Escaping High-order Saddles in Policy Optimization for Linear Quadratic Gaussian (LQG) Control

Yang Zheng[†1] Yue Sun[†2] Maryam Fazel[3] Na Li[4]

***Abstract*— First-order policy optimization has been widely used in reinforcement learning. It guarantees to find the optimal policy for the state-feedback linear quadratic regulator (LQR). However, the performance of policy optimization remains unclear for the linear quadratic Gaussian (LQG) control where the LQG cost has spurious suboptimal stationary points. In this paper, we introduce a novel perturbed policy gradient (PGD) method to escape a large class of bad stationary points (including high-order saddles). In particular, based on the specific structure of LQG, we introduce a novel reparameterization procedure that converts the iterate from a high-order saddle to a strict saddle, from which standard random perturbations in PGD can escape efficiently. We further characterize a class of high-order saddles that can be escaped by our algorithm.**

I. Introduction

In this paper, we revisit the linear quadratic Gaussian (LQG) control, one of the most fundamental problems in control theory, from a modern optimization view. In brief, we focus on a continuous-time linear time-invariant (LTI) system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + w(t), \\ y(t) &= Cx(t) + v(t), \end{aligned} \tag{1}$$

where $x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m, y(t) \in \mathbb{R}^p$ are the state, control input, and measurement (output) vector at time t, respectively, and $w(t)$, $v(t)$ are white Gaussian noises with intensity matrices $W \succeq 0$ and $V \succ 0$, respectively. The goal is to design a controller (i.e., policy) based on partial measurements $y(t)$ to minimize a quadratic cost

$$J(u) := \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_{t=0}^{T} \left(x^\mathsf{T} Q x + u^\mathsf{T} R u\right) dt\right]. \tag{2}$$

A special case is the linear quadratic regulator (LQR) [1], where we have direct access to the state x (i.e., $y(t) = x(t), v(t) = 0, \forall t \in \mathbb{R}$ in (1)). It is known that the optimal policy for the LQR is in the form of static state feedback $u(t) = Kx(t)$, where $K \in \mathbb{R}^{m\times n}$ is a constant matrix that can be obtained by solving a Riccati equation [2]. On the other hand, when the state is not directly observed, the policy that minimizes (2) is a dynamical controller of the form [3]

$$\begin{aligned} \dot{\xi}(t) &= A_\mathsf{K}\xi(t) + B_\mathsf{K} y(t), \\ u(t) &= C_\mathsf{K}\xi(t), \end{aligned} \tag{3}$$

where the optimal parameters $\mathsf{K}^* := (A_\mathsf{K}^*, B_\mathsf{K}^*, C_\mathsf{K}^*)$ can be obtained by solving two Riccati equations [3], [4] (see Section II-A). Algorithms for solving Riccati equations are well-studied, including iterative algorithms [5], algebraic solution methods [6], and semidefinite optimization [7]. All these methods are model-based and explicitly rely on the system model (1). Recently, policy gradient methods have achieved impressive results for many challenging problems [8]. These methods directly optimize the quadratic cost (2) as a function of the policy class $\mathsf{K} = (A_\mathsf{K}, B_\mathsf{K}, C_\mathsf{K})$ via gradient descent or its variants. They can be further made model-free, bypassing an explicit estimation of the model (1). The flexibility of model-free control has stimulated a growing interest in investigating foundations of policy gradient methods for classical control problems [9]–[20].

While it is guaranteed to obtain the optimal controller for LQR or LQG via classical model-based methods, such optimality guarantee is more difficult when using policy gradient methods since the cost (2) is typically nonconvex in the policy space. For LQR, recent work has shown that although the LQR cost is nonconvex, it is gradient dominant and coersive, and has a unique stationary point under very mild conditions, rendering the convergence of policy gradient to the globally optimal controller [9]–[12]. On the other hand, the LQG cost is neither gradient dominant nor coersive, and there may exist spurious saddle points [20], making it challenging for policy gradient to find the optimal controller.

Saddle points do not always destroy the performance of policy gradient methods. Suitable perturbed policy gradient methods are able to escape *strict saddle points* whose Hessian has at least one strictly negative eigenvalue [21]–[23]. However, it is shown in [20, Theorem 4.2] that the Hessian of the LQG cost at a saddle point can even degenerate to zero. We denote the saddle point whose Hessian does not give escaping directions as a *high-order saddle*. Perturbed policy gradient methods may thus get stuck and take an exponential number of iterations to escape high-order saddles [21]–[23].

All the (strict or high-order) saddle points of LQG discussed in [20] are due to a loss of *controllability* and/or *observability* for the controller $(A_\mathsf{K}, B_\mathsf{K}, C_\mathsf{K})$ in (3) (i.e., *non-minimal* controllers). Indeed, any stationary point corresponds to a full-order *minimal controller* cannot be saddle and it

[†] Y. Zheng and Y. Sun contributed to this work equally. The work of Y. Zheng is supported by NSF ECCS-2154650. The research of M. Fazel and Y. Sun was supported by NSF TRIPODS II 2023166, CCF 1839291, and CCF 2007036. The work of N. Li was supported by NSF CAREER ECCS-1553407, NSF AI institute 2112085, and ONR YIP N00014-19-1-2217.

[1] Y. Zheng (zhengy@eng.ucsd.edu) is with ECE Department, University of California San Diego, La Jolla, USA.

[2] Y. Sun (yuesun9308@gmail.com) worked on this project while he was with ECE Department, University of Washington, Seattle, USA.

[3] M. Fazel (mfazel@uw.edu) is with ECE Department, University of Washington, Seattle, USA.

[4] N. Li (nali@seas.harvard.edu) is with Department of Electrical Engineering and Applied Mathematics, Harvard University, USA.

is instead globally optimal [20]. Further, many intrigue landscape properties of LQG are brought by a classical notion of *similarity transformations* that induces a symmetry structure [20]. In this paper, we raise a natural question of whether this induced symmetry structure allows us to reveal more information about high-order saddles of LQG such that suitable perturbed policy gradient methods can escape those points. We provide a positive answer to this question.

