Convergence and Sample Complexity of Gradient Methods for the Model-Free Linear–Quadratic Regulator Problem

Hesameddin Mohammadi, Armin Zare, *Member, IEEE*, Mahdi Soltanolkotabi, and Mihailo R. Jovanović, *Fellow, IEEE*

***Abstract*—Model-free reinforcement learning attempts to find an optimal control action for an unknown dynamical system by directly searching over the parameter space of controllers. The convergence behavior and statistical properties of these approaches are often poorly understood because of the nonconvex nature of the underlying optimization problems and the lack of exact gradient computation. In this article, we take a step toward demystifying the performance and efficiency of such methods by focusing on the standard infinite-horizon linear–quadratic regulator problem for continuous-time systems with unknown state-space parameters. We establish exponential stability for the ordinary differential equation (ODE) that governs the gradient-flow dynamics over the set of stabilizing feedback gains and show that a similar result holds for the gradient descent method that arises from the forward Euler discretization of the corresponding ODE. We also provide theoretical bounds on the convergence rate and sample complexity of the random search method with two-point gradient estimates. We prove that the required simulation time for achieving ϵ-accuracy in the model-free setup and the total number of function evaluations both scale as $\log(1/\epsilon)$.**

***Index Terms*—Data-driven control, gradient descent, gradient-flow dynamics, linear–quadratic regulator (LQR), model-free control, nonconvex optimization, Polyak–Łojasiewicz inequality, random search method, reinforcement learning (RL), sample complexity.**

I. Introduction

IN MANY emerging applications, control-oriented models are not readily available, and classical approaches from optimal control may not be directly applicable. This challenge has led to the emergence of reinforcement learning (RL) approaches that often perform well in practice. Examples include learning complex locomotion tasks via neural network dynamics [1] and playing Atari games based on images using deep RL [2].

RL approaches can be broadly divided into model-based [3], [4] and model-free [5], [6]. While model-based RL uses data to obtain approximations of the underlying dynamics, its model-free counterpart prescribes control actions based on estimated values of a cost function without attempting to form a model. In spite of the empirical success of RL in a variety of domains, our mathematical understanding of it is still in its infancy, and there are many open questions surrounding convergence and sample complexity. In this article, we take a step toward answering such questions with a focus on the infinite-horizon linear–quadratic regulator (LQR) for continuous-time systems.

The LQR problem is the cornerstone of control theory. The globally optimal solution can be obtained by solving the Riccati equation, and efficient numerical schemes with provable convergence guarantees have been developed [7]. However, computing the optimal solution becomes challenging for large-scale problems, when prior knowledge is not available, or in the presence of structural constraints on the controller. This motivates the use of direct search methods for controller synthesis. Unfortunately, the nonconvex nature of this formulation complicates the analysis of first- and second-order optimization algorithms. To make matters worse, structural constraints on the feedback gain matrix may result in a disjoint search landscape limiting the utility of conventional descent-based methods [8]. Furthermore, in the model-free setting, the exact model (and hence the gradient of the objective function) is unknown so that only zeroth-order methods can be used.

In this article, we study convergence properties of gradient-based methods for the continuous-time LQR problem. In spite of the lack of convexity, we establish: 1) *exponential stability* of the ordinary differential equation (ODE) that governs the gradient-flow dynamics over the set of stabilizing feedback gains; and 2) *linear convergence* of the gradient descent algorithm with a

Manuscript received December 26, 2019; revised July 22, 2020 and March 15, 2021; accepted May 30, 2021. Date of publication June 8, 2021; date of current version April 26, 2022. The work of Hesameddin Mohammadi, Armin Zare, and Mihailo R. Jovanović was supported in part by the National Science Foundation (NSF) under Grants ECCS-1708906 and ECCS-1809833, and in part by the Air Force Office of Scientific Research (AFOSR) under Grant FA9550-16-1-0009. The work of Mahdi Soltanolkotabi was supported in part by the Packard Fellowship in Science and Engineering, in part by a Sloan Research Fellowship in Mathematics, in part by a Google Faculty Research Award, and in part by Awards from the NSF and the AFOSR Young Investigator Program. Recommended by Associate Editor N. Li. *(Corresponding author: Mihailo R. Jovanović.)*

Hesameddin Mohammadi, Mahdi Soltanolkotabi, and Mihailo R. Jovanović are with the Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: hesamedm@usc.edu; soltanol@usc.edu; mihailo@usc.edu).

Armin Zare is with the Department of Mechanical Engineering, University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: armin.zare@utdallas.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2021.3087455.

Digital Object Identifier 10.1109/TAC.2021.3087455

suitable stepsize. We employ a standard convex reparameterization for the LQR problem [9] to establish the convergence properties of gradient-based methods for the nonconvex formulation. In the model-free setting, we also examine convergence and sample complexity of the random search method [10] that attempts to emulate the behavior of gradient descent via gradient approximations resulting from objective function values. For the two-point gradient estimation setting, we prove linear convergence of the random search method and show that the total number of function evaluations and the simulation time required in our results to achieve ϵ-accuracy are proportional to $\log(1/\epsilon)$.

For the *discrete-time* LQR, global convergence guarantees were recently provided in [11] for gradient decent and the random search method with one-point gradient estimates. The authors established a bound on the sample complexity for reaching the error tolerance ϵ that requires a number of function evaluations that is at least proportional to $(1/\epsilon^4)\log(1/\epsilon)$. If one has access to the infinite-horizon cost values, the number of function evaluations for the random search method with one-point gradient estimates can be improved to $1/\epsilon^2$ [12]. In contrast, we focus on the *continuous-time* LQR and examine the two-point gradient estimation setting. The use of two-point gradient estimates reduces the required number of function evaluations to $1/\epsilon$ [12]. We significantly improve this result by showing that the required number of function evaluations is proportional to $\log(1/\epsilon)$. Similarly, the simulation time required in our results is proportional to $\log(1/\epsilon)$; this is in contrast to [11] that requires $\text{poly}(1/\epsilon)$ simulation time and to [12] that assumes an infinite simulation time. Furthermore, our convergence results hold both in terms of the error in the objective value and the optimization variable (i.e., the feedback gain matrix), whereas [11] and [12] only prove convergence in the objective value. We note that the literature on model-free RL is rapidly expanding, and recent extensions to Markovian jump linear systems [13], $\mathcal{H}_\infty$ robustness analysis through implicit regularization [14], learning distributed linear–quadratic problems [15], and output-feedback LQR [16] have been made.

Our presentation is structured as follows. In Section II, we revisit the LQR problem and present gradient-flow dynamics, gradient descent, and the random search algorithm. In Section III, we highlight the main results of this article. In Section IV, we utilize convex reparameterization of the LQR problem and establish exponential stability of the resulting gradient-flow dynamics and gradient descent method. In Section V, we extend our analysis to the nonconvex landscape of feedback gains. In Section VI, we quantify the accuracy of two-point gradient estimates, and in Section VII, we discuss convergence and sample complexity of the random search method. In Section VIII, we provide an example to illustrate our theoretical developments. Section IX concludes this article. Most technical details are relegated to the Appendixes.

Notation: We use $\text{vec}(M) \in \mathbb{R}^{mn}$ to denote the vectorized form of the matrix $M \in \mathbb{R}^{m\times n}$ obtained by concatenating the columns on top of each other. We use $\|M\|_F^2 = \langle M, M\rangle$ to denote the Frobenius norm, where $\langle X, Y\rangle := \text{trace}\,(X^TY)$ is the standard matricial inner product. We denote the largest singular value of linear operators and matrices by $\|\cdot\|_2$ and the spectral induced norm of linear operators by $\|\cdot\|_S$

$$\|\mathcal{M}\|_2 \;:=\; \sup_M \frac{\|\mathcal{M}(M)\|_F}{\|M\|_F},\; \|\mathcal{M}\|_S \;:=\; \sup_M \frac{\|\mathcal{M}(M)\|_2}{\|M\|_2}.$$

We denote by $\mathbb{S}^n \subset \mathbb{R}^{n\times n}$ the set of symmetric matrices. For $M \in \mathbb{S}^n$, $M \succ 0$ means M is positive definite and $\lambda_{\min}(M)$ is the smallest eigenvalue. We use $S^{d-1} \subset \mathbb{R}^d$ to denote the unit sphere of dimension $d-1$. We denote the expected value by $\mathbb{E}[\cdot]$ and probability by $\mathbb{P}(\cdot)$. To compare the asymptotic behavior of $f(\epsilon)$ and $g(\epsilon)$ as ϵ goes to 0, we use $f = O(g)$ (or, equivalently, $g = \Omega(f)$) to denote $\limsup_{\epsilon\to 0} f(\epsilon)/g(\epsilon) < \infty$, $f = \tilde{O}(g)$ to denote $f = O(g\log^k g)$ for some integer k, and $f = o(\epsilon)$ to signify $\lim_{\epsilon\to 0} f(\epsilon)/\epsilon = 0$.

II. Problem Formulation

The infinite-horizon LQR problem for continuous-time LTI systems is given by

$$\underset{x,u}{\text{minimize}} \quad \mathbb{E}\left[\int_0^\infty (x^T(t)Qx(t) + u^T(t)Ru(t))dt\right] \tag{1a}$$

$$\text{subject to} \quad \dot{x} = Ax + Bu, \;\; x(0) \sim \mathcal{D} \tag{1b}$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the control input, A and B are constant matrices of appropriate dimensions, Q and R are positive-definite matrices, and the expectation is taken over a random initial condition $x(0)$ with distribution $\mathcal{D}$. For a controllable pair (A, B), the solution to (1) is given by

$$u(t) \;=\; -K^\star x(t) \;=\; -R^{-1}B^TP^\star x(t) \tag{2a}$$

where $P^\star$ is the unique positive-definite solution to the algebraic Riccati equation (ARE)

$$A^TP^\star + P^\star A + Q - P^\star BR^{-1}B^TP^\star \;=\; 0. \tag{2b}$$

When the model is known, the LQR problem and the corresponding ARE can be solved efficiently via a variety of techniques [17]–[20]. However, these methods are not directly applicable in the model-free setting, i.e., when the matrices A and B are unknown. Exploiting the linearity of the optimal controller, we can alternatively formulate the LQR problem as a direct search for the optimal linear feedback gain, namely

$$\underset{K}{\text{minimize}}\; f(K) \tag{3a}$$

where

$$f(K) \;:=\; \begin{cases} \text{trace}\left((Q + K^TRK)X(K)\right), & K \in \mathcal{S}_K \\ \infty, & \text{otherwise.} \end{cases} \tag{3b}$$

Here, the function $f(K)$ determines the LQR cost in (1a) associated with the linear state-feedback law $u = -Kx$,

$$\mathcal{S}_K \;:=\; \{K \in \mathbb{R}^{m\times n} \,|\, A - BK \text{ is Hurwitz}\} \tag{3c}$$

is the set of stabilizing feedback gains and, for any $K \in \mathcal{S}_K$,

$$\begin{aligned} X(K) &:= \int_0^\infty \mathbb{E}\left[x(t)x^T(t)\right] \\ &= \int_0^\infty e^{(A-BK)t}\,\Omega\, e^{(A-BK)^Tt}\,dt \end{aligned} \tag{4a}$$

is the unique solution to the Lyapunov equation

$$(A - BK)X + X(A - BK)^T + \Omega \;=\; 0 \tag{4b}$$

and $\Omega := \mathbb{E}\left[x(0)x^T(0)\right]$. To ensure $f(K) = \infty$ for $K \notin \mathcal{S}_K$, we assume $\Omega \succ 0$. This assumption also guarantees $K \in \mathcal{S}_K$ if and only if the solution X to (4b) is positive definite.

In problem (3), the matrix K is the optimization variable, and A, B, $Q \succ 0$, $R \succ 0$, and $\Omega \succ 0$ are the problem parameters. This alternative formulation of the LQR problem has been studied for both continuous-time [7] and discrete-time systems [11], [21], and it serves as a building block for several important

control problems, including optimal static-output-feedback design [22], optimal design of sparse feedback gain matrices [23]–[26], and optimal sensor/actuator selection [27]–[29].

For all stabilizing feedback gains $K \in \mathcal{S}_K$, the gradient of the objective function is determined by [22], [23]

$$\nabla f(K) \;=\; 2(RK \;-\; B^T P(K))X(K). \tag{5}$$

Here, $X(K)$ is given by (4a) and

$$P(K) = \int_0^\infty e^{(A-BK)^T t}\,(Q + K^T R\,K)\,e^{(A-BK)t}\,dt \tag{6a}$$

is the unique positive-definite solution of

$$(A \;-\; BK)^T P \;+\; P(A \;-\; BK) \;=\; -Q \;-\; K^T R\,K. \tag{6b}$$

To simplify our presentation, for any $K \in \mathbb{R}^{m\times n}$, we define the closed-loop Lyapunov operator $\mathcal{A}_K\colon \mathbb{S}^n \to \mathbb{S}^n$ as

$$\mathcal{A}_K(X) \;:=\; (A \;-\; BK)X \;+\; X(A \;-\; BK)^T. \tag{7a}$$

For $K \in \mathcal{S}_K$, both $\mathcal{A}_K$ and its adjoint

$$\mathcal{A}_K^*(P) \;=\; (A \;-\; BK)^T P \;+\; P(A \;-\; BK) \tag{7b}$$

are invertible and $X(K)$ and $P(K)$ are determined by

$$X(K) \;=\; -\mathcal{A}_K^{-1}(\Omega),\; P(K) \;=\; -(\mathcal{A}_K^*)^{-1}(Q \;+\; K^T RK).$$

In this article, we first examine the global stability properties of the gradient-flow dynamics

$$\dot{K} \;=\; -\nabla f(K),\;\; K(0) \;\in\; \mathcal{S}_K \tag{GF}$$

associated with problem (3) and its discretized variant

$$K^{k+1} \;:=\; K^k \;-\; \alpha\,\nabla f(K^k),\;\; K^0 \;\in\; \mathcal{S}_K \tag{GD}$$

where $\alpha > 0$ is the stepsize. Next, we use this analysis as a building block to study the convergence of a search method based on random sampling [10], [30] for solving problem (3). As described in Algorithm 1, at each iteration, we form an empirical approximation $\overline{\nabla} f(K)$ to the gradient of the objective function via simulation of system (1b) for randomly perturbed feedback gains $K \pm U_i$, $i = 1, \ldots, N$, and update K via

$$K^{k+1} := K^k - \alpha\overline{\nabla} f(K^k),\quad K^0 \;\in\; \mathcal{S}_K. \tag{RS}$$

We note that the gradient estimation scheme in Algorithm 1 does not require knowledge of system matrices A and B in (1b) but only access to a simulation engine.

Algorithm 1: Two-Point Gradient Estimation.

Input: Feedback gain $K \in \mathbb{R}^{m\times n}$, state and control weight matrices Q and R, distribution $\mathcal{D}$, smoothing constant r, simulation time τ, number of random samples N.

for $i \;=\; 1, \ldots, N$ **do**

– Define perturbed feedback gains $K_{i,1} := K + rU_i$ and $K_{i,2} := K - rU_i$, where $\mathrm{vec}(U_i)$ is a random vector uniformly distributed on the sphere $\sqrt{mn}\,S^{mn-1}$.

– Sample an initial condition x_i from distribution $\mathcal{D}$.

– For $j \in \{1, 2\}$, simulate system (1b) up to time τ with the feedback gain $K_{i,j}$ and initial condition x_i to form

$$\hat{f}_{i,j} = \int_0^\tau (x^T(t)Qx(t) + u^T(t)Ru(t))\,dt.$$

end for

Output: The gradient estimate

$$\overline{\nabla} f(K) = \frac{1}{2rN}\sum_{i=1}^{N}\Big(\hat{f}_{i,1} - \hat{f}_{i,2}\Big)U_i.$$

III. Main Results

Optimization problem (3) is not convex [8]; see Appendix A for an example. The function $f(K)$, however, has two important properties: *uniqueness of the critical points* and the *compactness of sublevel sets* [31], [32]. Based on these, the LQR objective error $f(K) - f(K^\star)$ can be used as a maximal Lyapunov function (see [33] for a definition) to prove asymptotic stability of gradient-flow dynamics (GF) over the set of stabilizing feedback gains $\mathcal{S}_K$. However, this approach does not provide any guarantee on the rate of convergence, and additional analysis is necessary to establish exponential stability; see Section V for details.

