand the left hand side of the inequality above and define $\nabla^2_{L,P}\mathcal{L}(LP^{-1})[\Gamma_1, \Gamma_2]$ as

$$\begin{aligned} &\nabla^2_{L,P}\mathcal{L}(LP^{-1})[\Gamma_1, \Gamma_2] \\ :=\ &\nabla^2\mathcal{L}(LP^{-1})[\Gamma_1 G^{-2}, \Gamma_1] - 2\nabla^2\mathcal{L}(LP^{-1})[\Gamma_1, LP^{-3}\Gamma_2] \\ &- 2\langle\Gamma_1, \nabla\mathcal{L}(LP^{-1})P^{-1}\Gamma_2P^{-1}\rangle + 2\langle\Gamma_2, LP^{-1}\Gamma_2P^{-1}\nabla\mathcal{L}(LP^{-1})P^{-1}\rangle \\ &+ \nabla^2\mathcal{L}(LP^{-1})[LP^{-2}\Gamma_2, LP^{-2}\Gamma_2]. \end{aligned} \tag{57}$$

This is the directional Hessian of $\mathcal{L}$ with respect to (L, P) in direction (Γ_1, Γ_2). Thus, if $\nabla^2_{L,P}\mathcal{L}(LP^{-1})[\Gamma_1, \Gamma_2] > 0$ for all $(L, P) \in \mathcal{S}$ and all (Γ_1, Γ_2) such that $\mathcal{A}(\Gamma_2) + \mathcal{B}(\Gamma_1) = 0$ (which is a condition on nonconvex cost $\mathcal{L}$), we know that $f(L, P)$ is convex in L, P and the convex formulation can be made. □

F Experiments

Despite the convexity of the control problems, one can still run gradient descent on policy K. Figure 2 is a simple simulation for continuous time LQR. We denote K^* as the solution from Riccati equations, which is reviewed below as a classical method for finding the optimal controller. One can see that $\|K - K^*\|_F$ converges to 0. This motivates the study of the global convergence guarantee of the policy gradient method.

And moreover, for other control problems that can be solved by a similar parameterization while no papers have given the global convergence guarantee for the first order method, we propose the first proof of their global convergence. This can be verified numerically. For example, we simulate in Fig. 3 the zeroth order algorithm for minimizing $\mathcal{L}_2$ gain (Sec. 4.3). The zeroth order algorithm runs as the following: at time t, we randomly generate N matrices (entries are i.i.d. standard normal)

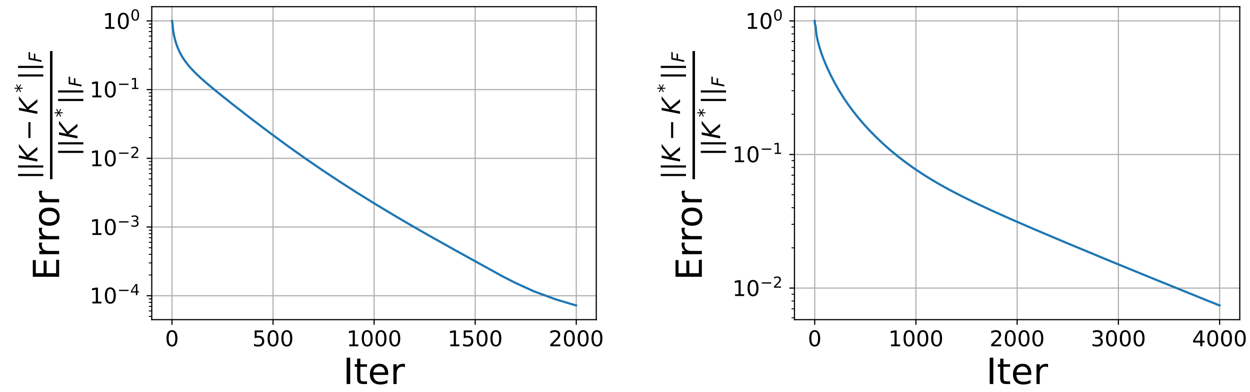

Figure 2: (a) Continuous time LQR, gradient descent with respect to K. $n = 8$, $p = 4$. A is block diagonal with 2×2 blocks, the first block is $\begin{bmatrix} -2 & 2 \\ 0 & -2 \end{bmatrix}$ and the other three blocks are $\begin{bmatrix} -6 & 4 \\ -8 & 2 \end{bmatrix}$. We initialize $K_0 = 0$ to make K_0 a stabilizing controller. The error $\|K - K^*\|_F \to 0$.
(b) Discrete time LQR, gradient descent with respect to K. $n = 8$, $p = 4$. A is block diagonal with 2×2 blocks, the first block is $\begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}/2$ and the other three blocks are $\begin{bmatrix} 3 & -2 \\ 4 & -1 \end{bmatrix}/2$. We initialize $K_0 = 0$ to make K_0 a stabilizing controller. The error $\|K - K^*\|_F \to 0$.

$\Delta_{t,i}K \in \mathbb{R}^{p \times n}$, $i = 1, ..., n$, and let $\epsilon > 0$ be a small number. Then

$$K_{t+1} \leftarrow K_t - \frac{\eta_t}{N} \sum_{i=1}^{N} (\mathcal{L}(K_t + \epsilon \Delta_{t,i}K) - \mathcal{L}(K_t))\Delta_{t,i}K.$$

The zeroth order method [45, 46] is convenient in model free setting, or model based setting when we cannot easily obtain a closed form of gradient. We can see it converges to the global optimum.

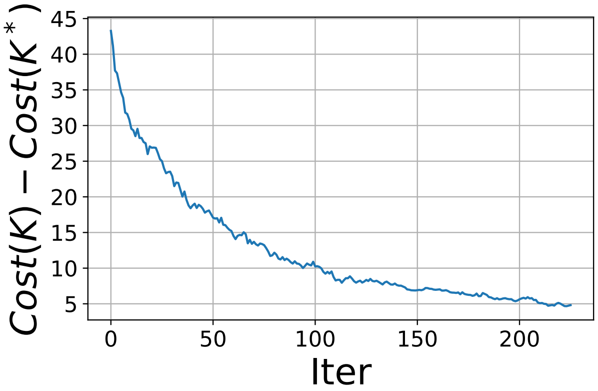

Figure 3: Minimizing $\mathcal{L}_2$ gain (See Sec. 4.3) by gradient descent with respect to K. $n = 8$, A is block diagonal with 2×2 blocks, the first block is $\begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$ and the other three blocks are $\begin{bmatrix} -3 & 2 \\ -4 & 1 \end{bmatrix}$. We initialize $K_0 = 0$ to make K_0 a stabilizing controller. The suboptimality of the cost goes to 0.