Learning the Globally Optimal Distributed LQ Regulator

Luca Furieri FURIERIL@CONTROL.EE.ETHZ.CH
Automatic Control Laboratory, ETH Zurich, Switzerland

Yang Zheng ZHENGY@G.HARVARD.EDU
School of Engineering and Applied Sciences, Harvard University, USA

Maryam Kamgarpour * MKAMGAR@CONTROL.EE.ETHZ.CH
Automatic Control Laboratory, ETH Zurich, Switzerland

Abstract

We study model-free learning methods for the output-feedback Linear Quadratic (LQ) control problem in finite-horizon subject to subspace constraints on the control policy. Subspace constraints naturally arise in the field of distributed control and present a significant challenge in the sense that standard model-based optimization and learning leads to intractable numerical programs in general. Building upon recent results in zeroth-order optimization, we establish model-free sample-complexity bounds for the class of distributed LQ problems where a local gradient dominance constant exists on any sublevel set of the cost function. We prove that a fundamental class of distributed control problems—commonly referred to as Quadratically Invariant (QI) problems—as well as others possess this property. To the best of our knowledge, our result is the first sample-complexity bound guarantee on learning globally optimal distributed output-feedback control policies.

1. Introduction

Recent years have witnessed significant attention and progress in controlling unknown dynamical systems solely based on system trajectory observations. This shift from classical control approaches to data-driven ones is motivated by the ever increasing complexity of critical emerging dynamical systems, whose mathematical models may be unreliable or simply not available (Hou and Wang, 2013). When it comes to learning an optimal control policy, the available approaches can be broadly divided into two categories. The first class of methods is denoted as *model-based*, where the historical system data is exploited to build an approximation of the nominal system and classical optimal robust control is then used on this system approximation. The second class of methods is denoted as *model-free*, where reinforcement learning is used to directly learn an optimal control policy based on the observed costs, without explicitly constructing a model for the system.

Model-free approaches tend to require more samples to achieve a policy of equivalent accuracy (Tu and Recht, 2018), but are inherently unaffected by the potential challenges of designing an optimal controller. Indeed, in large-scale dynamical systems, the control policy is often required to be *distributed*, in the sense that different controllers can only base their control policy on partial sensor measurements due to limited sensing capabilities, geographic distance or privacy concerns. Given such limitations, it has been known that the corresponding optimization problems are NP-hard in general (Papadimitriou and Tsitsiklis, 1986; Blondel and Tsitsiklis, 2000; Witsenhausen,

* This research was gratefully funded by the European Union ERC Starting Grant CONENE.

1968). Often, one can only derive a tractable approximation using convex relaxations (e.g. Fazelnia et al. (2016)) or restrictions (e.g. Furieri et al. (2019b)). The difficulties in solving model-based optimal control for large-scale systems motivate us to bypass numerical programming altogether and study the properties of model-free methods for distributed control.

For Linear Quadratic (LQ) control problems in infinite-horizon without additional constraints, the optimal policy can be derived with dynamic programming by solving a Riccati equation. For distributed control tasks, the optimal policy might not be linear in general (Witsenhausen, 1968) and even in those cases where an explicit solution can be computed (see e.g., Lamperski and Doyle (2015) and references within), the optimal controller requires several internal states and might admit a rather complicated formulation. Furthermore, when designing a *static* distributed controller in infinite-horizon, model-free methods are unlikely to find the globally optimal controller due to the feasible set being disconnected in general (Feng and Lavaei, 2019); for this setting, convergence to *local* optima was confirmed by Hassan-Moghaddam et al. (2019).

Motivated as per above, in this paper we consider model-free learning of globally optimal dynamic distributed controllers. We focus on the *finite-horizon* setup, where the feasible set is naturally connected because every control policy yields a finite closed-loop cost. Furthermore, in this setup we can 1) encode general *dynamic* time-varying linear policies in a relatively simple way, and 2) consider time-varying system dynamics.

Our contributions First, we provide a general framework for model-free learning of distributed dynamic linear policies in finite-horizon with uncertain initial state, process noise and noisy output observations. Second, our key contribution is to establish a property of local gradient dominance for a class of distributed control problems, including 1) all Quadratically Invariant (QI) problems (Rotkowitz and Lall, 2006) and 2) some non-QI problems. This local gradient dominance property is crucial for establishing model-free sample-complexity bounds using zeroth-order optimization; we base our corresponding analysis on the recent results of Malik et al. (2018), while adapting and extending relevant aspects.

