2021 60th IEEE Conference on Decision and Control (CDC)
December 13-15, 2021. Austin, Texas

On the lack of gradient domination for linear quadratic Gaussian problems with incomplete state information

Hesameddin Mohammadi, Mahdi Soltanolkotabi, and Mihailo R. Jovanović

Abstract— **Policy gradient algorithms in model-free reinforcement learning have been shown to achieve global exponential convergence for the Linear Quadratic Regulator problem despite the lack of convexity. However, extending such guarantees beyond the scope of standard LQR and full-state feedback has remained open. A key enabler for existing results on LQR is the so-called gradient dominance property of the underlying optimization problem that can be used as a surrogate for strong convexity. In this paper, we take a step further by studying the convergence of gradient descent for the Linear Quadratic Gaussian problem and demonstrate through examples that LQG does not satisfy the gradient dominance property. Our study shows the non-uniqueness of equilibrium points and thus disproves the global convergence of policy gradient methods for LQG.**

Index Terms— **Data-driven control, gradient dominance, gradient decent, nonconvex optimization, observer-based controller, policy gradient, reinforcement learning.**

I. Introduction

Modern reinforcement learning algorithms have shown great empirical performance in solving continuous control problems [1] with unknown dynamics. However, despite the recent surge in research, convergence and sample complexity of these methods are not yet fully understood. This has recently motivated a significant body of literature on data-driven control to focus on the Linear Quadratic Regulator (LQR) problem with unknown model parameters with the primary purpose of providing insight into the behavior and performance of RL algorithms in more challenging settings.

The LQR problem is the cornerstone of control theory. The globally optimal solution to LQR is given by a static linear feedback and, for problems with known models, the solution can be obtained by solving the celebrated Riccati equation using efficient numerical schemes with provable convergence guarantees [2]. In the data-driven setting, existing techniques are mainly divided into two categories, model-based [3] and model-free [4]. While model-based techniques use data to obtain approximations of the underlying dynamics, model-free methods directly search over the parameter space of controllers using the reward/cost values without attempting to form a model.

The work of H. Mohammadi and M. R. Jovanović is supported in part by the National Science Foundation (NSF) under Awards ECCS-1708906 and ECCS-1809833. The work of M. Soltanolkotabi is supported in part by the Packard Fellowship in Science and Engineering, a Sloan Research Fellowship in Mathematics, a Google Faculty Research Award, as well as Awards from NSF and AFOSR Young Investigator Program.

The authors are with the Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90089. Emails: {hesamedm, soltanol, mihailo}@usc.edu.

Among model-free approaches, simple random search, which emulates the behavior of gradient descent by forming estimates of the gradient via cost evaluations, has been shown to achieve sub-linear sample complexity for LQR [5]. This can be even further improved to a logarithmic complexity if one can access the so-called *two-point* gradient estimates [6], [7]. These results build on the fact that the gradient descent itself achieves linear convergence for both discrete [8] and continuous-time LQR problems [9] despite lack of convexity. A key enabler for these results is the so-called gradient dominance property of the underlying optimization problem that can be used as a surrogate for strong convexity [10].

In this paper, we take a step further by studying the convergence of gradient descent for the Linear Quadratic Gaussian (LQG) problem with incomplete state information. The separation principle states that the solution to the LQG problem is given by an observer-based controller, which consists of a Kalman filter and the corresponding LQR solution. This problem is also closely related to the output-feedback problem for distributed control, which is known to be fundamentally more challenging than LQR. In particular, the output-feedback problem has been shown to involve an optimization domain with exponential number of connected components [11], [12]. In contrast, the standard LQG problem allows for dynamic controllers and do not impose structural constraints on the controller.

Motivated by the convergence properties of gradient descent on LQR, we reformulate the LQG problem as a joint optimization of the control and observer feedback gains whose domain, unlike the output feedback problem is connected. We derive analytical expressions for the gradient of the LQG cost function with respect to gain matrices and demonstrate through examples that LQG does not satisfy the gradient dominance property. In particular, we show that, in addition to the global solution, the gradient vanishes at the origin for open-loop stable systems. Our study disproves global exponential convergence of policy gradient methods for LQG. The analysis of the optimization landscape of the LQG problem with unknown system parameters has also been recently provided in [13], where the authors relate the existence of multiple equilibrium points to the non-minimality of the controller transfer function.