A. Known Model

We first summarize our results for the case when the model is known. In spite of the nonconvex optimization landscape, we establish the exponential stability of gradient-flow dynamics (GF) for any stabilizing initial feedback gain $K(0)$. This result also provides an explicit bound on the rate of convergence to the LQR solution $K^\star$.

Theorem 1: For any initial stabilizing feedback gain $K(0) \in \mathcal{S}_K$, the solution $K(t)$ to gradient-flow dynamics (GF) satisfies the following:

a) $f(K(t)) \;-\; f(K^\star) \;\le\; e^{-\rho t}(f(K(0)) \;-\; f(K^\star))$;

b) $\|K(t) \;-\; K^\star\|_F^2 \;\le\; b\,e^{-\rho t}\,\|K(0) \;-\; K^\star\|_F^2$;

where the convergence rate ρ and constant b depend on $K(0)$ and the parameters of the LQR problem (3).

The proof of Theorem 1 along with explicit expressions for the convergence rate ρ and constant b is provided in Section V-A. Moreover, for a sufficiently small stepsize α, we show that gradient descent method (GD) also converges over $\mathcal{S}_K$ at a linear rate.

Theorem 2: For any initial stabilizing feedback gain $K^0 \in \mathcal{S}_K$, the iterates of gradient descent (GD) satisfy the following:

a) $f(K^k) \;-\; f(K^\star) \;\le\; \gamma^k(f(K^0) \;-\; f(K^\star))$;

b) $\|K^k \;-\; K^\star\|_F^2 \;\le\; b\,\gamma^k\,\|K^0 \;-\; K^\star\|_F^2$;

where the rate of convergence γ, stepsize α, and constant b depend on K^0 and the parameters of the LQR problem (3).

B. Unknown Model

We now turn our attention to the model-free setting. We use Theorem 2 to carry out the convergence analysis of the random search method (RS) under the following assumption on the distribution of initial condition.

Assumption 1: Let the distribution $\mathcal{D}$ of the initial conditions have independent identically distributed (i.i.d.) zero-mean unit-variance entries with bounded sub-Gaussian norm, i.e., for a random vector $v \in \mathbb{R}^n$ distributed according to $\mathcal{D}$, $\mathbb{E}[v_i] = 0$ and $\|v_i\|_{\psi_2} \le \kappa$, for some constant κ and $i = 1, \ldots, n$; see Appendix J for the definition of $\|\cdot\|_{\psi_2}$.

Our main convergence result holds under Assumption 1. Specifically, for a desired accuracy level $\epsilon > 0$, in Theorem 3, we establish that iterates of (RS) with constant stepsize (that does not depend on ϵ) reach accuracy level ϵ at a linear rate (i.e., in at most $O(\log(1/\epsilon))$ iterations) with high probability. Furthermore, the total number of function evaluations and the

simulation time required to achieve an accuracy level ϵ are proportional to $\log(1/\epsilon)$. This significantly improves the existing results for discrete-time LQR [11], [12] that require $O(1/\epsilon)$ function evaluations and $\mathrm{poly}(1/\epsilon)$ simulation time.

Theorem 3 (Informal): Let the initial condition $x_0 \sim \mathcal{D}$ of system (1b) obey Assumption 1. Also, let the simulation time τ and the number of samples N in Algorithm 1 satisfy

$$\tau \;\geq\; \theta_1 \log(1/\epsilon) \text{ and } N \;\geq\; c\left(1 \;+\; \beta^4\kappa^4\,\theta_1 \log^6 n\right) n$$

for some $\beta > 0$ and desired accuracy $\epsilon > 0$. Then, we can choose the smoothing parameter $r < \theta_3\sqrt{\epsilon}$ in Algorithm 1 and the constant stepsize α such that the random search method (RS) that starts from any initial stabilizing feedback gain $K^0 \in \mathcal{S}_K$ achieves $f(K^k) - f(K^\star) \leq \epsilon$ in at most

$$k \;\leq\; \theta_4 \log\left((f(K^0) \;-\; f(K^\star))/\epsilon\right)$$

iterations with probability not smaller than $1 - c'k(n^{-\beta} + N^{-\beta} + Ne^{-\frac{n}{8}} + e^{-c'N})$. Here, the positive scalars c and c' are absolute constants and $\theta_1, \ldots, \theta_4 > 0$ depend on K^0 and the parameters of the LQR problem (3).

The formal version of Theorem 3 along with a discussion of parameters θ_i and stepsize α is presented in Section VII.

IV. Convex Reparameterization

The main challenge in establishing the exponential stability of (GF) arises from nonconvexity of problem (3). Herein, we use a standard change of variables to reparameterize (3) into a convex problem, for which we can provide exponential stability guarantees for gradient-flow dynamics. We then connect the gradient flow on this convex reparameterization to its nonconvex counterpart and establish the exponential stability of (GF).

A. Change of Variables

The stability of the closed-loop system with the feedback gain $K \in \mathcal{S}_K$ in problem (3) is equivalent to the positive definiteness of the matrix $X(K)$ given by (4a). This condition allows for a standard change of variables $K = YX^{-1}$, for some $Y \in \mathbb{R}^{m\times n}$, to reformulate the LQR design as a convex optimization problem [9]. In particular, for any $K \in \mathcal{S}_K$ and the corresponding matrix X, we have

$$f(K) \;=\; h(X,Y) \;:=\; \mathrm{trace}\,(Q\,X + Y^T R\,Y X^{-1})$$

where $h(X,Y)$ is a jointly convex function of (X,Y) for $X \succ 0$. In the new variables, Lyapunov equation (4b) takes the affine form

$$\mathcal{A}(X) \;-\; \mathcal{B}(Y) \;+\; \Omega \;=\; 0 \tag{8a}$$

where $\mathcal{A}$ and $\mathcal{B}$ are the linear maps

$$\mathcal{A}(X) \;:=\; A\,X \;+\; X A^T,\; \mathcal{B}(Y) \;:=\; B\,Y \;+\; Y^T B^T. \tag{8b}$$

For an invertible map $\mathcal{A}$, we can express the matrix X as an affine function of Y

$$X(Y) \;=\; \mathcal{A}^{-1}(\mathcal{B}(Y) \;-\; \Omega) \tag{8c}$$

and bring the LQR problem into the convex form

$$\underset{Y}{\text{minimize}} \quad h(Y) \tag{9}$$

where $h(Y) := \{h(X(Y),Y),\, Y \in \mathcal{S}_Y; \infty, \text{otherwise}\}$ and $\mathcal{S}_Y := \{Y \in \mathbb{R}^{m\times n} \mid X(Y) \succ 0\}$ is the set of matrices Y that correspond to stabilizing feedback gains $K = YX^{-1}$. The set $\mathcal{S}_Y$ is open and convex because it is defined via a positive-definite condition imposed on the affine map $X(Y)$ in (8c). This positive-definite condition in $\mathcal{S}_Y$ is equivalent to the closed-loop matrix $A - B\,Y(X(Y))^{-1}$ being Hurwitz.

Remark 1: Although our presentation assumes invertibility of $\mathcal{A}$, this assumption comes without loss of generality. As shown in Appendix B, all results carry over to noninvertible $\mathcal{A}$ with an alternative change of variables $A = \hat{A} + BK^0$, $K = \hat{K} + K^0$, and $\hat{K} = \hat{Y}X^{-1}$, for some $K^0 \in \mathcal{S}_K$.

B. Smoothness and Strong Convexity of $h(Y)$

Our convergence analysis of gradient methods for problem (9) relies on the L-smoothness and μ-strong convexity of the function $h(Y)$ over its sublevel sets $\mathcal{S}_Y(a) := \{Y \in \mathcal{S}_Y \mid h(Y) \leq a\}$. These two properties were recently established in [29], where it was shown that over any sublevel set $\mathcal{S}_Y(a)$, the second-order term $\left\langle \tilde{Y}, \nabla^2 h(Y;\tilde{Y})\right\rangle$ in the Taylor series expansion of $h(Y+\tilde{Y})$ around $Y \in \mathcal{S}_Y(a)$ can be upper and lower bounded by quadratic forms $L\|\tilde{Y}\|_F^2$ and $\mu\|\tilde{Y}\|_F^2$ for some positive scalars L and μ. While an explicit form for the smoothness parameter L along with an existence proof for the strong convexity modulus μ was presented in [29], in Proposition 1, we establish an explicit expression for μ in terms of a and parameters of the LQR problem. This allows us to provide bounds on the convergence rate for gradient methods.

Proposition 1: Over any nonempty sublevel set $\mathcal{S}_Y(a)$, the function $h(Y)$ is L-smooth and μ-strongly convex with

$$L \;=\; \frac{2a\|R\|_2}{\nu}\left(1 \;+\; \frac{a\|\mathcal{A}^{-1}\mathcal{B}\|_2}{\sqrt{\nu\lambda_{\min}(R)}}\right)^2 \tag{10a}$$

$$\mu \;=\; \frac{2\lambda_{\min}(R)\lambda_{\min}(Q)}{a\,(1 \;+\; a^2\eta)^2} \tag{10b}$$

where the constants

$$\eta \;:=\; \frac{\|\mathcal{B}\|_2}{\lambda_{\min}(Q)\,\lambda_{\min}(\Omega)\sqrt{\nu\,\lambda_{\min}(R)}} \tag{10c}$$

$$\nu \;:=\; \frac{\lambda_{\min}^2(\Omega)}{4}\left(\frac{\|A\|_2}{\sqrt{\lambda_{\min}(Q)}} + \frac{\|B\|_2}{\sqrt{\lambda_{\min}(R)}}\right)^{-2} \tag{10d}$$

only depend on the problem parameters.

Proof: See Appendix C. ■

C. Gradient Methods Over $\mathcal{S}_Y$

The LQR problem can be solved by minimizing the convex function $h(Y)$, whose gradient is given by [29, Appendix C]

$$\nabla h(Y) \;=\; 2\,RY(X(Y))^{-1} \;-\; 2\,B^T W(Y) \tag{11a}$$

where $W(Y)$ is the solution to

$$A^T W \;+\; WA \;=\; (X(Y))^{-1}Y^T R\,Y(X(Y))^{-1} \;-\; Q. \tag{11b}$$

Using the strong convexity and smoothness properties of $h(Y)$ established in Proposition 1, we next show that the unique minimizer $Y^\star$ of the function $h(Y)$ is the exponentially stable equilibrium point of the gradient-flow dynamics over $\mathcal{S}_Y$

$$\dot{Y} \;=\; -\nabla h(Y), \quad Y(0) \;\in\; \mathcal{S}_Y. \tag{GFY}$$

Proposition 2: For any $Y(0) \in \mathcal{S}_Y$, the gradient-flow dynamics (GFY) are exponentially stable, i.e.,

$$\|Y(t) \;-\; Y^\star\|_F^2 \;\leq\; (L/\mu)\,e^{-2\,\mu\,t}\|Y(0) \;-\; Y^\star\|_F^2$$

where μ and L are the strong convexity and smoothness parameters of the function $h(Y)$ over the sublevel set $\mathcal{S}_Y(h(Y(0)))$.

Proof: The derivative of the Lyapunov function candidate $V(Y) := h(Y) - h(Y^\star)$ along the flow in (GFY) satisfies

$$\dot{V} \;=\; \Big\langle \nabla h(Y), \dot{Y} \Big\rangle \;=\; -\|\nabla h(Y)\|_F^2 \;\le\; -2\,\mu V. \qquad (12)$$

Inequality (12) is a consequence of the strong convexity of $h(Y)$, and it yields [34, Lemma 3.4]

$$V(Y(t)) \;\le\; e^{-2\,\mu\,t}\, V(Y(0)). \qquad (13)$$

Thus, for any $Y(0) \in \mathcal{S}_Y$, $h(Y(t))$ converges exponentially to $h(Y^\star)$. Moreover, since $h(Y)$ is μ-strongly convex and L-smooth, $V(Y)$ can be upper and lower bounded by quadratic functions, and the exponential stability of (GFY) over $\mathcal{S}_Y$ follows from Lyapunov theory [34, Th. 4.10]. ■

In Section V, we use the above result to prove exponential/linear convergence of gradient flow/descent for the nonconvex optimization problem (3). Before we proceed, we note that similar convergence guarantees can be established for the gradient descent method with a sufficiently small stepsize α

$$Y^{k+1} \;:=\; Y^k \;-\; \alpha\,\nabla h(Y^k), \quad Y^0 \;\in\; \mathcal{S}_Y. \qquad \text{(GY)}$$

Since the function $h(Y)$ is L-smooth over the sublevel set $\mathcal{S}_Y(h(Y^0))$, for any $\alpha \in [0, 1/L]$, the iterates Y^k remain within $\mathcal{S}_Y(h(Y^0))$. This property in conjunction with the μ-strong convexity of $h(Y)$ implies that Y^k converges to the optimal solution $Y^\star$ at a linear rate of $\gamma = 1 - \alpha\mu$.

V. Control Design With a Known Model

The asymptotic stability of (GF) is a consequence of the following properties of the LQR objective function [31], [32].

1) The function $f(K)$ is twice continuously differentiable over its open domain $\mathcal{S}_K$ and $f(K) \to \infty$ as $K \to \infty$ and/or $K \to \partial\mathcal{S}_K$.
2) The optimal solution $K^\star$ is the unique equilibrium point over $\mathcal{S}_K$, i.e., $\nabla f(K) = 0$ if and only if $K = K^\star$.

In particular, the derivative of the maximal Lyapunov function candidate $V(K) := f(K) - f(K^\star)$ along the trajectories of (GF) satisfies

$$\dot{V} \;=\; \Big\langle \nabla f(K), \dot{K} \Big\rangle \;=\; -\|\nabla f(K)\|_F^2 \;\le\; 0$$

where the inequality is strict for all $K \neq K^\star$. Thus, the Lyapunov theory [33] implies that, starting from any stabilizing initial condition $K(0)$, the trajectories of (GF) remain within the sublevel set $\mathcal{S}_K(f(K(0)))$ and asymptotically converge to $K^\star$.

Similar arguments were employed for the convergence analysis of the Anderson–Moore algorithm for output-feedback synthesis [31]. While [31] shows global asymptotic stability, it does not provide any information on the rate of convergence. In this section, we first demonstrate exponential stability of (GF) and prove Theorem 1. Then, we establish linear convergence of the gradient descent method (GD) and prove Theorem 2.

A. Gradient-Flow Dynamics: Proof of Theorem 1

We start our proof of Theorem 1 by relating the convex and nonconvex formulations of the LQR objective function. Specifically, in Lemma 1, we establish a relation between the gradients $\nabla f(K)$ and $\nabla h(Y)$ over the sublevel sets of the objective function $\mathcal{S}_K(a) := \{K \in \mathcal{S}_K \,|\, f(K) \le a\}$.

Lemma 1: For any stabilizing feedback gain $K \in \mathcal{S}_K(a)$ and $Y := KX(K)$, we have

$$\|\nabla f(K)\|_F \;\ge\; c\,\|\nabla h(Y)\|_F \qquad (14a)$$

where $X(K)$ is given by (4a), the constant c is determined by

$$c \;=\; \frac{\nu\sqrt{\nu\,\lambda_{\min}(R)}}{2\,a^2\,\|\mathcal{A}^{-1}\|_2\,\|B\|_2 \;+\; a\,\sqrt{\nu\,\lambda_{\min}(R)}} \qquad (14b)$$

and the scalar ν given by (10d) depends on the problem parameters.

Proof: See Appendix D. ■

Using Lemma 1 and the exponential stability of gradient-flow dynamics (GFY) over $\mathcal{S}_Y$, established in Proposition 2, we next show that (GF) is also exponentially stable. In particular, for any stabilizing $K \in \mathcal{S}_K(a)$, the derivative of $V(K) := f(K) - f(K^\star)$ along the gradient flow in (GF) satisfies

$$\dot{V} \;=\; -\|\nabla f(K)\|_F^2 \;\le\; -c^2\,\|\nabla h(Y)\|_F^2 \;\le\; -2\,\mu\,c^2\,V \qquad (15)$$

where $Y = KX(K)$ and the constants c and μ are provided in Lemma 1 and Proposition 1, respectively. The first inequality in (15) follows from (14a) and the second follows from $f(K) = h(Y)$ combined with $\|\nabla h(Y)\|_F^2 \ge 2\mu V$ (which, in turn, is a consequence of the strong convexity of $h(Y)$ established in Proposition 1).

