Proceedings of Machine Learning Research vol vvv:1–12, 2025

Extended Convex Lifting for Policy Optimization of Optimal and Robust Control

Yang Zheng ZHENGY@UCSD.EDU
Department of Electrical and Computer Engineering, University of California San Diego
Chih-Fan Pai CPAI@UCSD.EDU
Department of Electrical and Computer Engineering, University of California San Diego
Yujie Tang YUJIETANG@PKU.EDU.CN
Department of Industrial Engineering & Management, Peking University, China

Abstract

Many optimal and robust control problems are nonconvex and potentially nonsmooth in their policy optimization forms. In this paper, we introduce the Extended Convex Lifting (ECL) framework, which reveals *hidden convexity* in classical optimal and robust control problems from a modern optimization perspective. Our ECL framework offers a bridge between nonconvex policy optimization and convex reformulations. Despite non-convexity and non-smoothness, the existence of an ECL for policy optimization not only reveals that the policy optimization problem is equivalent to a convex problem, but also certifies a class of first-order *non-degenerate* stationary points to be globally optimal. We also show that this ECL framework can cover many benchmark control problems, including LQR, state-feedback and output-feedback $\mathcal{H}_\infty$ robust control. We believe that the ECL framework will be of independent interest for analyzing nonconvex problems beyond control.

1. Introduction

The classical optimal and robust control problems, including linear quadratic regulator (LQR), linear quadratic Gaussian (LQG) control, and $\mathcal{H}_\infty$ control, have been extensively studied (Kalman, 1963; Levine and Athans, 1970). It is well-known that almost all these problems are nonconvex in the space of controller (i.e., policy) parameters. Nevertheless, classical techniques based on controller re-parameterizations (Scherer and Weiland, 2015; Boyd et al., 1994) or Riccati equations (Zhou et al., 1996) have been established to characterize optimal or suboptimal controllers. These classical techniques do not optimize over the policy parameters directly, and often require an explicit system model. On the other hand, the optimization landscapes of optimal and robust control problems can also offer fruitful results, in which we view the control costs as functions of the policy parameters and study their analytical and geometrical properties (Lewis, 2007; Hu et al., 2023; Talebi et al., 2024). This perspective is naturally amenable for data-driven design paradigms such as reinforcement learning and learning-based control (Recht, 2019).

However, this policy optimization perspective for control generally leads to nonconvex and potentially nonsmooth problems. For example, the set of feedback gains K that stabilize the system $\dot{x} = Ax + Bu$ via $u = Kx$ is already nonconvex; if we consider output-feedback controller synthesis such as LQG, then the parameterized set of dynamic policies can even be disconnected (Tang et al., 2023). Furthermore, the LQG cost function may have spurious stationary points (Zheng et al., 2022), and there can be uncountably many globally optimal policies lying on a manifold induced by *similarity transformations* (Zheng et al., 2022; Tang et al., 2023; Kraisler and Mesbahi, 2024). In addition to non-convexity, non-smoothness may also arise when considering robust control problems. A typical performance measure for robust control is the $\mathcal{H}_\infty$ norm of certain closed-loop transfer function (Zhou et al., 1996), which is known to be both *nonconvex* and *nonsmooth* in the policy space (Apkarian and Noll, 2006; Lewis, 2007).

For nonconvex and nonsmooth optimization, it is generally very hard to derive theoretical guarantees for local search algorithms. On the other hand, a series of recent findings have revealed *benign nonconvex landscape properties* in benchmark control problems, including LQR (Fazel et al., 2018; Mohammadi et al., 2022; Fatkhullin and Polyak, 2021), risk-sensitive control (Zhang et al., 2021), LQG (Tang et al., 2023; Zheng et al., 2022; Duan et al., 2024), dynamic filtering (Umenberger et al., 2022; Zhang et al., 2023), $\mathcal{H}_\infty$ control (Hu and Zheng, 2022; Guo and Hu, 2022; Tang and Zheng, 2023), and distributed control (Furieri et al., 2020a). Many of these works leveraged the idea that the control problem under investigation admits suitable convex reformulations. However, these existing works are mostly on a case-by-case basis. Our work aims to provide a unified framework that explains the benign nonconvex landscape properties of these iconic control problems.

Our Contributions — Extended Convex Lifting (ECL)

We introduce a unified framework, called *Extended Convex Lifting* (ECL), to reveal *hidden convexity* in classical optimal and robust control problems from a modern optimization perspective. The core idea behind ECL stems from the existing results in control theory that, via a suitable *change of variables*, many optimal and robust control problems admit "convex reformulations" (Scherer and Weiland, 2015; Boyd et al., 1994). By fitting the change of variables into the ECL framework, we can analyze the nonconvex landscape of the corresponding policy optimization problem by exploiting its hidden convexity.

Figure 1 provides a schematic illustration of ECL. Specifically, we begin with the epigraph of the objective $J(\mathsf{K})$, and *lift* certain subset of its closure to a higher dimensional set $\mathcal{L}_{\text{lft}}$ by incorporating Lyapunov variables in the change of variables. Then, we construct a smooth bijection Φ that maps $\mathcal{L}_{\text{lft}}$ to some $\mathcal{F}_{\text{cvx}} \times \mathcal{G}_{\text{aux}}$ in which $\mathcal{F}_{\text{cvx}}$ is convex. This bijection essentially encodes the change of variables for the corresponding control problem. In many control problems, $\mathcal{F}_{\text{cvx}}$ can be represented by LMIs, while $\mathcal{G}_{\text{aux}}$ accounts for similarity transformations of output-feedback policies.

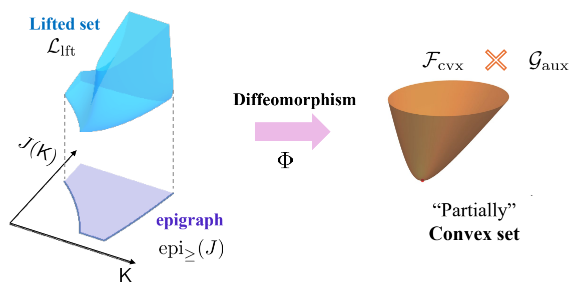

Figure 1: A schematic illustration of ECL.

Despite non-convexity and non-smoothness, the existence of an ECL not only shows that the policy optimization problem is equivalent to a convex problem (Theorem 2.1) but also identifies a class of *non-degenerate* stationary points to be globally optimal (Theorem 3.1). Our ECL framework covers many iconic control problems; many recent results on global optimality of (non-degenerate) stationary points, such as LQR (Fazel et al., 2018; Mohammadi et al., 2022), LQG (Tang et al., 2023), state-feedback $\mathcal{H}_\infty$ control (Guo and Hu, 2022), output-feedback $\mathcal{H}_\infty$ control (Tang and Zheng, 2023), are special cases once the corresponding ECL is constructed.