In particular, we first show that any stationary point after model reduction remains to be stationary. This gives a classification of the stationary points: all bad (suboptimal or saddle) stationary points after model reduction become lower-order and form new stationary points with the same LQG cost. We then reveal an intriguing transfer function $\mathbf{G}(s)$ at any stationary point $(A_\mathsf{K}, B_\mathsf{K}, C_\mathsf{K})$: 1) if $(A_\mathsf{K}, B_\mathsf{K}, C_\mathsf{K})$ is globally optimal, the function $\mathbf{G}(s)$ is identically zero, $\forall s \in \mathbb{C}$; 2) if $\mathbf{G}(s)$ is not identically zero, we can perturb $(A_\mathsf{K}, B_\mathsf{K}, C_\mathsf{K})$ to get a new stationary point with the same LQG cost, which is a strict saddle with probability one. Standard perturbed policy gradient (PGD) methods [21], [22] can thus escape this new strict saddle. We emphasize that our PGD method include perturbations on two parts: 1) a novel structural perturbation on the stationary point $(A_\mathsf{K}, B_\mathsf{K}, C_\mathsf{K})$; 2) a standard random perturbation on gradients [22]. This combination enables escaping a large class of bad stationary points (including high-order saddles) in LQG problems.

The rest of this paper is organized as follows. We present the problem statement in Section II. Our main results on characterizing stationary points and Hessians are presented in Section III. Section IV shows empirical performance of our perturbed policy gradient method. We conclude the paper in Section V. Technical proofs and auxiliary computations are postponed to our report [24].

II. Preliminaries and Problem Statement

A. Review of LQG control

The classical LQG control problem is defined as

$$\begin{aligned} \min_{u(t)} \quad & J(u) \\ \text{subject to} \quad & (1), \end{aligned} \tag{4}$$

where $J(u)$ is defined in (2) with $Q \succeq 0$ and $R \succ 0$. In (4), the input $u(t)$ depends on all past observation $y(\tau)$ with $\tau < t$. We make the following standard assumption.

Assumption 1. *(A, B) and $(A, W^{1/2})$ are controllable, and (C, A) and $(Q^{1/2}, A)$ are observable.*

The optimal solution to (4) is a dynamical controller in the form of (3), in which $\xi(t) \in \mathbb{R}^q$ is the controller internal state, and $A_\mathsf{K} \in \mathbb{R}^{q\times q}$, $B_\mathsf{K} \in \mathbb{R}^{q\times p}$, $C_\mathsf{K} \in \mathbb{R}^{m\times q}$ specify the dynamics of the controller. While q can be any positive number, one does not have to use $q > n$ and the optimal controller has $q = n$, given by algebraic Riccati equations (AREs) [3, Thm. 14.7]. Precisely, let P, S be the unique positive semidefinite solutions to the following AREs

$$\begin{aligned} AP + PA^\mathsf{T} - PC^\mathsf{T}V^{-1}CP + W = 0, \\ A^\mathsf{T}S + SA - SBR^{-1}B^\mathsf{T}S + Q = 0. \end{aligned} \tag{5}$$

Then, the parameters of an optimal controller to (4) are

$$A_\mathsf{K}^\star = A - BM - LC, \;\; B_\mathsf{K}^\star = L, \;\; C_\mathsf{K}^\star = -M \tag{6}$$

where $L = PC^\mathsf{T}V^{-1}$, $M = R^{-1}B^\mathsf{T}S$. The optimal solution $(A_\mathsf{K}^\star, B_\mathsf{K}^\star, C_\mathsf{K}^\star)$ is not unique in the state-space domain. Any similarity transformation leads to another equivalent optimal controller (they correspond to the same transfer function in the frequency domain).

B. Problem Statement

In this paper, we embrace the spirit of [9]–[12], [20] and view the LQG problem (4) from a modern optimization perspective. We consider the policy class $(A_\mathsf{K}, B_\mathsf{K}, C_\mathsf{K})$ in (3), and the closed-loop matrix becomes

$$A_\mathrm{cl} := \begin{bmatrix} A & BC_\mathsf{K} \\ B_\mathsf{K}C & A_\mathsf{K} \end{bmatrix} \in \mathbb{R}^{(n+q)\times(n+q)}. \tag{7}$$

The set of internally stabilizing policies [3, Chapter 13] is

$$\mathcal{C}_q := \left\{ \mathsf{K} = \begin{bmatrix} 0 & C_\mathsf{K} \\ B_\mathsf{K} & A_\mathsf{K} \end{bmatrix} \in \mathbb{R}^{(m+q)\times(p+q)} \,\middle|\, (7) \text{ is stable} \right\}.$$

Let $J_q(\mathsf{K}) : \mathcal{C}_q \to \mathbb{R}$ denote the corresponding LQG cost (2) for each stabilizing policy in $\mathcal{C}_q$. It is known [20, Lemmas 2.3 & 2.4] that this function $J_q(\mathsf{K})$ is real analytic on $\mathcal{C}_q$ and admits efficient computation.