Related work Thanks to its well-understood solution structure and its properties, the LQ problem has enjoyed significant attention in the line of work on model-based learning, originating from classical system identification (see Ljung (2010) for a nice overview). A non-asymptotic analysis was provided by Fiechter (1997) and significantly refined by Dean et al. (2017), and sublinear regret results for online model-based methods were recently obtained by Dean et al. (2018); Abbasi-Yadkori and Szepesvári (2011); Abeille and Lazaric (2018). Still assuming full sensor information, Mania et al. (2019) exploited Riccati perturbation theory to analyse the output-feedback case and Dean et al. (2019) included safety constraints on states and inputs. The literature on model-free learning has recently been attracting significant research interest starting from the works of Fazel et al. (2018) and Abbasi-Yadkori et al. (2018). Related to our work is Fazel et al. (2018), which showed that for the state-feedback LQ problem without an information structure, a standard policy-gradient method is guaranteed to converge to the global optimum and established sample-complexity bounds that scaled with $\tilde{\mathcal{O}}(\epsilon^{-4})$, where ϵ is the suboptimality gap. This bound was improved to $\tilde{\mathcal{O}}(\epsilon^{-2})$ in Malik et al. (2018), at the expense of a constant probability of success, for a discounted LQ cost function. Furthermore, similar convergence properties were shown for robust control tasks without an information structure; we refer the reader to Gravell et al. (2019) for the case of multiplicative noise and to Zhang et al. (2019) for $\mathcal{H}_\infty$ robustness guarantees.

To the best of our knowledge, global convergence for distributed control problems, where a subspace constraint is imposed on the control policy, has not been studied from a model-free perspective. A related problem has been addressed with a model-based approach in Fattahi et al. (2019), where the authors extended the method of Dean et al. (2017) by adding subspace constraints on the *closed-loop responses*. In general, a sparse closed-loop response does not lead to a sparse controller implementation that is exclusively based on measuring the outputs, and vice-versa (see Zheng et al. (2020) for details on this aspect). The resulting framework is thus not directly comparable with the one considered in this paper. We also note that the work of Fattahi et al. (2019) restricts the analysis to state-feedback, whereas we consider noisy output-feedback.

2. Background and Problem Statement

Notation: We use $\mathbb{R}$ and $\mathbb{N}$ to denote the set of real numbers and integers, respectively. We write $M = \text{blkdg}(M_1, \dots, M_n)$ to denote a block-diagonal matrix with $M_1, \dots, M_n$ on its diagonal block entries. The Kronecker product between $M \in \mathbb{R}^{m\times n}$ and $P \in \mathbb{R}^{p\times q}$ is denoted as $M \otimes P \in \mathbb{R}^{mp\times nq}$. Given $K \in \mathbb{R}^{m\times n}$, $\text{vec}(K) \in \mathbb{R}^{mn}$ is a column vector that stacks the columns of K. We define the inverse operator $\text{vec}^{-1} : \mathbb{R}^{mn} \to \mathbb{R}^{m\times n}$ that maps a vector into a matrix (the matrix dimension shall be clear in the context). The Euclidean norm of a vector $v \in \mathbb{R}^n$ is denoted by $\|v\|_2^2 = v^\mathsf{T} v$ and the Frobenius norm of a matrix $M \in \mathbb{R}^{m\times n}$ is denoted by $\|M\|_F^2 = \text{Trace}(M^\mathsf{T} M)$. For a symmetric matrix M, we write $M \succ 0$ (resp. $M \succeq 0$) if and only if it is positive definite (resp. positive semidefinite). We say that $x \sim \mathcal{D}$ if the random variable $x \in \mathbb{R}^n$ is distributed according to $\mathcal{D}$. Given a binary matrix $X \in \{0, 1\}^{m\times n}$, we define the associated *sparsity subspace* as

$$\text{Sparse}(X) := \{Y \in \mathbb{R}^{m\times n} |\ Y_{i,j} = 0 \ \text{ if } X_{i,j} = 0, i = 1, \dots, m, j = 1, \dots, n\ \}.$$

The set $\mathbb{S}_r \subseteq \mathbb{R}^d$ denotes the shell of radius $r > 0$ in $\mathbb{R}^d$, that is $\mathbb{S}_r = \{z \in \mathbb{R}^d |\ \|z\|_2 = r\}$. A zero block of dimension $m \times n$ is denoted as $0_{m\times n}$.