The rest of the paper is structured as follows. In Section II, we formulate the LQG problem and provide background information. In Section III, we derive an analytical expression for the gradient. In Section IV, we discuss the lack of gradient domination and non-uniqueness of equilibrium points. We present numerical experiments in Section V and finally provide concluding remarks in Section VI.

II. Linear Quadratic Gaussian

Consider the stochastic LTI system

$$\dot{x} \;=\; Ax \,+\, Bu \,+\, w, \quad y \;=\; Cx \,+\, v \tag{1a}$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the control input, $y(t) \in \mathbb{R}^p$ is the measured output, A, B, and C are constant matrices, and $w(t)$ and $v(t)$ are independent zero-mean Gaussian white noise processes with covariance functions $\mathbb{E}[w(t)w^T(\tau)] = \delta(t-\tau)\Sigma_w$ and $\mathbb{E}[v(t)v^T(\tau)] = \delta(t-\tau)\Sigma_v$. Here, δ is the Dirac delta (impulse) function and we assume $\Sigma_w, \Sigma_v \succ 0$ are positive definite matrices. The Linear Quadratic Gaussian (LQG) problem associated with system (1a) is given by

$$\underset{u(t)\in\mathcal{Y}(t)}{\text{minimize}} \;\; \lim_{t\,\to\,\infty} \;\mathbb{E}\left[x^T(t)Q\,x(t) \,+\, u^T(t)R\,u(t)\right] \tag{1b}$$

where Q and R are positive definite matrices and $\mathcal{Y}(t)$ is the set of functions that depend only on the available information up to time t, i.e., the measured outputs $y(s)$ with $s \le t$.

A. Separation principle

It is well-known that if the pair (A, B) is controllable and (A, C) is observable, the solution to (1) is given by an observer-based controller of the form

$$\begin{aligned} \dot{\hat{x}} \;&=\; A\,\hat{x} \,+\, B\,u \,-\, L(\hat{y} \,-\, y) \\ \hat{y} \;&=\; C\,\hat{x}, \quad u \;=\; -K\hat{x} \end{aligned} \tag{2}$$

where $\hat{x}(t) \in \mathbb{R}^n$ is the state estimate, and $L \in \mathbb{R}^{n\times p}$ and $K \in \mathbb{R}^{m\times n}$ are the observer and controller feedback gain matrices, respectively [2], [14]. The separation principle states that the optimal gains $K^\star$ and $L^\star$ correspond to solutions to two decoupled problems associated with (1), namely the linear quadratic regulator

$$\underset{K}{\text{minimize}} \;\; \lim_{t\,\to\,\infty} \;\mathbb{E}\left[x^T(t)Qx(t) \,+\, u^T(t)Ru(t)\right] \tag{3}$$

subject to (1a) with the full-state feedback $u = -Kx$, and the Kalman filter, which seeks to

$$\underset{L}{\text{minimize}} \;\; \lim_{t\,\to\,\infty} \;\mathbb{E}\left[\|e(t)\|^2\right] \tag{4a}$$

subject to the error dynamics

$$\dot{e} \,=\, (A - LC)\,e \,-\, L\,v \,+\, w \tag{4b}$$

where $e := x - \hat{x}$ is the state estimation error. The solutions to these two problems (and also to the original LQG problem) are given by

$$K^\star \;=\; R^{-1}B^TP_c^\star, \quad L^{\star T} \;=\; \Sigma_v^{-1}CX_o^\star \tag{5}$$

where $P_c^\star$ and $X_o^\star$ are the unique solutions to the decoupled pair of Algebraic Riccati Equations (ARE)

$$\begin{aligned} A^TP_c^\star \,+\, P_c^\star A + Q \,-\, P_c^\star BR^{-1}B^TP_c^\star \;&=\; 0 \\ AX_o^\star \,+\, X_o^\star A^T \,+\, \Sigma_w \,-\, X_o^\star C^T\Sigma_v^{-1}CX_o^\star \;&=\; 0. \end{aligned}$$