Lemma 1. *Fix $q \in \mathbb{N}$ such that $\mathcal{C}_q \neq \varnothing$. Given $\mathsf{K} \in \mathcal{C}_q$, we have*

$$J_q(\mathsf{K}) = \mathrm{tr}\left(Q_{\mathrm{cl},\mathsf{K}} X_\mathsf{K}\right) = \mathrm{tr}\left(W_{\mathrm{cl},\mathsf{K}} Y_\mathsf{K}\right), \tag{8}$$

where X_K and Y_K are the unique positive semidefinite solutions to the following Lyapunov equations

$$A_\mathrm{cl} X_\mathsf{K} + X_\mathsf{K} A_\mathrm{cl}^\mathsf{T} + W_{\mathrm{cl},\mathsf{K}} = 0, \tag{9a}$$

$$A_\mathrm{cl}^\mathsf{T} Y_\mathsf{K} + Y_\mathsf{K} A_\mathrm{cl} + Q_{\mathrm{cl},\mathsf{K}} = 0, \tag{9b}$$

where A_cl is defined in (7) and

$$Q_{\mathrm{cl},\mathsf{K}} := \begin{bmatrix} Q & 0 \\ 0 & C_\mathsf{K}^\mathsf{T} R C_\mathsf{K} \end{bmatrix}, \; W_{\mathrm{cl},\mathsf{K}} := \begin{bmatrix} W & 0 \\ 0 & B_\mathsf{K} V B_\mathsf{K}^\mathsf{T} \end{bmatrix}.$$

Lemma 1 works for stabilizing controllers of any order q. In this paper, we are mainly interested in characterizing the full-order case $\mathcal{C}_n$. Now, given the state dimension n, we can formulate the LQG problem (4) into a constrained optimization problem

$$\begin{aligned} \min_{\mathsf{K}} \quad & J_n(\mathsf{K}) \\ \text{subject to} \quad & \mathsf{K} \in \mathcal{C}_n. \end{aligned} \tag{10}$$

An important notion of dynamical controllers is *minimality*: a controller $(A_\mathsf{K}, B_\mathsf{K}, C_\mathsf{K})$ is minimal if $(A_\mathsf{K}, B_\mathsf{K})$ is controllable and $(C_\mathsf{K}, A_\mathsf{K})$ is observable. As revealed in [20], the optimization landscape of (10) is more complicated than that of LQR: 1) the feasible region $\mathcal{C}_n$ can have at most two disconnected components; 2) the cost function $J_n(\mathsf{K})$ is not coersive and not gradient dominant, and it can have suboptimal saddle points [20, Theorem 4.2]. Two nice features are 1) all sub-optimal saddle points correspond to *non-minimal*

controllers and 2) all stationary points that correspond to *minimal controllers* in $\mathcal{C}_n$ are globally optimal [20, Theorem 4.3].

Naive policy gradient methods can thus get stuck around sub-optimal saddle points. In this paper, we aim to provide further landscape characterizations of (10) and introduce a perturbed policy gradient method to escape bad stationary points of (10). In particular, we first show that any stationary point of the LQG problem (10) after model reduction remain to be stationary, and then characterize the second-order behavior of $J_n(\mathsf{K})$ on a non-minimal stationary point. This motivates the design of our perturbed policy gradient method.

III. Stationary Points and Their Hessians

The LQG problem (4) has an inherent symmetry structure induced by the notion of similarity transformation. Let GL_q denote the set of $q \times q$ invertible matrices. Given $q \geq 1$ such that $\mathcal{C}_q \neq \varnothing$, the following map $\mathscr{T}_q : \mathrm{GL}_q \times \mathcal{C}_q \to \mathcal{C}_q$ represents similarity transformations

$$\begin{aligned}\mathscr{T}_q(T, \mathsf{K}) &:= \begin{bmatrix} I_m & 0 \\ 0 & T \end{bmatrix} \begin{bmatrix} 0 & C_{\mathsf{K}} \\ B_{\mathsf{K}} & A_{\mathsf{K}} \end{bmatrix} \begin{bmatrix} I_p & 0 \\ 0 & T \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 0 & C_{\mathsf{K}} T^{-1} \\ T B_{\mathsf{K}} & T A_{\mathsf{K}} T^{-1} \end{bmatrix}.\end{aligned} \tag{11}$$

It is well-known that similarity transformations do not change the behavior of dynamical controllers and thus the LQG cost (2) is invariant with respect to $\mathscr{T}_q(T, \mathsf{K})$, i.e., we have $J_q(\mathsf{K}) = J_q(\mathscr{T}_q(T, \mathsf{K})), \forall \mathsf{K} \in \mathcal{C}_q, T \in \mathrm{GL}_q$.

A. Classification of stationary points

The symmetry via similarity transformations brings rich and complicated landscape properties. Here, we show that the underlying symmetry also allows a classification of stationary points of LQG (4). The lemma below gives an explicit relationship among the gradients of $J_q(\mathsf{K})$ at K and $\mathscr{T}_q(T, \mathsf{K})$.

Lemma 2 ([20, Lemma 4.3]). *Let* $\mathsf{K} = \begin{bmatrix} 0 & C_{\mathsf{K}} \\ B_{\mathsf{K}} & A_{\mathsf{K}} \end{bmatrix} \in \mathcal{C}_q$. *For any* $T \in \mathrm{GL}_q$, *we have*

$$\nabla J_q|_{\mathscr{T}_q(T,\mathsf{K})} = \begin{bmatrix} I_m & 0 \\ 0 & T^{-\mathsf{T}} \end{bmatrix} \cdot \nabla J_q|_{\mathsf{K}} \cdot \begin{bmatrix} I_p & 0 \\ 0 & T^{\mathsf{T}} \end{bmatrix}. \tag{12}$$

As expected, a direct consequence of Lemma 2 is that a stationary point K of J_q remains to be stationary over $\mathcal{C}_q$ after any similarity transformation. We can further derive a classification of the stationary points of J_n over the set of full-order controllers $\mathcal{C}_n$.

