Optimization Landscape of Gradient Descent for Discrete-time Static Output Feedback

Jingliang Duan, Jie Li, Shengbo Eben Li, Lin Zhao

Abstract—In this paper, we analyze the optimization landscape of gradient descent methods for static output feedback (SOF) control of discrete-time linear time-invariant systems with quadratic cost. The SOF setting can be quite common, for example, when there are unmodeled hidden states in the underlying process. We first establish several important properties of the SOF cost function, including coercivity, L-smoothness, and M-Lipschitz continuous Hessian. We then utilize these properties to show that the gradient descent is able to converge to a stationary point at a dimension-free rate. Furthermore, we prove that under some mild conditions, gradient descent converges linearly to a local minimum if the starting point is close to one. These results not only characterize the performance of gradient descent in optimizing the SOF problem, but also shed light on the efficiency of general policy gradient methods in reinforcement learning.

I. INTRODUCTION

Reinforcement learning (RL) has achieved human-level performance in a variety of challenging domains, from games to robotic control [1]–[5]. For continuous control settings, policy gradient based RL algorithms such as DDPG [3], TRPO [6], PPO [7], SAC [8], DSAC [4], are widely used to find the nearly optimal parameterized policy function. Despite their success in various applications, the theoretical understandings of their performance and computational complexity are rather limited, even in the most basic settings such as linear systems.

Recent literature investigated the optimization landscape and the global convergence of policy gradient methods when applied to the linear quadratic regulator (LQR). It is well known that directly optimizing the LQR cost with respect to the linear policy is generally a nonconvex optimization. The domain of stabilizing linear policy can be nonconvex with non-smooth boundaries. As a pioneering work, Fazel et al. (2018) demonstrated the gradient dominance and the almost smoothness properties of the LQR cost in a discrete-time setting, leading to the global convergence at a linear rate for policy gradient methods [9]. This work has inspired a line of subsequent studies that examined similar properties through the lens of policy gradient methods for discrete-time and continuous-time LQR [10]–[13], finite-horizon noisy LQR [14], LQR tracking [15], Markovian jump LQR [16], distributed LQR [17], linear H₂ control with H∞ constraints [18], as well as linear quadratic Gaussian problem [19], [20].

The aforementioned literature mainly focuses on the full state feedback or dynamic output feedback control, whereas the convergence of the gradient method for static output feedback (SOF) LQR is barely studied. SOF is one of the most important open questions in control theory, which considers a linear time-invariant system with quadratic performance in the case when only some outputs (linear functions of the state) instead of the complete state are available [21]. Note that state feedback LQR can be regarded as a special case of SOF. Compared with state feedback LQR, the optimization landscape for SOF LQR is much more complicated. The domain of stabilizing SOF control gains can be disconnected and the stationary points in each component of the domain can be local minima, saddle points, or even local maxima [13], [22]–[24]. The recent work [13] analyzed the convergence rate to stationary points, which however only focused on continuous-time linear systems.

This paper aims to analyze the optimization landscape of the discrete-time SOF. The contributions are summarized as follows.

1) We establish several important properties of the SOF cost function, including coercivity, L-smoothness, and M-Lipschitz continuous Hessian. These results enable further convergence analysis of policy gradient for discrete-time SOF.

2) For the partially observed case, results on convergence to stationary points and the corresponding dimension-free convergence rate are formally established. Moreover, utilizing the Lipschitz continuity of Hessian, we demonstrate that under mild conditions, gradient descent converges linearly to a local minimum if the starting point is close to one.

3) For the fully observed case, we show that gradient descent converges to the global optimal controller at a linear rate if an initial stabilizing controller is used. Compared with the results given in [9], we provide a more explicit characterization of computational complexity.

The remainder of this paper is organized as follows. In Section III, we describe some preliminaries of SOF. Section IV derives the formula of the gradient and establishes several key properties of the SOF cost function. Section V shows the convergence results of gradient descent for SOF, and Section VI concludes this article.

Notation: ∥X∥ and ρ(X) denote the spectral norm and spectral radius of a matrix X, respectively; Tr(X), σ min(X), and λ min(X) denote the trace, minimal singular value,
and minimal eigenvalue of a square matrix, respectively. \( \partial X \) denotes the boundary of a set \( X \). \( \text{vec}(X) \) denotes the vectorization of matrix \( X \). \( X \succ Y \) and \( X \succeq Y \) represent that \( X - Y \) is positive definite and positive semidefinite, respectively.

II. PRELIMINARIES

Consider the discrete-time time-invariant linear system

\[
\begin{align*}
x_{t+1} &= Ax_t + Bu_t, \\
y_t &= Cx_t,
\end{align*}
\]

(1)

with state \( x \), output \( y \), and matrices \( A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m}, C \in \mathbb{R}^{d \times n} \). Without loss of generality, we assume \( C \) has full row rank. The objective of infinite-horizon LQR performance is given by

\[
E_{x_0 \sim \mathcal{D}} \left[ \sum_{t=0}^{\infty} (x_t^T Q x_t + u_t^T R u_t) \right],
\]

(2)

where initial state \( x_0 \) is assumed to be randomly distributed according to a distribution \( \mathcal{D} \), and \( Q \in \mathbb{R}^{n \times n} \) and \( R \in \mathbb{R}^{m \times m} \) are positive definite matrices.

The static output feedback is described by the form

\[
u_t = -K y_t,
\]

(3)

where \( K \in \mathbb{R}^{m \times d} \). Then the closed loop system can be written as

\[
x_{t+1} = (A - BKC)x_t,
\]

(4)

and the objective function becomes

\[
J(K) = E_{x_0 \sim \mathcal{D}} \left[ \sum_{t=0}^{\infty} x_t^T (Q + C^T K^T RKC)x_t \right].
\]

(5)

We assume that a stabilizing controller exists. We denote the feasible set which contains all stabilizing \( K \) as \( \mathcal{K} \), that is

\[
\mathcal{K} = \{ K \in \mathbb{R}^{m \times d} : \rho(A - BKC) < 1 \}.
\]

Define \( P_K \) as the solution to the Lyapunov equation

\[
P_K = Q + C^T K^T RKC + (A - BKC)^T P_K (A - BKC),
\]

(6)

and then the cost function can be expressed by

\[
J(K) = E_{x_0 \sim \mathcal{D}} x_0^T P_K x_0 = \text{Tr}(P_K X_0),
\]

(7)

where

\[
X_0 := E_{x_0 \sim \mathcal{D}} x_0 x_0^T.
\]

Define the state correlation matrix as

\[
\Sigma_K := E_{x_0 \sim \mathcal{D}} \sum_{t=0}^{\infty} x_t x_t^T,
\]

then it is straightforward that

\[
\mu := \sigma_{\text{min}}(X_0) \leq \sigma_{\text{min}}(\Sigma_K).
\]

It is also well-known that \( \Sigma_K \) satisfies the following Lyapunov equation

\[
\Sigma_K = X_0 + (A - BKC) \Sigma_K (A - BKC)^T.
\]

(8)

III. GRADIENT AND COST PROPERTIES

In this section, we first derive the formula of the gradient and Hessian, and then establish several key properties of the SOF cost.