Now, since the sublevel set $\mathcal{S}_K(a)$ is invariant with respect to (GF), following [34, Lemma 3.4], inequality (15) guarantees that system (GF) converges exponentially in the objective value with rate $\rho = 2\mu c^2$. This concludes the proof of part (a) in Theorem 1. In order to prove part (b), we use the following lemma, which connects the errors in the objective value and the optimization variable.

Lemma 2: For any stabilizing feedback gain K, the objective function $f(K)$ in problem (3) satisfies

$$f(K) \;-\; f(K^\star) \;=\; \operatorname{trace}\Big((K - K^\star)^T R\,(K - K^\star)\,X(K)\Big)$$

where $K^\star$ is the optimal solution and $X(K)$ is given by (4a).

Proof: See Appendix D. ■

From Lemma 2 and part (a) of Theorem 1, we have

$$\begin{aligned}\|K(t) \;-\; K^\star\|_F^2 \;&\le\; \frac{f(K(t)) \;-\; f(K^\star)}{\lambda_{\min}(R)\,\lambda_{\min}(X(K(t)))} \\ &\le\; e^{-\rho\,t}\,\frac{f(K(0)) \;-\; f(K^\star)}{\lambda_{\min}(R)\,\lambda_{\min}(X(K(t)))} \\ &\le\; b'\,e^{-\rho\,t}\,\|K(0) \;-\; K^\star\|_F^2\end{aligned}$$

where $b' := \|R\|_2\|X(K(0))\|_2/(\lambda_{\min}(R)\lambda_{\min}(X(K(t))))$. Here, the first and third inequalities follow from basic properties of the matrix trace combined with Lemma 2 applied with $K = K(t)$ and $K = K(0)$, respectively. The second inequality follows from part (a) of Theorem 1.

Finally, to upper bound parameter b', we use Lemma 16 presented in Appendix K that provides the lower and upper bounds $\nu/a \le \lambda_{\min}(X(K))$ and $\|X(K)\|_2 \le a/\lambda_{\min}(Q)$ on the matrix $X(K)$ for any $K \in \mathcal{S}_K(a)$, where the constant ν is given by (10d). Using these bounds and the invariance of $\mathcal{S}_K(a)$ with respect to (GF), we obtain

$$b' \;\le\; b \;:=\; \frac{a^2\,\|R\|_2}{\nu\,\lambda_{\min}(R)\,\lambda_{\min}(Q)} \qquad (16)$$

which completes the proof of part (b).

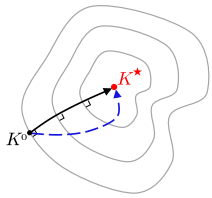


Fig. 1. Trajectories $K(t)$ of (GF) (solid black) and $K_{\rm ind}(t)$ resulting from (18) (dashed blue) along with the level sets of the function $f(K)$.

Remark 2 (Gradient domination): Expression (15) implies that the objective function $f(K)$ over any given sublevel set $\mathcal{S}_K(a)$ satisfies the Polyak–Łojasiewicz (PL) condition [35]

$$\|\nabla f(K)\|_F^2 \;\geq\; 2\,\mu_f\,(f(K) \,-\, f(K^\star)) \tag{17}$$

with parameter $\mu_f := \mu\, c^2$, where μ and c are functions of a that are given by (10b) and (14b), respectively. This condition is also known as gradient dominance, and it has been recently used to show convergence of gradient-based methods for *a discrete-time* LQR problem [11].

B. Geometric Interpretation

The solution $Y(t)$ to gradient-flow dynamics (GFY) over the set $\mathcal{S}_Y$ induces the trajectory

$$K_{\rm ind}(t) \;:=\; Y(t)(X(Y(t)))^{-1} \tag{18}$$

over the set of stabilizing feedback gains $\mathcal{S}_K$, where the affine function $X(Y)$ is given by (8c). The induced trajectory $K_{\rm ind}(t)$ can be viewed as the solution to the differential equation

$$\dot{K} \;=\; g(K) \tag{19a}$$

where $g\colon \mathcal{S}_K \to \mathbb{R}^{m\times n}$ is given by

$$g(K) \;:=\; \big(K\mathcal{A}^{-1}(\mathcal{B}(\nabla h(Y(K)))) - \nabla h(Y(K))\big)\,(X(K))^{-1}. \tag{19b}$$

Here, the matrix $X = X(K)$ is given by (4a) and $Y(K) = KX(K)$. System (19) is obtained by differentiating both sides of (18) with respect to time t and applying the chain rule. Fig. 1 illustrates an induced trajectory $K_{\rm ind}(t)$ and a trajectory $K(t)$ resulting from gradient-flow dynamics (GF) that starts from the same initial condition.

Moreover, using the definition of $h(Y)$, we have

$$h(Y(t)) \;=\; f(K_{\rm ind}(t)). \tag{20}$$

Thus, the exponential decay of $h(Y(t))$ established in Proposition 2 implies that f decays exponentially along the vector field g, i.e., for $K_{\rm ind}(0) \neq K^\star$, we have

$$\frac{f(K_{\rm ind}(t)) \,-\, f(K^\star)}{f(K_{\rm ind}(0)) \,-\, f(K^\star)} \;=\; \frac{h(Y(t)) \,-\, h(Y^\star)}{h(Y(0)) \,-\, h(Y^\star)} \;\leq\; e^{-2\,\mu\, t}.$$

This inequality follows from inequality (13), where μ denotes the strong convexity modulus of the function $h(Y)$ over the sublevel set $\mathcal{S}_Y(h(Y(0)))$ (see Proposition 1). Herein, we provide a geometric interpretation of the exponential decay of f under the trajectories of (GF) that is based on the relation between the vector fields g and $-\nabla f$.

Differentiating both sides of (20) with respect to t yields

$$\|\nabla h(Y)\|^2 = \langle -\nabla f(K), g(K)\rangle\,. \tag{21}$$

Thus, for each $K \in \mathcal{S}_K$, the inner product between the vector fields $-\nabla f(K)$ and $g(K)$ is nonnegative. However, this is not sufficient to ensure exponential decay of f along (GF). To address this challenge, our proof utilizes inequality (14a) in Lemma 1. Based on (21), (14a) can be equivalently restated as

$$\frac{\|-\nabla f(K)\|_F}{\|\Pi_{-\nabla f(K)}(g(K))\|_F} \;=\; \frac{\|\nabla f(K)\|_F^2}{\langle -\nabla f(K), g(K)\rangle} \;\geq\; c^2$$

where $\Pi_b(a)$ denotes the projection of a onto b. Thus, Lemma 1 ensures that the ratio between the norm of the vector field $-\nabla f(K)$ associated with gradient-flow dynamics (GF) and the norm of the projection of $g(K)$ onto $-\nabla f(K)$ is uniformly lower bounded by a positive constant. This lower bound is the key geometric feature that allows us to deduce exponential decay of f along the vector field $-\nabla f$ from the exponential decay of the vector field g.

C. Gradient Descent: Proof of Theorem 2

Given the exponential stability of gradient-flow dynamics (GF) established in Theorem 1, the convergence analysis of gradient descent (GD) amounts to finding a suitable stepsize α. Lemma 3 provides a Lipschitz continuity parameter for $\nabla f(K)$, which facilitates finding such a stepsize.

Lemma 3: Over any nonempty sublevel set $\mathcal{S}_K(a)$, the gradient $\nabla f(K)$ is Lipschitz continuous with parameter

$$L_f := \frac{2a\|R\|_2}{\lambda_{\min}(Q)} + \frac{8a^3\|B\|_2}{\lambda_{\min}^2(Q)\lambda_{\min}(\Omega)} \times \left(\frac{\|B\|_2}{\lambda_{\min}(\Omega)} + \frac{\|R\|_2}{\sqrt{\nu\lambda_{\min}(R)}}\right)$$

where ν given by (10d) depends on the problem parameters.

Proof: See Appendix D. ∎

Let $K_\alpha := K - \alpha\nabla f(K)$, $\alpha \geq 0$ parameterize the half-line starting from $K \in \mathcal{S}_K(a)$ with $K \neq K^\star$ along $-\nabla f(K)$, and let us define the scalar $\beta_m := \max \beta$ such that $K_\alpha \in \mathcal{S}_K(a)$, for all $\alpha \in [0, \beta]$. The existence of β_m follows from the compactness of $\mathcal{S}_K(a)$ [31]. We next show that $\beta_m \geq 2/L_f$.

For the sake of contradiction, suppose $\beta_m < 2/L_f$. From the continuity of $f(K_\alpha)$ with respect to α, it follows that $f(K_{\beta_m}) = a$. Moreover, since $-\nabla f(K)$ is a descent direction of the function $f(K)$, we have $\beta_m > 0$. Thus, for $\alpha \in (0, \beta_m]$,

$$f(K_\alpha) \,-\, f(K) \;\leq\; -\frac{\alpha(2 \,-\, L_f\alpha)}{2}\,\|\nabla f(K)\|_F^2 \;<\; 0.$$

Here, the first inequality follows from the L_f-smoothness of $f(K)$ over $\mathcal{S}_K(a)$ (Descent Lemma [36, eq. (9.17)]) and the second inequality follows from $\nabla f(K) \neq 0$ in conjunction with $\beta_m \in (0, 2/L_f)$. This implies $f(K_{\beta_m}) < f(K) \leq a$, which contradicts $f(K_{\beta_m}) = a$. Thus, $\beta_m \geq 2/L_f$.

We can now use induction on k to show that, for any stabilizing initial condition $K^0 \in \mathcal{S}_K(a)$, the iterates of (GD) with $\alpha \in [0, 2/L_f]$ remain in $\mathcal{S}_K(a)$ and satisfy

$$f(K^{k+1}) \,-\, f(K^k) \;\leq\; -\frac{\alpha(2 \,-\, L_f\alpha)}{2}\,\|\nabla f(K^k)\|_F^2. \tag{22}$$

Inequality (22) in conjunction with the PL condition (17) evaluated at K^k guarantees linear convergence for gradient descent (GD) with the rate $\gamma \leq 1 - \alpha\,\mu_f$ for all $\alpha \in (0, 1/L_f]$, where μ_f is the PL parameter of the function $f(K)$. This completes the proof of part (a) of Theorem 2.

Using part (a) and Lemma 2, we can make a similar argument to what we used for the proof of Theorem 1 to establish part (b) with constant b in (16). We omit the details for brevity.

Remark 3: Using our results, it is straightforward to show linear convergence of $K^{k+1} = K^k - \alpha H_1^k \nabla f(K^k) H_2^k$ with $K^0 \in \mathcal{S}_K$ and small enough stepsize, where H_1^k and H_2^k are uniformly upper- and lower-bounded positive-definite matrices. In particular, the Kleinman iteration [17] is recovered for $\alpha = 0.5$, $H_1^k = R^{-1}$, and $H_2^k = (X(K^k))^{-1}$. Similarly, convergence of gradient descent may be improved by choosing $H_1^k = I$ and $H_2^k = (X(K^k))^{-1}$. In this case, the corresponding update direction provides the continuous-time variant of the so-called *natural gradient* for discrete-time systems [37].

VI. Bias and Correlation in Gradient Estimation

In the model-free setting, we do not have access to the gradient $\nabla f(K)$ and the random search method (RS) relies on the gradient estimate $\overline{\nabla} f(K)$ resulting from Algorithm 1. According to [11], achieving $\|\overline{\nabla} f(K) - \nabla f(K)\|_F \leq \epsilon$ may take $N = \Omega(1/\epsilon^4)$ samples using one-point gradient estimates. Our computational experiments (not included in this article) also suggest that to achieve $\|\overline{\nabla} f(K) - \nabla f(K)\|_F \leq \epsilon$, N must scale as $\text{poly}\,(1/\epsilon)$ even when a two-point gradient estimate is used. To avoid this poor sample complexity, in our proof, we take an alternative route and give up on the objective of controlling the gradient estimation error. By exploiting the problem structure, we show that with a linear number of samples $N = \tilde{O}(n)$, where n is the number of states, the estimate $\overline{\nabla} f(K)$ concentrates with *high probability* when projected to the direction of $\nabla f(K)$.

Our proof strategy allows us to significantly improve upon the existing literature both in terms of the required function evaluations and simulation time. Specifically, using the random search method (RS), the total number of function evaluations required in our results to achieve an accuracy level ϵ is proportional to $\log(1/\epsilon)$ compared to at least $(1/\epsilon^4)\log(1/\epsilon)$ in [11] and $1/\epsilon$ in [12]. Similarly, the simulation time that we require to achieve an accuracy level ϵ is proportional to $\log(1/\epsilon)$; this is in contrast to $\text{poly}\,(1/\epsilon)$ simulation times in [11] and infinite simulation time in [12].

Algorithm 1 produces a biased estimate $\overline{\nabla} f(K)$ of the gradient $\nabla f(K)$. Herein, we first introduce an unbiased estimate $\widehat{\nabla} f(K)$ of $\nabla f(K)$ and establish that the distance $\|\widehat{\nabla} f(K) - \overline{\nabla} f(K)\|_F$ can be readily controlled by choosing a large simulation time τ and an appropriate smoothing parameter r in Algorithm 1; we call this distance the estimation bias. Next, we show that with $N = \tilde{O}(n)$ samples, the unbiased estimate $\widehat{\nabla} f(K)$ becomes highly correlated with $\nabla f(K)$. We exploit this fact in our convergence analysis.

A. Bias in Gradient Estimation due to Finite Simulation Time

We first introduce an unbiased estimate of the gradient that is used to quantify the bias. For any $\tau \geq 0$ and $x_0 \in \mathbb{R}^n$, let

$$f_{x_0,\tau}(K) \;:=\; \int_0^\tau \left(x^T(t) Q x(t) \,+\, u^T(t) R u(t)\right) dt$$

denote the τ-truncated version of the LQR objective function associated with system (1b) with the initial condition $x(0) = x_0$ and feedback law $u = -Kx$ for all $K \in \mathbb{R}^{m\times n}$. Note that for any $K \in \mathcal{S}_K$ and $x(0) = x_0 \in \mathbb{R}^n$, the infinite-horizon cost

$$f_{x_0}(K) \;:=\; f_{x_0,\infty}(K) \tag{23a}$$

exists and it satisfies $f(K) = \mathbb{E}_{x_0}[f_{x_0}(K)]$. Furthermore, the gradient of $f_{x_0}(K)$ is given by [cf. (5)]

$$\nabla f_{x_0}(K) \;=\; 2\left(RK \,-\, B^T P(K)\right) X_{x_0}(K) \tag{23b}$$

where $X_{x_0}(K) = -\mathcal{A}_K^{-1}(x_0 x_0^T)$ is determined by the closed-loop Lyapunov operator in (7) and $P(K) = -(\mathcal{A}_K^*)^{-1}(Q + K^T R K)$. Note that the gradients $\nabla f(K)$ and $\nabla f_{x_0}(K)$ are linear in $X(K) = -\mathcal{A}_K^{-1}(\Omega)$ and $X_{x_0}(K)$, respectively. Thus, for any zero-mean random initial condition $x(0) = x_0$ with covariance $\mathbb{E}[x_0 x_0^T] = \Omega$, the linearity of the closed-loop Lyapunov operator $\mathcal{A}_K$ implies

$$\mathbb{E}_{x_0}[X_{x_0}(K)] \;=\; X(K),\; \mathbb{E}_{x_0}[\nabla f_{x_0}(K)] \;=\; \nabla f(K).$$

Let us define the following three estimates of the gradient:

$$\begin{aligned}
\overline{\nabla} f(K) &:= \frac{1}{2rN} \sum_{i=1}^N \left(f_{x_i,\tau}(K + rU_i) - f_{x_i,\tau}(K - rU_i)\right) U_i \\
\widetilde{\nabla} f(K) &:= \frac{1}{2rN} \sum_{i=1}^N \left(f_{x_i}(K + rU_i) - f_{x_i}(K - rU_i)\right) U_i \\
\widehat{\nabla} f(K) &:= \frac{1}{N} \sum_{i=1}^N \langle \nabla f_{x_i}(K), U_i \rangle\, U_i
\end{aligned} \tag{24}$$

where $U_i \in \mathbb{R}^{m\times n}$ are i.i.d. random matrices with $\text{vec}(U_i)$ uniformly distributed on the sphere $\sqrt{mn}\, S^{mn-1}$ and $x_i \in \mathbb{R}^n$ are i.i.d. initial conditions sampled from distribution $\mathcal{D}$. Here, $\widetilde{\nabla} f(K)$ is the infinite-horizon version of the output $\overline{\nabla} f(K)$ of Algorithm 1 and $\widehat{\nabla} f(K)$ provides an unbiased estimate of $\nabla f(K)$. To see this, note that by the independence of U_i and x_i, we have

$$\mathbb{E}_{x_i,U_i}\left[\text{vec}(\widehat{\nabla} f(K))\right] \;=\; \mathbb{E}_{U_1}[\langle \nabla f(K), U_1\rangle \,\text{vec}(U_1)] \;=$$

$$\mathbb{E}_{U_1}[\text{vec}(U_1)\text{vec}(U_1)^T]\text{vec}(\nabla f(K)) \;=\; \text{vec}(\nabla f(K))$$

and thus $\mathbb{E}[\widehat{\nabla} f(K)] = \nabla f(K)$. Here, we have utilized the fact that for the uniformly distributed random variable $\text{vec}(U_1)$ over the sphere $\sqrt{mn}\, S^{mn-1}$, $\mathbb{E}_{U_1}[\text{vec}(U_1)\text{vec}(U_1)^T] = I$.