We point out that our ECL framework is more general than Tang et al. (2023); Umenberger et al. (2022); Guo and Hu (2022); Sun and Fazel (2021); Mohammadi et al. (2022), in the sense that it can directly handle both state-feedback and output-feedback policies, as well as smooth and nonsmooth cost functions. More importantly, our ECL framework naturally classifies *degenerate* and *non-degenerate* policies, which reflects the subtleties between strict and non-strict LMIs in control. Due to the page limit, we omit the proofs of most of the results in this paper. Detailed proofs, relevant discussions, and examples can be found in our extended reports (Zheng et al., 2023, 2024).

2. Extended Convex Lifting (ECL) for Benign Non-convexity

2.1. A Motivating Example

We first provide a motivating example. Consider an LTI system $\dot{x}(t) = Ax(t)+Bu(t)+w(t)$ with $A = \begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, where $w(t)$ is a white Gaussian noise with $\mathbb{E}[w(t)w(\tau)] = 4\delta(t-\tau)I$. We aim to design a state-feedback policy $u(t) = Kx(t)$ with $K = \begin{bmatrix} k_1 & k_2 \end{bmatrix} \in \mathbb{R}^{1\times 2}$ to minimize the LQR cost $\lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\int_0^T \left(x_1(t)^2+x_2(t)^2+u(t)^2\right)dt\right]$. Standard calculation shows the cost equals to

$$J(K) = \frac{1-2k_2+3k_2^2-2k_2^3-2k_1^2k_2}{k_2^2-1}, \qquad k_1 \in \mathbb{R},\ k_2 < -1. \tag{1}$$

Following the classical change of variables $Y = KX$ where X solves the Lyapunov equation $(A+BK)X + X(A+BK)^{\mathsf{T}} + 4I_2 = 0$, we obtain the nonlinear mapping

$$Y = \begin{bmatrix} y_1 & y_2 \end{bmatrix} = g(K) := \begin{bmatrix} \frac{k_1}{1-k_2} & \frac{2k_2-k_1^2-2k_2^2}{k_2^2-1} \end{bmatrix}, \qquad \forall\, k_1 \in \mathbb{R},\ k_2 < -1.$$

One can check that g is invertible, and the cost function after applying this mapping becomes

$$h(Y) = J(g^{-1}(Y)) = -y_2-1+Y\begin{bmatrix} 1 & y_1 \\ y_1 & -y_2-2 \end{bmatrix}^{-1} Y^{\mathsf{T}}, \quad \text{for } Y \text{ such that } \begin{bmatrix} 1 & y_1 \\ y_1 & -y_2-2 \end{bmatrix} \succ 0.$$

By standard techniques in convex analysis, one can show that $h(Y)$ is convex. Thanks to the smooth bijection g, minimizing $J(K)$ is now equivalent to minimizing the convex function $h(Y)$, and any stationary point of $J(K)$ is globally optimal.

This motivating example demonstrates how one can utilize a proper change of variables to certify the global optimality of stationary points via convex analysis for policy optimization. In the next subsection, we propose a general framework for nonconvex policy optimization that can cover a much wider range of benchmark problems with convex reformulations.

2.2. The Extended Convex Lifting (ECL) Framework

Consider a policy optimization problem where the objective is $f : \mathcal{D} \to \mathbb{R}$, with $\mathcal{D} \subseteq \mathbb{R}^d$ being its domain. To study the landscape of f, we resort to its strict and non-strict epigraphs defined by

$$\operatorname{epi}_>(f) := \{(x,\gamma) \in \mathcal{D}\times\mathbb{R} \mid \gamma > f(x)\}, \qquad \operatorname{epi}_\geq(f) := \{(x,\gamma) \in \mathcal{D}\times\mathbb{R} \mid \gamma \geq f(x)\}.$$

Definition 2.1 (Extended Convex Lifting) *Suppose $f : \mathcal{D} \to \mathbb{R}$ is continuous. We say that the tuple $(\mathcal{L}_{\text{lft}}, \mathcal{F}_{\text{cvx}}, \mathcal{G}_{\text{aux}}, \Phi)$ is an* ECL *of f, if the following conditions hold:*

1. *$\mathcal{L}_{\text{lft}} \subseteq \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^{d_\xi}$ is a lifted set with an extra variable $\xi \in \mathbb{R}^{d_\xi}$, such that the canonical projection of $\mathcal{L}_{\text{lft}}$ onto the first $d+1$ coordinates, given by $\pi_{x,\gamma}(\mathcal{L}_{\text{lft}}) = \{(x,\gamma) : \exists \xi \in \mathbb{R}^{d_\xi} \text{ s.t. } (x,\gamma,\xi) \in \mathcal{L}_{\text{lft}}\}$, satisfies*
$$\operatorname{epi}_>(f) \subseteq \pi_{x,\gamma}(\mathcal{L}_{\text{lft}}) \subseteq \operatorname{cl}\operatorname{epi}_\geq(f). \tag{2a}$$
2. *$\mathcal{F}_{\text{cvx}} \subseteq \mathbb{R}\times\mathbb{R}^{d_1}$ is a convex set, $\mathcal{G}_{\text{aux}} \subseteq \mathbb{R}^{d_2}$ is an auxiliary set, and Φ is a C^2 diffeomorphism from $\mathcal{L}_{\text{lft}}$ to $\mathcal{F}_{\text{cvx}} \times \mathcal{G}_{\text{aux}}$.*[1]

1. We allow $d_2 = 0$, in which case we adopt the convention $\mathcal{G}_{\text{aux}} = \{0\}$ and identify $\mathcal{F}_{\text{cvx}} \times \{0\}$ with $\mathcal{F}_{\text{cvx}}$.

3. *For any* $(x, \gamma, \xi) \in \mathcal{L}_{\text{lft}}$, *we have*

$$\Phi(x, \gamma, \xi) = (\gamma, \zeta_1, \zeta_2) \quad \text{and} \quad (\gamma, \zeta_1) \in \mathcal{F}_{\text{cvx}} \tag{2b}$$

for some $\zeta_1 \in \mathbb{R}^{d_1}$ *and* $\zeta_2 \in \mathcal{G}_{\text{aux}}$ *(i.e., the mapping* Φ *directly outputs* γ *in the first component).*

The notion of ECL gives extensive flexibility by 1) adding an extra variable to *lift* the epigraph to a higher dimension, 2) relaxing the projection of the lifted set to sit between the strict epigraph and the closure of the non-strict epigraph (*chain of inclusion*), and 3) introducing an auxiliary set $\mathcal{G}_{\text{aux}}$ to extend the convex image under a diffeomorphism. The extra variable ξ often corresponds to Lyapunov variables, and $\mathcal{G}_{\text{aux}}$ is often related to similarity transformations of dynamic policies.