A. Gradient and Second Derivative

**Lemma 1** (Policy Gradient Expression). For \( \forall K \in \mathcal{K} \), the policy gradient is

\[
\nabla J(K) = 2E_K \Sigma_K C^T,
\]

(9)

where \( E_K = (R + B^T P_K B)KC - B^T P_K A \).

**Proof.** Define the value function of state \( x_t \) as

\[
V_K(x_t) := x_t^T P_K x_t.
\]

From (6), it follows that

\[
V_K(x_0) = x_0^T P_K x_0 = x_0^T (Q + C^T K^T RKC) x_0
\]

\[
+ x_0^T (A - BKC) P_K (A - BKC) x_0
\]

\[
= x_0^T (Q + C^T K^T RKC) x_0 + V_K((A - BKC) x_0).
\]

Taking the gradient of \( V_K(x_0) \) w.r.t. \( K \), one has

\[
\nabla V_K(x_0) = 2RKCx_0 x_0^T C^T - 2B^T P_K (A - BKC) x_0 x_0^T C^T
\]

\[
+ x_0^T \nabla P_K x_0 \bigg|_{x_0 = (A - BKC) x_0}
\]

\[
= 2[(R + B^T P_K B)KC - B^T P_K A] x_0 x_0^T C^T
\]

\[
+ \nabla V_K(x_0) \bigg|_{x_0 = (A - BKC) x_0}
\]

\[
- 2[(R + B^T P_K B)KC - B^T P_K A] \sum_{t=0}^{\infty} (x_t x_t^T) C^T.
\]

Then, we can observe that

\[
\nabla J(K) = E_{x_0 \sim \mathcal{D}} \nabla V_K(x_0) = 2E_K \Sigma_K C^T.
\]

\[\square\]

The SOF cost \( J(K) \) is twice differentiable. To avoid tensors, we analyze the Hessian of the SOF cost \( J(K) \) applied to a nonzero direction \( Z \in \mathbb{R}^{m \times d} \), which is given as

\[
\nabla^2 J(K)[Z, Z] := \frac{d^2}{d\eta^2} \bigg|_{\eta=0} J(K + \eta Z)
\]

\[
= E_{x_0 \sim \mathcal{D}} x_0^T \frac{d^2}{d\eta^2} \bigg|_{\eta=0} P_K + \eta Z x_0
\]

(10)

\[
= \text{Tr} \left( \frac{d^2}{d\eta^2} \bigg|_{\eta=0} P_K + \eta Z X_0 \right).
\]

**Lemma 2.** For \( \forall K \in \mathcal{K} \), the Hessian of the SOF cost \( J(K) \) applied to a direction \( Z \in \mathbb{R}^{m \times d} \) satisfies

\[
\nabla^2 J(K)[Z, Z] = 2 \text{Tr}(C^T Z^T (R + B^T P_K B) Z \Sigma_K)
\]

\[
- 4 \text{Tr}((BZC)^T P_K [Z](A - BKC) \Sigma_K).
\]

(11)
where
\[ P'_K[Z] = \sum_{j=0}^{\infty} (A - BKC)^{\top}J(C^T Z^\top E_K + E_K^T ZC)(A - BKC)^j. \tag{12} \]

**Proof.** Denote \( P'_K[Z] = \frac{d}{dn} P_{K+nZ}. \) From (6), we have
\[ P'_K[Z] = C^T Z^\top RK + C^T K^\top RZC \]
\[ - (BZC)^\top P_K (A - BKC) - (A - BKC)^\top P_K BZC \]
\[ + (A - BKC)^\top P_K' [Z] (A - BKC) \]
\[ = \sum_{j=0}^{\infty} (A - BKC)^{\top}J(C^T Z^\top E_K + E_K^T ZC)(A - BKC)^j. \tag{13} \]

Furthermore, its second derivative can be derived as
\[ \frac{d^2}{dn^2} P_{K+nZ} = C^T Z^\top ((R + B^TP_K B)ZC - B^TP_K' [Z] (A - BKC)) \]
\[ + ((R + B^TP_K B)ZC - B^TP_K' [Z] (A - BKC))^\top ZC \]
\[ - (BZC)^\top P_K' [Z] (A - BKC) \]
\[ - (A - BKC)^\top P_K' [Z] BZC \]
\[ + (A - BKC)^\top \frac{d^2}{dn^2} P_{K+nZ} (A - BKC) \]
\[ = \sum_{j=0}^{\infty} (A - BKC)^{\top}J S_1 (A - BKC)^j, \tag{14} \]

where
\[ S_1 := 2(C^T Z^\top (R + B^TP_K B)ZC \]
\[ - (BZC)^\top P_K' [Z] (A - BKC) \]
\[ - (A - BKC)^\top P_K' [Z] BZC). \]

Finally, using (10), we can show that
\[ \nabla^2 J(K) [Z, Z] = \text{Tr}(\sum_{j=0}^{\infty} (A - BKC)^{\top}J S_1 (A - BKC)^j X_0) \]
\[ = \text{Tr}(S_1 \sum_{j=0}^{\infty} (A - BKC)^j X_0 (A - BKC)^{\top}J) \]
\[ = \text{Tr}(S_1 \Sigma_K) \]
\[ = 2\text{Tr}(C^T Z^\top (R + B^TP_K B)ZC \Sigma_K) \]
\[ - 4\text{Tr}((BZC)^\top P_K' [Z] (A - BKC) \Sigma_K). \]

**B. Cost Function Properties**

Given the analytical policy gradient, before moving forward to analyze the performance of gradient-based methods in solving the SOF problem, we first need to establish several key properties of the SOF cost. The intermediate lemmas required by the property analysis are provided in Appendix A.

**Lemma 3 (Coercivity).** The cost function \( J(K) \) is coercive in the sense that for any sequence \( \{K_i\}_{i=1}^{\infty} \subseteq \mathbb{K} \)
\[ J(K_i) \to +\infty, \quad \text{if} \quad \|K_i\| \to +\infty \quad \text{or} \quad K_i \to K \in \partial \mathbb{K}. \]

**Proof.** From (7), we can show that
\[ J(K_i) = \mathbb{E}_{x_0 \sim \mathcal{D}} \left( \sum_{t=0}^{\infty} x_t^\top (Q + C^T K_i^\top RK_i) x_t \right) \]
\[ = \text{Tr}((Q + C^T K_i^\top RK_i) \Sigma_K) \]
\[ \geq \mu \sigma_{\min}(Q) \sigma_{\min}(C^T) \|K_i\|^2 \]
\[ \geq \mu \sigma_{\min}(Q) \sigma_{\min}(C)^2 \|K_i\|^2, \]
which directly leads to that \( J(K_i) \to +\infty \) as \( \|K_i\| \to +\infty \).