1) Local Boundedness of the Function $f(K)$: An important requirement for the gradient estimation scheme in Algorithm 1 is the stability of the perturbed closed-loop systems, i.e., $K \pm rU_i \in \mathcal{S}_K$; violating this condition leads to an exponential growth of the state and control signals. Moreover, this condition is necessary and sufficient for $\widetilde{\nabla} f(K)$ to be well defined. In Proposition 3, we establish a radius, within which any perturbation of $K \in \mathcal{S}_K$ remains stabilizing.

Proposition 3: For any stabilizing feedback gain $K \in \mathcal{S}_K$, we have $\{\hat{K} \in \mathbb{R}^{m\times n} \,|\, \|\hat{K} - K\|_2 < \zeta\} \subset \mathcal{S}_K$, where $\zeta := \lambda_{\min}(\Omega)/(2\,\|B\|_2\,\|X(K)\|_2)$ and $X(K)$ is given by (4a).

Proof: See Appendix E. ∎

If we choose the parameter r in Algorithm 1 to be smaller than ζ, then the sample feedback gains $K \pm rU_i$ are all stabilizing. In this article, we further require that the parameter r is small enough so that $K \pm rU_i \in \mathcal{S}_K(2a)$ for all $K \in \mathcal{S}_K(a)$. Such an upper bound on r is provided in the next lemma.

Lemma 4: For any $U \in \mathbb{R}^{m\times n}$ with $\|U\|_F \leq \sqrt{mn}$ and $K \in \mathcal{S}_K(a)$, $K + r(a)U \in \mathcal{S}_K(2a)$, where $r(a) := \tilde{c}/a$ for some positive constant $\tilde{c}$ that depends on the problem data.

Proof: See Appendix E. ∎

Note that for any $K \in \mathcal{S}_K(a)$, and $r \leq r(a)$ in Lemma 4, $\widetilde{\nabla} f(K)$ is well defined because $K + rU_i \in \mathcal{S}_K(2a)$ for all i.

2) Bounding the Bias: Herein, we establish an upper bound on the difference between the output $\overline{\nabla} f(K)$ of Algorithm 1 and the unbiased estimate $\widehat{\nabla} f(K)$ of the gradient $\nabla f(K)$. This is accomplished by bounding the difference between these two quantities and $\widetilde{\nabla} f(K)$ through the use of the triangle inequality

$$\|\widehat{\nabla} f(K) - \overline{\nabla} f(K)\|_F \leq \|\widetilde{\nabla} f(K) - \overline{\nabla} f(K)\|_F + \|\widehat{\nabla} f(K) - \widetilde{\nabla} f(K)\|_F. \quad (25)$$

The first term on the right-hand side of (25) arises from a bias caused by the finite simulation time in Algorithm 1. The next proposition quantifies an upper bound on this term.

Proposition 4: For any $K \in \mathcal{S}_K(a)$, the output of Algorithm 1 with parameter $r \leq r(a)$ (given by Lemma 4) satisfies

$$\|\widetilde{\nabla} f(K) - \overline{\nabla} f(K)\|_F \leq \frac{\sqrt{mn}\max_i \|x_i\|^2}{r} \kappa_1(2a)\, e^{-\kappa_2(2a)\tau}$$

where $\kappa_1(a) > 0$ is a degree-5 polynomial and $\kappa_2(a) > 0$ is inversely proportional to a, and they are given by (46).

Proof: See Appendix F. ∎

Although small values of r may result in a large error $\|\widetilde{\nabla} f(K) - \overline{\nabla} f(K)\|_F$, the exponential dependence of the upper bound in Proposition 4 on the simulation time τ implies that this error can be readily controlled by increasing τ. In the next proposition, we handle the second term in (25).

Proposition 5: For any $K \in \mathcal{S}_K(a)$ and $r \leq r(a)$ (given by Lemma 4), we have

$$\|\widehat{\nabla} f(K) - \widetilde{\nabla} f(K)\|_F \leq \frac{(rmn)^2}{2} \ell(2a) \max_i \|x_i\|^2$$

where the function $\ell(a) > 0$ is a degree-4 polynomial, and it is given by (49).

Proof: See Appendix G. ∎

The third derivatives of the functions $f_{x_i}(K)$ are utilized in the proof of Proposition 5. It is also worth noting that unlike $\overline{\nabla} f(k)$ and $\widetilde{\nabla} f(K)$, the unbiased gradient estimate $\widehat{\nabla} f(K)$ is independent of the parameter r. Thus, Proposition 5 provides a quadratic upper bound on the estimation error in terms of r.

B. Correlation Between Gradient and Gradient Estimate

As mentioned earlier, one approach to analyzing convergence for the random search method in (RS) is to control the gradient estimation error $\overline{\nabla} f(K) - \nabla f(K)$ by choosing a large number of samples N. For the one-point gradient estimation setting, this approach was taken in [11] for the discrete-time LQR (and in [38] for the continuous-time LQR) and has led to an upper bound on the required number of samples for reaching ϵ-accuracy that grows at least proportionally to $1/\epsilon^4$. Alternatively, our proof exploits the problem structure and shows that with a linear number of samples $N = \tilde{O}(n)$, where n is the number of states, the gradient estimate $\widehat{\nabla} f(K)$ concentrates with *high probability* when projected to the direction of $\nabla f(K)$. In particular, in Propositions 7 and 8, we show that the following events occur with high probability for some positive scalars μ_1, μ_2,

$$\mathsf{M}_1 := \left\{ \left\langle \widehat{\nabla} f(K), \nabla f(K) \right\rangle \geq \mu_1 \|\nabla f(K)\|_F^2 \right\} \quad (26a)$$

$$\mathsf{M}_2 := \left\{ \|\widehat{\nabla} f(K)\|_F^2 \leq \mu_2 \|\nabla f(K)\|_F^2 \right\}. \quad (26b)$$

To justify the definitions of these events, we first show that if they both take place, then the unbiased estimate $\widehat{\nabla} f(K)$ can be used to decrease the objective error by a geometric factor.

Proposition 6 (Approximate GD): If the matrix $G \in \mathbb{R}^{m\times n}$ and the feedback gain $K \in \mathcal{S}_K(a)$ are such that

$$\langle G, \nabla f(K)\rangle \geq \mu_1 \|\nabla f(K)\|_F^2 \quad (27a)$$

$$\|G\|_F^2 \leq \mu_2 \|\nabla f(K)\|_F^2 \quad (27b)$$

for some positive scalars μ_1 and μ_2, then $K - \alpha G \in \mathcal{S}_K(a)$ for all $\alpha \in [0, \mu_1/(\mu_2 L_f)]$, and

$$f(K - \alpha G) - f(K^\star) \leq \gamma\,(f(K) - f(K^\star))$$

with $\gamma = 1 - \mu_f \mu_1 \alpha$. Here, L_f and μ_f are the smoothness and the PL parameters of the function f over $\mathcal{S}_K(a)$.

Proof: See Appendix H. ∎

Remark 4: The fastest convergence rate guaranteed by Proposition 6, $\gamma = 1 - \mu_f \mu_1^2/(L_f \mu_2)$, is achieved with the stepsize $\alpha = \mu_1/(\mu_2 L_f)$. This rate bound is tight in the sense that if $G = c\nabla f(K)$, for some $c > 0$, we recover the standard convergence rate $\gamma = 1 - \mu_f/L_f$ of gradient descent.

We next quantify the probability of the events M_1 and M_2. In our proofs, we exploit modern nonasymptotic statistical analysis of the concentration of random variables around their average. While in Appendix J, we set notation and provide basic definitions of key concepts, we refer the reader to a recent book [39] for a comprehensive discussion. Herein, we use c, c', c'', etc., to denote positive absolute constants.

1) Handling M_1: We first exploit the problem structure to confine the dependence of $\widehat{\nabla} f(K)$ on the random initial conditions x_i into a zero-mean random vector. In particular, for any $K \in \mathcal{S}_K$ and $x_0 \in \mathbb{R}^n$,

$$\nabla f(K) = E\,X,\ \nabla f_{x_0}(K) = E\,X_{x_0}$$

where $E := 2(RK - B^T P(K)) \in \mathbb{R}^{m\times n}$ is a fixed matrix, $X = -\mathcal{A}_K^{-1}(\Omega)$, and $X_{x_0} = -\mathcal{A}_K^{-1}(x_0 x_0^T)$. This allows us to represent the unbiased estimate $\widehat{\nabla} f(K)$ of the gradient as

$$\widehat{\nabla} f(K) = \frac{1}{N}\sum_{i=1}^N \langle E X_{x_i}, U_i\rangle\, U_i = \widehat{\nabla}_1 + \widehat{\nabla}_2 \quad (28a)$$

$$\widehat{\nabla}_1 = \frac{1}{N}\sum_{i=1}^N \langle E(X_{x_i} - X), U_i\rangle\, U_i \quad (28b)$$

$$\widehat{\nabla}_2 = \frac{1}{N}\sum_{i=1}^N \langle \nabla f(K), U_i\rangle\, U_i. \quad (28c)$$

Note that $\widehat{\nabla}_2$ does not depend on the initial conditions x_i. Moreover, from $\mathbb{E}[X_{x_i}] = X$ and the independence of X_{x_i} and U_i, we have $\mathbb{E}[\widehat{\nabla}_1] = 0$ and $\mathbb{E}[\widehat{\nabla}_2] = \nabla f(K)$.

In Lemma 5, we show that $\left\langle \widehat{\nabla}_1, \nabla f(K) \right\rangle$ can be made arbitrary small with a large number of samples N. This allows us to analyze the probability of the event M_1 in (26).

Lemma 5: Let $U_1, \ldots, U_N \in \mathbb{R}^{m\times n}$ be i.i.d. random matrices with each $\mathrm{vec}(U_i)$ uniformly distributed on the sphere $\sqrt{mn}\, S^{mn-1}$, and let $X_1, \ldots, X_N \in \mathbb{R}^{n\times n}$ be i.i.d. random matrices distributed according to $\mathcal{M}(xx^T)$. Here, $\mathcal{M}$ is a linear operator and $x \in \mathbb{R}^n$ is a random vector, whose entries are i.i.d., zero-mean, unit-variance, sub-Gaussian random variables with sub-Gaussian norm less than κ. For any fixed matrix $E \in \mathbb{R}^{m\times n}$

and positive scalars δ and β, if

$$N \geq C\,(\beta^2\kappa^2/\delta)^2\,(\|\mathcal{M}^*\|_2 + \|\mathcal{M}^*\|_S)^2\,n\log^6 n \qquad (29)$$

then, with probability not smaller than $1 - C'N^{-\beta} - 4Ne^{-\frac{n}{8}}$,

$$\left|\frac{1}{N}\sum_{i=1}^{N}\langle E\,(X_i - X)\,, U_i\rangle\,\langle EX, U_i\rangle\right| \leq \delta\|EX\|_F\|E\|_F$$

where $X := \mathbb{E}[X_1] = \mathcal{M}(I)$.

Proof: See Appendix I. ∎

In Lemma 6, we show that $\left\langle \widehat{\nabla}_2, \nabla f(K)\right\rangle$ concentrates with high probability around its average $\|\nabla f(K)\|_F^2$.

Lemma 6: Let $U_1, \ldots, U_N \in \mathbb{R}^{m\times n}$ be i.i.d. random matrices with each $\mathrm{vec}(U_i)$ uniformly distributed on the sphere $\sqrt{mn}\,S^{mn-1}$. Then, for any $W \in \mathbb{R}^{m\times n}$ and $t \in (0, 1]$,

$$\mathbb{P}\left\{\frac{1}{N}\sum_{i=1}^{N}\langle W, U_i\rangle^2 < (1-t)\|W\|_F^2\right\} \leq 2\,e^{-cNt^2}.$$

Proof: See Appendix I. ∎

In Proposition 7, we use Lemmas 5 and 6 to address M_1.

Proposition 7: Under Assumption 1, for any stabilizing feedback gain $K \in \mathcal{S}_K$ and positive scalar β, if

$$N \geq C_1\,\frac{\beta^4\kappa^4}{\lambda_{\min}^2(X)}\left(\|(\mathcal{A}_K^*)^{-1}\|_2 + \|(\mathcal{A}_K^*)^{-1}\|_S\right)^2 n\log^6 n$$

then the event M_1 in (26) with $\mu_1 := 1/4$ satisfies $\mathbb{P}(\mathsf{M}_1) \geq 1 - C_2N^{-\beta} - 4Ne^{-\frac{n}{8}} - 2e^{-C_3N}$.

Proof: We use Lemma 5 with $\delta := \lambda_{\min}(X)/4$ to show that, with probability not smaller than $1 - C'N^{-\beta} - 4Ne^{-n/8}$,

$$\begin{aligned}\left|\left\langle \widehat{\nabla}_1, \nabla f(K)\right\rangle\right| &\leq \delta\,\|EX\|_F\,\|E\|_F \\ &\leq \frac{1}{4}\,\|EX\|_F^2 \;=\; \frac{1}{4}\,\|\nabla f(K)\|_F^2.\end{aligned} \qquad (30a)$$

Furthermore, Lemma 6 with $t := 1/2$ implies that

$$\left\langle \widehat{\nabla}_2, \nabla f(K)\right\rangle \;\geq\; \frac{1}{2}\,\|\nabla f(K)\|_F^2 \qquad (30b)$$

holds with probability not smaller than $1 - 2e^{-cN}$. Since $\widehat{\nabla} f(K) = \widehat{\nabla}_1 + \widehat{\nabla}_2$, we can use a union bound to combine (30a) and (30b). This together with a triangle inequality completes the proof. ∎

2) Handling M_2: In Lemma 7, we quantify a high probability upper bound on $\|\widehat{\nabla}_1\|_F/\|\nabla f(K)\|$. This lemma is analogous to Lemma 5, and it allows us to analyze the probability of the event M_2 in (26).

Lemma 7: Let X_i and U_i with $i = 1, \ldots, N$ be random matrices defined in Lemma 5, $X := \mathbb{E}[X_1]$, and let $N \geq c_0 n$. Then, for any $E \in \mathbb{R}^{m\times n}$ and positive scalar β,

$$\frac{1}{N}\,\|\sum_{i=1}^{N}\langle E\,(X_i - X)\,, U_i\rangle\,U_i\|_F \;\leq\; c_1\beta\,\kappa^2(\|\mathcal{M}^*\|_2 + \|\mathcal{M}^*\|_S)\|E\|_F\sqrt{mn}\,\log n$$

with probability not smaller than $1 - c_2(n^{-\beta} + Ne^{-\frac{n}{8}})$.

Proof: See Appendix J. ∎

In Lemma 8, we quantify a high-probability upper bound on $\|\widehat{\nabla}_2\|_F/\|\nabla f(K)\|$.

Lemma 8: Let $U_1, \ldots, U_N \in \mathbb{R}^{m\times n}$ be i.i.d. random matrices with $\mathrm{vec}(U_i)$ uniformly distributed on the sphere $\sqrt{mn}\,S^{mn-1}$ and let $N \geq Cn$. Then, for any $W \in \mathbb{R}^{m\times n}$,

$$\mathbb{P}\left\{\frac{1}{N}\,\|\sum_{j=1}^{N}\langle W, U_j\rangle\,U_j\|_F > C'\sqrt{m}\|W\|_F\right\} \leq 2Ne^{-\frac{mn}{8}} + 2e^{-\hat{c}N}.$$

Proof: See Appendix J. ∎

In Proposition 8, we use Lemmas 7 and 8 to address M_2.