B. Characterization based on gain matrices

In this paper, we analyze the LQG problem as optimization of feedback gain matrices K and L. In particular, the closed-loop dynamics in (1a) and (2) can be jointly described by

$$\dot{\xi} \;=\; A_{\rm cl}\,\xi \,+\, \mu \tag{6}$$

where $\xi := \left[\begin{array}{cc} x^T & e^T \end{array}\right]^T \in \mathbb{R}^{2n}$ consists of the state and error signals, $\mu := \left[\begin{array}{cc} w^T & w^T - v^TL^T \end{array}\right]^T$ is white noise, and the closed-loop matrix $A_{\rm cl}$ is given by

$$A_{\rm cl} \;:=\; \left[\begin{array}{cc} A - BK & BK \\ 0 & A - LC \end{array}\right]. \tag{7}$$

The closed-loop representation (6) allows us to reformulate the LQG problem as an optimization over the set $\mathcal{S}_c \times \mathcal{S}_o$ of stabilizing gain matrices, where

$$\begin{aligned} \mathcal{S}_c \;&:=\; \{K \in \mathbb{R}^{m\times n} \,|\, A \,-\, BK \text{ is Hurwitz}\} \\ \mathcal{S}_o \;&:=\; \{L \in \mathbb{R}^{n\times p} \,|\, A \,-\, LC \text{ is Hurwitz}\}. \end{aligned} \tag{8}$$

In particular, the LQG problem in (1b) amounts to

$$\underset{K,L}{\text{minimize}} \;\; f(K,L) \;:=\; \langle \Omega, X\rangle \tag{9}$$

where $X = \lim_{t\to\infty} \mathbb{E}\left[\xi(t)\xi^T(t)\right]$ is the steady-state covariance matrix associated with closed-loop system (6) and it can be determined by solving the algebraic Lyapunov equation

$$A_{\rm cl}\,X \;+\; XA_{\rm cl}^T \;+\; \Sigma \;=\; 0. \tag{10}$$

Here, the positive semi-definite matrices Ω, Σ are given by

$$\Omega \;:=\; \left[\begin{array}{cc} Q + K^TRK & -K^TRK \\ -K^TRK & K^TRK \end{array}\right] \tag{11a}$$

$$\Sigma \;:=\; \left[\begin{array}{cc} \Sigma_w & \Sigma_w \\ \Sigma_w & \Sigma_w + L\Sigma_vL^T \end{array}\right]. \tag{11b}$$

The matrix Ω accounts for the weight matrices in the cost function (1b) and the matrix Σ determines the covariance function $\Sigma\delta(t-\tau)$ of μ.

III. Gradient method

In this section, we introduce the gradient method on the LQG objective function over the set of stabilizing gain matrices $\mathcal{S}_c \times \mathcal{S}_o$ and discuss its convergence properties.

Lemma 1: For any stabilizing pair of gain matrices $(K, L) \in \mathcal{S}_c \times \mathcal{S}_o$, the gradient of the LQG objective function f in (9) is given by

$$\begin{aligned} \nabla_K f(K,L) \;&= 2(RK - B^T\hat{P}_1)\hat{X}_1 \,-\, 2B^T\hat{P}_2\hat{X}_2^T \\ \nabla_L f(K,L) \;&= 2P_3(L\Sigma_v \,-\, X_3C^T) \,-\, 2P_2^TX_2C^T \end{aligned}$$

where the matrices

$$\begin{aligned} X \;&=\; \left[\begin{array}{cc} X_1 & X_2 \\ X_2^T & X_3 \end{array}\right], \quad \hat{X} \;=\; \left[\begin{array}{cc} \hat{X}_1 & \hat{X}_2 \\ \hat{X}_2^T & \hat{X}_3 \end{array}\right] \\ P \;&=\; \left[\begin{array}{cc} P_1 & P_2 \\ P_2^T & P_3 \end{array}\right], \quad \hat{P} \;=\; \left[\begin{array}{cc} \hat{P}_1 & \hat{P}_2 \\ \hat{P}_2^T & \hat{P}_3 \end{array}\right] \end{aligned} \tag{12}$$