From (6), \( P_K \) can be rewritten as
\[ P_K = \sum_{j=0}^{\infty} (A - BKC)^{\top}J (Q + C^T K^\top RK)(A - BKC)^j. \tag{15} \]

Therefore, we have
\[ J(K_i) = \]
\[ \text{Tr}(\sum_{j=0}^{\infty} (A - BKC)^{\top}J (Q + C^T K_i^\top RK_i) (A - BKC)^j X_0) \]
\[ \geq \mu \sigma_{\min}(Q) \sum_{t=0}^{\infty} \| (A - BKC)^j \|^2 \]
\[ \geq \mu \sigma_{\min}(Q) \sum_{t=0}^{\infty} \rho (A - BKC)^j \]
\[ \geq \mu \sigma_{\min}(Q) \sum_{t=0}^{\infty} \rho (A - BKC)^j \]
\[ = \mu \sigma_{\min}(Q) \frac{1 - \rho (A - BKC)^\infty}{1 - \rho (A - BKC)^2}. \]

Since \( \rho (A - BKC) = 1 \) when \( K \in \partial \mathbb{K} \), by continuity of the \( \rho (A - BKC) \), we have \( \rho (A - BKC) \to 1 \) as \( K_i \to K \in \partial \mathbb{K} \). Thereby, for every \( \epsilon > 0 \), there exists some \( N(\epsilon) \in \mathbb{N} \) such that \( 1 - \rho (A - BKC) < \epsilon \) for all \( i \geq N(\epsilon) \). That is \( 1 > \rho (A - BKC) > 1 - \epsilon \) for \( i \geq N(\epsilon) \). Thereby, \( J(K_i) \) is bounded below by
\[ J(K_i) \geq \frac{\mu \sigma_{\min}(Q)}{1 - (1 - \epsilon)^2}. \]

It thus follows that \( J(K_i) \to +\infty \) as \( K_i \to \partial \mathbb{K} \). This completes the proof of Lemma 3.

With the coercive property in place, we can easily show the compactness on the sublevel of the SOF cost.

**Lemma 4 (Compactness on the sublevel set).** The sublevel set \( \mathbb{K}_\alpha = \{ K \in \mathbb{K} : J(K) \leq \alpha \} \) is compact for \( \forall \alpha \geq J(K^*) \).

**Proof.** With the coercive property in place, by [25, Proposition 11.12], we know that \( \mathbb{K}_\alpha \) is bounded. Since \( J(K) \) is
continuous on $K$, the set $\mathbb{K}_\alpha$ is also closed, which completes the proof.

Based on the compactness of the sublevel set, we can finally show that under certain conditions, the decrease of the cost ensures the line segment between two adjacent iterates stays inside $\mathbb{K}_\alpha$.

**Lemma 5** (Smoothness on the sublevel set). For any $K \in \mathbb{K}_\alpha$, we have $\|\nabla^2 J(\text{vec}(K))\| \leq L$, where the smoothness constant $L$ is

$$
L = 2\|C\|^2 \left(\|R\| + \|B\|^2 (1 + \frac{2\zeta_1}{\alpha\|B\|} \frac{\alpha}{\mu} \frac{\alpha}{\sigma_{\text{min}}(Q)}) \right),
$$

with

$$
\zeta_1 = \frac{1}{\sigma_{\text{min}}(Q)} \left( 1 + \|C\|^2 \|B\|^2 \frac{\alpha}{\mu} + \|C\|^2 \|R\| \right) - 1. \quad (16)
$$

This means that for any $K, K' \in \mathbb{K}_\alpha$ satisfying $K + \delta(K' - K) \in \mathbb{K}_\alpha$, $\forall \delta \in [0, 1]$, the following inequality holds

$$
J(K') \leq J(K) + \text{Tr}(\nabla J(K) \nabla J(K')^\top) + \frac{L}{2} \|K' - K\|^2. \quad (17)
$$

**Proof.** From $\text{(10)}$, applying the Taylor series expansion about direction $Z$, we can show that

$$
\nabla^2 J(K)[Z, Z] = \text{vec}(Z) \nabla^2 J(\text{vec}(K)) \text{vec}(Z).
$$

Since $\nabla^2 J(\text{vec}(K))$ is symmetric, one has

$$
\|\nabla^2 J(\text{vec}(K))\| = \sup_{\|Z\|=1} |\text{vec}(Z) \nabla^2 J(\text{vec}(K)) \text{vec}(Z)|
= \sup_{\|Z\|=1} \|\nabla^2 J(K)[Z, Z]\|. \quad (18)
$$

Using $\text{(11)}$, we further have

$$
\|\nabla^2 J(\text{vec}(K))\|
\leq 2 \sup_{\|Z\|=1} |\text{Tr}(C^\top Z^\top (R + B^T P_K B) Z \Sigma_K)|
+ 4 \sup_{\|Z\|=1} |\text{Tr}((B Z C)^\top P_K^\top (A - B K C) \Sigma_K)|
= 2q_1 + 4q_2. \quad (19)
$$

Therefore, we only need to analyze the upper bounds of $q_1$ and $q_2$, separately. Firstly, we have

$$
q_1 \leq \sup_{\|Z\|=1} (\|C^\top Z^\top (R + B^T P_K B) Z\| \text{Tr}(\Sigma_K))
\leq \sup_{\|Z\|=1} (\|C\|^2 \|Z\|^2 \|R + B^T P_K B\| \text{Tr}(\Sigma_K))
\leq \sup_{\|Z\|=1} (\|C\|^2 \|Z\|^2 \|R\| + \|B\|^2 \|P_K\| \text{Tr}(\Sigma_K))
\leq \|C\|^2 (\|R\| + \|B\|^2 \frac{J(K)}{\mu} \frac{J(K)}{\sigma_{\text{min}}(Q)}),
$$

where the last step follows from Lemma $9$ in the appendix.

Next, by the Cauchy-Schwarz inequality, we have

$$
q_2 \leq \sup_{\|Z\|=1} (\|B Z C\| \|P_K^\top (A - B K C) \Sigma_K^{-1/2} F\| \|\Sigma_K^{-1/2} F\|)
\leq \sup_{\|Z\|=1} (\|C\|^2 \|Z\|^2 \|B\| \|P_K\| \|F\|)
\leq \|C\|^2 \|B\| \text{Tr}(\Sigma_K) \sup_{\|Z\|=1} (\|P_K\| \|F\|)
\leq \|C\|^2 \|B\| \text{Tr}(\Sigma_K) \sup_{\|Z\|=1} (\|P_K\|),
$$

where the last step follows from Lemma $9$. Combining $20$ and $21$, the only thing left is to show the following bound holds:

$$
\sup_{\|Z\|=1} \|P_K\| \leq \zeta_1 \|P_K\|,
$$

where $\zeta_1$ is as given by $16$.