Theorem 1. *Let* $\mathsf{K} = \begin{bmatrix} 0 & C_{\mathsf{K}} \\ B_{\mathsf{K}} & A_{\mathsf{K}} \end{bmatrix} \in \mathcal{C}_n$ *be a stationary point of LQG (4), and let* $\hat{\mathsf{K}} = \begin{bmatrix} 0 & \hat{C}_{\mathsf{K}} \\ \hat{B}_{\mathsf{K}} & \hat{A}_{\mathsf{K}} \end{bmatrix} \in \mathcal{C}_q$ *be a minimal realization of* K, *where* $q \leq n$ *is the order of its minimal realization. Then, the following dynamical controller with any stable matrix* $\Lambda \in \mathbb{R}^{(n-q)\times(n-q)}$

$$\tilde{\mathsf{K}} = \left[\begin{array}{c|cc} 0 & \hat{C}_{\mathsf{K}} & 0 \\ \hline \hat{B}_{\mathsf{K}} & \hat{A}_{\mathsf{K}} & 0 \\ 0 & 0 & \Lambda \end{array}\right] \in \mathcal{C}_n \tag{13}$$

is a stationary point of (4). If $q = n$ *(i.e.* K *itself is minimal), then* K *is globally optimal.*

The fact of $\tilde{\mathsf{K}}$ (13) being stationary of $J_n(\mathsf{K})$ seems to be expected, since $\tilde{\mathsf{K}}$ and K correspond to the same transfer function in the frequency domain and K is in a higher dimensional space $\mathcal{C}_n$ than $\mathcal{C}_q$. The technical proof is not difficult, which combines the classical Kalman decomposition with Lemma 2 and a result [20, Theorem 4.1]. We provide the details in [24, Appendix A]. The second part that K is globally optimal if $q = n$ has been proved in [20, Theorem 4.3].

Theorem 1 shows all stationary points that correspond to non-minimal controllers admit a *standard* parameterization as we defined in (13), which splits the controller state $\xi \in \mathbb{R}^n$ into 1) the controllable/observable (associated with $\hat{A}_{\mathsf{K}}, \hat{B}_{\mathsf{K}}, \hat{C}_{\mathsf{K}}$ blocks) part and 2) non-controllable/non-observable (associated with Λ) part. Furthermore, Theorem 1 indicates that all *bad* stationary points of (4) after model reduction are in the same form of (13). Thus, policy gradient methods only need to escape those bad saddle points of the form (13). This motivates our results in the next section.

Remark 1 (Non-minimal globally optimal controllers). A non-minimal controller in the form of (13) might still be globally optimal; See Example 2 below. This happens when the solutions $(A^\star_{\mathsf{K}}, B^\star_{\mathsf{K}}, C^\star_{\mathsf{K}})$ from the Riccati equations (5) is not minimal, i.e. $(A^\star_{\mathsf{K}}, B^\star_{\mathsf{K}})$ is uncontrollable or $(C^\star_{\mathsf{K}}, A^\star_{\mathsf{K}})$ is unobservable or both. We conjecture that a random LQG instance should have $(A^\star_{\mathsf{K}}, B^\star_{\mathsf{K}}, C^\star_{\mathsf{K}})$ (6) being minimal with probability one. An exact characterization is left for future work. □

B. Hessian of stationary points

Once a policy gradient method reaches a stationary point, if the stationary point corresponds to a minimal controller, it has found a globally optimal solution to (4). If the stationary point does not correspond to a minimal controller, we can bring it into the form of (13), for which we have the following characterization of its hessian.

Theorem 2. *Consider a stationary point of* $J_n(\mathsf{K})$ *over* $\mathcal{C}_n$ *of the form*

$$\tilde{\mathsf{K}} = \left[\begin{array}{c|cc} 0 & \hat{C}_{\mathsf{K}} & 0 \\ \hline \hat{B}_{\mathsf{K}} & \hat{A}_{\mathsf{K}} & 0 \\ 0 & 0 & \Lambda \end{array}\right] \in \mathcal{C}_n, \tag{14}$$

with $\hat{A}_{\mathsf{K}} \in \mathbb{R}^{q\times q}, \hat{B}_{\mathsf{K}} \in \mathbb{R}^{q\times p}, \hat{C}_{\mathsf{K}} \in \mathbb{R}^{m\times q}$, *stable* $\Lambda \in \mathbb{R}^{(n-q)\times(n-q)}$ *and* $q \leq n$. *Let* $X_{\mathrm{op}} \in \mathbb{S}^{n+q}_+$ *and* $Y_{\mathrm{op}} \in \mathbb{S}^{n+q}_+$ *be the unique positive semidefinite solutions to the Lyapunov equations (9a) and (9b) with* $\hat{\mathsf{K}} = \begin{bmatrix} 0 & \hat{C}_{\mathsf{K}} \\ \hat{B}_{\mathsf{K}} & \hat{A}_{\mathsf{K}} \end{bmatrix} \in \mathcal{C}_q$, *respectively. Define a transfer function of size* $p \times m$

$$\mathbf{G}(s) := C_{\mathrm{cl}}(sI - A_{\mathrm{cl}}^{\mathsf{T}})^{-1} B_{\mathrm{cl}}. \tag{15}$$

where A_{cl} *is defined in (7) with the* $\hat{\mathsf{K}}$ *above, and* $C_{\mathrm{cl}} := \bar{C} X_{\mathrm{op}} + V \bar{B}_{\mathsf{K}}^{\mathsf{T}}$, $B_{\mathrm{cl}} := Y_{\mathrm{op}} \bar{B} + \bar{C}_{\mathsf{K}}^{\mathsf{T}} R$, *with*

$$\bar{C} = \begin{bmatrix} C & 0 \end{bmatrix} \in \mathbb{R}^{p\times(n+q)}, \bar{B} = \begin{bmatrix} B \\ 0 \end{bmatrix} \in \mathbb{R}^{(n+q)\times m}, \tag{16}$$

$$\bar{C}_{\mathsf{K}} = \begin{bmatrix} 0 & \hat{C}_{\mathsf{K}} \end{bmatrix} \in \mathbb{R}^{m\times(n+q)}, \bar{B}_{\mathsf{K}} = \begin{bmatrix} 0 \\ \hat{B}_{\mathsf{K}} \end{bmatrix} \in \mathbb{R}^{(n+q)\times p}.$$

The following statements hold.