The interested reader may wonder why we need such a peculiar chain of inclusion in (2a). A simpler and more straightforward requirement for the lifting process might be

$$\pi_{x,\gamma}(\mathcal{L}_{\text{lft}}) = \text{epi}_{\geq}(f). \tag{3}$$

Evidently, (2a) includes (3) as a special case. In Section 4, we will present an ECL for output-feedback $\mathcal{H}_\infty$ control where the more general (2a) is necessary, which is largely due to the intricacy between strict and non-strict LMIs in the convex reformulations of control problems. This intricacy is important for global optimality, but has been less emphasized before since classical results often focused on suboptimal controller design; see (Zheng et al., 2023, 2024) for details. An immediate benefit of ECL is that we can reformulate the minimization of $f(x)$ over $x \in \mathcal{D}$ as a convex problem.

Theorem 2.1 *Let* $f : \mathcal{D} \to \mathbb{R}$ *be continuous and equipped with an* ECL $(\mathcal{L}_{\text{lft}}, \mathcal{F}_{\text{cvx}}, \mathcal{G}_{\text{aux}}, \Phi)$. *Then, we have* $\inf_{x \in \mathcal{D}} f(x) = \inf_{(\gamma,\zeta_1) \in \mathcal{F}_{\text{cvx}}} \gamma$.

Thanks to the diffeomorphism Φ, the proof is straightforward but requires careful reasoning about (non-)strict epigraphs; see Zheng et al., 2024, Theorem 3.1 for a detailed proof. In Theorem 2.1, the function f can be nonsmooth and nonconvex, but the existence of an ECL reveals its hidden convexity in the sense that optimizing $f(x)$ over $x \in \mathcal{D}$ is equivalent to a convex problem. Theorem 2.1 provides the rationale behind convex re-parameterizations of many control problems (Scherer and Weiland, 2015; Boyd et al., 1994). Note that we only guarantee an infimum instead of a minimum in Theorem 2.1 (the infimum may not always be achieved in $\mathcal{H}_\infty$ control).

3. Non-degenerate Policies and Global Optimality[2]

In addition to convex reformulation as shown in Theorem 2.1, the existence of an ECL can further reveal *global optimality of certain first-order stationary points* for the potentially nonconvex and nonsmooth function f. This allows us to optimize $f(x)$ by direct local search without knowing the particular form of the ECL, which is particularly important for learning-based model-free control.

Before proceeding, we re-emphasize that the chain of inclusion (2a) is critical to the construction of ECL in many control problems. This chain of inclusion allows existence of points $(x, f(x))$ that are not covered by $\pi_{x,\gamma}(\mathcal{L}_{\text{lft}})$, as well as members of $\pi_{x,\gamma}(\mathcal{L}_{\text{lft}})$ that are only accumulation points of $\text{epi}_{\geq}(f)$. We introduce the notion of *(non-)degeneracy* to characterize the former type of points.

Definition 3.1 (Non-degenerate points) *Let* $f : \mathcal{D} \to \mathbb{R}$ *be a continuous function equipped with an* ECL $(\mathcal{L}_{\text{lft}}, \mathcal{F}_{\text{cvx}}, \mathcal{G}_{\text{aux}}, \Phi)$. *A point* $x \in \mathcal{D}$ *is called non-degenerate if* $(x, f(x)) \in \pi_{x,\gamma}(\mathcal{L}_{\text{lft}})$, *otherwise degenerate. The set of non-degenerate points in* $\mathcal{D}$ *will be denoted by* $\mathcal{D}_{\text{nd}}$.

2. Some materials rely on the notion of *Clarke subdifferential*, which extends subdifferential to nonconvex nonsmooth functions. We refer the readers to Clarke (1990) for details, or Zheng et al. (2023, Appendix B) for a brief review.

In Section 4.1, we will show that all stabilizing state-feedback policies for LQR and $\mathcal{H}_\infty$ control are non-degenerate using standard ECL constructions. We will also see that for output-feedback $\mathcal{H}_\infty$ control, the set of non-degenerate policies defined in our previous work (Tang and Zheng, 2023) corresponds to the set of non-degenerate points per Definition 3.1 under the ECL framework.

We now present one main technical result of this paper, which provides global optimality certificates for stationary points that are non-degenerate.

Theorem 3.1 *Let $f : \mathcal{D} \to \mathbb{R}$ be a subdifferentially regular[3] function defined on an open domain $\mathcal{D} \subseteq \mathbb{R}^d$, and let $(\mathcal{L}_{\mathrm{lft}}, \mathcal{F}_{\mathrm{cvx}}, \mathcal{G}_{\mathrm{aux}}, \Phi)$ be an* ECL *of f. If $x^* \in \mathcal{D}_{\mathrm{nd}}$ is a Clarke stationary point, i.e., $0 \in \partial f(x^*)$, then x^* is a global minimizer of $f(x)$ over $\mathcal{D}$.*

The proof of Theorem 3.1 has a strong geometric intuition, but the details are technically involved, which are given in Zheng et al., 2024, Section 3.3. Theorem 3.1 guarantees that *non-degenerate stationarity implies global optimality* for any subdifferentially regular function with an ECL. Subdifferentially regular functions are a very large class of functions, covering all optimal and robust control problems discussed in Section 4.

We now provide a corollary considering the case when (3) holds for the ECL.

Corollary 3.1 *Let $f : \mathcal{D} \to \mathbb{R}$ be a subdifferentially regular function defined on an open domain $\mathcal{D} \subseteq \mathbb{R}^d$, and let $(\mathcal{L}_{\mathrm{lft}}, \mathcal{F}_{\mathrm{cvx}}, \mathcal{G}_{\mathrm{aux}}, \Phi)$ be an* ECL *of f. If (3) holds, then*

1. *All points $x \in \mathcal{D}$ are non-degenerate.*
2. *Any Clarke stationary point is a global minimizer of $f(x)$ over $x \in \mathcal{D}$.*

As mentioned before, many state-feedback and full-order output-feedback controller synthesis problems are nonconvex in their natural forms but admit "convex reformulations" in terms of LMIs using a suitable change of variables. We argue that our notion of ECL presents a unified treatment for many of these convex reformulations. In Section 4, we will present some ECL construction details for benchmark optimal and robust control problems. We point out that, despite the wide use of convex reformulations in control, exact constructions of ECL require special care, especially for output-feedback control problems (since strict vs. non-strict inequalities are quite subtle).