We will prove the above inequality by showing that $P_K^\top [Z] \preceq \zeta_1 P_K$. Given $6$ and $12$, if $(C^\top Z^\top E_K + E_K^\top Z C) \preceq \zeta_1 (Q + C^\top K^\top R K C)$ for some $\zeta_1 \in \mathbb{R}^+$, we immediately have $P_K^\top [Z] \preceq \zeta_1 P_K$. Therefore, it remains to find $\zeta_1$. It follows from $6$ that

$$
C^\top Z^\top E_K + E_K^\top Z C
= C^\top Z^\top R K C + C^\top K^\top Z C^\top R K C
+ (A - B K C)^\top P_K (A - B K C) + (B Z C)^\top P_K B Z C
= P_K - Q + C^\top Z^\top R K C + (B Z C)^\top P_K B Z C
\preceq (\|P_K\| + \|C\|^2 \|R\| + \|C\|^2 \|B\| \|Z\| \|Q\| \sigma_{\text{min}}(Q)) - Q.
$$

Therefore, we proved that $P_K^\top [Z] \preceq \zeta_1 P_K$, which directly leads to

$$
q_2 \leq \zeta_1 \|C\|^2 \|B\|^2 \|P_K\| \frac{J(K)}{\sigma_{\text{min}}(Q)}
\leq \zeta_1 \|C\|^2 \|B\|^2 \frac{J(K)^2}{\mu \sigma_{\text{min}}(Q)}. \quad (23)
$$

Combining $19$, $20$, and $23$, for $K \in \mathbb{K}_\alpha$, we finally have

$$
\|\nabla^2 J(\text{vec}(K))\|
\leq 2\|C\|^2 (\|R\| + \|B\|^2 \frac{\alpha}{\mu} \frac{\alpha}{\sigma_{\text{min}}(Q)} + 4\zeta_1 \|C\|^2 \|B\| \frac{\alpha^2}{\mu \sigma_{\text{min}}(Q)}
\leq 2\|C\|^2 (\|R\| + \|B\|^2 (1 + \frac{2\zeta_1}{\|C\|^2 \|B\|^2} \frac{\alpha}{\mu} \frac{\alpha}{\sigma_{\text{min}}(Q)})
\equiv \frac{\alpha}{\mu} \sigma_{\text{min}}(Q). \quad (24)
$$

If $J(K)$ is globally $L$-smooth, it is well known that gradient descent can converge to a stationary point with
a dimension-free iteration complexity [26], [27]. However, Lemma 5 only shows that $\|\nabla^2 J(\text{vec}(K))\| \leq L$ for $K \in \mathbb{K}_n$. Since the sublevel set $\mathbb{K}_n$ can be nonconvex with a non-smooth boundary, or even disconnected [13], the inequality (17) cannot be directly applied to any $K, K' \in \mathbb{K}_n$. Hence, the convergence analysis is non-trivial.

**Lemma 6** (Performance difference lemma). Suppose $K, K' \in \mathbb{K}$. It holds that:

$$J(K') - J(K) = 2 \text{Tr}(\Sigma_K (K'C - KC)^T E_K) + \text{Tr}(\Sigma_K (K'C - KC)^T (R + B_\delta^T P_K B)(K'C - KC)).$$

We provide a brief proof of Lemma 6 in Appendix B. The performance difference lemma, also known as almost cost and $K$ following inequality holds optimal SOF policy, and $K$ is symmetric, we have $J(K) = \frac{1}{2} \|R + B_\delta^T P_K B\|_F^2$. Hence, the convergence analysis is non-trivial.

**Lemma 7** (Gradient domination). Define the set $\mathbb{C}$ as $\mathbb{C} = \{C \in \mathbb{R}^{n \times n} : \text{rank}(C) = n\}$. Let $K^*$ be the global optimal SOF policy, and $J_*^*$ be the global optimal cost of the corresponding state feedback LQR. Suppose $K$ has finite cost and $\mu > 0$. It holds that

$$J(K) - J(*^*) \leq \frac{\|\Sigma_K\| \|\nabla J(K)\|_F^2}{4\mu^2 \sigma_{\min}(C)^2 \sigma_{\min}(R)}, \forall C \in \mathbb{C}. \quad (25)$$

Moreover, we have the following lower bound

$$J(K) - J_*^* \geq \frac{\mu \text{Tr}(E_K^T P_K E_K)}{\|R + B_\delta^T P_K B\|}, \forall C \in \mathbb{C}. \quad (26)$$

These results can be proved in a similar way to [9, Lemma 3] where $C = I_n$ (see Appendix C for details). The gradient domination condition indicates that there are no stationary points other than the global minimum, which is crucial to achieve global convergence for policy gradient [26], [28]. However, when rank($C$) $< n$, gradient dominance does not hold. Therefore, in the latter case, it is hard to expect anything better than converging to a stationary point.

**Lemma 8** (M-Lipschitz continuous Hessian). For any $K, K' \in \mathbb{K}_n$ satisfying $K + \delta(K' - K) \in \mathbb{K}_n, \forall \delta \in [0, 1]$, the following inequality holds

$$\|\nabla^2 J(\text{vec}(K)) - \nabla^2 J(\text{vec}(K'))\| \leq M \|K - K'\|,$$

where

$$M = \frac{2\alpha^2}{\mu \sigma_{\min}(Q)} \|B\|_F \|C|((2\zeta_1 + \zeta_2)\|B\|_F \|C| + 2\zeta_3 + \zeta_4),$$

and $\zeta_{1,2,3,4}$ are defined in Lemma 5 and 7.

**Proof.** Similar to (13), since $\nabla^2 J(\text{vec}(K)) - \nabla^2 J(\text{vec}(K'))$ is symmetric, we have

$$\|\nabla^2 J(\text{vec}(K)) - \nabla^2 J(\text{vec}(K'))\| = \sup_{\|Z\|_F = 1} \|\nabla^2 J(K)|Z, Z| - \nabla^2 J(K')|Z, Z|.$$

By (10), we define

$$g(\delta) := \nabla^2 J(K + \delta(K' - K))|Z, Z|,$$

and denote $K := K + \delta(K' - K)$, $\Delta K := K' - K$, and $P''_K[Z] := \frac{\partial^2}{\partial \eta^2} P_K + \eta Z \cdot Then, from (14), one has

$$g'(\delta) = \mathbb{E}_{\epsilon_0 \sim \mathcal{D}} \left( \int_0^1 \frac{\partial^3}{\partial \eta^3} P_K + \delta(\Delta K) \right) dx.$$
By plugging (29) into (27), we finally complete the proof.

The Lipschitz continuity of the Hessian can be used to obtain a stronger result of convergence to a local minimum for nonconvex optimization under mild assumptions [27]. Besides, this property has also been widely used in [29]–[31] to achieve efficient strict saddle point escape.

IV. CONVERGENCE ANALYSIS

In this section, we prove several novel convergence results of the gradient descent method for SOF LQR. Since \( J(K) \) is nonconvex, the properties of the cost function obtained in Section II-B will play a key role in the convergence analysis.

A. Convergence to Stationary Points

We consider the following gradient descent update rule

\[
K_{i+1} = K_i - \eta \nabla J(K_i),
\]

where \( K_0 \in \mathbb{R} \). By Lemma 5, if we can ensure that the line segment between \( K_i \) and \( K_i - \eta \nabla J(K_i) \) is in \( \mathbb{R}^d \), the convergence of policy descent (30) for SOF can be proved by invoking a similar analysis to the case of global \( L \)-smoothness. Before presenting the main theorem, we introduce the following definition for convenience.