1) *If $\tilde{\mathsf{K}}$ in (14) is globally optimal in $\mathcal{C}_n$, then the function $\mathbf{G}(s)$ in (15) is identically zero $\forall s \in \mathbb{C}$.*
2) *If $\mathbf{G}(s)$ in (15) is not a zero function, then $\tilde{\mathsf{K}}$ is a strict saddle point (the Hessian of $J_n(\mathsf{K})$ at $\tilde{\mathsf{K}}$ is indefinite) with probability one when randomly choosing a stable and symmetric $\Lambda \in \mathbb{S}^{n-q}$.*
3) *Let $\mathcal{Z}$ be the set of zeros of $\mathbf{G}(s)$, i.e., $\mathcal{Z} = \{s \in \mathbb{C} \mid \mathbf{G}(s) = 0\}$. Given a stable and symmetric $\Lambda \in \mathbb{S}^{n-q}$, let $\mathrm{eig}(-\Lambda)$ denote the set of (distinct) eigenvalues of $-\Lambda$. If $\mathrm{eig}(-\Lambda) \nsubseteq \mathcal{Z}$, then the Hessian of $J_n(\mathsf{K})$ at $\tilde{\mathsf{K}}$ is indefinite.*

Proof. Statements 1) and 2) are direct consequences of Statement 3). We give simple arguments below.

3) $\Rightarrow$ 1): If $\tilde{\mathsf{K}}$ in (14) is globally optimal in $\mathcal{C}_n$, then the Hessian of $J_n(\mathsf{K})$ at $\tilde{\mathsf{K}}$ must be positive semidefinite. If $\mathbf{G}(s)$ is not identically zero, then its zero set $\mathcal{Z}$ is a set of finite points due to the fundamental theorem of algebra[1]. Then, there exists a symmetric $\Lambda \in \mathbb{S}^{n-q}$ such that $\mathrm{eig}(-\Lambda) \nsubseteq \mathcal{Z}$, and thus its Hessian at $\tilde{\mathsf{K}}$ is indefinite. This is contradicted with $\tilde{\mathsf{K}}$ being globally optimal.

3) $\Rightarrow$ 2): If $G(s)$ is not an identically zero function, then its zero set $\mathcal{Z}$ is a set of finite points. When choosing a stable and symmetric $\Lambda \in \mathbb{S}^{n-q}$ randomly, we have $\mathrm{eig}(-\Lambda) \nsubseteq \mathcal{Z}$ holds with probability one. Thus, $\tilde{\mathsf{K}}$ is a strict saddle point with probability one.

The proof of Statement 3) exploits the bilinear property of the Hessian and the non-controllable/non-observable property to identify a two-by-two hessian block

$$\begin{bmatrix} \mathrm{Hess}_{\tilde{\mathsf{K}}}(\Delta^{(1)}, \Delta^{(1)}) & \mathrm{Hess}_{\tilde{\mathsf{K}}}(\Delta^{(1)}, \Delta^{(2)}) \\ \mathrm{Hess}_{\tilde{\mathsf{K}}}(\Delta^{(1)}, \Delta^{(2)}) & \mathrm{Hess}_{\tilde{\mathsf{K}}}(\Delta^{(2)}, \Delta^{(2)}) \end{bmatrix} \in \mathbb{S}^2$$

in which the diagonal entries are always zero. Using the Hessian calculation in [20, Lemma 4.3], we then prove that if $\mathrm{eig}(-\Lambda) \nsubseteq \mathcal{Z}$, then the off-diagonal entries are non-zero. The Hessian of $J_n(\cdot)$ at $\tilde{\mathsf{K}}$ is thus indefinite. Details are presented in our extended report [24, Appendix B]. □

Our Theorem 2 includes the recent result [20, Theorem 4.2] as a special case in which the authors only consider a zero controller $\mathsf{K} = 0$. Our main proof in [24, Appendix B], however, is motivated by that in [20, Theorem 4.2] with more complicated and careful calculations.

If the transfer function $\mathbf{G}(s)$ is not identically zero, then $\tilde{\mathsf{K}}$ in (14) is a strict saddle point with probability one when randomly choosing Λ. Thus, we can apply the perturbed policy gradient method for "escaping saddle" [22], so that the policy gradient iterations do not get stuck around these sub-optimal saddle points. We note that when $\mathbf{G}(s)$ is not identically zero, $\tilde{\mathsf{K}}$ in (14) may still have a zero Hessian (i.e., high-order saddle) if Λ is chosen such that $\mathrm{eig}(-\Lambda) \subseteq \mathcal{Z}$; an explicit example is given Example 3 below. Therefore, our proposed perturbed policy gradient method for the LQG problem (4) includes perturbations on Λ as well as on the gradients. More details are given in Section IV.

[1]Every non-zero, single-variable, degree n polynomial with complex coefficients has, counted with multiplicity, exactly n complex roots.

Remark 2 (Sufficiency of $\mathbf{G}(s) \equiv 0$ for global optimality and its interpretation). Theorem 2 holds with $q = n$, so $\mathbf{G}(s) \equiv 0, \forall s \in \mathbb{C}$ is also true when K comes from the Riccati equations. In this case, we expect that $\mathbf{G}(s)$ in (15) should have a nice control-theoretic interpretation. It is interesting to further investigate whether $\mathbf{G}(s) \equiv 0, \forall s \in \mathbb{C}$ is sufficient (or some other suitable conditions are needed) to certify the global optimality of $\tilde{\mathsf{K}}$. □

Example 1. We first consider the famous Doyle's LQG example [25], which has system matrices

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \ B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \ C = \begin{bmatrix} 1 & 0 \end{bmatrix},$$

and performance weights

$$W = 5\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \ V = 1, \ Q = 5\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \ R = 1.$$

The globally optimal LQG controller from (6) is

$$A^\star_{\mathsf{K}} = \begin{bmatrix} -4 & 1 \\ -10 & -4 \end{bmatrix}, \ \hat{B}_{\mathsf{K}} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}, \hat{C}_{\mathsf{K}} = \begin{bmatrix} -5 & -5 \end{bmatrix}.$$