Remark 3.1 (Degenerate points and saddles) *By (2a), $\pi_{x,\gamma}(\mathcal{L}_{\mathrm{lft}})$ may not cover the whole non-strict epigraph. As a result, Theorem 3.1 does not provide global optimality guarantees for degenerate stationary points $x \in \mathcal{D} \backslash \mathcal{D}_{\mathrm{nd}}$ since they cannot be covered by convex parameterization.[4] Suboptimal saddle points for f might exist even when equipped with an* ECL*. Indeed, it has been revealed that LQG policy optimization has strictly sub-optimal saddle points (Tang et al., 2023, Thereom 5) (Zheng et al., 2022, Theorem 2), which are all degenerate per Definition 3.1.*

4. Applications in Optimal and Robust Control

In this section, we present the ECL constructions for several benchmark optimal and robust control problems; Theorems 2.1 and 3.1 can then be directly applied to their policy optimization formulations. Due to page limit, we omit the mathematical justifications of these ECL constructions, and refer interested readers to Zheng et al. (2023, 2024) for detailed proofs.

3. Subdifferential regularity allows us to relate the Clarke subdifferential with ordinary directional derivatives.

4. In this sense, some classical LMI formulations are not "equivalent" convex parameterizations for original control problems, especially in output-feedback cases. This subtle point has been less emphasized in classical literature since most of them focus on suboptimal policies (Scherer et al., 1997; Scherer and Weiland, 2015; Boyd et al., 1994).

Table 1: ECL construction for LQR and state-feedback $\mathcal{H}_\infty$ control.

	LQR	State-feedback $\mathcal{H}_\infty$
$\mathcal{L}_{\rm lft}$	$\mathcal{L}_{\tt LQR} = \left\{(K,\gamma,X) \,\middle\vert\, \begin{array}{l} X\succ 0, \mathbb{A}_{K,X}+W=0 \\ \gamma \geq \operatorname{tr}\left((Q+K^{\mathsf T}RK)X\right)\end{array}\right\}$	$\mathcal{L}_\infty = \{(K,\gamma,P) \mid \mathbb{L}_\infty \preceq 0,\, P\succ 0\}$, $\mathbb{L}_\infty := \begin{bmatrix} \mathbb{A}^t_{K,P} & PB_w & Q^{1/2} & K^{\mathsf T}R^{1/2} \\ B_w^{\mathsf T}P & -\gamma I & 0 & 0 \\ Q^{1/2} & 0 & -\gamma I & 0 \\ R^{1/2}K & 0 & 0 & -\gamma I \end{bmatrix}$
$\mathcal{F}_{\rm cvx}$	$\mathcal{F}_{\tt LQR} = \left\{(\gamma,Y,X) \,\middle\vert\, \begin{array}{l} X\succ 0, Y\in\mathbb{R}^{m\times n}, \mathbb{A}_{X,Y}+W=0 \\ \gamma \geq \operatorname{tr}\left(QX+X^{-1}Y^{\mathsf T}RY\right)\end{array}\right\}$	$\mathcal{F}_\infty = \{(\gamma,Y,X) \mid X\succ 0, Y\in\mathbb{R}^{m\times n}, \mathbb{F}_\infty \preceq 0\}$, $\mathbb{F}_\infty := \begin{bmatrix} \mathbb{A}_{X,Y} & B_w & XQ^{1/2} & Y^{\mathsf T}R^{1/2} \\ B_w^{\mathsf T} & -\gamma I & 0 & 0 \\ Q^{1/2}X & 0 & -\gamma I & 0 \\ R^{1/2}Y & 0 & 0 & -\gamma I \end{bmatrix}$
Φ	$\Phi_{\tt LQR}(K,\gamma,X) = (\gamma, KX, X)$	$\Phi_\infty(K,\gamma,P) = (\gamma, KP^{-1}, P^{-1})$

Notations: $\mathbb{A}_{K,X} := (A+BK)X + P(A+BK)^{\mathsf T}$; $\mathbb{A}^t_{K,P} := (A+BK)^{\mathsf T}P + P(A+BK)$;
$\mathbb{A}_{X,Y} := AX + BY + (AX+BY)^{\mathsf T}$.

4.1. State-Feedback Policy Optimization

Consider a continuous-time LTI system

$$\dot{x}(t) = Ax(t) + Bu(t) + B_w w(t), \qquad z(t) = \begin{bmatrix} Q^{1/2} \\ 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ R^{1/2} \end{bmatrix} u(t), \tag{4}$$

where $x(t) \in \mathbb{R}^n$ is the state variable, $u(t) \in \mathbb{R}^m$ is the control input, and $w(t) \in \mathbb{R}^n$ is the disturbance on the system process. We introduce $B_w \in \mathbb{R}^{n\times n}$ for a unified treatment of LQR and $\mathcal{H}_\infty$ control in this section. $z(t)$ represents the performance signal, where $Q \succeq 0$ and $R \succ 0$. We assume (A,B) is controllable and $(Q^{1/2}, A)$ is observable.

For both LQR and $\mathcal{H}_\infty$ control, their cost values depend on B_w only via $B_w B_w^{\mathsf T}$. We thus define $W := B_w B_w^{\mathsf T}$, and assume $B_w = W^{1/2}$ without loss of generality. We also assume that $W \succ 0$. We consider the class of state-feedback policies of the form $u(t) = Kx(t)$, with $K \in \mathbb{R}^{m\times n}$, to regulate $z(t)$ under the influence of $w(t)$. The set of stabilizing state-feedback policies, parameterized by K, is then $\mathcal{K} := \{K \in \mathbb{R}^{m\times n} \mid \max_i \operatorname{Re} \lambda_i(A+BK) < 0\}$.

4.1.1. LINEAR QUADRATIC REGULATOR (LQR)

In the LQR problem, $w(t)$ is assumed to be white Gaussian noise with unit intensity. The policy optimization formulation for LQR is then (see, e.g., Mohammadi et al. (2022)),

$$\min_{K\in\mathbb{R}^{m\times n}} \; J_{\tt LQR}(K) := \operatorname{tr}\left[(Q+K^{\mathsf T}RK)X_K\right] \quad \text{s.t. } K \in \mathcal{K}, \tag{5}$$

where X_K is the positive semidefinite solution to the Lyapunov equation $(A+BK)X_K + X_K(A+BK) + W = 0$. It is known that $J_{\tt LQR}(K)$ is smooth and nonconvex, but has a unique stationary point (which is globally optimal) and is gradient dominated on any sublevel set (Mohammadi et al., 2022). These nice landscape properties are closely related to the hidden convexity of (5).