2.1. The LQ Optimal Control Problem Subject To Subspace Constraints

We consider time-varying linear systems in discrete-time

$$x_{t+1} = A_t x_t + B_t u_t + w_t\,, \quad y_t = C_t x_t + v_t\,, \tag{1}$$

where $x_t \in \mathbb{R}^n$ is the system state at time t affected by process noise $w_t \sim \mathcal{D}_w$ with $x_0 = \mu_0 + \delta_0$, $\delta_0 \sim \mathcal{D}_{\delta_0}$, $y_t \in \mathbb{R}^p$ is the observed output at time t affected by measurement noise $v_t \sim \mathcal{D}_v$, and $u_t \in \mathbb{R}^m$ is the control input at time t to be designed. We assume that the distributions $\mathcal{D}_w, \mathcal{D}_{\delta_0}$ $\mathcal{D}_v$ are bounded, have zero mean and variances of $\Sigma_w, \Sigma_{\delta_0}, \Sigma_v \succ 0$ respectively. Boundedness of the disturbances is a reasonable assumption in physical applications and it is commonly exploited to simplify the analysis of model-free methods (Fazel et al., 2018; Malik et al., 2018)[1]. We consider the evolution of (1) in finite-horizon for $t = 0, \dots N$, where $N \in \mathbb{N}$. By defining the matrices

$$\mathbf{A} = \text{blkdg}(A_0, \dots, A_N), \quad \mathbf{B} = \begin{bmatrix}\text{blkdg}(B_0, \dots, B_{N-1}) \\ 0_{n\times mN}\end{bmatrix}, \quad \mathbf{C} = \text{blkdg}(C_0, \dots, C_N)\,,$$

and the vectors $\mathbf{x} = \begin{bmatrix}x_0^\mathsf{T} & \dots & x_N^\mathsf{T}\end{bmatrix}^\mathsf{T}$, $\mathbf{y} = \begin{bmatrix}y_0^\mathsf{T} & \dots & y_N^\mathsf{T}\end{bmatrix}^\mathsf{T}$, $\mathbf{u} = \begin{bmatrix}u_0^\mathsf{T} & \dots & u_{N-1}^\mathsf{T}\end{bmatrix}^\mathsf{T}$, $\mathbf{w} = \begin{bmatrix}x_0^\mathsf{T} & w_0^\mathsf{T} & \dots & w_{N-1}^\mathsf{T}\end{bmatrix}^\mathsf{T}$ and $\mathbf{v} = \begin{bmatrix}v_0^\mathsf{T} & \dots & v_N^\mathsf{T}\end{bmatrix}^\mathsf{T}$, and the block-down shift matrix

$$\mathbf{Z} = \begin{bmatrix}0_{1\times N} & 0 \\ I_N & 0_{N\times 1}\end{bmatrix} \otimes I_n\,,$$

1. Malik et al. (2018) noted that extension to sub-Gaussian disturbances is possible; we leave this case to future work.

we can write the system (1) compactly as $\mathbf{x} = \mathbf{ZAx} + \mathbf{ZBu} + \mathbf{w}$, $\mathbf{y} = \mathbf{Cx} + \mathbf{v}$, leading to

$$\mathbf{x} = \mathbf{P}_{11}\mathbf{w} + \mathbf{P}_{12}\mathbf{u}\,, \quad \mathbf{y} = \mathbf{Cx} + \mathbf{v}, \tag{2}$$

where $\mathbf{P}_{11} = (I - \mathbf{ZA})^{-1}$ and $\mathbf{P}_{12} = (I - \mathbf{ZA})^{-1}\mathbf{ZB}$. In this paper, we consider linear output-feedback policies $u_t = K_{t,0}y_0 + K_{t,1}y_1, + \dots, K_{t,t}y_t, t = 0, 1, \dots, N-1$. More compactly

$$\mathbf{u} = \mathbf{Ky}, \quad \mathbf{K} \in \mathcal{K}\,, \tag{3}$$

where $\mathcal{K}$ is a subspace in $\mathbb{R}^{mN\times p(N+1)}$ that $1)$ ensures causality of $\mathbf{K}$ by setting to 0 those entries that correspond to future outputs, $2)$ can enforce a time-varying spatio-temporal information structure for distributed control. The presence of these information constraints presents a significant challenge for optimal distributed control; we refer to Furieri and Kamgarpour (2019b) for details.