The Hessian $J_2(\mathsf{K})$ at $\mathsf{K}^\star = \left[\begin{array}{c|c} 0 & \hat{C}_{\mathsf{K}} \\ \hline \hat{B}_{\mathsf{K}} & A^\star_{\mathsf{K}} \end{array}\right] \in \mathcal{C}_2$ is positive semidefinite and has eigenvalues $\lambda_1 = 8.1111 \times 10^5, \lambda_2 = 6\,133.9, \lambda_3 = 131.2, \lambda_4 = 6.36, \lambda_5 = \cdots = \lambda_8 = 0$ (see [24, Appendix C] for details). Four zero eigenvalues are expected due to the symmetry induced by the similarity transformation [20, Lemma 4.6]. We further compute the matrices in (16) (their values can be found in [24, Appendix C]), and we have

$$\begin{aligned} &(\bar{C}X_{\mathrm{op}} + V\bar{B}_{\mathsf{K}}^{\mathsf{T}})(sI - A_{\mathrm{cl}}^{\mathsf{T}})^{-1}Y_{\mathrm{op}}\bar{B} \\ =&\frac{-12.5s^3 - 604.2s^2 - 1712.5s - 566.7}{s^4 + 6s^3 + 11s^2 + 6s + 1}, \end{aligned}$$

and

$$\begin{aligned} &(\bar{C}X_{\mathrm{op}} + V\bar{B}_{\mathsf{K}}^{\mathsf{T}})(sI - A_{\mathrm{cl}}^{\mathsf{T}})^{-1}\bar{C}_{\mathsf{K}}^{\mathsf{T}}R \\ =&\frac{12.5s^3 + 604.2s^2 + 1712.5s + 566.7}{s^4 + 6s^3 + 11s^2 + 6s + 1}. \end{aligned}$$

Thus, we have

$$\mathbf{G}(s) = (\bar{C}X_{\mathrm{op}} + V\bar{B}_{\mathsf{K}}^{\mathsf{T}})(sI - A_{\mathrm{cl}}^{\mathsf{T}})^{-1}(Y_{\mathrm{op}}\bar{B} + \bar{C}_{\mathsf{K}}^{\mathsf{T}}R) \equiv 0.$$

This result that $\mathbf{G}(s)$ being identically zero is expected from Theorem 2 since $\mathsf{K}^\star$ is globally optimal. □

We then consider [20, Example 7] for which the globally optimal LQG controller is non-minimal in $\mathcal{C}_n$.

Example 2. Consider an LQG instance with matrices

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \ B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \ C = \begin{bmatrix} 1 & -1 \end{bmatrix}$$

and performance weights

$$W = \begin{bmatrix} 1 & -1 \\ -1 & 16 \end{bmatrix}, \ V = 1, \ Q = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}, \ R = 1.$$

The globally optimal controller from (6) is given by

$$A_{\mathsf{K}}^\star = \begin{bmatrix} -3 & 0 \\ 5 & -4 \end{bmatrix}, B_{\mathsf{K}}^\star = \begin{bmatrix} 1 \\ -4 \end{bmatrix}, C_{\mathsf{K}}^\star = \begin{bmatrix} -2 & 0 \end{bmatrix}.$$

It is easy to verify that $(C_{\mathsf{K}}^*, A_{\mathsf{K}}^*)$ is not observable. The Hessian of $J_2(\mathsf{K})$ at $\mathsf{K}^\star \in \mathcal{C}_2$ is positive semidefinite with eigenvalues as $\lambda_1 = 581.5529, \lambda_2 = 7.1879, \lambda_3 = 0.2592, \lambda_4 = \cdots = \lambda_8 = 0$. (See [24, Appendix C] for details). Four zero eigenvalues are expected, due to the symmetry by similarity transformations, and the other zero is caused by the unobservablility of $(C_{\mathsf{K}}^*, A_{\mathsf{K}}^*)$. Consider two reduced-order controllers

$$\mathsf{K}_1 = \left[\begin{array}{c|cc} 0 & -2 & \\ \hline 1 & -3 & \end{array}\right] \in \mathcal{C}_1, \quad \mathsf{K}_2 = \left[\begin{array}{c|cc} 0 & 0.5 & \\ \hline -4 & -3 & \end{array}\right] \in \mathcal{C}_1,$$

both of which are globally optimal. Thus, the following two full-order controllers

$$\tilde{\mathsf{K}}_1 = \left[\begin{array}{c|cc} 0 & -2 & 0 \\ \hline 1 & -3 & 0 \\ 0 & 0 & \Lambda \end{array}\right], \quad \tilde{\mathsf{K}}_2 = \left[\begin{array}{c|cc} 0 & 0.5 & 0 \\ \hline -4 & -3 & 0 \\ 0 & 0 & \Lambda \end{array}\right],$$

are globally optimal as well. From Theorem 2, we expect $\mathbf{G}(s) \equiv 0$ for both $\tilde{\mathsf{K}}_1$ and $\tilde{\mathsf{K}}_2$. For both of them, we can compute (details are in Appendix D) that

$$(\bar{C}X_{\mathrm{op}} + V\bar{B}_{\mathsf{K}}^{\mathsf{T}})(sI - A_{\mathrm{cl}}^{\mathsf{T}})^{-1}Y_{\mathrm{op}}\bar{B} = \frac{26.5s + 56.5}{(s+1)^2}$$
$$(\bar{C}X_{\mathrm{op}} + V\bar{B}_{\mathsf{K}}^{\mathsf{T}})(sI - A_{\mathrm{cl}}^{\mathsf{T}})^{-1}\bar{C}_{\mathsf{K}}^{\mathsf{T}}R = -\frac{26.5s + 56.5}{(s+1)^2}.$$

Thus, we have the expected result from Theorem 2 that $\mathbf{G}(s) = (\bar{C}X_{\mathrm{op}} + V\bar{B}_{\mathsf{K}}^{\mathsf{T}})(sI - A_{\mathrm{cl}}^{\mathsf{T}})^{-1}(Y_{\mathrm{op}}\bar{B} + \bar{C}_{\mathsf{K}}^{\mathsf{T}}R) \equiv 0$. □

Finally, we consider an LQG problem with a high-order saddle point. This high-order saddle point is predicted in Theorem 2 and [20, Theorem 4.2].