Definition 1. For a differentiable function \( J(\cdot) \), we say \( K \) is a first-order stationary point if \( \|\nabla J(K)\|_F = 0 \); we also say \( K \) is an \( \epsilon \)-first-order stationary point if \( \|\nabla J(K)\|_F \leq \epsilon \).

Theorem 1. Suppose \( J(K_0) = \alpha, \mu > 0 \), and \( L \) is as defined in Lemma 5. If we run gradient descent (30) with any step size \( \eta \leq 1/L, J(K_i) \) is monotone decreasing, and the algorithm will output an \( \epsilon \)-first-order stationary point within the following number of iterations

\[
N \geq \frac{\|\Sigma_K\|}{2\eta \mu^2 \sigma_{\min}(C)^2 \sigma_{\min}(R)} \log\left( \frac{J(K_0) - J^*_s}{\epsilon_f} \right).
\]

Proof. Firstly, we define \( \mathbb{R}_0^d := \{ K \in \mathbb{R} : J(K) < \alpha \} \), and denote its complement as \( \mathbb{R}_0^d \), which is closed. From Lemma 5, given \( \phi \in (0, 1) \), there exists \( \epsilon > 0 \) such that

\[
\|\nabla^2 J(\text{vec}(K))\| \leq (1 + \phi) L \quad \text{for} \quad K \in \mathbb{R}_0^d + \epsilon.C.
\]

It is clear that \( \mathbb{R}_0^d \cap (\mathbb{R}_0^d + C) = \emptyset \).

Since \( \mathbb{R}_0^d \) is compact by Lemma 4, the distance between \( \mathbb{R}_0^d \) and \( (\mathbb{R}_0^d + C) \), defined as \( \omega := \inf\{\|x - y\| : x \in \mathbb{R}_0^d, y \in (\mathbb{R}_0^d + C)\} \), must be strictly positive. Let \( \tau \in \mathbb{R}_+ \) be the value such that \( \|K - \tau \nabla J(K)\| \leq (1 + \phi) L \). It follows from \( \tau \leq \omega / \|\nabla J(K)\| \) that the line segment between \( K \) and \( K - \tau \nabla J(K) \) is in \( \mathbb{R}_0^d + \epsilon.C. \). Since \( \|\nabla^2 J(\text{vec}(K))\| \leq (1 + \phi) L \), by (31), we can show that

\[
J(K - \tau \nabla J(K)) \leq J(K) - \tau(1 - (1 + \phi) L/2\tau)\|\nabla J(K)\|_F^2.
\]

Since \( \tau \leq 2/(1 + \phi) L \), we have \( J(K - \tau \nabla J(K)) \leq J(K) \). Thereby, \( K - \tau \nabla J(K) \in \mathbb{R}_0^d \). It is then obvious that the line segment \( [K, K - \tau \nabla J(K)] \in \mathbb{R}_0^d \) is compact. By Lemma 5, if we can ensure that \( [K, K - \tau \nabla J(K)] \in \mathbb{R}_0^d \). By induction, we can show that the line segment \( [K, K - \tau \nabla J(K)] \in \mathbb{R}_0^d \). Therefore, \( [K, K - \eta \nabla J(K)] \in \mathbb{R}_0^d \). Applying (32) again, we have

\[
J(K - \eta \nabla J(K)) \leq J(K) - \eta(1 - (1 + \phi) L/2\tau)\|\nabla J(K)\|_F^2
\]

where the last step utilizes the fact that \( \eta \leq 1/L \). Note that the boundary 1/L is chosen because it leads to the fastest convergence rate.

Since \( J(K_0) = \alpha \), according to (32), \( K_1 \) generated by (30) is in \( \mathbb{R}_0^d \). By induction, we can show that

\[
J(K_{i+1}) - J(K_i) \leq \frac{\eta^2}{2}\|\nabla J(K_i)\|_F^2
\]

holds for any \( i \), which also means that \( [K_i, K_{i+1}] \in \mathbb{R}_0^d \) for any \( i \). Let us sum (33) for \( i = 0, \ldots, N \). We can obtain

\[
\frac{\eta^2}{2} \sum_{i=0}^{N} \|\nabla J(K_i)\|_F^2 \leq J(K_0) - J(K_{N+1}) \leq J(K_0) - J(K^*_s).
\]

Therefore, it is clear that

\[
\min_{0 \leq i \leq N} \|\nabla J(K_i)\|_F^2 \leq \frac{2(J(K_0) - J(K^*_s))}{\eta N} \leq \frac{2\alpha}{\eta N}.
\]

Therefore, the policy descent will output an \( \epsilon \)-first-order stationary point within the following number of iterations:

\[
N \geq \frac{2\alpha}{\eta \mu^2 \sigma_{\min}(C)^2 \sigma_{\min}(R)} \log\left( \frac{J(K_0) - J^*_s}{\epsilon_f} \right).
\]

If \( C \in \mathbb{C} \), by (33) and (25), we further have

\[
J(K_{i+1}) - J(K_i) \leq \frac{2\mu^2 \sigma_{\min}(C)^2 \sigma_{\min}(R)}{\|\Sigma_{K^*}\|} \log\left( \frac{J(K_0) - J^*_s}{\epsilon_f} \right),
\]

which directly leads to

\[
J(K_{i+1}) - J^*_s \leq \left[ 1 - \frac{2\mu^2 \sigma_{\min}(C)^2 \sigma_{\min}(R)}{\|\Sigma_{K^*}\|} \right] J(K_0) - J^*_s.
\]

Therefore, when \( C \in \mathbb{C} \), for \( \eta \geq 2/[(1 + \phi) L] \), the gradient descent (30) enjoys that \( J(K_{i+1}) - J^*_s \leq \epsilon_f \).
of the convergence rate, which provides a baseline to the model-free settings. In the latter case, existing policy learning techniques, such as the zeroth-order optimization approach, provide effective ways to obtain an unbiased estimation of $\nabla J(K)$ from data [32], [33]. The work in [9] showed that with enough samples and roll-out length, the zeroth-order optimization approach could estimate the gradient within a desired accuracy. The same can be said of discrete-time STO as well.