The distributed Linear Quadratic (LQ) optimal control problem in finite-horizon is:

$$\textbf{Problem } LQ_{\mathcal{K}}: \qquad \min_{\mathbf{K}\in\mathcal{K}} \quad J(\mathbf{K})\,,$$

where the cost $J(\mathbf{K})$ is defined as

$$J(\mathbf{K}) := \mathbb{E}_{\mathbf{w},\mathbf{v}}\left[\sum_{t=0}^{N-1}\left(y_t^\mathsf{T} M_t y_t + u_t^\mathsf{T} R_t u_t\right) + y_N^\mathsf{T} M_N y_N\right], \tag{4}$$

and $M_t \succeq 0$ and $R_t \succ 0$ for every t. We denote the optimal value of problem $LQ_{\mathcal{K}}$ as $J^\star$. By rearranging (2)-(3), it can be observed that $J(\mathbf{K})$ is in general a non-convex multivariate polynomial in the entries of $\mathbf{K}$; see Appendix A of our Arxiv report Furieri et al. (2019a) for an explicit expression of $J(\mathbf{K})$ and some useful properties. Note that $LQ_{\mathcal{K}}$ is a constrained problem over the subspace $\mathcal{K}$; it is convenient to observe that $LQ_{\mathcal{K}}$ is actually equivalent to an unconstrained problem.

Lemma 1 *Let $d \in \mathbb{N}$ be the dimension of $\mathcal{K}$, and the columns of $P \in \mathbb{R}^{mpN(N+1)\times d}$ be a basis of the subspace $\{vec(\mathbf{K}) |\ \forall \mathbf{K} \in \mathcal{K}\}$. Define the function $f : \mathbb{R}^d \to \mathbb{R}$ as $f(z) := J(vec^{-1}(Pz))$. Then, $LQ_{\mathcal{K}}$ is equivalent to the* unconstrained *problem*[2]

$$\min_{z\in\mathbb{R}^d} f(z)\,. \tag{5}$$

Proof Since the columns of P are a basis of $\mathcal{K}$, we have 1) $\forall \mathbf{K} \in \mathcal{K},\ \exists z \in \mathbb{R}^d$ such that $\text{vec}(\mathbf{K}) = Pz$ and 2) $\forall z \in \mathbb{R}^d,\ \text{vec}^{-1}(Pz) \in \mathcal{K}$. Hence, (5) is equivalent to $LQ_{\mathcal{K}}$. ■

The function $f(z)$ is generally a non-convex multivariate polynomial in $z \in \mathbb{R}^d$ which may possess multiple local-minima, thus preventing global convergence of model-free algorithms. Furthermore, as opposed to the standard LQ problem without subspace constraints, one cannot in general exploit a tractable reformulation or Riccati-based solutions and apply model-based learning as per e.g. Dean et al. (2017); Mania et al. (2019). Fortunately, $f(z)$ admits a unique global minimum if it is *gradient dominated i.e.*, $\mu(f(z) - J^\star) \le \|\nabla f(z)\|_2^2\ ,\ \forall z \in \mathbb{R}^d$ for some $\mu > 0$ (Karimi et al., 2016). Gradient dominance has been proved for the standard LQ problem in infinite horizon without subspace constraints (Fazel et al., 2018; Gravell et al., 2019). Inspired by these recent results, we explore conditions under which $LQ_{\mathcal{K}}$ admits a gradient dominance constant, to be exploited for model-free learning of globally optimal distributed controllers.

2. Throughout this paper, $J(\mathbf{K})$ is reserved for the LQ cost function in (4) and $f(z)$ is reserved for the equivalent cost function $f(z) := J(\text{vec}^{-1}(Pz))$.