Proposition 8: Let Assumption 1 hold. Then, for any $K \in \mathcal{S}_K$, scalar $\beta > 0$, and $N \geq C_4 n$, the event M_2 in (26) with

$$\mu_2 := C_5\left(\beta\kappa^2\,\frac{\|(\mathcal{A}_K^*)^{-1}\|_2 + \|(\mathcal{A}_K^*)^{-1}\|_S}{\lambda_{\min}(X)}\sqrt{mn}\log n + \sqrt{m}\right)^2$$

satisfies $\mathbb{P}(\mathsf{M}_2) \geq 1 - C_6(n^{-\beta} + Ne^{-\frac{n}{8}} + e^{-C_7N})$.

Proof: We use Lemma 7 to show that, with probability at least $1 - c_2(n^{-\beta} + Ne^{-\frac{n}{8}})$, $\widehat{\nabla}_1$ satisfies

$$\begin{aligned}\|\widehat{\nabla}_1\|_F \;&\leq\; c_1\beta\kappa^2(\|(\mathcal{A}_K^*)^{-1}\|_2 + \|(\mathcal{A}_K^*)^{-1}\|_S)\|E\|_F\sqrt{mn}\,\log n \;\leq \\ &c_1\beta\kappa^2\,\frac{\|(\mathcal{A}_K^*)^{-1}\|_2 + \|(\mathcal{A}_K^*)^{-1}\|_S}{\lambda_{\min}(X)}\,\|\nabla f(K)\|_F\sqrt{mn}\,\log n.\end{aligned}$$

Furthermore, we can use Lemma 8 to show that, with probability not smaller than $1 - 2Ne^{-\frac{mn}{8}} - 2e^{-\hat{c}N}$, $\widehat{\nabla}_2$ satisfies

$$\|\widehat{\nabla}_2\|_F \;\leq\; C'\sqrt{m}\|\nabla f(K)\|_F. \qquad \text{(VI-B2)}$$

Now, since $\widehat{\nabla} f(K) = \widehat{\nabla}_1 + \widehat{\nabla}_2$, we can use a union bound to combine the last two inequalities. This together with a triangle inequality completes the proof. ∎

VII. Model-Free Control Design

In this section, we prove a more formal version of Theorem 3.

Theorem 4: Consider the random search method (RS) that uses the gradient estimates of Algorithm 1 for finding the optimal solution $K^\star$ of LQR problem (3). Let the initial condition x_0 obey Assumption 1 and let the simulation time τ, the smoothing constant r, and the number of samples N satisfy $\tau \geq \theta'(a)\log(1/(r\epsilon))$, $r < \min\{r(a), \theta''(a)\sqrt{\epsilon}\}$, and

$$N \geq c_1(1 + \beta^4\kappa^4\,\theta(a)\log^6 n)\,n \qquad (31)$$

for some $\beta > 0$ and a desired accuracy $\epsilon > 0$. Then, for any initial condition $K^0 \in \mathcal{S}_K(a)$, (RS) with the constant stepsize $\alpha \leq 1/(32\mu_2(a)L_f)$ achieves $f(K^k) - f(K^\star) \leq \epsilon$ with probability not smaller than $1 - kp - 2kNe^{-n}$ in at most

$$k \;\leq\; \log\left((f(K^0) - f(K^\star))/\epsilon\right)/\log\left(1/(1 - \mu_f(a)\alpha/8)\right)$$

iterations. Here, $p := c_2(n^{-\beta} + N^{-\beta} + Ne^{-\frac{n}{8}} + e^{-c_3\,N})$; $\mu_2 := c_4(\sqrt{m} + \beta\kappa^2\theta(a)\sqrt{mn}\log n)^2$; $c_1, \ldots, c_4$ are positive absolute constants; μ_f and L_f are the PL and smoothness parameters of the function f over the sublevel set $\mathcal{S}_K(a)$; θ, θ', θ'' are positive functions that depend only on the parameters of the LQR problem; and $r(a)$ is given by Lemma 4.

Proof: The proof combines Propositions 4–8. We first show that for any $r \leq r(a)$ and $\tau > 0$,

$$\|\overline{\nabla} f(K) - \widehat{\nabla} f(K)\|_F \;\leq\; \sigma \qquad (32)$$

with probability not smaller than $1 - 2Ne^{-n}$, where

$$\sigma := c_5(\kappa^2 + 1)\left(\frac{n\sqrt{m}}{r}\kappa_1(2a)e^{-\kappa_2(2a)\tau} + \frac{r^2\,m^2n^{\frac{5}{2}}}{2}\ell(2a)\right).$$

Here, $r(a)$, $\kappa_i(a)$, and $\ell(a)$ are positive functions that are given by Lemma 4, (46), and (49), respectively.

Under Assumption 1, the vector $v \sim \mathcal{D}$ satisfies [39, eq. (3.3)], $\mathbb{P}\{\|v\| \leq c_5(\kappa^2+1)\sqrt{n}\} \geq 1-2e^{-n}$. Thus, for the random initial conditions $x_1, \ldots, x_N \sim \mathcal{D}$, we can apply the union bound (Boole's inequality) to obtain

$$\mathbb{P}\left\{\max_i \|x_i\| \;\leq\; c_5(\kappa^2+1)\sqrt{n}\right\} \;\geq\; 1 - 2Ne^{-n}. \quad (33)$$

Now, we combine Propositions 4 and 5 to write

$$\|\overline{\nabla} f(K) - \widehat{\nabla} f(K)\|_F \;\leq$$
$$\left(\frac{\sqrt{mn}}{r}\,\kappa_1(2a)e^{-\kappa_2(2a)\tau} + \frac{(rmn)^2}{2}\,\ell(2a)\right)\max_i \|x_i\|^2 \leq \sigma.$$

The first inequality is obtained by combining Propositions 4 and 5 through the use of the triangle inequality, and the second inequality follows from (33). This completes the proof of (32).

Let $\theta(a)$ be a uniform upper bound on $(\|(\mathcal{A}_K^*)^{-1}\|_2 + \|(\mathcal{A}_K^*)^{-1}\|_S)/\lambda_{\min}(X) \leq \theta(a)$, for all $K \in \mathcal{S}_K(a)$; see Appendix L for a discussion on $\theta(a)$. Since the number of samples satisfies (31), for any given $K \in \mathcal{S}_K(a)$, we can combine Propositions 7 and 8 with a union bound to show that

$$\left\langle \widehat{\nabla} f(K), \nabla f(K) \right\rangle \;\geq\; \mu_1 \|\nabla f(K)\|_F^2 \quad (34a)$$

$$\|\widehat{\nabla} f(K)\|_F^2 \;\leq\; \mu_2 \|\nabla f(K)\|_F^2 \quad (34b)$$

holds with probability not smaller than $1-p$, where $\mu_1 = 1/4$, and μ_2 and p are determined in the statement of the theorem.

Without loss of generality, let us assume that the initial error satisfies $f(K^0) - f(K^\star) > \epsilon$. We next show that

$$\left\langle \overline{\nabla} f(K^0), \nabla f(K^0) \right\rangle \;\geq\; \frac{\mu_1}{2}\|\nabla f(K^0)\|_F^2 \quad (35a)$$

$$\|\overline{\nabla} f(K^0)\|_F^2 \;\leq\; 4\mu_2 \|\nabla f(K^0)\|_F^2 \quad (35b)$$

holds with probability not smaller than $1 - p - 2Ne^{-n}$.

Since the function f is gradient dominant over the sublevel set $\mathcal{S}_K(a)$ with parameter μ_f, combining $f(K^0) - f(K^\star) > \epsilon$ and (17) yields $\|\nabla f(K^0)\|_F \geq \sqrt{2\mu_f \epsilon}$. Also, let the positive scalars $\theta'(a)$ and $\theta''(a)$ be such that for any pair of τ and r satisfying $\tau \geq \theta'(a)\log(1/(r\epsilon))$ and $r < \min\{r(a), \theta''(a)\sqrt{\epsilon}\}$, the upper bound σ in (32) becomes smaller than $\sigma \leq \sqrt{2\mu_f\epsilon}\,\min\{\mu_1/2, \sqrt{\mu_2}\}$. The choice of θ' and θ'' with the above property is straightforward using the definition of σ. Combining $\|\nabla f(K^0)\|_F \geq \sqrt{2\mu_f\epsilon}$ and $\sigma \leq \sqrt{2\mu_f\epsilon}\,\min\{\mu_1/2, \sqrt{\mu_2}\}$ yields

$$\sigma \;\leq\; \|\nabla f(K^0)\|_F\,\min\{\mu_1/2, \sqrt{\mu_2}\}. \quad (36)$$

Using the union bound, we have

$$\left\langle \overline{\nabla} f(K^0), \nabla f(K^0) \right\rangle$$
$$= \left\langle \widehat{\nabla} f(K^0), \nabla f(K^0) \right\rangle + \left\langle \overline{\nabla} f(K^0) - \widehat{\nabla} f(K^0), \nabla f(K^0) \right\rangle$$
$$\overset{(a)}{\geq} \mu_1 \|\nabla f(K^0)\|_F^2 \;-\; \|\overline{\nabla} f(K^0) - \widehat{\nabla} f(K^0)\|_F \|\nabla f(K^0)\|_F$$
$$\overset{(b)}{\geq} \mu_1 \|\nabla f(K^0)\|_F^2 \;-\; \sigma\|\nabla f(K^0)\|_F \;\overset{(c)}{\geq}\; \frac{\mu_1}{2}\|\nabla f(K^0)\|_F^2$$

with probability not smaller than $1 - p - 2Ne^{-n}$. Here, (a) follows from combining (34a) and the Cauchy–Schwarz inequality,

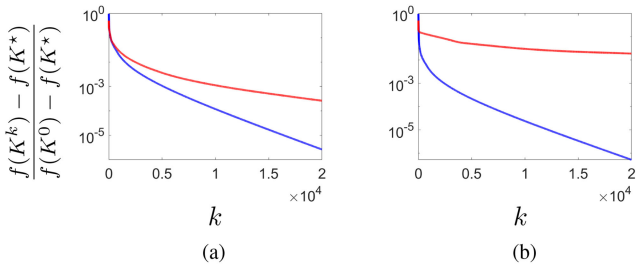


Fig. 2. Convergence curves for gradient descent (blue) over the set $\mathcal{S}_K$ and gradient descent (red) over the set $\mathcal{S}_Y$ with (a) $s = 10$ and (b) 20 masses.

(b) follows from (32), and (c) follows from (36). Moreover

$$\|\overline{\nabla} f(K^0)\|_F \overset{(a)}{\leq} \|\widehat{\nabla} f(K^0)\|_F + \|\overline{\nabla} f(K^0) - \widehat{\nabla} f(K^0)\|_F$$
$$\overset{(b)}{\leq} \sqrt{\mu_2}\|\nabla f(K^0)\|_F + \sigma \overset{(c)}{\leq} 2\sqrt{\mu_2}\|\nabla f(K^0)\|_F$$

where (a) follows from the triangle inequality, (b) from (32), and (c) from (36). This completes the proof of (35).

Inequality (35) allows us to apply Proposition 6 and obtain with probability not smaller than $1 - p - 2Ne^{-n}$ that for the stepsize $\alpha \leq \mu_1/(8\mu_2 L_f)$, we have $K^1 \in \mathcal{S}_K(a)$ and also $f(K^1) - f(K^\star) \leq \gamma(f(K^0) - f(K^\star))$, with $\gamma = 1 - \mu_f\mu_1\alpha/2$, where L_f is the smoothness parameter of the function f over $\mathcal{S}_K(a)$. Finally, using the union bound, we can repeat this procedure via induction to obtain that for some

$$k \;\leq\; \frac{1}{\log(1/\gamma)}\,\log\frac{f(K^0) - f(K^\star)}{\epsilon}$$

the error satisfies $f(K^k) - f(K^\star) \leq \gamma^k(f(K^0) - f(K^\star)) \leq \epsilon$ with probability not smaller than $1 - kp - 2kNe^{-n}$. ■

Remark 5: For the failure probability in Theorem 4 to be negligible, the problem dimension n needs to be large. Moreover, to account for the conflicting term $Ne^{-n/8}$ in the failure probability, we can require a crude exponential bound $N \leq e^{n/16}$ on the sample size. We also note that although Theorem 4 only guarantees convergence in the objective value, similar to the proof of Theorem 1, we can use Lemma 2 that relates the error in optimization variable, K, and the error in the objective function, $f(K)$, to obtain convergence guarantees in the optimization variable as well.

Remark 6: Theorem 4 requires the lower bound on the simulation time τ in (31) to ensure that, for any desired accuracy ϵ, the smoothing constant r satisfies $r \geq (1/\epsilon)\,e^{-\tau/\theta'(a)}$. As we demonstrate in the proof, this requirement accounts for the bias that arises from a finite value of τ. Since this form of bias can be readily controlled by increasing τ, the above lower bound on r does not contradict the upper bound $r = O(\sqrt{\epsilon})$ required by Theorem 4. Finally, we note that letting $r \to 0$ can cause large bias in the presence of other sources of inaccuracy in the function approximation process.

VIII. Computational Experiments

We consider a mass–spring–damper system with s masses, where we set all mass, spring, and damping constants to unity. In state-space representation (1b), the state $x = [\,p^T\; v^T\,]^T$ contains

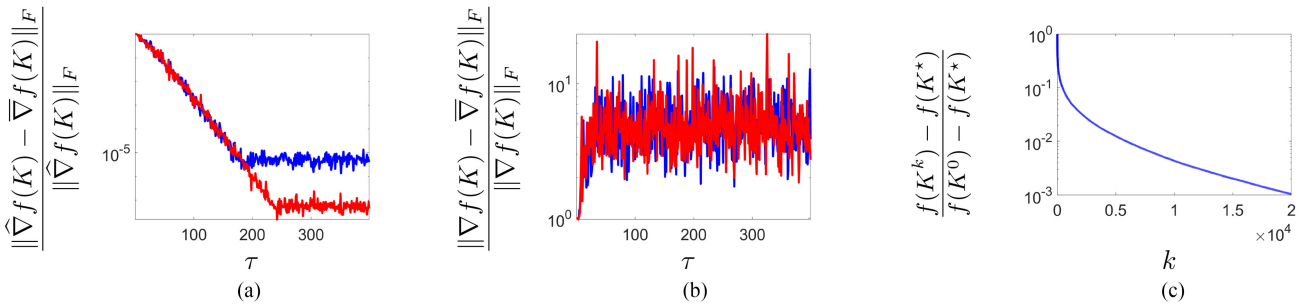


Fig. 3. (a) Bias in gradient estimation and (b) total error in gradient estimation as functions of the simulation time τ. The blue and red curves correspond to two values of the smoothing parameter $r = 10^{-4}$ and $r = 10^{-5}$, respectively. (c) Convergence curve of the random search method (RS).

the position and velocity vectors, and the dynamic and input matrices are given by

$$A = \begin{bmatrix} 0 & I \\ -T & -T \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ I \end{bmatrix}$$

where 0 and I are $s \times s$ zero and identity matrices, and T is a Toeplitz matrix with 2 on the main diagonal and -1 on the first super and subdiagonals.

A. Known Model

To compare the performance of gradient descent methods (GD) and (GY) on K and Y, we solve the LQR problem with $Q = I + 100\, e_1 e_1^T$, $R = I + 1000\, e_4 e_4^T$, and $\Omega = I$ for $s \in \{10,\ 20\}$ masses (i.e., $n = 2s$ state variables), where e_i is the ith unit vector in the standard basis of $\mathbb{R}^n$.

Fig. 2 illustrates the convergence curves for both algorithms with a stepsize selected using a backtracking procedure that guarantees stability of the closed-loop system. Both algorithms were initialized with $Y^0 = K^0 = 0$. Even though Fig. 2 suggests that gradient decent/flow on $\mathcal{S}_K$ converges faster than that on $\mathcal{S}_Y$, this observation does not hold in general.

B. Unknown Model

To illustrate our results on the accuracy of the gradient estimation in Algorithm 1 and the efficiency of our random search method, we consider the LQR problem with Q and R equal to identity for $s = 10$ masses (i.e., $n = 20$ state variables). We also let the initial conditions x_i in Algorithm 1 be standard normal and use $N = n = 2s$ samples.