B. Convergence to Local Minima

**Theorem 2.** Let $\omega$ denote the distance between $K_\beta$ and $(K_\alpha)^c$. Suppose $\beta < \alpha$, i.e., $K_\beta \subset K_\alpha$. Set $L$ and $M$ as described in Lemma 5 and 8, respectively. Assume the initial point $K_0$ be close to a local minimum $K^* \in K_\beta$ with positive definite Hessian, i.e., $l = \lambda_{\min}(\nabla^2 J(\text{vec}(K^*))) > 0$. Define $\bar{r} := 2l/M$. If $r_0 = \|K_0 - K^*\|_F < \bar{r}$ and $r_0/\|r - r_0\| \leq \omega$, then the gradient descent with a fixed step-size $\eta \leq 1/L$ converges with the following rate:

$$\|K_t - K^*\|_F \leq \frac{\bar{r}r_0}{\bar{r} - r_0} \left( \frac{1}{1 + \eta t} \right)^i.$$

**Proof.** Given a certain local minimum $K^*$, define set $K^* = \{K \in \mathbb{R}^{m \times n} : \|K - K^*\|_F \leq \bar{r}r_0/(\bar{r} - r_0)\}$. Since both $K_\beta$ and $(K_\alpha)^c$ are compact, suppose $\bar{r}r_0/(\bar{r} - r_0) \leq \omega$, it is clear that $K^* \subset K_\alpha$. For any $K, K' \in K^*$, we can show that $K + \delta(K' - K) \in K^* \subset K_\alpha$, $\forall \delta \in [0, 1]$. Then, from Lemma 5 and 8 it is straightforward that for any $K, K' \in K^*$, it holds:

$$\|\nabla J(K') - \nabla J(K)\|_F \leq L\|K' - K\|,$$

and

$$\|\nabla^2 J(\text{vec}(K')) - \nabla^2 J(\text{vec}(K))\|_F \leq M\|K' - K\|.$$

The end of the proof can be completed by directly applying [27, Theorem 1.2.4] because we are able to use $L$-smoothness and $M$-Hessian Lipschitz along the entire descent domain.

Theorem 2 implies that gradient descent will converge to the local minimum at a linear rate if the starting point is sufficiently close to one. Note that the above is linear convergence in terms of the sequence and not the function value.

C. Influence of the Initial Distribution

**Proposition 1.** Denote the stationary points of SOF as $K^*$. If $1 \leq \text{rank}(C) < n$, $K^*$ depends on the initial state distribution $\mathcal{D}$, unless $K^* = K^*_s$, where $K^*_s$ is the optimal control gain of the state feedback LQR (i.e., $C = I_n$).

**Proof.** By (23) in Lemma 7 we have $\|E_{K^*}\|_F = 0$, for $C \in \mathbb{C}$. From (33), one has

$$u_t = -K y_t = -KC x_t,$$

which means $KC$ can be deemed as a state-feedback gain.

By the gradient dominance property, we know the stationary point $K^*_s$ of state feedback LQR (when $C = I_n$) is unique, it is straightforward that

$$\|E_K\|_F = 0 \text{ iff } KC = K^*_s.$$

Note that if rank($C$) < $n$, such an output feedback $K$ that makes $KC = K^*_s$ hold may not exist.

The classical control theory [34, Lemma 2] shows that $K^*_s$ has a unique positive definite solution $\Sigma_K$ if $K \in \mathbb{R}$ and $X_0 \succ 0$. Given two different initial distributions $\mathcal{D}' \neq \mathcal{D}$, it follows from (8) that

$$\Sigma_{K^*_s} \neq \Sigma_K$$

for all possible distribution $\mathcal{D}'$ due to the impact of $\mathcal{D}'$ on $\Sigma_{K^*_s}$. This indicates that the stationary point of SOF when $C \notin \mathbb{C}$ depends on the initial state distribution $\mathcal{D}$, that is, different initial distribution will lead to a different stationary point, unless $K^* = K^*_s$.

Proposition 1 suggests that in practice, the initial state distribution should be chosen to be close to the actual application scenario to minimize the expected cost.

V. CONCLUSION

In this paper, we have analyzed the optimization landscape of the gradient descent method for the static output feedback (SOF) linear quadratic control. First, we established several important properties of the SOF cost function, such as coercivity, $L$-smoothness, and $M$-Lipschitz continuous Hessian. Based on these results, we showed that the gradient descent method converges to the global optimal controller at a linear rate in the fully observed case. Novel results on convergence (and rate of convergence) to stationary points were obtained in the partially observed case. The dependence of the stationary points on the initial state distribution is characterized explicitly. We further demonstrated that under mild assumptions, gradient descent converges linearly to a local minimum if the starting point is close to one for partially observed SOF. Our results provided a baseline for evaluating the sample complexities of other model-free gradient descent methods in reinforcement learning, where only estimated gradients are utilized.

ACKNOWLEDGMENT

We would like to acknowledge Yang Zheng and Shengbo Eben Li for their valuable suggestions.
APPENDIX

A. Intermediate Lemmas

Lemma 9. The upper bound of $\|P_K\|$ and $\|\Sigma_K\|$ are given by

$$\|P_K\| \leq \frac{J(K)}{\mu}, \quad \|\Sigma_K\| \leq \text{Tr}(\Sigma_K) \leq \frac{J(K)}{\sigma_{\min}(Q)}.$$ 

Proof. From (37), one has

$$J(K) = \text{Tr}(P_K X_0) \geq \|P_K\|.$$ 

Then, the first claim can be directly obtained. Similarly, $J(K)$ can also be lower bounded by

$$J(K) = \text{Tr}((Q + C^T K^T RKC) \Sigma_K) \geq \sigma_{\min}(Q) \text{Tr}(\Sigma_K) \geq \sigma_{\min}(Q) \|\Sigma_K\|,$$

which leads to the second claim.

Lemma 10. For any $K \in \mathbb{K}$, it holds that

$$\|KC\| \leq q_3 := \frac{\sqrt{\|R\| + \|B\|^2 \sigma^2 / \mu}}{\sqrt{\mu \sigma_{\min}(R)}} + \frac{\|B\| \|A\| \alpha}{\mu \sigma_{\min}(R)}.$$ 

Proof. First, we can observe that

$$\|KC\| = \|(R + B^T P_K B)^{-1} (R + B^T P_K B)KC\| \leq \|(R + B^T P_K B)^{-1} \| (R + B^T P_K B)KC\| \leq \frac{1}{\sigma_{\min}(R)} \|(R + B^T P_K B)KC - B^T P_K A + B^T P_K A\| \leq \|E_K\| + \|B^T P_K A\| + \frac{\text{Tr}(E_K^T E_K)}{\sigma_{\min}(R)} \leq \frac{\text{Tr}(E_K^T E_K) + \|B^T P_K A\|}{\sigma_{\min}(R)}.$$ 

From (26), we know that

$$\text{Tr}(E_K^T E_K) \leq \frac{\|R + B^T P_K B\|}{\mu} (J(K) - J_\Sigma^*) \leq \frac{\|R + B^T P_K B\|}{\mu} J(K).$$

Therefore, we finally have

$$\|KC\| \leq \frac{\sqrt{\|R\| + \|B\|^2 \|J(K)\|^2 / \mu}}{\sqrt{\mu \sigma_{\min}(R)}} + \frac{\|B\| \|A\| \alpha}{\mu \sigma_{\min}(R)} \leq \frac{\|R\| + \|B\|^2 \|J(K)\|^2 / \mu}{\sqrt{\mu \sigma_{\min}(R)}} + \frac{\|B\| \|A\| \alpha}{\mu \sigma_{\min}(R)}.$$ 

Lemma 11. Define $\gamma := \max_{K \in \mathbb{K}} \|A - BK\|$. Let

$$g(\delta) := \nabla^2 \frac{\|R + B^T P_K B\|}{\sigma_{\min}(R)} J(K + \delta (K' - K)[Z, Z].$$