are the unique solutions to the Lyapunov equations

$$A_{\rm cl} X \,+\, X A_{\rm cl}^T \,+\, \Sigma \,=\, 0 \tag{13a}$$

$$\hat{A}_{\rm cl} \hat{X} \,+\, \hat{X} \hat{A}_{\rm cl}^T \,+\, \hat{\Sigma} \,=\, 0 \tag{13b}$$

$$A_{\rm cl}^T P \,+\, P A_{\rm cl} \,+\, \Omega \,=\, 0 \tag{13c}$$

$$\hat{A}_{\rm cl}^T \hat{P} \,+\, \hat{P} \hat{A}_{\rm cl} \,+\, \hat{\Omega} \,=\, 0. \tag{13d}$$

Here, the matrices $A_{\rm cl}$, and Ω and Σ are given by (7) and (11), respectively, and

$$\hat{A}_{\rm cl} \;:=\; \begin{bmatrix} A - BK & LC \\ 0 & A - LC \end{bmatrix} \tag{14a}$$

$$\hat{\Omega} \;:=\; \begin{bmatrix} Q + K^T R K & Q \\ Q & Q \end{bmatrix} \tag{14b}$$

$$\hat{\Sigma} \;:=\; \begin{bmatrix} L\Sigma_v L^T & -L\Sigma_v L^T \\ -L\Sigma_v L^T & \Sigma_w + L\Sigma_v L^T \end{bmatrix}. \tag{14c}$$

Proof: See the appendix. ■

Using the explicit formula of the gradient in Lemma 1, the gradient descent method over the set of stabilizing gain matrices $\mathcal{S}_c \times \mathcal{S}_o$ follows the update rule

$$\begin{aligned} K^{k+1} \;&:=\; K^k \,-\, \alpha \nabla_K f(K^k, L^k), \quad K^0 \,\in\, \mathcal{S}_c \\ L^{k+1} \;&:=\; L^k \,-\, \alpha \nabla_L f(K^k, L^k), \quad\; L^0 \,\in\, \mathcal{S}_o \end{aligned} \tag{GD}$$

where $\alpha > 0$ is the stepsize.

A. Non-separability of gradients

For the LQG problem, unlike the optimal solution that satisfies the separation principle, we observe from Lemma 1 that the gradient is not separable as $\nabla_K f$ and $\nabla_L f$ depend on both L and K. To provide more insight, let us examine the value of gradient over two special subsets of the domain $\mathcal{S}_c \times \mathcal{S}_o$, namely $\mathcal{S}_c \times \{L^\star\}$, where $L^\star$ is the optimal Kalman gain, and $\{K^\star\} \times \mathcal{S}_o$, where $K^\star$ is the optimal control feedback gain in (5).

1) Optimal observer gain $L = L^\star$: In this case, from (5) and the corresponding Riccati equation, it follows that

$$L\Sigma_v \;=\; X_o^\star C^T \tag{15}$$

where $X_o^\star$ is the unique positive definite solution to the Lyapunov equation

$$(A \,-\, LC)X_o^\star \,+\, X_o^\star (A \,-\, LC)^T \;=\; -\Sigma_w - L\Sigma_v L^T.$$

Expanding (13a) and (13b), we observe that X_3 and $\hat{X}_3$ also satisfy the above Lyapunov equation. Thus, since $A - LC$ is Hurwitz, it follows that

$$X_o^\star \;=\; X_3 \;=\; \hat{X}_3. \tag{16}$$

In addition, combining equations (13b), (15), and (16) yields

$$(A - BK)\hat{X}_2 \,+\, \hat{X}_2 (A - LC)^T \;=\; 0. \tag{17}$$

Now, since $K \in \mathcal{S}_c$ and $L \in \mathcal{S}_o$, we obtain that $\hat{X}_2 = 0$. Form this equation in conjunction with (15) and (16), we obtain that the following terms in the gradient vanish

$$B^T \hat{P}_2 \hat{X}_2^T \;=\; 0, \quad P_3(L\Sigma_v \,-\, X_3 C^T) \;=\; 0 \tag{18a}$$

and thus the gradient simplifies to

$$\begin{aligned} \nabla_K f(K, L^\star) \;&= 2(RK - B^T \hat{P}_1)\hat{X}_1 \\ \nabla_L f(K, L^\star) \;&=\; -\, 2 P_2^T X_2 C^T. \end{aligned}$$