3. Local Gradient Dominance for QI Problems and Beyond

It is well-known since the work of Rotkowitz and Lall (2006) that problem $LQ_{\mathcal{K}}$ can be equivalently transformed into a strongly convex program if and only if QI holds, that is

$$\mathbf{K}\mathbf{C}\mathbf{P}_{12}\mathbf{K} \in \mathcal{K}, \quad \forall \mathbf{K} \in \mathcal{K}\,. \tag{6}$$

We refer to Appendix B of our Arxiv report Furieri et al. (2019a) for a detailed discussion of the QI property. In model-free learning, one directly investigates whether $LQ_{\mathcal{K}}$ possesses favourable properties for convergence (such as gradient dominance) rather than convexifying through a system-dependent change of variables. Our main contribution is to prove a *local* gradient dominance property for 1) the class of all QI instances of $LQ_{\mathcal{K}}$ 2) other non-QI instances of $LQ_{\mathcal{K}}$.

Theorem 2 *Let $\mathcal{K}$ be QI with respect to $\mathbf{CP}_{12}$, i.e.,* (6) *holds. For any $\delta > 0$ and initial value $z_0 \in \mathbb{R}^d$, define the sublevel set $\mathcal{G}_{10\delta^{-1}} = \{z \in \mathbb{R}^d \mid \; f(z) - J^\star \leq 10\delta^{-1}\Delta_0\}$, where $\Delta_0 := f(z_0) - J^\star$ is the initial optimality gap. Then, the following statements hold.*

1. *$\mathcal{G}_{10\delta^{-1}}$ is compact.*
2. *$f(z)$ has a unique stationary point.*
3. *$f(z)$ admits a local gradient dominance constant $\mu_\delta > 0$ over $\mathcal{G}_{10\delta^{-1}}$, that is*

$$\mu_\delta(f(z) - J^\star) \leq \|\nabla f(z)\|_2^2 \;, \; \forall z \in \mathcal{G}_{10\delta^{-1}}\,. \tag{7}$$

The proof of Theorem 2 is reported in Appendix B of our Arxiv report Furieri et al. (2019a). In other words, QI guarantees existence of a gradient dominance constant μ_δ which is "global" on $\mathcal{G}_{10\delta^{-1}}$, for any $\delta > 0$. By inspection of (7), for every $\delta > 0$, the only stationary point contained in $\mathcal{G}_{10\delta^{-1}}$ is the global optimum, since whenever $\nabla f(z) = 0$, we have $f(z) = J^\star$.

We remark that the property (7) is weaker than the more common global gradient dominance; we present a simple instance of $LQ_{\mathcal{K}}$ satisfying (7) in Appendix B of our Arxiv report Furieri et al. (2019a). We will show in Section 4 that (7) is sufficient for global convergence of model-free algorithms. Furthermore, diverse classes of non-QI $LQ_{\mathcal{K}}$ that yet are convex in $\mathbf{K}$ have been found in Lessard and Lall (2010); Shin and Lall (2011), and more recently in Furieri and Kamgarpour (2019a). For completeness, we report an explicit example in Appendix B of our Arxiv report. Finally, notice that $\mathcal{K}$ typically enforces a sparsity pattern for $\mathbf{K}$. Therefore, the QI property (6) can be checked without knowing the specific system dynamics, but only using the knowledge of the sparsity pattern of $\mathbf{CP}_{12}$ (see Furieri and Kamgarpour (2019b) for example). This is a realistic assumption for dynamical systems that are distributed by nature.

4. Learning the Globally Optimal Constrained Control Policy

Here, we derive sample-complexity bounds for model-free learning of globally optimal distributed controllers for the problems identified in Section 3. Our analysis technique is founded on recent zeroth-order optimization results (Malik et al., 2018; Fazel et al., 2018); we extend the derived bounds on the gradient estimates to include noise on the initial state, process noise and measurement noise. Furthermore, our analysis hinges on the observation that local gradient dominance is sufficient to guarantee the sample-complexity bounds in our framework, whereas Malik et al. (2018); Fazel et al. (2018) used a global one.