The ECL construction for LQR policy optimization is presented in the second column of Table 1. Specifically, the construction consists of three steps:

Step 1: Lifting. We first define the lifted set $\mathcal{L}_{\mathtt{LQR}}$, as delineated in the second row of Table 1. By the theory of linear quadratic control, we have $K \in \mathcal{K}$ and $\gamma \geq J_{\mathtt{LQR}}(K)$ if and only if there exists X such that $(K, \gamma, X) \in \mathcal{L}_{\mathtt{LQR}}$. This further implies that $\pi_{K,\gamma}(\mathcal{L}_{\mathtt{LQR}}) = \mathrm{epi}_{\geq}(J_{\mathtt{LQR}})$.

Step 2: Convex set. We define the convex set $\mathcal{F}_{\mathtt{LQR}}$ as given in the third row of Table 1, and let the auxiliary set be $\mathcal{G}_{\mathtt{LQR}} = \{0\}$. The first three constraints in the definition of $\mathcal{F}_{\mathtt{LQR}}$ are obviously convex, and the convexity of the last inequality follows by the joint convexity of $\mathrm{tr}(QX + X^{-1}Y^{\mathsf{T}}RY)$ with respect to (X, Y).

Step 3: Diffeomorphism. We employ the classical change of variables $Y = KX$ and define $\Phi_{\mathtt{LQR}}(K, \gamma, X) = (\gamma, KX, X)$, where (KX, X) represents the variable ζ_1 in ECL. This mapping naturally satisfies (2b). We do not need ζ_2 here as no similarity transformation exists. One can check by standard calculation that $\mathcal{F}_{\mathtt{LQR}} = \Phi_{\mathtt{LQR}}(\mathcal{L}_{\mathtt{LQR}})$, and that $\Phi_{\mathtt{LQR}}$ admits an inverse on $\mathcal{F}_{\mathtt{LQR}}$ given by $\Phi_{\mathtt{LQR}}^{-1}(\gamma, Y, X) = (YX^{-1}, \gamma, X)$. Also, $\Phi_{\mathtt{LQR}}$ is a C^{∞} diffeomorphism between $\mathcal{L}_{\mathtt{LQR}}$ and $\mathcal{F}_{\mathtt{LQR}}$.

Consequently, $(\mathcal{L}_{\mathtt{LQR}}, \mathcal{F}_{\mathtt{LQR}}, \{0\}, \Phi_{\mathtt{LQR}})$ is an ECL of $J_{\mathtt{LQR}}(K)$ in (5). One key step in the construction is the utilization of the classical change of variables $Y = KX$ or equivalently $K = YX^{-1}$. Our ECL framework then immediately implies the following well-known results:

1. Theorem 2.1 justifies that the LQR (5) can be reformulated as a convex program:[5]
$$\min_{K \in \mathcal{K}} \; J_{\mathtt{LQR}}(K) = \min_{(\gamma, Y, X) \in \mathcal{F}_{\mathtt{LQR}}} \gamma,$$
and their optimal solutions K^* and (γ^*, Y^*, X^*) are related by $K^* = Y^* X^{*-1}$ and $\gamma^* = J_{\mathtt{LQR}}(K^*)$. The policy optimization for LQR can be viewed as a convex problem in disguise.
2. Any stationary point $K^{\star}$ of $J_{\mathtt{LQR}}$ is globally optimal, confirmed by $\pi_{K,\gamma}(\mathcal{L}_{\mathtt{LQR}}) = \mathrm{epi}_{\geq}(J_{\mathtt{LQR}})$ and Corollary 3.1.

We mention that existing literature has further proved that $J_{\mathtt{LQR}}(K)$ has a unique stationary point, is coercive, and is L-smooth and gradient dominated over any sublevel set (Mohammadi et al., 2022). These properties are fundamental to establishing global convergence of direct policy search and their model-free extensions for solving LQR (Malik et al., 2020; Mohammadi et al., 2022). It would be interesting to investigate how to refine our ECL framework to cover these properties.

4.1.2. State-Feedback $\mathcal{H}_\infty$ Control

In state-feedback $\mathcal{H}_\infty$ control, we consider $w(t)$ as adversarial disturbance with bounded energy, and the goal is to minimize the maximum energy gain from the disturbance $w(t)$ to the performance signal $z(t)$. It is a standard result in robust control that the state-feedback $\mathcal{H}_\infty$ control problem can be formulated as
$$\inf_{K \in \mathbb{R}^{m \times n}} \; J_\infty(K) := \|\mathbf{T}_{zw}(K, s)\|_{\mathcal{H}_\infty} \quad \text{s.t. } K \in \mathcal{K}, \tag{6}$$
where $\mathbf{T}_{zw}(K, s)$ is the transfer matrix from $w(t)$ to $z(t)$ when the policy $u(t) = Kx(t)$ is applied, and $\|\cdot\|_{\mathcal{H}_\infty}$ denotes the $\mathcal{H}_\infty$ norm. Note that the infimum of (6) may not be attainable.

The $\mathcal{H}_\infty$ policy optimization problem (6) is nonconvex and nonsmooth, but it admits a convex reformulation (Scherer and Weiland, 2015). It has been recently revealed in Guo and Hu (2022) that

5. The infima for both the original policy optimization and the convex reformulation can be achieved, due to the coerciveness of $J_{\mathtt{LQR}}(K)$ and the compactness of $\{(Y, X) \mid (\gamma, Y, X) \in \mathcal{F}_{\mathtt{LQR}}\}$ for any given $\gamma > 0$.

for the discrete-time version of (6), any Clarke stationary point is globally optimal. Our aim here is to construct an ECL for (6), which then allows us to draw conclusions from Theorems 2.1 and 3.1 directly. The construction process is very similar to the LQR case, and is delineated in the third column of Table 1. One key difference is that the $\mathcal{H}_\infty$ case relies on the bounded real lemmas.

Specifically, our ECL construction consists of the following steps:

Step 1: Lifting. We define the lifted set $\mathcal{L}_\infty = \big\{(K,\gamma,P) \;\big|\; \mathbb{L}_\infty \preceq 0,\ P \succ 0\big\}$, where P is an extra Lyapunov variable, and the matrix $\mathbb{L}_\infty$ is given in the second row of Table 1.

Step 2: Convex set. We define a convex set $\mathcal{F}_\infty = \big\{(\gamma,Y,X) \;\big|\; X \succ 0, Y \in \mathbb{R}^{m\times n}, \mathbb{F}_\infty \preceq 0\big\}$, with $\mathbb{F}_\infty$ defined in the third row of Table 1. The auxiliary set is $\mathcal{G}_\infty = \{0\}$.

Step 3: Diffeomorphism. We employ the classical change of variables $Y = KP^{-1}, X = P^{-1}$ and introduce the mapping $\Phi_\infty(K,\gamma,P) = (\gamma, KP^{-1}, P^{-1})$, where (KP^{-1}, P^{-1}) represents the variable ζ_1. Similar to the LQR case, no auxiliary variable ζ_2 is needed.