Fig. 3(a) illustrates the dependence of the relative error $\|\widehat{\nabla} f(K) - \overline{\nabla} f(K)\|_F / \|\widehat{\nabla} f(K)\|_F$ on the simulation time τ for $K = 0$ and two values of the smoothing parameter $r = 10^{-4}$ (blue) and $r = 10^{-5}$ (red). We observe an exponential decrease in error for small values of τ. In addition, the error does not pass a saturation level, which is determined by r. We also see that, as r decreases, this saturation level becomes smaller. These observations are in harmony with our theoretical developments; in particular, combining Propositions 4 and 5 through the use of the triangle inequality yields

$$\|\widehat{\nabla} f(K) - \overline{\nabla} f(K)\|_F \leq \left(\frac{\sqrt{mn}}{r}\, \kappa_1(2a)\, e^{-\kappa_2(2a)\tau} + \frac{r^2\, m^2 n^2}{2}\, \ell(2a) \right) \max_i \|x_i\|^2.$$

This upper bound clearly captures the exponential dependence of the bias on the simulation time τ as well as the saturation level that depends quadratically on the smoothing parameter r.

In Fig. 3(b), we demonstrate the dependence of the total relative error $\|\nabla f(K) - \overline{\nabla} f(K)\|_F / \|\nabla f(K)\|_F$ on the simulation time τ for two values of the smoothing parameter $r = 10^{-4}$ (blue) and $r = 10^{-5}$ (red), resulting from the use of $N = n$ samples. We observe that the distance between the approximate gradient and the true gradient is rather large. This is exactly why prior analysis of sample complexity and simulation time is subpar to our results. In contrast to the existing results, which rely on the use of the estimation error shown in Fig. 3(b), our analysis shows that the simulated gradient $\overline{\nabla} f(K)$ is close to the gradient estimate $\widehat{\nabla} f(K)$. While $\widehat{\nabla} f(K)$ is not close to the true gradient $\nabla f(K)$, it is highly correlated with it. This is sufficient for establishing convergence guarantees, and it allows us to significantly improve upon existing results [11], [12] in terms of sample complexity and simulation time reducing both to $O(\log(1/\epsilon))$.

Finally, Fig. 3(c) demonstrates linear convergence of the random search method (RS) with stepsizes $\alpha = 10^{-4}, r = 10^{-5}$, and $\tau = 200$ in Algorithm 1, as established in Theorem 4. In this experiment, we implemented Algorithm 1 using the ode45 and trapz subroutines in MATLAB to numerically integrate the state/input penalties with the corresponding weight matrices Q and R. However, our theoretical results only account for an approximation error that arises from a finite simulation horizon. Clearly, employing empirical ODE solvers and numerical integration may introduce additional errors in our gradient approximation that require further scrutiny.

IX. Conclusion

We prove exponential/linear convergence of gradient flow/descent algorithms for solving the continuous-time LQR problem based on a nonconvex formulation that directly searches for the controller. A salient feature of our analysis is that we relate the gradient-flow dynamics associated with this nonconvex formulation to that of a convex reparameterization. This allows us to deduce convergence of the nonconvex approach from its convex counterpart. We also establish a bound on the sample complexity of the random search method for solving the continuous-time LQR problem that does not require the knowledge of system parameters. We have recently proved similar result for the discrete-time LQR problem [40].

Our ongoing research directions include: 1) providing theoretical guarantees for the convergence of gradient-based methods

for sparsity-promoting and structured control synthesis; and 2) extension to nonlinear systems via successive linearization techniques.

APPENDIX

A. Lack of Convexity of Function f

See [41].

B. Invertibility of the Linear Map $\mathcal{A}$

See [41].

C. Proof of Proposition 1

The second-order term in the Taylor series expansion of $h(Y+\tilde{Y})$ around Y is given by [29, Lemma 2]

$$\left\langle \tilde{Y}, \nabla^2 h(Y;\tilde{Y}) \right\rangle \;=\; 2\,\|R^{\frac{1}{2}}(\tilde{Y} \,-\, K\tilde{X})X^{-\frac{1}{2}}\|_F^2 \qquad (37)$$

where $\tilde{X}$ is the unique solution to $\mathcal{A}(\tilde{X}) = \mathcal{B}(\tilde{Y})$. We show that this term is upper and lower bounded by $L\|\tilde{Y}\|_F^2$ and $\mu\|\tilde{Y}\|_F^2$, where L and μ are given by (10a) and (10b), respectively. The proof for the upper bound is borrowed from [29, Lemma 1]; we include it for completeness. We repeatedly use the bounds on the variables presented in Lemma 16 (see Appendix K).

Smoothness: For any $Y \in \mathcal{S}_Y(a)$ and $\tilde{Y}$ with $\|\tilde{Y}\|_F = 1$,

$$\begin{aligned}\left\langle \tilde{Y}, \nabla^2 h(Y;\tilde{Y}) \right\rangle \;&=\; 2\|R^{\frac{1}{2}}(\tilde{Y} \,-\, K\tilde{X})X^{-\frac{1}{2}}\|_F^2 \\ &\le\; 2\|R\|_2\|X^{-1}\|_2\|\tilde{Y} \,-\, K\mathcal{A}^{-1}\mathcal{B}(\tilde{Y})\|_F^2 \\ &\le\; \frac{2\,\|R\|_2}{\lambda_{\min}(X)}\left(\|\tilde{Y}\|_F \,+\, \|K\|_2\|\mathcal{A}^{-1}\mathcal{B}\|_2\|\tilde{Y}\|_F\right)^2 \\ &\le\; \frac{2a\|R\|_2}{\nu}\left(1 \,+\, \frac{a\|\mathcal{A}^{-1}\mathcal{B}\|_2}{\sqrt{\nu\lambda_{\min}(R)}}\right)^2 =:\; L.\end{aligned}$$

Here, the first and second inequalities are obtained from the definition of the 2-norm in conjunction with the triangle inequality, and the third inequality follows from (58b) and (58c). This completes the proof of smoothness.

Strong convexity: Using the positive definiteness of matrices R and X, the second-order term (37) can be lower bounded by

$$\left\langle \tilde{Y}, \nabla^2 h(Y;\tilde{Y}) \right\rangle \;\ge\; 2\lambda_{\min}(R)\|H\|_F^2/\|X\|_2 \qquad (38)$$

where $H := \tilde{Y} - K\tilde{X}$. Next, we show that

$$\|H\|_F/\|\tilde{X}\|_F \;\ge\; \lambda_{\min}(\Omega)\lambda_{\min}(\Omega)/(a\|\mathcal{B}\|_2). \qquad (39)$$

We substitute $H + K\tilde{X}$ for $\tilde{Y}$ in $\mathcal{A}(\tilde{X}) = \mathcal{B}(\tilde{Y})$ to obtain

$$\Gamma \;=\; B\,H \,+\, H^T B^T \qquad (40)$$

where $\Gamma := \mathcal{A}_K(\tilde{X})$. The closed-loop stability implies that $\tilde{X} = \mathcal{A}_K^{-1}(\Gamma)$, and from (40), we have

$$\|H\|_F \;\ge\; \|\Gamma\|_F/\|\mathcal{B}\|_2. \qquad (41)$$

This allows us to use Lemma 18, presented in Appendix L, to write $a\|\Gamma\|_F \ge \lambda_{\min}(\Omega)\lambda_{\min}(Q)\|\tilde{X}\|_F$. This inequality in conjunction with (41) yields (39).

Next, we derive upper bound on $\|\tilde{Y}\|_F$, as follows:

$$\begin{aligned}\|\tilde{Y}\|_F \;&=\; \|H \,+\, K\tilde{X}\|_F \;\le\; \|H\|_F \,+\, \|K\|_F\|\tilde{X}\|_F \\ &\le\; \|H\|_F\left(1 \,+\, a^2\eta\right) \qquad (42)\end{aligned}$$

where η is given by (10c) and the second inequality follows from (58d) and (39). Finally, inequalities (38) and (42) yield

$$\begin{aligned}\frac{\left\langle \tilde{Y}, \nabla^2 f(Y;\tilde{Y}) \right\rangle}{\|\tilde{Y}\|_F^2} \;&\ge\; \frac{2\lambda_{\min}(R)\|H\|_F^2}{\|X\|_2\|\tilde{Y}\|_F^2} \\ &\ge\; \frac{2\lambda_{\min}(R)}{\|X\|_2(1 \,+\, a^2\eta)^2} \;\ge\; \frac{2\lambda_{\min}(R)\lambda_{\min}(Q)}{a(1 \,+\, a^2\eta)^2} \;=:\; \mu \qquad (43)\end{aligned}$$

where the last inequality follows from (58a).

D. Proofs for Section V

Proof of Lemma 1: The gradients are given by $\nabla f(K) = EX$ and $\nabla h(Y) = E + 2B^T(P - W)$, where $E := 2(RK - B^TP)$, P is determined by (6a), and W is the solution to (11b). Subtracting (11b) from (6b) yields $A^T(P-W) + (P-W)A = -\frac{1}{2}(K^TE + E^TK)$, which, in turn, leads to $\|P - W\|_F \le \|\mathcal{A}^{-1}\|_2\|K\|_F\|E\|_F \le a\|\mathcal{A}^{-1}\|_2\|E\|_F/\sqrt{\nu\lambda_{\min}(R)}$, where the second inequality follows from (58d) in Appendix K. Thus, by applying the triangle inequality to $\nabla h(Y)$, we obtain $\|\nabla h(Y)\|_F/\|E\|_F \le 1 + 2a\|\mathcal{A}^{-1}\|_2\|B\|_2/\sqrt{\nu\lambda_{\min}(R)}$. Moreover, using the lower bound (58c) on $\lambda_{\min}(X)$, we have $\|\nabla f(K)\|_F = \|EX\|_F \ge (\nu/a)\|E\|_F$. Combining the last two inequalities completes the proof.

Proof of Lemma 2: For any pair of stabilizing feedback gains K and $\hat{K} := K + \tilde{K}$, we have [31, eq. (2.10)], $f(\hat{K}) - f(K) = \mathrm{trace}(\tilde{K}^T(R(K+\hat{K}) - 2B^T\hat{P})X)$, where $X = X(K)$ and $\hat{P} = P(\hat{K})$ are given by (4a) and (6a), respectively. Letting $\hat{K} = K^\star$ in this equation and using the optimality condition $B^T\hat{P} = R\hat{K}$ completes the proof.

Proof of Lemma 3: See [41].

E. Proofs for Section VI-A1

We first present a technical lemma.

Lemma 9: Let the matrices F, $X \succ 0$, and $\Omega \succ 0$ satisfy

$$FX \,+\, XF^T \,+\, \Omega \;=\; 0. \qquad (44)$$

Then, the matrix $F + \Delta$ is Hurwitz for all Δ that satisfy $\|\Delta\|_2 < \lambda_{\min}(\Omega)/(2\|X\|_2)$.

Proof: See [41, Appendix E].

Proof of Proposition 3: For any feedback gain $\hat{K}$ such that $\|\hat{K} - K\|_2 < \zeta$, the closed-loop matrix $A - B\hat{K}$ satisfies $\|A - B\hat{K} - (A - BK)\|_2 \le \|K - \hat{K}\|_2\|B\|_2 < \zeta\|B\|_2$. This bound on the distance between the closed-loop matrices $A - BK$ and $A - B\hat{K}$ allows us to apply Lemma 9 with $F := A - BK$ and $X := X(K)$ to complete the proof.

We next present a technical lemma.

Lemma 10: For any $K \in \mathcal{S}_K$ and $\hat{K} \in \mathbb{R}^{m\times n}$ such that $\|\hat{K} - K\|_2 < \delta$, with

$$\delta \;:=\; \frac{1}{4\,\|B\|_F}\,\min\left\{\frac{\lambda_{\min}(\Omega)}{\mathrm{trace}\,(X(K))},\, \frac{\lambda_{\min}(Q)}{\mathrm{trace}\,(P(K))}\right\}$$

the feedback gain matrix $\hat{K} \in \mathcal{S}_K$, and

$$\|X(\hat{K}) \,-\, X(K)\|_F \;\le\; \epsilon_1\|\hat{K} - K\|_2 \qquad (45a)$$

$$\|P(\hat{K}) - P(K)\|_F \leq \epsilon_2 \|\hat{K} - K\|_2 \tag{45b}$$

$$\|\nabla f(\hat{K}) - \nabla f(K)\|_F \leq \epsilon_3 \|\hat{K} - K\|_2 \tag{45c}$$

$$|f(\hat{K}) - f(K)| \leq \epsilon_4 \|\hat{K} - K\|_2 \tag{45d}$$

where $X(K)$ and $P(K)$ are given by (4a) and (6a), respectively. Furthermore, the parameters ϵ_i, which only depend on K and problem data, are given by $\epsilon_1 := \|X(K)\|_2/\delta, \epsilon_2 := 2\,\mathrm{trace}(P)(2\,\|P\|_2\|B\|_F + (\delta + 2\|K\|_2)\|R\|_F)/\lambda_{\min}(Q), \epsilon_4 := \epsilon_2\|\Omega\|_F$, and $\epsilon_3 := 2(\epsilon_1\|K\|_2 + 2\|X(K)\|_2)\|R\|_F + 2\epsilon_1(\|P(K)\|_2 + 2\epsilon_2\|X(K)\|_2)\|B\|_F$.

Proof: See [41, Appendix E].

Proof of Lemma 4: For any $K \in \mathcal{S}_K(a)$, we can use the bounds provided in Appendix K to show that $c_1/a \leq \delta$ and $\epsilon_4 \leq c_2 a^2$, where δ and ϵ_4 are given in Lemma 10, and each c_i is a positive constant that depends on the problem data. Now, Lemma 10 implies that $f(K + r(a)U) - f(K) \leq \epsilon_4 r(a)\|U\|_2 \leq a$, where $r(a) := \min\{c_1, 1/c_2\}/(a\sqrt{mn})$. This inequality together with $f(K) \leq a$ completes the proof.

F. Proof of Proposition 4

Lemma 11 establishes an exponentially decaying upper bound on the difference between $f_{x_0}(K)$ and $f_{x_0,\tau}(K)$ over any sublevel set $\mathcal{S}_K(a)$ of the LQR objective function $f(K)$.

Lemma 11: For any $K \in \mathcal{S}_K(a)$ and $v \in \mathbb{R}^n$,

$$|f_v(K) - f_{v,\tau}(K)| \leq \|v\|^2 \kappa_1(a) e^{-\kappa_2(a)\tau}$$

where

$$\kappa_1(a) := \left(\|Q\|_F + \frac{a^2\|R\|_2}{\nu\lambda_{\min}(R)}\right)\frac{a^3}{\nu\lambda_{\min}(\Omega)\lambda^2_{\min}(Q)} \tag{46a}$$

$$\kappa_2(a) := \lambda_{\min}(\Omega)\lambda_{\min}(Q)/a \tag{46b}$$

and the constant ν is given by (10d).

Proof: See [41, Appendix F].

Proof of Proposition 4: Since $K \in \mathcal{S}_K(a)$ and $r \leq r(a)$, Lemma 4 implies that $K \pm rU_i \in \mathcal{S}_K(2a)$. Thus, $f_{x_i}(K \pm rU_i)$ is well defined for $i = 1, \ldots, N$, and $\widetilde{\nabla} f(K) - \overline{\nabla} f(K) = \frac{1}{2rN}(\sum_i (f_{x_i}(K + rU_i) - f_{x_i,\tau}(K + rU_i))U_i - \sum_i (f_{x_i}(K - rU_i) - f_{x_i,\tau}(K - rU_i))U_i)$. Furthermore, since $K \pm rU_i \in \mathcal{S}_K(2a)$, we can use triangle inequality and apply Lemma 11, $2N$ times, to bound each term individually and obtain

$$\|\widetilde{\nabla} f(K) - \overline{\nabla} f(K)\|_F \leq (\sqrt{mn}/r) \max_i \|x_i\|^2 \kappa_1(2a) e^{-\kappa_2(2a)\tau}$$

where we used $\|U_i\|_F = \sqrt{mn}$. This completes the proof.