Remark 1: As we demonstrate in the proof of Lemma 1, for any stabilizing gains L and K, the matrix $\hat{X}_2$ is given by

$$\hat{X}_2 \;=\; \lim_{t \to \infty} \mathbb{E}\left[e(t) \hat{x}^T(t) \right].$$

Thus, the equality $\hat{X}_2 = 0$ can be directly established using the orthogonality principle which states that the optimal estimator is orthogonal to the estimation error.

2) Optimal control gain $K = K^\star$: Similar to the previous case, from (5) and the corresponding Riccati equation, it follows that

$$RK \;=\; B^T P_c^\star$$

where $P_c^\star$ is the unique positive definite solution to the Lyapunov equation

$$(A \,-\, BK)P_c^\star \,+\, P_c^\star (A \,-\, BK)^T \;=\; -Q \,-\, K^T R K.$$

Combining this equations with (13c) and (13d) yields $\hat{P}_1 = P_c^\star$ and $P_2 = 0$. Thus, we have

$$(RK - B^T \hat{P}_1)\hat{X}_1 \;=\; 0, \quad P_2^T X_2 C^T \;=\; 0 \tag{18b}$$

which yields

$$\begin{aligned} \nabla_K f(K^\star, L) \;&=\; -\, 2 B^T \hat{P}_2 \hat{X}_2^T \\ \nabla_L f(K^\star, L) \;&= 2 P_3 (L\Sigma_v \,-\, X_3 C^T). \end{aligned}$$

We observe that $\nabla_K f(K^\star, L)$ and $\nabla_L f(K, L^\star)$ do not vanish and thus the sets $\mathcal{S}_c \times \{L^\star\}$ and $\{K^\star\} \times \mathcal{S}_o$ are not invariant with respect to the gradient descent method. Therefore, unlike the optimal solutions, the gradient of the LQG objective function may not be decoupled.

IV. Lack of gradient domination

Recently, it has been shown that the gradient descent method achieves linear convergence for the LQR problem with full-state feedback in both discrete [8] and continuous-time [9] settings. These results build on the key observation that the full-state feedback LQR cost in (3) as a function of the feedback gains, denoted by $g(K)$, satisfies the Polyak-Łojasiewicz (PL) condition over its sub-levelsets, i.e.

$$\|\nabla g(K)\|_F^2 \;\geq\; \mu_g \,(g(K) \,-\, g(K^\star)) \tag{19}$$

for some constant $\mu_g > 0$. The PL condition, also known as gradient dominance, can be used as a surrogate to strong convexity to ensure convergence of gradient descent at a linear rate even for nonconvex problems. This observation raises the question of whether the LQG problem is also gradient dominant.

In addition, it has been recently shown that the set of stabilizing gains for the case of static output feedback, i.e. $u = -Ky,\ y = Cx$ consists of multiple connected components and local minima [12], which hinders the convergence of local search algorithms. However, in contrast to the static output feedback problem, the joint optimization of the controller and observer feedback gains for the LQG, as studied in this paper, involves the connected domain $\mathcal{S}_c \times \mathcal{S}_o$.

We now demonstrate that despite connectivity of the optimization domain, this formulation yet suffers from the existence of non-optimal equilibrium points and thus lack of gradient domination.

A. *Non-uniqueness of critical points*

The nonconvexity of the function f suggests the possibility of having multiple critical points $\nabla f(K,L) = 0$. In this section, we demonstrate that this is in fact the case by providing two of such points for the LQG problem in the general form. This should be compared and contrasted to the full-state feedback LQR problem which, despite nonconvexity, has been shown to have a unique critical point.