Example 3. Consider an LQG instance with an open-loop stable system, in which the problem data are

$$A = \begin{bmatrix} -0.5 & 0 \\ 0.5 & -1 \end{bmatrix}, \; B = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \; C = \begin{bmatrix} -\frac{1}{6} & \frac{11}{12} \end{bmatrix},$$

with weight matrices $W = Q = I_2, \; V = R = 1$. Since this example is open-loop stable, [20, Theorem 4.2] guarantees that $\tilde{\mathsf{K}} = \left[\begin{array}{c|cc} 0 & 0 & \\ \hline 0 & \Lambda & \end{array}\right] \in \mathcal{C}_2$ with any stable $\Lambda \in \mathbb{R}^{2\times 2}$ is a stationary point. At this controller, we can compute that the transfer function in (15) is $\mathbf{G}(s) = \dfrac{5(2s-1)}{108(2s^2+3s+1)}$. The zero set $\mathcal{Z} = \{0.5\}$ contains a single value. For any stable Λ with $\mathrm{eig}(-\Lambda) \not\subseteq \mathcal{Z}$, the Hessian is indefinite by Theorem 2. For instance, with $\Lambda = -\mathrm{diag}(0.5, 0.1)$, the Hessian is indefinite with eigenvalues $\lambda_1 = 0.0561, \lambda_2 = -0.0561, \lambda_i = 0, i = 3, \ldots, 8$ (see [24, Appendix C] for details). However, we can check that if $\Lambda = -0.5I_2$, (i.e. $A_{\mathsf{K}} = -0.5I_2, \; B_{\mathsf{K}} = 0, \; C_{\mathsf{K}} = 0$), its Hessian is degenerated to zero, implying that it is a high-order saddle. Our proposed perturbed gradient descent algorithm in the next section can escape this type of high-order saddles efficiently. □

Algorithm 1 Perturbed policy gradient

Require: 1) Loss $J(\mathsf{K})$ with its gradient. 2) Thresholds g_{th}, ι. 3) Constant T, τ, step size η. 4) Function $\lambda_{\mathrm{Han,min}}(\mathsf{K})$ that returns the minimum Hankel singular value of the stable part in K. 5) Function reduce_order(K) that finds the approximate order of K.

1: Set $t = 0$, $t_{\mathrm{perturb}} = -\tau - 1$ and initialize a stabilizing controller K_0.
2: **while** $t \le T$ **do**
3: **if** $\|\nabla J(\mathsf{K}_t)\| \le g_{\mathrm{th}}$ and $\lambda_{\mathrm{Han,min}}(\mathsf{K}_t) \ge \iota$ **then**
4: **output** K
5: **else if** $\|\nabla J(\mathsf{K}_t)\| \le g_{\mathrm{th}}$ and $\lambda_{\mathrm{Han,min}}(\mathsf{K}_t) \le \iota$ and $t - t_{\mathrm{perturb}} > \tau$ **then**
6: $\hat{\mathsf{K}}_t, q_t \leftarrow$ reduce_order(K_t) where q_t is the order after model reduction;
7: $\Lambda_t \leftarrow \lambda I_{n-q_t}$ with $\lambda < 0$ randomly selected;
8: $\mathsf{K}_t \leftarrow \mathrm{diag}(\hat{\mathsf{K}}_t, \Lambda_t)$ as in (14) (Theorem 2);
9: $\mathsf{K}_t \leftarrow \mathsf{K}_t + \xi_t$ with ξ_t uniformly sampled from $\mathbb{B}_{\mathsf{K}_t}(r)$;
10: $t_{\mathrm{perturb}} \leftarrow t$;
11: **end if**
12: $\mathsf{K}_{t+1} \leftarrow \mathsf{K}_t - \eta\nabla J(\mathsf{K}_t)$;
13: $t \leftarrow t + 1$;
14: **end while**

IV. Perturbed Policy Gradient Method

Inspired by Theorems 1 and 2, we introduce a novel perturbed policy gradient method that combines a structural perturbation on Λ in (14) with a standard perturbation on gradients [21], [22]. Numerical results confirm that our perturbed policy gradient method can escape high-order saddles efficiently.

Our method combines the standard perturbed gradient descent [22, Algorithm 2] with an additional oracle of random structural perturbation on Λ. Our perturbed policy gradient descent is listed in Algorithm 1. We note that Algorithm 1 is a *prototype* algorithm in the sense that some quantities (e.g., model reduction, gradient and Hessian Lipschitz constants, step size) of the LQG problem require more investigation. Convergence conditions and further quantitative analysis of our algorithm are also left for future work. Algorithm 1 can escape a large class of (but not all) high-order saddles at which $\mathbf{G}(s)$ in (15) is not identically zero. When Algorithm 1 terminates, it is likely to produce an approximately global minimum or return a point at which the transfer function $\mathbf{G}(s)$ in (15) is close to zero (quantitative analysis is left for future work as well). In the later case, the point may not be globally optimal, and this is related to the sufficiency of $\mathbf{G}(s) \equiv 0$ for global optimality in Remark 2.

We implement Algorithm 1, and consider Example 3 for numerical comparison with three other algorithms: 1) Vanilla policy gradient; 2) Standard perturbed policy gradient [22] (with no perturbation on dynamics Λ, i.e., no Lines 6-8 in Algorithm 1); 3) Perturb the dynamics Λ but with no perturbation on gradients (i.e., no Line 9 in Algorithm 1.).