For the construction above, we have the following results.

Proposition 4.1 *Consider the state-feedback $\mathcal{H}_\infty$ policy optimization problem (6), where (A,B) is controllable, B_w has full row rank, and $Q \succ 0, R \succ 0$.*

1. *For any $K \in \mathbb{R}^{m\times n}$ and $\gamma \in \mathbb{R}$, we have $K \in \mathcal{K}$ and $\gamma \geq J_\infty(K)$ if and only if there exists P such that $(K,\gamma,P) \in \mathcal{L}_\infty$. This further implies $\pi_{K,\gamma}(\mathcal{L}_\infty) = \mathrm{epi}_{\geq}(J_\infty)$.*
2. *The mapping Φ_∞ is a C^∞ diffeomorphism between the lifted set $\mathcal{L}_\infty$ and the convex set $\mathcal{F}_\infty$.*

The proof is not very difficult, but one needs to be careful about some technical subtleties in (non)-strict Riccti inequalities. The details are provided in Zheng et al. (2024, Appendix C.4). Proposition 4.1 guarantees that $(\mathcal{L}_\infty, \mathcal{F}_\infty, \{0\}, \Phi_\infty)$ is an ECL of $J_\infty(K)$ in (6). For this ECL, we further have the following nice results, which are consequences of Theorem 2.1 and Corollary 3.1, and the fact that $J_\infty(K)$ is sudifferentially regular.

Theorem 4.1 *Under the conditions of Proposition 4.1, the following statements hold.*

1. *Problem (6) is equivalent to the convex problem $\inf_{(\gamma,Y,X)\in\mathcal{F}_\infty} \gamma$.*
2. *All stabilizing policies $K \in \mathcal{K}$ are non-degenerate with respect to the* ECL $(\mathcal{L}_\infty, \mathcal{F}_\infty, \{0\}, \Phi_\infty)$.
3. *Any Clarke stationary point of (6) is globally optimal.*

We note that the global optimality of Clarke stationary points for (6) has not been reported before. This is the continuous-time counterpart of the discrete-time result in Guo and Hu (2022). We also note that the infimum of (6) may not be achieved, in which case the Clarke stationary point does not exist. An explicit SISO example is provided in Zheng et al. (2024, Appendix C.2).

Remark 4.1 *The diffeomorphisms $\Phi_{\mathtt{LQR}}$ and Φ_∞ are essentially in the same form and follow from the classical change of variable $K = YX^{-1}$ (Khargonekar and Rotea, 1991; Boyd et al., 1994; Bernussou et al., 1989). This change of variable is able to linearize many bilinear matrix inequalities that appear in state-feedback control problems, most of which are related to the Lyapunov inequality $(A+BK)X + X(A+BK)^{\mathsf{T}} \prec 0$ (see Boyd et al., 1994, Chapter 7 for a historical perspective). As we will see in the next subsection, the linearization for dynamic output-feedback policies turns out to be much more complicated, and we will utilize the techniques in Scherer et al. (1997); Scherer and Weiland (2015) to construct* ECL*s for output-feedback $\mathcal{H}_\infty$ control.* □

4.2. Output-Feedback Policy Optimization

In this subsection, we show that ECL is also applicable to policy optimization with dynamic output-feedback policies. Due to space limitation, we only treat the output-feedback $\mathcal{H}_\infty$ control; see Zheng et al. (2023, 2024) for LQG control.

Consider an LTI system with partial observations,

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

where $y(t) \in \mathbb{R}^p$ is the vector of measured outputs available for feedback control, and $w(t) \in \mathbb{R}^n, v(t) \in \mathbb{R}^p$ are the disturbances on the system process and measurement at time t. We define $W := B_w B_w^\mathsf{T}$, $V := D_v D_v^\mathsf{T}$, and, without loss of generality, assume $B_w = W^{1/2}$ and $D_v = V^{1/2}$. We adopt the same performance signal $z(t)$ in (4). The following assumption is standard.

Assumption 4.1 *(A, B) and $(A, W^{1/2})$ are controllable, and (C, A) and $(Q^{1/2}, A)$ are observable. Moreover, $Q \succeq 0, R \succ 0, W \succeq 0, V \succ 0$.*

To properly regulate $z(t)$, we consider full-order dynamic output-feedback policies of the form

$$\begin{aligned} \dot{\xi}(t) &= A_\mathsf{K}\xi(t) + B_\mathsf{K} y(t), \\ u(t) &= C_\mathsf{K}\xi(t) + D_\mathsf{K} y(t). \end{aligned} \qquad \mathsf{K} = \begin{bmatrix} D_\mathsf{K} & C_\mathsf{K} \\ B_\mathsf{K} & A_\mathsf{K} \end{bmatrix} \in \mathbb{R}^{(m+n)\times(p+n)}, \tag{7}$$

where $\xi(t) \in \mathbb{R}^n$ is the internal state, and $A_\mathsf{K}, B_\mathsf{K}, C_\mathsf{K}$ and D_K are matrices of proper dimensions that specify the policy dynamics. We parameterize dynamic policies by K. By closing the feedback loop, we can represent the transfer matrix from the disturbance $d(t) = \begin{bmatrix} w^\mathsf{T}(t) & v^\mathsf{T}(t) \end{bmatrix}^\mathsf{T}$ to the performance signal $z(t)$ in the following form:

$$\mathbf{T}_{zd}(\mathsf{K}, s) = C_{\mathrm{cl}}(\mathsf{K}) \left(sI - A_{\mathrm{cl}}(\mathsf{K})\right)^{-1} B_{\mathrm{cl}}(\mathsf{K}) + D_{\mathrm{cl}}(\mathsf{K}),$$

where $A_{\mathrm{cl}}(\mathsf{K}), B_{\mathrm{cl}}(\mathsf{K}), C_{\mathrm{cl}}(\mathsf{K})$ and $D_{\mathrm{cl}}(\mathsf{K})$ are certain matrix-valued functions that characterize the state-space model of the closed-loop system (see Zheng et al., 2023, Appendix A.5 for details).