The zeroth-order optimization literature is quite rich, see for instance the works of Balasubramanian and Ghadimi (2018); Nesterov and Spokoiny (2017); Ghadimi and Lan (2013) and references therein. The key idea of such algorithms is to sample noisy function values of f generated by an *oracle*, based on which an approximated gradient $\widehat{\nabla f}$ is estimated and standard gradient descent is applied to optimize over z. While Malik et al. (2018) proposed an analysis for two-point evaluation oracles that allow for tighter sample-complexity bounds, we notice that in many control applications one cannot control or predict the noise affecting each separate measurement. We will thus focus on the one-point evaluation oracle setup according to the Algorithm 1 below.

Algorithm 1 Model-free learning of distributed controllers

1: Input: z_0, number of iterations T, stepsize $\eta > 0$ and smoothing radius $r > 0$.
2: **for** $i = 0, \dots, T-1$ **do**
3: Sample $u \sim \text{Unif}(\mathbb{S}_r)$, let nature "choose" disturbances $\delta_0 \sim \mathcal{D}_{\delta_0}$, $w_t \sim \mathcal{D}_w$ for all $t = 0, \dots, N-1$, $v_t \sim \mathcal{D}_v$ for all $t = 0, \dots, N$.
4: Apply $\hat{\mathbf{u}} = \text{vec}^{-1}[P(z_i + u)]\hat{\mathbf{y}}$ iteratively using (1) and store the resulting trajectories $\hat{\mathbf{y}}, \hat{\mathbf{u}}$.
5: Compute $\hat{f} = \hat{\mathbf{y}}^\mathsf{T}\text{blkdg}(M_0, \dots, M_N)\hat{\mathbf{y}} + \hat{\mathbf{u}}^\mathsf{T}\text{blkdg}(R_0, \dots R_{N-1})\hat{\mathbf{u}}$ and $\widehat{\nabla f} = \hat{f}\frac{d}{r^2}u$.
6: $z_{i+1} \leftarrow z_i - \eta\widehat{\nabla f}$.
7: **end for**
8: **return** $\mathbf{K}_T = \text{vec}^{-1}(Pz_T)$.

In Algorithm 1, the observed cost $\hat{f}$ can be regarded as the output of a one-point evaluation oracle. Indeed, we have $\mathbb{E}_{\mathbf{w},\mathbf{v}}[\hat{f}] = f(z_i + u)$ by definition, and each observation $\hat{f}$ is affected by a different noise sequence. The value $\widehat{\nabla f}$ can be interpreted as a noisy estimate[3] of the gradient $\nabla f(z_i)$. We now turn to the convergence analysis.

4.1. Sample-complexity bounds

Our sample-complexity analysis holds under two main assumptions on the function f.

Assumption 1. For any $\delta > 0$ and initial value $z_0 \in \mathbb{R}^d$, the sublevel set $\mathcal{G}_{10\delta^{-1}}$ of f is compact.

Assumption 2. For any $\delta > 0$, the function f admits a local gradient dominance constant μ_δ over $\mathcal{G}_{10\delta^{-1}}$ as per (7).

In Section 3 we have provided our main result about verifying that both assumptions hold for 1) all QI control problems and 2) some instances of non-QI problems, therefore establishing a novel fundamental connection between distributed control and zeroth-order optimization. As is common in zeroth-order analysis, we also verify Lipschitzness and smoothness of f on its sublevel sets.

Lemma 3 *Let $\delta > 0$ and Assumption 1 hold. Then, there exist $\rho_0 > 0$, and $L_\delta, M_\delta > 0$, such that*

$$|f(z') - f(z)| \le L_\delta \left\|z' - z\right\|_2 , \quad \left\|\nabla f(z') - \nabla f(z)\right\|_2 \le M_\delta \left\|z' - z\right\|_2 , \tag{8}$$

for every $z', z \in \mathcal{G}_{10\delta^{-1}}$ such that $\|z' - z\|_2 \le \rho_0$.

3. Technically, $\mathbb{E}[\widehat{\nabla f}] = \nabla f_r(z_i)$, where $f_r(z_i) = \mathbb{E}_u[f(z_i + u)]$, with u is taken uniformly at random over $\mathbb{S}_r$; see, e.g., (Malik et al., 2018, Lemma 6) for details.

Proof We know that f is a multivariate polynomial. Since $\mathcal{G}_{10\delta^{-1}}$ is compact, it suffices to note that ∇f is a vector of polynomials and that polynomials are bounded on any compact set. ■

We are now ready to present the sample-complexity result. Its proof is reported in Appendix C of our Arxiv report Furieri et al. (2019a).