1) Global minimizer: The most obvious critical point is the unique global minimizer of f, which is given by (5). To verify this, note that for the optimal gains $L^\star$ $K^\star$, we have (18a) and (18b), respectively. Using these equations, and the form of gradient in Lemma 1, it immediately follows that $\nabla f(K^\star, L^\star) = 0$.

2) The origin for stable systems: To find another critical point, let us assume for simplicity that the system is open-loop stable. We next show that the origin $(K,L) = (0,0)$ is also a critical point, i.e., $\nabla f(0,0) = 0$.

For $(K,L) = (0,0)$, from (13b) it follows that $\hat{X}_1 = \hat{X}_2 = 0$. In addition, from (13c), it follows that $P_2 = P_3 = 0$. Combining these equalities and the form of gradient in Lemma 1 ensures $\nabla f(0,0) = 0$.

The existence of the sub-optimal critical point $(K,L) = (0,0)$ also implies that gradient domination may not hold for the LQG problem.

V. An Example

We consider the mass-spring-damper system in Figure 1 with s masses to demonstrate the performance of gradient descent given by (GD) on the LQG problem over the set $\mathcal{S}_c \times \mathcal{S}_o$ of stabilizing gains. We set all spring and damping constants as well as masses to unity. In state-space representation (1a), the state vector $x = [p^T\ v^T]^T$ contains the position and velocity of masses and the measured output $y = p$ is the position only. In this example, the dynamic, input, and output matrices are given by

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

where 0 and I are zero and identity matrices of suitable size, and T is a Toeplitz matrix with 2 on the main diagonal, -1 on the first super and sub-diagonals, and 0 elsewhere.

We solve the LQG problem with $Q = \Sigma_w = I$, $R = \Sigma_v = I$ for $s = 50$ masses, i.e., $n = 2s$ state variables. The

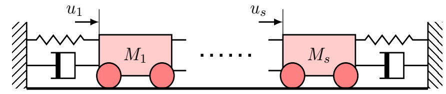

Fig. 1. Mass-spring-damper system.

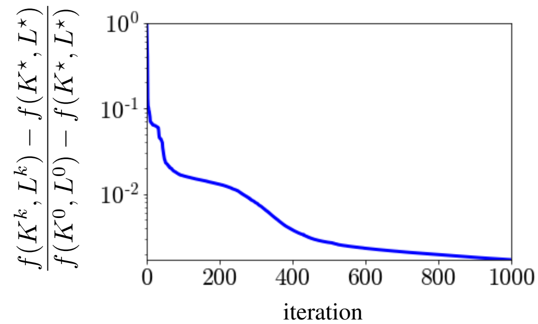

Fig. 2. Convergence curve of gradient descent for $s = 50$.

algorithm was initialized with scaled matrices of all ones $K^0 = (L^0)^T = 10^{-5}\mathbb{1}$. Figure 2 illustrates the convergence curves of gradient descent with a stepsize selected using a backtracking-based procedure initialized with $\alpha_0 = 10^{-3}$ that guarantees stability of the feedback loop and ensures descent. The optimal solution $K^\star$, $L^\star$ is obtained using (5) and the corresponding Riccati equations.

VI. Concluding Remarks

Motivated by the recent results on the global exponential convergence of policy gradient algorithms for the model-free LQR problem, in this paper we studied the standard LQG problem as optimization over controller and observer feedback gains. We present an explicit formulae for the gradient and demonstrate that for open-loop stable systems, in addition to the unique global minimizer, the origin is also a critical point for the LGQ problem, thus disproving the gradient dominance property. Numerical experiments for the convergence of gradient descent are also provided. Our work is ongoing to identify conditions under which gradient decent can solve the LQG problem at a linear rate.