Now consider a standard output-feedback $\mathcal{H}_\infty$ policy optimization problem,

$$\inf_{\mathsf{K}} \;\; J_{\infty,n}(\mathsf{K}) := \|\mathbf{T}_{zd}(\mathsf{K}, s)\|_{\mathcal{H}_\infty} \qquad \text{s.t. } \mathsf{K} \in \mathcal{C}_n, \tag{8}$$

with $\mathcal{C}_n$ denoting the set of internally stabilizing dynamic policies $\mathcal{C}_n = \left\{\mathsf{K} \mid A_{\mathrm{cl}}(\mathsf{K}) \text{ is stable}\right\}$. It is known that the policy optimization problem (8) is nonsmooth and nonconvex, which has complicated landscape properties. Our ECL construction for $J_{\infty,n}(\mathsf{K})$ is based on the change of variables given in Scherer et al. (1997); Scherer and Weiland (2015), and the details are as follows:

Step 1: Lifting. We first introduce the lifted set $\mathcal{L}_{\infty,\mathrm{d}}$ by

$$\mathcal{L}_{\infty,\mathrm{d}} = \left\{ (\mathsf{K}, \gamma, P) \;\middle|\; \begin{array}{l} \mathsf{K} \in \mathbb{R}^{(m+n)\times(p+n)}, \; \gamma \in \mathbb{R}, \; P \succ 0, \; \det P_{12} \neq 0, \\ \begin{bmatrix} A_{\mathrm{cl}}(\mathsf{K})^\mathsf{T} P + P A_{\mathrm{cl}}(\mathsf{K}) & P B_{\mathrm{cl}}(\mathsf{K}) & C_{\mathrm{cl}}(\mathsf{K})^\mathsf{T} \\ B_{\mathrm{cl}}(\mathsf{K})^\mathsf{T} P & -\gamma I & D_{\mathrm{cl}}(\mathsf{K})^\mathsf{T} \\ C_{\mathrm{cl}}(\mathsf{K}) & D_{\mathrm{cl}}(\mathsf{K}) & -\gamma I \end{bmatrix} \preceq 0 \end{array} \right\}. \tag{9}$$

where P_{12} denotes the $n \times n$ submatrix of P corresponding to the first n rows and last n columns. The extra variable P plays the role of the lifting variable in ECL.

Step 2: Convex and auxiliary sets. We let the convex set be

$$\mathcal{F}_{\infty,\mathrm{d}} = \left\{ (\gamma, \Lambda, X, Y) \;\middle|\; \gamma \in \mathbb{R},\ \Lambda \in \mathbb{R}^{(m+n)\times(p+n)}, \begin{bmatrix} X & I_n \\ I_n & Y \end{bmatrix} \succ 0, \mathscr{M}(\gamma, \Lambda, X, Y) \preceq 0 \right\},$$

where (Λ, X, Y) corresponds to ζ_1, and $\mathscr{M}(\gamma, \Lambda, X, Y)$ is an affine operator whose definition is omitted here due to space limitation. The auxiliary set is $\mathrm{GL}_n = \{T \in \mathbb{R}^{n\times n} \mid \det T \neq 0\}$.
Step 3: Diffeomorphism. We define the mapping $\Phi_{\infty,\mathrm{d}}$ by

$$\Phi_{\infty,\mathrm{d}}(\mathsf{K}, \gamma, P) = \left(\gamma, \begin{bmatrix} D_{\mathsf{K}} & \Phi_F \\ \Phi_H & \Phi_M \end{bmatrix}, (P^{-1})_{11}, P_{11}, P_{12}\right), \qquad (\mathsf{K}, \gamma, P) \in \mathcal{L}_{\infty,\mathrm{d}},$$

where $\Phi_M = P_{12}B_{\mathsf{K}}C(P^{-1})_{11} + P_{11}BC_{\mathsf{K}}(P^{-1})_{21} + P_{11}(A + BD_{\mathsf{K}}C)(P^{-1})_{11} + P_{12}A_{\mathsf{K}}(P^{-1})_{21}$, $\Phi_H = P_{11}BD_{\mathsf{K}} + P_{12}B_{\mathsf{K}}$, and $\Phi_F = D_{\mathsf{K}}C(P^{-1})_{11} + C_{\mathsf{K}}(P^{-1})_{21}$.

The following proposition justifies that $(\mathcal{L}_{\infty,\mathrm{d}}, \mathcal{F}_{\infty,\mathrm{d}}, \mathrm{GL}_n, \Phi_{\infty,\mathrm{d}})$ is an `ECL` for $J_{\infty,n}(\mathsf{K})$. The proof is technically involved and is given in the report Zheng et al., 2024, Appendix D; the main difficulty lies in that (3) does not hold anymore, and we need to establish the chain of inclusion (2a).

Proposition 4.2 *Under Assumption 4.1, we have i)* $\mathrm{epi}_{>}(J_{\infty,n}) \subseteq \pi_{\mathsf{K},\gamma}(\mathcal{L}_{\infty,\mathrm{d}}) \subseteq \mathrm{cl}\ \mathrm{epi}_{\geq}(J_{\infty,n})$. *ii) The mapping* $\Phi_{\infty,\mathrm{d}}$ *is a* C^∞ *diffeomorphism from* $\mathcal{L}_{\infty,\mathrm{d}}$ *to* $\mathcal{F}_{\infty,\mathrm{d}} \times \mathrm{GL}_n$.

Then, by Theorems 2.1 and 3.1, we get the following corollary.

Corollary 4.1 *Under Assumption 4.1, the output-feedback* $\mathcal{H}_\infty$ *policy optimization problem* (8) *is equivalent to a convex problem in the sense that* $\inf_{\mathsf{K}\in\mathcal{C}_n} J_{\infty,n}(\mathsf{K}) = \inf_{(\gamma,\Lambda,X,Y)\in\mathcal{F}_{\infty,\mathrm{d}}} \gamma$*. Furthermore, for a Clarke stationary point* $\mathsf{K} \in \mathcal{C}_n$ *(i.e.,* $0 \in \partial J_{\infty,n}(\mathsf{K})$*), if* K *is non-degenerate in the sense of Definition 3.1, then it is globally optimal for* (8).

We can now see that our `ECL` framework covers the output-feedback $\mathcal{H}_\infty$ policy optimization problem as a special case, providing global optimality certificates for non-degenerate Clarke stationary points. Note that the equivalence to the convex reformulation is essentially the same as (Scherer and Weiland, 2015, Chapter 4.2.3), but the global optimality of non-degenerate $\mathcal{H}_\infty$ policies cannot be derived from (Scherer and Weiland, 2015, Chapter 4.2.3) due to its use of strict LMIs. Finally, we remark that our `ECL` can also cover a class of distributed control problems under the condition of quadratic invariance (Furieri et al., 2020a,b); see Zheng et al., 2024, Section 4.4 for details.

5. Conclusion

This paper introduced the `ECL` framework to reveal hidden convexity in nonconvex and potentially nonsmooth policy optimization for optimal and robust control. Particularly, we have shown that, with the existence of an `ECL` for nonconvex policy optimization, all *non-degenerate* stationary policies are globally optimal. We have built explicit `ECL`s for LQR, state feedback $\mathcal{H}_\infty$ control, and dynamic output-feedback $\mathcal{H}_\infty$ control. We hope the `ECL` framework will be useful for analyzing nonconvex problems in other areas beyond control.