G. Proof of Proposition 5

We first establish bounds on the smoothness parameter of $\nabla f(K)$. For $J \in \mathbb{R}^{m\times n}$, $v \in \mathbb{R}^n$, and $f_v(K)$ given by (23a), let $j_v(K) := \langle J, \nabla^2 f_v(K; J)\rangle$ denote the second-order term in the Taylor series expansion of $f_v(K + J)$ around K. Following similar arguments as in [42, eq. (2.3)] leads to $j_v(K) = 2\,\mathrm{trace}\,(J^T(RJ - 2B^T D)X_v)$, where X_v and D are the solutions to

$$\mathcal{A}_K(X_v) = -vv^T \tag{47a}$$

$$\mathcal{A}^*_K(D) = J^T(B^T P - RK) + (B^T P - RK)^T J \tag{47b}$$

and P is given by (6a). The following lemma provides an analytical expression for the gradient $\nabla j_v(K)$.

Lemma 12: For any $v \in \mathbb{R}^n$ and $K \in \mathcal{S}_K$, $\nabla j_v(K) = 4(B^T W_1 X_v + (RJ - B^T D)W_2 + (RK - B^T P)W_3)$, where W_i are the solutions to the linear equations

$$\mathcal{A}^*_K(W_1) = J^T R J - J^T B^T D - DBJ \tag{48a}$$

$$\mathcal{A}_K(W_2) = BJX_v + X_v J^T B^T \tag{48b}$$

$$\mathcal{A}_K(W_3) = BJ\,W_2 + W_2\,J^T B^T. \tag{48c}$$

Proof: See [41].

We next establish a bound on $\|\nabla j_v(K)\|_F$.

Lemma 13: Let $K, K' \in \mathbb{R}^{m\times n}$ be such that the line segment $K + t(K' - K)$ with $t \in [0, 1]$ belongs to $\mathcal{S}_K(a)$ and let $J \in \mathbb{R}^{m\times n}$ and $v \in \mathbb{R}^n$ be fixed. Then, the function $j_v(K)$ satisfies $|j_v(K_1) - j_v(K_2)| \leq \ell(a)\|J\|^2_F\|v\|^2\|K_1 - K_2\|_F$, where $l(a)$ is a positive function given by

$$\ell(a) := ca^2 + c'a^4 \tag{49}$$

and c and c' are positive scalars that depend only on problem data.

Proof: We show that the gradient $\nabla j_v(K)$ given by Lemma 12 is upper bounded by $\|\nabla j_v(K)\|_F \leq \ell(a)\|J\|^2_F\|v\|^2$. Applying Lemma 18 on (47), the bounds in Lemma 16, and the triangle inequality, we have $\|X_v\|_F \leq c_1 a\|v\|^2$ and $\|D\|_F \leq c_2 a^2\|J\|_F$, where c_1 and c_2 are positive constants that depend on problem data. We can use the same technique to bound the norms of W_i in (48), $\|W_1\|_F \leq (c_3 a + c_4 a^3)\|J\|^2_F, \|W_2\|_F \leq c_5 a^2\|v\|^2\|J\|_F, \|W_3\|_F \leq c_6 a^3\|v\|^2\|J\|^2_F$, where $c_3, \ldots, c_6$ are positive constants that depend on problem data. Combining these bounds with the Cauchy–Schwarz and triangle inequalities applied to $\nabla f_v(K)$ completes the proof. ■

Proof of Proposition 5: Since $r \leq r(a)$, Lemma 4 implies that $K \pm sU \in \mathcal{S}_K(2a)$ for all $s \leq r$. Also, the mean-value theorem implies that, for any $U \in \mathbb{R}^{m\times n}$ and $v \in \mathbb{R}^n$, $f_v(K \pm rU) = f_v(K) \pm r\langle \nabla f_v(K), U\rangle + (r^2/2)\langle U, \nabla^2 f_v(K \pm s_\pm U; U)\rangle$, where $s_\pm \in [0, r]$ are constants that depend on K and U. Now, if $\|U\|_F = \sqrt{mn}$, the above identity allows us to write

$$(f_v(K + rU) - f_v(K - rU))/(2r) - \langle \nabla f_v(K), U\rangle = \frac{r}{4}\left(\langle U, \nabla^2 f_v(K + s_+U; U)\rangle - \langle U, \nabla^2 f_v(K - s_-U; U)\rangle\right) \leq \frac{r}{4}(s_+ + s_-)\|U\|^3_F\,\ell(2a)\,\|v\|^2 \leq (r^2 mn\sqrt{mn}/2)\,\ell(2a)\,\|v\|^2$$

where the first inequality follows from Lemma 13. Combining this inequality with the triangle inquality applied to the definition of $\widehat{\nabla} f(K) - \widetilde{\nabla} f(K)$ completes the proof.

H. Proof of Proposition 6

From inequality (27a), it follows that G is a descent direction of the function $f(K)$. Thus, we can use the descent lemma [36, eq. (9.17)] to show that $K^+ := K - \alpha G$ satisfies

$$f(K^+) - f(K) \leq (L_f\alpha^2/2)\,\|G\|^2_F - \alpha\langle \nabla f(K), G\rangle \tag{50}$$

for any α for which the line segment between K^+ and K lies in $\mathcal{S}_K(a)$. Using (27), for any $\alpha \in [0, 2\mu_1/(\mu_2 L_f)]$, we have

$$(L_f\alpha^2/2)\,\|G\|_F^2 \;-\; \alpha\,\langle \nabla f(K), G\rangle \;\le\; (\alpha\,(L_f\mu_2\alpha - 2\mu_1)/2)\,\|\nabla f(K)\|_F^2 \;\le\; 0 \quad (51)$$

and the right-hand side of inequality (50) is nonpositive for $\alpha \in [0, 2\mu_1/(\mu_2 L_f)]$. Thus, we can use the continuity of the function $f(K)$ along with inequalities (50) and (51) to conclude that $K^+ \in \mathcal{S}_K(a)$ for all $\alpha \in [0, 2\mu_1/(\mu_2 L_f)]$, and $f(K^+) - f(K) \le (\alpha\,(L_f\mu_2\alpha - 2\mu_1)/2)\,\|\nabla f(K)\|_F^2$. Combining this inequality with the PL condition (17), it follows that, for any $\alpha \in [0, c_1/(c_2 L_f)]$, $f(K^+) - f(K) \le -(\mu_1\alpha/2)\,\|\nabla f(K)\|_F^2 \le -\mu_f\mu_1\alpha\,(f(K) - f(K^\star))$. Subtracting $f(K^\star)$ and rearranging terms complete the proof.

I. Proofs of Section VI-B1

We first present two technical results. Lemma 14 extends [43, Th. 3.2] on the norm of Gaussian matrices presented in Appendix J to random matrices with uniform distribution on the sphere $\sqrt{mn}\,S^{mn-1}$.

Lemma 14: Let $E \in \mathbb{R}^{m\times n}$ be a fixed matrix, and let $U \in \mathbb{R}^{m\times n}$ be a random matrix with $\mathrm{vec}(U)$ uniformly distributed on the sphere $\sqrt{mn}\,S^{mn-1}$. Then, for any $s \ge 1$ and $t \ge 1$, we have $\mathbb{P}(\mathsf{B}) \le 2e^{-s^2 q - t^2 n} + e^{-mn/8}$, where $\mathsf{B} := \{\|E^T U\|_2 > c'(s\|E\|_F + t\sqrt{n}\|E\|_2)\}$, and $q := \|E\|_F^2/\|E\|_2^2$ is the stable rank of E.

Proof: For a matrix G with i.i.d. standard normal entries, we have $\|E^T U\|_2 \sim \sqrt{mn}\|E^T G\|_2/\|G\|_F$. Let the constant κ be the ψ_2-norm of the standard normal random variable, and let us define two auxiliary events, $\mathsf{C}_1 := \{\sqrt{mn} > 2\|G\|_F\}$ and $\mathsf{C}_0 := \{\sqrt{mn}\,\|E^T G\|_2 > 2c\kappa^2\|G\|_F(s\|E\|_F + t\sqrt{n}\|E\|_2)\}$. For $c' := 2c\kappa^2$, we have $\mathbb{P}(\mathsf{B}) = \mathbb{P}(\mathsf{C}_0) \le \mathbb{P}(\mathsf{C}_1 \cup \mathsf{A}) \le \mathbb{P}(\mathsf{C}_1) + \mathbb{P}(\mathsf{A})$, where $\mathsf{A} := \{\|E^T G\|_2 > c\kappa^2(s\|E\|_F + t\sqrt{n}\|E\|_2)\}$. Here, the first inequality follows from $\mathsf{C}_0 \subset \mathsf{C}_1 \cup \mathsf{A}$ and the second follows from the union bound. Now, since $\|\cdot\|_F$ is Lipschitz continuous with parameter 1, from the concentration of Lipschitz functions of standard normal Gaussian vectors [39, Th. 5.2.2], it follows that $\mathbb{P}(C_1) \le e^{-mn/8}$. This in conjunction with [43, Th. 3.2] completes the proof. ■

Lemma 15: In the setting of Lemma 14, we have $\mathbb{P}\{\|E^T U\|_F > 2\sqrt{n}\,\|E\|_F\} \le e^{-n/2}$.

Proof: We begin by observing that $\|E^T U\|_F = \|\mathrm{vec}(E^T U)\|_F = \|(I \otimes E^T)\mathrm{vec}(U)\|_F$, where $\otimes$ denotes the Kronecker product. Thus, it is easy to verify that $\|E^T U\|_F$ is a Lipschitz continuous function of U with parameter $\|I \otimes E^T\|_2 = \|E\|_2$. Now, from the concentration of Lipschitz functions of uniform random variables on the sphere $\sqrt{mn}\,S^{mn-1}$ [39, Th. 5.1.4], for all $t > 0$, we have $\mathbb{P}\{\|E^T U\|_F > \sqrt{\mathbb{E}[\|E^T U\|_F^2]} + t\} \le e^{-t^2/(2\|E\|_2^2)}$. Now, since $\mathbb{E}\,[\|E^T U\|_F^2] = \mathbb{E}\,[\|(I \otimes E^T)\,\mathrm{vec}(U)\|_F^2] = \mathbb{E}\,[\mathrm{trace}\,((I \otimes E^T)\mathrm{vec}(U)\mathrm{vec}(U)^T(I \otimes E))] = \mathrm{trace}\,((I \otimes E^T)(I \otimes E)) = n\|E\|_F^2$, we can rewrite the last inequality for $t = \sqrt{n}\|E\|_F$ to obtain

$$\mathbb{P}\,\{\|E^T U\|_F > 2\sqrt{n}\,\|E\|_F\} \;\le\; e^{-n\|E\|_F^2/(2\|E\|_2^2)} \;\le\; e^{-n/2}$$

where the last inequality follows from $\|E\|_F \ge \|E\|_2$. ■

Proof of Lemma 5: We define the auxiliary events $\mathsf{D}_i := \{\|\mathcal{M}^*(E^T U_i)\|_2 \le c\sqrt{n}\,\|M^*\|_S\|E\|_F\} \cap \{\|\mathcal{M}^*(E^T U_i)\|_F \le 2\sqrt{n}\|M^*\|_2\|E\|_F\}$ for $i = 1, \ldots, N$. Since $\|\mathcal{M}^*(E^T U_i)\|_2 \le \|\mathcal{M}^*\|_S\|E^T U_i\|_2$ and $\|\mathcal{M}^*(E^T U_i)\|_F \le \|\mathcal{M}^*\|_2\|E^T U_i\|_F$, we have $\mathbb{P}(\mathsf{D}_i) \ge \mathbb{P}\left(\{\|E^T U_i\|_2 \le c\sqrt{n}\|E\|_F\} \cap \{\|E^T U_i\|_F \le 2\sqrt{n}\|E\|_F\}\right)$. Applying Lemmas 14 and 15 to the right-hand side of the above events together with the union bound yield $\mathbb{P}(\mathsf{D}_i^c) \le 2e^{-n} + e^{-mn/8} + e^{-n/2} \le 4e^{-n/8}$, where D_i^c is the complement of D_i. This, in turn, implies that

$$\mathbb{P}(\mathsf{D}^c) \;=\; \mathbb{P}(\bigcup_{i=1}^N \mathsf{D}_i^c) \;\le\; \sum_{i=1}^N \mathbb{P}(\mathsf{D}_i^c) \;\le\; 4\,Ne^{-\frac{n}{8}} \quad (52)$$

where $\mathsf{D} := \cap_i \mathsf{D}_i$. We can now use the conditioning identity to bound the failure probability

$$\begin{aligned}\mathbb{P}\{|a| > b\} &= \mathbb{P}\left\{|a| > b \,\middle|\, \mathsf{D}\right\}\mathbb{P}(\mathsf{D}) + \mathbb{P}\left\{|a| > b \,\middle|\, \mathsf{D}^c\right\}\mathbb{P}(\mathsf{D}^c)\\ &\le \mathbb{P}\left\{|a| > b \,\middle|\, \mathsf{D}\right\}\mathbb{P}(\mathsf{D}) + \mathbb{P}(\mathsf{D}^c)\\ &= \mathbb{P}\left\{|a\,\mathbb{1}_{\mathsf{D}}| > b\right\} + \mathbb{P}(\mathsf{D}^c)\\ &\le \mathbb{P}\left\{|a\,\mathbb{1}_{\mathsf{D}}| > b\right\} + 4Ne^{-n/8} \quad (53)\end{aligned}$$

where $a := (1/N)\sum_i \langle E(X_i - X), U_i\rangle\,\langle EX, U_i\rangle$, $b := \delta\|EX\|_F\|E\|_F$, and $\mathbb{1}_{\mathsf{D}}$ is the indicator function of D. It is now easy to verify that $\mathbb{P}\{|a\,\mathbb{1}_{\mathsf{D}}| > b\} \le \mathbb{P}\{|Y| > b\}$, where $Y := (1/N)\sum_i Y_i$, $Y_i := \langle E(X_i - X), U_i\rangle\,\langle EX, U_i\rangle\,\mathbb{1}_{\mathsf{D}_i}$. The rest of the proof uses the $\psi_{1/2}$-norm of Y to establish an upper bound on $\mathbb{P}\{|Y| > b\}$.

Since Y_i are linear in the zero-mean random variables $X_i - X$, we have $\mathbb{E}[Y_i|U_i] = 0$. Thus, the law of total expectation yields $\mathbb{E}[Y_i] = \mathbb{E}[\mathbb{E}[Y_i|U_i]] = 0$ and Talagrand's inequality presented in Appendix J implies that

$$\|Y\|_{\psi_{1/2}} \;\le\; (c'/\sqrt{N})(\log N)\max_i \|Y_i\|_{\psi_{1/2}}. \quad (54)$$

Now, using the standard properties of the ψ_α-norm, we have

$$\begin{aligned}\|Y_i\|_{\psi_{1/2}} &\le c''\|\,\langle E\,(X_i - X)\,, U_i\rangle\,\mathbb{1}_{\mathsf{D}_i}\|_{\psi_1}\|\,\langle EX, U_i\rangle\,\|_{\psi_1}\\ &\le c'''\|\,\langle E\,(X_i - X)\,, U_i\rangle\,\mathbb{1}_{\mathsf{D}_i}\|_{\psi_1}\|EX\|_F \quad (55)\end{aligned}$$

where the second inequality follows from [39, Th. 3.4.6]

$$\|\,\langle EX, U_i\rangle\,\|_{\psi_1} \;\le\; \|\,\langle EX, U_i\rangle\,\|_{\psi_2} \;\le\; c_0\|EX\|_F. \quad (56)$$

We can now use $\langle E(X_i - X), U_i\rangle = \langle X_i - X, E^T U_i\rangle = \langle \mathcal{M}(x_i x_i^T), E^T U_i\rangle - \langle \mathcal{M}(I), E^T U_i\rangle = x_i^T\mathcal{M}^*(E^T U_i)x_i - \mathrm{trace}\,(\mathcal{M}^*(E^T U_i))$ to bound the right-hand side of (55). This identity allows us to use the Hanson–Write inequality to upper bound the conditional probability

$$\mathbb{P}\left\{|\langle E\,(X_i - X)\,, U_i\rangle| > t \;\middle|\; U_i\right\} \;\le\; 2e^{-\hat{c}\min\{\frac{t^2}{\kappa^4\|\mathcal{M}^*(E^T U_i)\|_F^2}, \frac{t}{\kappa^2\|\mathcal{M}^*(E^T U_i)\|_2}\}}.$$