The globally optimal controller from (6) for the LQG

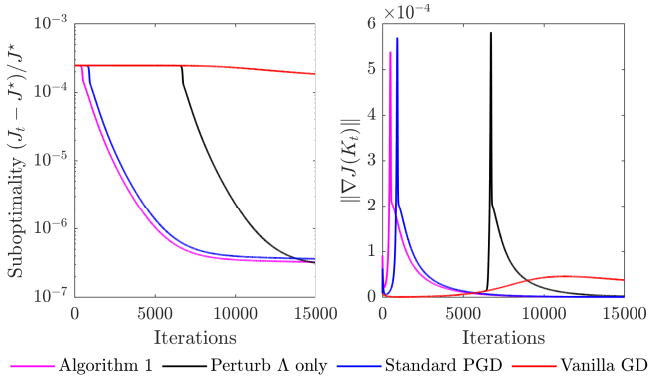

Fig. 1: Comparison of different perturbed and Vanilla policy gradient (PG) methods: Our Algorithm 1, Vanilla GD, standard PGD in [22] (with no perturbation on dynamics Λ), and PGD with perturbation on dynamics Λ only. These algorithms all start from the same point (17) near a high-order saddle, and applied fixed step-size gradient descent iterations. Left: suboptimality $\frac{J(\mathsf{K}_t)-J^\star}{J^\star}$; Right: norm of gradients $\|\nabla J(\mathsf{K}_t)\|$.

instance in Example 3 is

$$A_{\mathsf{K}}^\star = \begin{bmatrix} -1.10 & 0.13 \\ 1.19 & -1.64 \end{bmatrix},\ \hat{B}_{\mathsf{K}} = \begin{bmatrix} 0.11 \\ 0.45 \end{bmatrix},\ \hat{C}_{\mathsf{K}} = [0.62\ -0.22]\,.$$

To illustrate the performance of different algorithms, we initialize the controller at

$$A_{\mathsf{K},0} = -0.5 I_2,\ B_{\mathsf{K},0} = \begin{bmatrix} 0 \\ 0.01 \end{bmatrix},\ C_{\mathsf{K},0} = [0, -0.01]\,. \quad (17)$$

As discussed in Example 3, this initial point is close to a high-order saddle $A_{\mathsf{K}} = -0.5 I_2, B_{\mathsf{K}} = 0, C_{\mathsf{K}} = 0$. We add a perturbation (on dynamics Λ, gradients, or both) to the first iteration and run gradient descent with an fixed step size.

The results are shown in Figure 1: the left sub-figure shows the suboptimality gap, and the right one shows the norm of gradients at each iteration. Our Algorithm 1 implements both perturbations: 1) identifying an one-dimensional Λ as in the standard form (14) and change it randomly, and 2) randomly perturb all variables with a small quantity 0.01. As shown in Figure 1, our Algorithm 1 can escape this high-order saddle faster than the other three algorithms, including the standard PGD in [22] (no perturbation on dynamics Λ was applied).

V. Conclusions

We have proposed a novel PGD algorithm (cf. Algorithm 1) to escape high-order saddles of LQG. Our PGD algorithm combines the inherent structure of LQG control with standard perturbation on gradients. We have shown the structure of all stationary points after model reduction (cf. Theorem 1). We have also introduced a reparameterization procedure with an intriguing transfer function $\mathbf{G}(s)$ at any stationary point (cf. Theorem 2). If $\mathbf{G}(s) \not\equiv 0$, we can certify that the high-order saddle can be made as a strict saddle by the reparameterization. Numerical simulations confirmed that Algorithm 1 combining the reparameterization with random perturbation on gradients can accelerate the speed of escaping high-order saddles. Ongoing and future directions include quantitative analysis of Algorithm 1. We are also interested in the sufficiency of $\mathbf{G}(s) \equiv 0$ (or other conditions are needed) for global optimality of LQG (see Remark 2).

References

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

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

[3] K. Zhou, J. C. Doyle, K. Glover *et al.*, *Robust and optimal control*. Prentice hall New Jersey, 1996, vol. 40.

[4] D. Bertsekas, *Dynamic programming and optimal control: Volume I*. Athena scientific, 2012, vol. 1.

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

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

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

[8] B. Recht, "A tour of reinforcement learning: The view from continuous control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 2, pp. 253–279, 2019.

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

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

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

[12] Y. Sun and M. Fazel, "Learning optimal controllers by policy gradient: Global optimality via convex parameterization," in *60th IEEE Conference on Decision and Control*. IEEE, 2021, pp. 4576–4581.

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

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

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

[16] J. Umenberger, M. Simchowitz, J. C. Perdomo, K. Zhang, and R. Tedrake, "Globally convergent policy search over dynamic filters for output estimation," *arXiv preprint arXiv:2202.11659*, 2022.

[17] J. Duan, W. Cao, Y. Zheng, and L. Zhao, "On the optimization landscape of dynamical output feedback linear quadratic control," *arXiv preprint arXiv:2201.09598*, 2022.

[18] 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.

[19] B. Hu and Y. Zheng, "Connectivity of the feasible and sublevel sets of dynamic output feedback control with robustness constraints," *arXiv preprint arXiv:2203.11177*, 2022.

[20] Y. Zheng, Y. Tang, and N. Li, "Analysis of the optimization landscape of linear quadratic gaussian (LQG) control," *arXiv preprint arXiv:2102.04393*, 2021.

[21] R. Ge, F. Huang, C. Jin, and Y. Yuan, "Escaping from saddle points—online stochastic gradient for tensor decomposition," in *Conference on learning theory*. PMLR, 2015, pp. 797–842.

[22] C. Jin, R. Ge, P. Netrapalli, S. M. Kakade, and M. I. Jordan, "How to escape saddle points efficiently," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1724–1732.

[23] Y. Carmon and J. C. Duchi, "Gradient descent efficiently finds the cubic-regularized non-convex newton step," *arXiv:1612.00547*, 2016.

[24] Y. Zheng, Y. Sun, M. Fazel, and N. Li, "Escaping high-order saddles in policy optimization for Linear Quadratic Gaussian (LQG) control," Technical report, https://arxiv.org/pdf/2204.00912.pdf, 2022.

[25] J. C. Doyle, "Guaranteed margins for LQG regulators," *IEEE Transactions on automatic Control*, vol. 23, no. 4, pp. 756–757, 1978.