Appendix

To obtain $\nabla_L f(K,L)$, we use the Taylor series expansion of $f(K, L+\tilde{L})$ around (K,L) and collect first-order terms. From (9), we have

$$\begin{aligned} f(K,L+\tilde{L}) - f(K,L) \ &\approx\ \left\langle \nabla_L f(K,L), \tilde{L} \right\rangle \\ &= \left\langle \Omega, \tilde{X} \right\rangle \end{aligned} \tag{20a}$$

where $\tilde{X}$ is the unique solution to

$$\begin{aligned} &A_{\mathrm{cl}}\tilde{X} + \tilde{X}A_{\mathrm{cl}}^T = -\tilde{A}_{\mathrm{cl}}X - X\tilde{A}_{\mathrm{cl}}^T - \tilde{\Sigma} \\ &= \begin{bmatrix} 0 & X_2 C^T \tilde{L}^T \\ \tilde{L}CX_2^T & \tilde{L}CX_3 + X_3 C^T \tilde{L}^T \end{bmatrix} - \tilde{\Sigma} =: \Phi \end{aligned} \tag{20b}$$

Here, the first equality is obtained by differentiating Lyapunov equation (10), and the second follows by noting that

$$\tilde{A}_{\rm cl} \;=\; \left[\begin{array}{cc} 0 & 0 \\ 0 & -\tilde{L}C \end{array}\right], \quad \tilde{\Sigma} \;=\; \left[\begin{array}{cc} 0 & 0 \\ 0 & \tilde{L}\Sigma_v L^T + L\Sigma_v \tilde{L}^T \end{array}\right].$$

Using the adjoint identity and (20), we obtain that

$$\left\langle \nabla_L f(K,L), \tilde{L} \right\rangle \;=\; \langle -\Phi, P \rangle$$

where P is given by (13c). Rearranging terms completes the proof for $\nabla_L f(K,L)$.

In order to obtain $\nabla_K f(K,L)$, we use a slightly different representation of the objective function. In particular, if we let $\hat{\xi} := \left[\begin{array}{cc} \hat{x}^T & e^T \end{array}\right]^T$, it is easy to verify that the closed-loop system satisfies

$$\dot{\hat{\xi}} \;=\; \hat{A}_{\rm cl}\,\hat{\xi} \,+\, \hat{\mu}$$

where the closed-loop matrix $\hat{A}_{\rm cl}$ is given by (14a) and $\hat{\mu} = \left[\begin{array}{cc} v^T L^T & w^T - v^T L^T \end{array}\right]^T$. Furthermore, it is straightforward to verify that for any stabilizing gain matrices $K \in \mathcal{S}_c$ and $L \in \mathcal{S}_o$, the LQG cost in (1b) is given by

$$f(K,L) \;:=\; \left\langle \hat{\Omega}, \hat{X} \right\rangle \qquad (21)$$

where $\hat{X} = \lim_{t\to\infty} \mathbb{E}\left[\hat{\xi}(t)\hat{\xi}^T(t)\right]$ is the unique solution to the algebraic Lyapunov equation (13b) and and the matrices $\hat{\Omega}$ and $\hat{\Sigma}$ are given by (14). Now, using this representation, the same technique as in the first part of the proof can be used to obtain $\nabla_L f(K,L)$. This completes the proof.

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 *IEEE Int Conf. Robot. Autom.*, 2018, pp. 7559–7566.

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

[3] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "On the sample complexity of the linear quadratic regulator," *Found. Comput. Math.*, pp. 1–47, 2017.

[4] 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. PMLR, 2019, pp. 3108–3117.

[5] 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, p. 1–51, 2019.

[6] H. Mohammadi, M. Soltanolkotabi, and M. R. Jovanović, "On the linear convergence of random search for discrete-time LQR," *IEEE Control Syst. Lett.*, 2020, in press; doi:10.1109/LCSYS.2020.3006256.

[7] 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," *IEEE Trans. Automat. Control*, 2021, doi:10.1109/TAC.2021.3087455; also arXiv:1912.11899.

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

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

[10] H. Karimi, J. Nutini, and M. Schmidt, "Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition," in *In European Conference on Machine Learning*, 2016, pp. 795–811.

[11] J. Ackermann, "Parameter space design of robust control systems," *IEEE Trans. Automat. Control*, vol. 25, no. 6, pp. 1058–1072, 1980.

[12] H. Feng and J. Lavaei, "On the exponential number of connected components for the feasible set of optimal decentralized control problems," in *Proceedings of the 2019 American Control Conference*, 2019, pp. 1430–1437.

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

[14] K. J. Åström, *Introduction to stochastic control theory*. Academic Press, New York, 1970.