Thus, we have

$$\begin{aligned}&\mathbb{P}\,\{|\langle E\,(X_i - X)\,, U_i\rangle\,\mathbb{1}_{\mathsf{D}_i}| > t\}\\ &\quad= \mathbb{E}_{U_i}\left[\mathbb{1}_{\mathsf{D}_i}\,\mathbb{E}_{x_i}\left[\mathbb{1}_{\{|\langle E(X_i - X), U_i\rangle| > t\}}\right]\right]\\ &\quad= \mathbb{E}_{U_i}\left[\mathbb{1}_{\mathsf{D}_i}\mathbb{P}\left\{|\langle E\,(X_i - X)\,, U_i\rangle| > t \;\middle|\; U_i\right\}\right]\end{aligned}$$

$$\leq \mathbb{E}_{U_i}\left[\mathbb{1}_{\mathsf{D}_i} 2e^{-\hat{c}\min\{\frac{t^2}{\kappa^4\|\mathcal{M}^*(E^TU_i)\|_F^2}, \frac{t}{\kappa^2\|\mathcal{M}^*(E^TU_i)\|_2}\}}\right]$$
$$\leq 2e^{-\hat{c}\min\{\frac{t^2}{4n\kappa^4\|\mathcal{M}^*\|_2^2\|E\|_F^2}, \frac{t}{c\sqrt{n}\kappa^2\|\mathcal{M}^*\|_S\|E\|_F}\}}$$

where the definition of D_i was used to obtain the last inequality. The above tail bound implies [44, Lemma 11] that

$$\|\langle E(X_i - X), U_i\rangle \mathbb{1}_{\mathsf{D}_i}\|_{\psi_1} \leq \tilde{c}\kappa^2\sqrt{n}(\|\mathcal{M}^*\|_2 + \|\mathcal{M}^*\|_S)\|E\|_F. \quad (57)$$

Using (29), it is easy to obtain the lower bound on the number of samples, $N \geq C'(\beta^2\kappa^2/\delta)^2(\|\mathcal{M}^*\|_2 + \|\mathcal{M}^*\|_S)^2 n\log^6 N$. We can now combine (54), (55), and (57) to obtain

$$\|Y\|_{\psi_{1/2}} \leq C'\kappa^2\frac{\sqrt{n}\log N}{\sqrt{N}}(\|\mathcal{M}^*\|_2 + \|\mathcal{M}^*\|_S)\|E\|_F\|EX\|_F \leq \frac{\delta}{\beta^2\log^2 N}\|E\|_F\|EX\|_F$$

where the last inequality follows from the above lower bound on N. Combining this inequality and

$$\mathbb{P}\{|\xi| > t\|\xi\|_{\psi_\alpha}\} \leq c_\alpha e^{-t^\alpha}, t := \delta\|E\|_F\|EX\|_F/\|Y\|_{\psi_{1/2}}$$

(presented Appendix J) yields $\mathbb{P}\{|Y| > \delta\|E\|_F\|EX\|_F\} \leq 1/N^\beta$, which completes the proof.

Proof of Lemma 6: See [41].

J. Proofs for Section VI-B2 and Probabilistic Toolbox

See [41].

K. Bounds on Optimization Variables

Building on [31], in Lemma 16, we provide useful bounds on K, $X = X(K)$, $P = P(K)$, and $Y = KX(K)$.

Lemma 16: Over the sublevel set $\mathcal{S}_K(a)$ of the LQR objective function $f(K)$, we have

$$\operatorname{trace}(X) \leq a/\lambda_{\min}(Q) \quad (58a)$$
$$\|Y\|_F \leq a/\sqrt{\lambda_{\min}(R)\lambda_{\min}(Q)} \quad (58b)$$
$$\nu/a \leq \lambda_{\min}(X) \quad (58c)$$
$$\|K\|_F \leq a/\sqrt{\nu\lambda_{\min}(R)} \quad (58d)$$
$$\operatorname{trace}(P) \leq a/\lambda_{\min}(\Omega) \quad (58e)$$

where the constant ν is given by (10d).

Proof: See [41].

L. Bound on the Norm of the Inverse Lyapunov Operator

Lemma 17 provides an upper bound on the norm of the inverse Lyapunov operator for stable LTI systems.

Lemma 17: For any Hurwitz matrix $F \in \mathbb{R}^{n\times n}$, the linear map $\mathcal{F}: \mathbb{S}^n \to \mathbb{S}^n$, $\mathcal{F}(W) := \int_0^\infty e^{Ft}We^{F^Tt}dt$ is well defined, and for any $\Omega \succ 0$,

$$\|\mathcal{F}\|_2 \leq \operatorname{trace}(\mathcal{F}(I)) \leq \operatorname{trace}(\mathcal{F}(\Omega))/\lambda_{\min}(\Omega). \quad (59)$$

Proof: See [41]. ■

We next use Lemma 17 to establish a bound on the norm of the inverse of the closed-loop Lyapunov operator $\mathcal{A}_K$ over the sublevel sets of the LQR objective function $f(K)$.

Lemma 18: For any $K \in \mathcal{S}_K(a)$, the closed-loop Lyapunov operator $\mathcal{A}_K$ given by (7) satisfies $\|\mathcal{A}_K^{-1}\|_2 = \|(\mathcal{A}_K^*)^{-1}\|_2 \leq a/\lambda_{\min}(\Omega)\lambda_{\min}(Q)$.

Proof: Applying Lemma 17 with $F = A - BK$ yields $\|\mathcal{A}_K^{-1}\|_2 = \|(\mathcal{A}_K^*)^{-1}\|_2 \leq \operatorname{trace}(X)/\lambda_{\min}(\Omega)$. Combining this inequality with (58a) completes the proof. ■

Parameter $\theta(a)$ in Theorem 4: See [41].

REFERENCES

[1] A. Nagabandi, G. Kahn, R. Fearing, and S. Levine, "Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning," in *Proc. IEEE Int Conf. Robot. Autom.*, 2018, pp. 7559–7566.

[2] V. Mnih *et al.*, "Playing Atari with deep reinforcement learning," 2013, *arXiv:1312.5602*.

[3] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "On the sample complexity of the linear quadratic regulator," *Found. Comput. Math.*, vol. 20, pp. 633–679, 2020.

[4] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht, "Learning without mixing: Towards a sharp analysis of linear system identification," in *Proc*. Mach. Learn. Res., 2018, pp. 439–473.

[5] D. Bertsekas, "Approximate policy iteration: A survey and some new methods," *J. Control Theory Appl.*, vol. 9, no. 3, pp. 310–335, 2011.

[6] Y. Abbasi-Yadkori, N. Lazic, and C. Szepesvári, "Model-free linear quadratic control via reduction to expert prediction," in *Proc*. Mach. Learn. Res., vol. 89, 2019, pp. 3108–3117.

[7] B. Anderson and J. Moore, *Optimal Control: Linear Quadratic Methods*. New York, NY, USA: Prentice-Hall, 1990.

[8] J. Ackermann, "Parameter space design of robust control systems," *IEEE Trans. Autom. Control*, vol. AC-25, no. 6, pp. 1058–1072, Dec. 1980.

[9] G. E. Dullerud and F. Paganini, *A Course in Robust Control Theory: A Convex Approach*. New York, NY, USA: Springer-Verlag, 2000.

[10] H. Mania, A. Guy, and B. Recht, "Simple random search of static linear policies is competitive for reinforcement learning," in *Proc. Int. Conf. Neural Inf. Process.*, 2018, vol. 31, pp. 1800–1809.

[11] M. Fazel, R. Ge, S. M. Kakade, and M. Mesbahi, "Global convergence of policy gradient methods for the linear quadratic regulator," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1467–1476.

[12] D. Malik, A. Panajady, K. Bhatia, K. Khamaru, P. L. Bartlett, and M. J. Wainwright, "Derivative-free methods for policy optimization: Guarantees for linear-quadratic systems," *J. Mach. Learn. Res.*, vol. 51, pp. 1–51, 2019.

[13] J. P. Jansch-Porto, B. Hu, and G. E. Dullerud, "Convergence guarantees of policy optimization methods for Markovian jump linear systems," in *Proc. Amer. Control Conf.*, 2020, pp. 2882–2887.

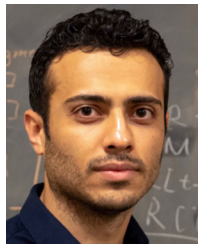
[14] K. Zhang, B. Hu, and T. Başar, "Policy optimization for $\mathcal{H}_2$ linear control with $\mathcal{H}_\infty$ robustness guarantee: Implicit regularization and global convergence," in *Proc. 2nd Conf. Learn. Dyn. Control*, 2020, vol. 120, pp. 179–190.

[15] L. Furieri, Y. Zheng, and M. Kamgarpour, "Learning the globally optimal distributed LQ regulator," in *Proc. 2nd Conf. Learn. Dyn. Control*, 2020, vol. 120, pp. 287–297.

[16] I. Fatkhullin and B. Polyak, "Optimizing static linear feedback: Gradient method," 2020, *arXiv:2004.09875*.

[17] D. Kleinman, "On an iterative technique for Riccati equation computations," *IEEE Trans. Autom. Control*, vol. AC-13, no. 1, pp. 114–115, Feb. 1968.

[18] S. Bittanti, A. J. Laub, and J. C. Willems, *The Riccati Equation*. Berlin, Germany: Springer-Verlag, 2012.
[19] P. L. D. Peres and J. C. Geromel, "An alternate numerical solution to the linear quadratic problem," *IEEE Trans. Autom. Control*, vol. 39, no. 1, pp. 198–202, Jan. 1994.
[20] V. Balakrishnan and L. Vandenberghe, "Semidefinite programming duality and linear time-invariant systems," *IEEE Trans. Autom. Control*, vol. 48, no. 1, pp. 30–41, Jan. 2003.
[21] J. Bu, A. Mesbahi, M. Fazel, and M. Mesbahi, "LQR through the lens of first order methods: discrete-time case," 2019, *arXiv:1907.08921*.
[22] W. S. Levine and M. Athans, "On the determination of the optimal constant output feedback gains for linear multivariable systems," *IEEE Trans. Autom. Control*, vol. AC-15, no. 1, pp. 44–48, Feb. 1970.
[23] F. Lin, M. Fardad, and M. R. Jovanović, "Augmented Lagrangian approach to design of structured optimal state feedback gains," *IEEE Trans. Autom. Control*, vol. 56, no. 12, pp. 2923–2929, Dec. 2011.
[24] M. Fardad, F. Lin, and M. R. Jovanović, "Sparsity-promoting optimal control for a class of distributed systems," in *Proc. Amer. Control Conf.*, 2011, pp. 2050–2055.
[25] F. Lin, M. Fardad, and M. R. Jovanović, "Design of optimal sparse feedback gains via the alternating direction method of multipliers," *IEEE Trans. Autom. Control*, vol. 58, no. 9, pp. 2426–2431, Sep. 2013.
[26] M. R. Jovanović and N. K. Dhingra, "Controller architectures: Tradeoffs between performance and structure," *Eur. J. Control*, vol. 30, pp. 76–91, 2016.
[27] B. Polyak, M. Khlebnikov, and P. Shcherbakov, "An LMI approach to structured sparse feedback design in linear control systems," in *Proc. Eur. Control Conf.*, 2013, pp. 833–838.
[28] N. K. Dhingra, M. R. Jovanović, and Z. Q. Luo, "An ADMM algorithm for optimal sensor and actuator selection," in *Proc. 53rd IEEE Conf. Decis. Control*, 2014, pp. 4039–4044.
[29] 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 Trans. Autom. Control*, vol. 65, no. 8, pp. 3441–3456, Aug. 2020.
[30] B. Recht, "A tour of reinforcement learning: The view from continuous control," *Annu. Rev. Control Robot. Auton. Syst.*, vol. 2, pp. 253–279, 2019.
[31] H. T. Toivonen, "A globally convergent algorithm for the optimal constant output feedback problem," *Int. J. Control*, vol. 41, no. 6, pp. 1589–1599, 1985.
[32] T. Rautert and E. W. Sachs, "Computational design of optimal output feedback controllers," *SIAM J. Optim*, vol. 7, no. 3, pp. 837–852, 1997.
[33] A. Vannelli and M. Vidyasagar, "Maximal Lyapunov functions and domains of attraction for autonomous nonlinear systems," *Automatica*, vol. 21, no. 1, pp. 69–80, 1985.
[34] H. K. Khalil, *Nonlinear Systems*. New York, NY, USA: Prentice-Hall, 1996.
[35] H. Karimi, J. Nutini, and M. Schmidt, "Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition," in *Proc. Eur. Conf. Mach. Learn.*, 2016, pp. 795–811.
[36] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[37] S.-I. Amari, "Natural gradient works efficiently in learning," *Neural Comput.*, vol. 10, no. 2, pp. 251–276, 1998.
[38] H. Mohammadi, M. Soltanolkotabi, and M. R. Jovanović, "Random search for learning the linear quadratic regulator," in *Proc. Amer. Control Conf.*, 2020, pp. 4798–4803.
[39] R. Vershynin, *High-Dimensional Probability: An Introduction With Applications in Data Science*. Cambridge, U.K.: Cambridge Univ. Press, 2018.
[40] H. Mohammadi, M. Soltanolkotabi, and M. R. Jovanović, "On the linear convergence of random search for discrete-time LQR," *IEEE Control Syst. Lett.*, vol. 5, no. 3, pp. 989–994, Jul. 2021.
[41] H. Mohammadi, A. Zare, M. Soltanolkotabi, and M. R. Jovanović, "Convergence and sample complexity of gradient methods for the model-free linear quadratic regulator problem," 2019, *arXiv:1912.11899*.
[42] H. T. Toivonen and P. M. Mäkilä, "Newton's method for solving parametric linear quadratic control problems," *Int. J. Control*, vol. 46, no. 3, pp. 897–911, 1987.
[43] M. Rudelson and R. Vershynin, "Hanson-Wright inequality and sub-Gaussian concentration," *Electron. Commun. Probab.*, vol. 18, pp. 1–9, 2013.
[44] M. Soltanolkotabi, A. Javanmard, and J. D. Lee, "Theoretical insights into the optimization landscape of over-parameterized shallow neural networks," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 742–769, Feb. 2019.



Hesameddin Mohammadi received the B.Sc. degree from the Sharif University of Technology, Tehran, Iran, in 2015, and the M.Sc. degree from Arizona State University, Tempe, AZ, USA, in 2017, both in mechanical engineering. He is currently working toward the Ph.D. degree with the Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, USA.

His primary research interests include large-scale optimization, control, and inference problems.

Armin Zare (Member, IEEE) received the B.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2010, and the Ph.D. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, MN, USA, in 2016.

He is currently an Assistant Professor of mechanical engineering with the University of Texas at Dallas, Richardson, TX, USA. From 2017 to 2019, he was a Postdoctoral Research Associate with the Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, USA.

Dr. Zare was a recipient of the Doctoral Dissertation Fellowship with the University of Minnesota in 2015 and a finalist for the Best Student Paper Award at the American Control Conference in 2014.

Mahdi Soltanolkotabi received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2014.

He is currently an Assistant Professor with the Ming Hsieh Department of Electrical and Computer Engineering and Computer Science (by courtesy), University of Southern California, Los Angeles, CA, where he holds an Andrew and Erna Viterbi Early Career Chair. From 2014 to 2015, he was a Postdoctoral Researcher with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA.

Dr. Soltanolkotabi is the recipient of the Information Theory Society Best Paper Award, the Packard Fellowship in Science and Engineering, a Sloan Research Fellowship, an National Science Foundation CAREER Award, an Air Force Office of Scientific Research Young Investigator Award, and a Google Faculty Research Award.

Mihailo R. Jovanović (Fellow, IEEE) received the Ph.D. degree in mechanical engineering from the University of California at Santa Barbara, Santa Barbara, CA, USA, in 2004.

He is currently a Professor with the Ming Hsieh Department of Electrical and Computer Engineering and the Founding Director of the Center for Systems and Control, University of Southern California, Los Angeles, CA. He was a Faculty Member with the Department of Electrical and Computer Engineering, University of Minnesota, Twin Cities, MN, USA, from 2004 to 2017, and has held visiting positions with Stanford University, the Institute for Mathematics and its Applications, the Simons Institute for the Theory of Computing, and the University of Belgrade.

Dr. Jovanović received the CAREER Award from the National Science Foundation in 2007, the George S. Axelby Outstanding Paper Award from the IEEE Control Systems Society in 2013, and the Distinguished Alumnus Award from the Department of Mechanical Engineering, University of California at Santa Barbara, in 2014. Papers of his students were finalists for the Best Student Paper Award at the American Control Conference in 2007 and 2014. He is a Fellow of the American Physical Society.