Acknowledgment

Y. Zheng and C. Pai are supported by NSF ECCS-2154650, NSF CMMI-2320697, and NSF CAREER 2340713. Y. Tang is supported by NSFC through grants 72301008 and 72431001.

References

Pierre Apkarian and Dominikus Noll. Nonsmooth H_∞ synthesis. *IEEE Trans. Autom. Control*, 51 (1):71–86, 2006.

J. Bernussou, P. L. D. Peres, and J. C. Geromel. A linear programming oriented procedure for quadratic stabilization of uncertain systems. *Systems & Control Letters*, 13(1):65–72, 1989.

Stephen Boyd et al. *Linear Matrix Inequalities in System and Control Theory*. SIAM, 1994.

Frank H. Clarke. *Optimization and Nonsmooth Analysis*. SIAM, 1990.

Jingliang Duan et al. On the optimization landscape of dynamic output feedback linear quadratic control. *IEEE Trans. Autom. Control*, 69(2):920–935, 2024.

Ilyas Fatkhullin and Boris Polyak. Optimizing static linear feedback: Gradient method. *SIAM Journal on Control and Optimization*, 59(5):3887–3911, 2021.

Maryam Fazel et al. Global convergence of policy gradient methods for the linear quadratic regulator. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1467–1476, 2018.

Luca Furieri, Yang Zheng, and Maryam Kamgarpour. Learning the globally optimal distributed LQ regulator. In *Proceedings of the 2nd Conference on Learning for Dynamics and Control*, volume 120 of *Proceedings of Machine Learning Research*, pages 287–297, 2020a.

Luca Furieri et al. Sparsity invariance for convex design of distributed controllers. *IEEE Transactions on Control of Network Systems*, 7(4):1836–1847, 2020b.

Xingang Guo and Bin Hu. Global convergence of direct policy search for state-feedback $\mathcal{H}_\infty$ robust control: A revisit of nonsmooth synthesis with Goldstein subdifferential. In *Advances in Neural Information Processing Systems*, volume 35, pages 32801–32815, 2022.

Bin Hu and Yang Zheng. Connectivity of the feasible and sublevel sets of dynamic output feedback control with robustness constraints. *IEEE Control Systems Letters*, 7:442–447, 2022.

Bin Hu et al. Toward a theoretical foundation of policy optimization for learning control policies. *Annual Review of Control, Robotics, and Autonomous Systems*, 6:123–158, 2023.

Rudolf E. Kalman. The theory of optimal control and the calculus of variations. In Richard Bellman, editor, *Mathematical Optimization Techniques*. University of California Press, 1963.

Pramod P. Khargonekar and Mario A. Rotea. Mixed H_2/H_∞ control: A convex optimization approach. *IEEE Trans. Autom. Control*, 36(7):824–837, 1991.

Spencer Kraisler and Mehran Mesbahi. Output-feedback synthesis orbit geometry: Quotient manifolds and LQG direct policy optimization. *arXiv preprint arXiv:2403.17157*, 2024.

William Levine and Michael Athans. On the determination of the optimal constant output feedback gains for linear multivariable systems. *IEEE Trans. Autom. Control*, 15(1):44–48, 1970.

Adrian S. Lewis. Nonsmooth optimization and robust control. *Annual Reviews in Control*, 31(2): 167–177, 2007.

Dhruv Malik et al. Derivative-free methods for policy optimization: Guarantees for linear quadratic systems. *Journal of Machine Learning Research*, 21(202):1–51, 2020.

H. Mohammadi et al. Convergence and sample complexity of gradient methods for the model-free linear–quadratic regulator problem. *IEEE Trans. Autom. Control*, 67(5):2435–2450, 2022.

Benjamin Recht. A tour of reinforcement learning: The view from continuous control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:253–279, 2019.

Carsten Scherer and Siep Weiland. Linear Matrix Inequalities in Control. Lecture Notes, 2015. https://www.imng.uni-stuttgart.de/mst/files/LectureNotes.pdf.

Carsten Scherer, Pascal Gahinet, and Mahmoud Chilali. Multiobjective output-feedback control via LMI optimization. *IEEE Trans. Autom. Control*, 42(7):896–911, 1997.

Yue Sun and Maryam Fazel. Learning optimal controllers by policy gradient: Global optimality via convex parameterization. In *Proceedings of the 60th IEEE Conference on Decision and Control (CDC)*, pages 4576–4581, 2021.

Shahriar Talebi, Yang Zheng, Spencer Kraisler, Na Li, and Mehran Mesbahi. Policy optimization in control: Geometry and algorithmic implications. *arXiv preprint arXiv:2406.04243*, 2024.

Yujie Tang and Yang Zheng. On the global optimality of direct policy search for nonsmooth $\mathcal{H}_\infty$ output-feedback control. In *Proceedings of the 62nd IEEE Conference on Decision and Control (CDC)*, pages 6148–6153, 2023.

Yujie Tang, Yang Zheng, and Na Li. Analysis of the optimization landscape of Linear Quadratic Gaussian (LQG) control. *Mathematical Programming*, 202:399–444, 2023.

Jack Umenberger et al. Globally convergent policy search for output estimation. In *Advances in Neural Information Processing Systems*, volume 35, pages 22778–22790, 2022.

Kaiqing Zhang, Bin Hu, and Tamer Başar. Policy optimization for $\mathcal{H}_2$ linear control with $\mathcal{H}_\infty$ robustness guarantee: Implicit regularization and global convergence. *SIAM Journal on Control and Optimization*, 59(6):4081–4109, 2021.

Xiangyuan Zhang, Bin Hu, and Tamer Başar. Learning the Kalman filter with fine-grained sample complexity. In *2023 American Control Conference (ACC)*, 2023.

Yang Zheng, Chih-Fan Pai, and Yujie Tang. Benign nonconvex landscapes in optimal and robust control, Part I: Global optimality. *arXiv preprint arXiv:2312.15332*, 2023.

Yang Zheng, Chih-Fan Pai, and Yujie Tang. Benign nonconvex landscapes in optimal and robust control, Part II: Extended convex lifting. *arXiv preprint arXiv:2406.04001*, 2024.

Yang Zheng et al. Escaping high-order saddles in policy optimization for Linear Quadratic Gaussian (LQG) control. In *61st IEEE Conference on Decision and Control*, pages 5329–5334, 2022.

Kemin Zhou, John C. Doyle, and Keith Glover. *Robust and Optimal Control*. Prentice Hall, 1996.