Denote $\bar{K} := K + \delta (K' - K), \Delta K := K' - K, \text{ and } P_K[Z] := \frac{\partial^2}{\partial \delta^2} P_K[Z]_{\eta=0}.$ Suppose $K, \bar{K} \in \mathbb{K}$, it holds that: $\frac{\partial^2 P_K[Z]}{\partial \delta^2} \leq \zeta_2 \Delta K\|P_K\|, \frac{\partial P_K[Z]}{\partial \delta} \leq \zeta_3 \Delta K\|P_K\|, \text{ and } P_K[Z] \leq \zeta_4 P_K.$

The expression of $\zeta_2, \zeta_3, \text{ and } \zeta_4$ are expressed as

$$\zeta_2 = \frac{2}{\sigma_{\min}(Q)} \left(\|C\| \|R\| \|q_3 + \gamma \|C\| \|B\| \|A\| \alpha \mu\right),$$

and

$$\zeta_3 = \frac{2}{\sigma_{\min}(Q)} \left(\|C\| \|R\| + \|C\| \|B\| (\|C\| \|B\| + \zeta_1 \gamma + \zeta_2 \gamma) \alpha \mu\right),$$

and

$$\zeta_4 = \frac{2}{\sigma_{\min}(Q)} \left(\|C\| \|R\| + \|C\| \|B\| (\|C\| \|B\| + \zeta_1 \gamma + \zeta_2 \gamma) \alpha \mu\right).$$

Proof. First, from (12), it is clear that

$$\frac{\partial P_K}{\partial \delta} = \sum_{j=0}^{\infty} (A - B\bar{K}C)^T \Delta K^T E_K$$

Then, we can observe that

$$C^T \Delta K^T E_K + E_K^T \Delta KC = C^T \Delta K^T R\bar{K}C + C^T \bar{K}^T R\Delta KC = -C^T \Delta K^T B^T P_K (A - B\bar{K}C) - (A - B\bar{K}C)^T P_K B R\Delta KC \leq 2 \|C\| \|R\| \|\bar{K}C\| \|\Delta K\| I + 2 \|C\| \|B\| \|P_K\| \|A - B\bar{K}C\| \|\Delta K\| I \leq 2 \|C\| \|R\| \|q_3 + \gamma \|C\| \|B\| \|A\| \alpha \mu \|\Delta K\| \|Q\| \sigma_{\min}(Q),$$

where the last step follows from Lemma 10. Therefore, according to (13), we have $\frac{\partial^2 P_K[Z]}{\partial \delta^2} \leq \zeta_2 \Delta K\|P_K\|.$

Next, we will prove that $\frac{\partial P_K[Z]}{\partial \delta} \leq \zeta_3 \Delta K\|P_K\|.$ Based on (13), we get

$$\frac{\partial P_K[Z]}{\partial \delta} = \sum_{j=0}^{\infty} (A - B\bar{K}C)^T S_3 (A - B\bar{K}C)^T,$$

where

$$S_3 := C^T Z^T (R + B^T P_K B) \Delta KC - (B\bar{K}C)^T \frac{\partial P_K}{\partial \delta} (A - B\bar{K}C) + C^T \Delta K^T (R + B^T P_K B) ZC - (A - B\bar{K}C)^T \frac{\partial P_K}{\partial \delta} BZC - (B\Delta KC)^T P_K[Z] (A - B\bar{K}C) - (A - B\bar{K}C)^T P_K[Z] B \Delta KC.$$
We can also show that
\[
S_3 \leq 2(||C||^2 ||Z|| ||R + B^T P_K B|| ||\Delta K|| \\
+ ||B|| ||Z|| ||C|| ||A - B K C|| \frac{\partial P_K}{\partial \delta} || \\
+ ||B|| ||Z|| ||C|| ||A - B K C|| ||P'_K [Z]|| ||\Delta K|| I
\]
\[
\leq 2(||C||^2 ||R|| + ||C||^2 ||B||^2 ||P_K|| + \zeta_2 \gamma ||B|| ||C|| ||P_K|| \\
+ \zeta_1 \gamma ||B|| ||C|| ||P_K|| ||\Delta K|| I
\]
\[
\leq 2(||C||^2 ||R|| + ||C|| ||B|| |||C|| ||B|| \\
+ \zeta_1 \gamma \frac{\alpha}{\mu} ||\Delta K|| \frac{Q}{\sigma_{\min}(Q)}.
\]
Therefore, we get \(\frac{\partial P_K [Z]}{\partial \delta} \leq \zeta_3 ||\Delta K|| P_K\).

Similarly, for \(P'_K [Z]\), from (14), we can show that
\[
S_1 \leq 2(||C||^2 ||Z|| ||R + B^T P_K B|| \\
+ ||B|| ||Z|| ||C|| ||A - B K C|| ||P'_K [Z]|| I
\]
\[
\leq 2(||C||^2 ||R|| + ||C||^2 ||B||^2 ||P_K|| + \zeta_1 \gamma ||B|| ||C|| ||P_K|| I
\]
\[
\leq 2(||C||^2 ||R|| + ||C|| ||B|| |||C|| ||B|| + \zeta_1 \gamma \frac{\alpha}{\mu} ||\Delta K|| \frac{Q}{\sigma_{\min}(Q)}.
\]
So, it is clear \(P''_K [Z] \leq \zeta_4 P_K\), which completes the proof.

**B. Proof of Lemma [7]**

**Proof.** Let \(x'_t\) and \(u'_t\) be the state and action sequences generated by \(K'\), and \(c'_t = x'_t^T Q x'_t + u'_t^T R u'_t\). Then, one has
\[
J(K') - J(K) = \mathbb{E}_{x_0 \sim D} \left[ \sum_{t=0}^{\infty} c'_t - V_K(x_0) \right]
\]
\[
= \mathbb{E}_{x_0 \sim D} \left[ \sum_{t=0}^{\infty} (c'_t + V_K(x'_t) - V_K(x'_t)) - V_K(x_0) \right]
\]
\[
= \mathbb{E}_{x_0 \sim D} \left[ \sum_{t=0}^{\infty} (c'_t + V_K(x'_t - 1) - V_K(x'_t)) \right],
\]
where the last step utilizes the fact that \(x_0 = x'_0\).

Let \(A_K(x_t, K') = c_t + V_K(x_{t+1}) - V_K(x_t)\), which can be expanded as
\[
A_K(x_t, K') = x_t^T (Q + C^T K' R K'C) x_t \\
+ x_t^T (A - B K'C)^T P_K (A - B K'C) x_t - V_K(x_t)
\]
\[
= x_t^T (Q + (K'C - K C + K C)^T R (K'C - K C + K C)) x_t \\
+ x_t^T (A - B (K'C - K C + K C)^T P_K (A - B (K'C - K C + K C)) x_t
\]
\[
= 2 x_t^T (K'C - K K C)^T (R + B^T P_K B) K C - B^T P_K A) x_t \\
+ x_t^T (K'C - K K C)^T (R + B^T P_K B) (K'C - K C) x_t
\]
\[
= 2 x_t^T (K'C - K K C)^T E_K x_t \\
+ x_t^T (K'C - K K C)^T (R + B^T P_K B) (K'C - K C) x_t.
\]
Then, we get
\[
J(K') - J(K) = \mathbb{E}_{x_0 \sim D} \left[ \sum_{t=0}^{\infty} A_K(x'_t, K') \right]
\]
\[
= \mathbb{E}_{x_0 \sim D} \left[ \sum_{t=0}^{\infty} \left( 2 \text{Tr} (x'_t x'_t^T (K'C - K C)^T E_K) + \text{Tr} (x'_t x'_t^T (R + B^T P_K B) (K'C - K C)) \right) \right]
\]
\[
= 2 \text{Tr} (\Sigma_K (K'C - K C)^T E_K) + \text{Tr} (\Sigma_K (K'C - K C)^T (R + B^T P_K B) (K'C - K C)).
\]

**C. Proof of Lemma [7]**

**Proof.** Denote \(\mathcal{L}_K = C \Sigma_K C^T\). Let \(X = (R + B^T P_K B)^{-1} E_K \Sigma_K C^T \mathcal{L}_K^{-1}\). From Lemma [6] we find that
\[
J(K') - J(K) = 2 \text{Tr} (\Sigma_K (K'C - K C)^T E_K) \\
+ \text{Tr} (\Sigma_K (K'C - K C)^T (R + B^T P_K B) (K'C - K C)) \\
- \text{Tr} (\Sigma_K C^T (\Delta K + X)^T (R + B^T P_K B) (\Delta K + X))
\]
\[
\geq - \text{Tr} (\Sigma_K C^T (R + B^T P_K B)^{-1} E_K \Sigma_K C^T),
\]
where \(\Delta K = K' - K\) and the equality holds when \(K' = K\).

Then, one has
\[
J(K) - J(K^*) \leq \text{Tr} (\mathcal{L}_K^{-1} C \Sigma_K E_K (R + B^T P_K B)^{-1} E_K \Sigma_K C^T) \\
\leq ||\Sigma_K C^T \mathcal{L}_K^{-1} C \Sigma_K|| \text{Tr} (E_K (R + B^T P_K B)^{-1} E_K) \\
\leq ||\Sigma_K|| ||\Sigma_K|| \text{Tr} (E_K (R + B^T P_K B)^{-1} E_K) \\
\leq \frac{||\Sigma_K||}{\sigma_{\min}(R)} \text{Tr} (E_K E_K),
\]
(38)

From (39), it follows that
\[
\| \nabla J(K) \|_F^2 = 4 \text{Tr} (C \Sigma_K E_K E_K \Sigma_K C^T) \\
\geq 4 \mu^2 \sigma_{\min}(C)^2 \text{Tr} (E_K E_K), \forall C \in \mathbb{C}.
\]
(39)

By (38) and (39), it holds that
\[
J(K) - J(K^*) \leq \frac{||\Sigma_K||}{4 \mu^2 \sigma_{\min}(C)^2} \text{Tr} (E_K E_K), \forall C \in \mathbb{C}.
\]
(40)

Suppose \(K'\) satisfies that \(K' = K - X\), then we get
\[
J(K) - J(K^*) \geq J(K) - J(K') = \text{Tr} (\mathcal{L}_K^{-1} C \Sigma_K E_K (R + B^T P_K B)^{-1} E_K \Sigma_K C^T).
\]
Therefore, we further have
\[
J(K) - J(K^*) \geq \frac{\mu \text{Tr} (E_K E_K)}{||R + B^T P_K B||}, \forall C \in \mathbb{C}.
\]
(41)
In addition, when $C \in \mathbb{C}$, since we can always identity the state $x$ by $x = (C^T C)^{-1} C y$, it is clear that $J(K^*) = J^*_s$ for every $C \in \mathbb{C}$. By replacing $J(K^*)$ in (40) and (41) with $J^*_s$, we finally complete the proof. \hfill \square

References

[1] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al., “Human-level control through deep reinforcement learning,” Nature, vol. 518, no. 7540, p. 529, 2015.

[2] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lncotet, et al., “Mastering the game of go with deep neural networks and tree search,” Nature, vol. 529, no. 7587, p. 484, 2016.

[3] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous control with deep reinforcement learning,” in 4th International Conference on Learning Representations (ICLR 2016), (San Juan, Puerto Rico), 2016.

[4] J. Duan, Y. Guan, S. E. Li, Y. Ren, Q. Sun, and B. Cheng, “Distributional soft actor-critic: Off-policy reinforcement learning for addressing value estimation errors,” IEEE Transactions on Neural Networks and Learning Systems, 2021.

[5] J. Duan, S. E. Li, Y. Guan, Q. Sun, and B. Cheng, “Hierarchical reinforcement learning for self-driving decision-making without reliance on labelled driving data,” IET Intelligent Transport Systems, vol. 14, no. 3, pp. 297–305, 2020.

[6] J. Schulman, S. Levine, P. Abbeel, M. I. Jordan, and P. Moritz, “Trust region policy optimization,” in Proceedings of the 32nd International Conference on Machine Learning. (ICML 2015), (Lille, France), pp. 1889–1897, 2015.

[7] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” arXiv preprint arXiv:1707.06347, 2017.

[8] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” in Proceedings of the 35th International Conference on Machine Learning (ICML 2018), (Stockholm, Sweden), pp. 1861–1870, PMLR, 2018.

[9] M. Fazel, R. Ge, S. Kakade, and M. Mesbahi, “Global convergence of policy gradient methods for the linear quadratic regulator,” in International Conference on Machine Learning, pp. 1467–1476, PMLR, 2018.

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

[11] J. Bhandari and D. Russo, “Global optimality guarantees for policy gradient methods,” arXiv preprint arXiv:1906.01786, 2019.

[12] H. Mohammadi, A. Zare, M. Sohtanolkotabi, and M. R. Jovanović, “Global exponential convergence of gradient methods over the non-convex landscape of the linear quadratic regulator,” in 2019 IEEE 58th Conference on Decision and Control (CDC), pp. 7474–7479, IEEE, 2019.

[13] I. Fatkhullin and B. Polyak, “Optimizing static linear feedback: Gradient method,” SIAM Journal on Control and Optimization, vol. 59, no. 5, pp. 3887–3911, 2021.

[14] B. Hassibi, R. Xu, and H. Yang, “Policy gradient methods for the noisy linear quadratic regulator over a finite horizon,” SIAM Journal on Control and Optimization, vol. 59, no. 5, pp. 3359–3391, 2021.

[15] Z. Ren, A. Zhong, and N. Li, “LQR with tracking: A zeroth-order approach and its global convergence,” in 2021 American Control Conference (ACC), pp. 2562–2568, IEEE, 2021.

[16] J. P. Jansch-Porto, B. Hu, and G. Dullerud, “Policy optimization for markovian jump linear quadratic control: Gradient-based methods and global convergence,” arXiv preprint arXiv:2011.11852, 2020.

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

[18] K. Zhang, B. Hu, and T. Basar, “Policy optimization for H2 linear control with H∞ robustness guarantee: Implicit regularization and global convergence,” in Learning for Dynamics and Control, pp. 179–190, PMLR, 2020.

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