Theorem 4 *Let Assumptions 1 and 2 hold, and consider Algorithm 1. Let $\eta > 0$ and $r > 0$ be selected according to*

$$\eta \leq \min\left\{\frac{\epsilon\mu_\delta\delta^3 r^2}{16000 M_\delta d^2 D^2 f(z_0)^2}, \frac{1}{2M_\delta}, \frac{\rho_0 r\delta}{20dDf(z_0)}\right\},$$
$$r \leq \min\left\{\frac{\min\left(\frac{1}{2M_\delta}, \frac{\rho_0}{L_\delta}\right)\mu_\delta}{2M_\delta}\sqrt{\frac{\delta\epsilon}{40}}, \frac{1}{2M_\delta}\sqrt{\frac{\epsilon\mu_\delta\delta}{5}}, \rho_0, \frac{10\delta^{-1}f(z_0)}{L_\delta}\right\},$$

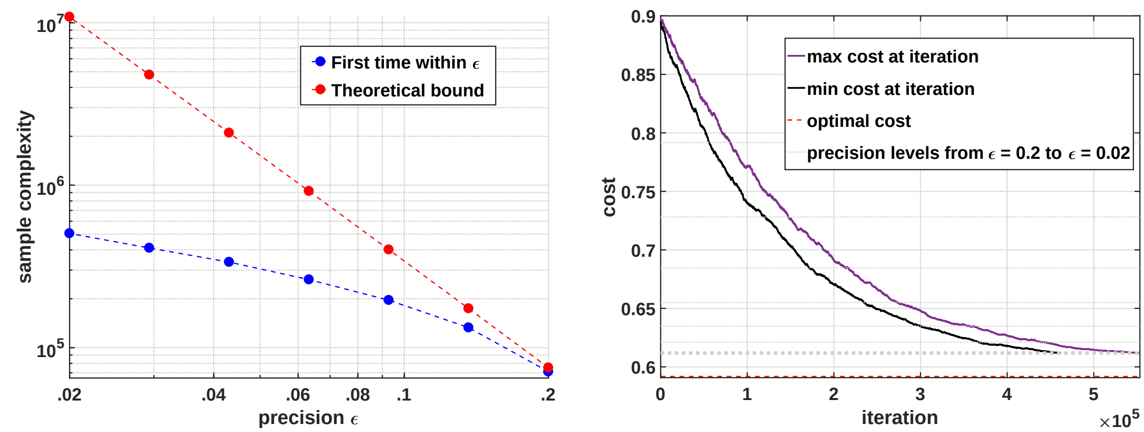
where $\rho_0 > 0$, μ_δ is the local gradient dominance constant of $f(z)$ associated with $\mathcal{G}_{10\delta^{-1}}$, L_δ, M_δ are the local Lipschitzness and smoothness constants described in Lemma 3, and $D = \max\left(\frac{W^2}{\lambda_{\mathbf{w}}}, \frac{V^2}{\lambda_{\mathbf{v}}}\right)$, with W the value such that $\|\mathbf{w}\|_2 \leq W$ for all $\delta_0 \sim \mathcal{D}_{\delta_0}$, $w_0, \dots, w_{N-1} \sim \mathcal{D}_w$, V the value such that $\|\mathbf{v}\|_2 \leq V$ for all $v_0, \dots, v_N \sim \mathcal{D}_v$, and $\lambda_{\mathbf{w}}$ and $\lambda_{\mathbf{v}}$ are the minimum eigenvalues of $\mathbb{E}[\mathbf{w}\mathbf{w}^\mathsf{T}]$ and $\mathbb{E}[\mathbf{v}\mathbf{v}^\mathsf{T}]$ respectively. Then for any $\epsilon > 0$ and $0 < \delta < 1$ such that $\epsilon\log(\frac{4\Delta_0}{\delta\epsilon}) \leq \frac{16}{\delta}\Delta_0$, running Algorithm 1 with $T = \frac{4}{\eta\mu_\delta}\log(\frac{4\Delta_0}{\delta\epsilon})$ iterations yields a distributed *control policy $\mathbf{K}_T \in \mathcal{K}$ such that*

$$J(\mathbf{K}_T) - J^\star \leq \epsilon\,,$$

with probability greater than $1-\delta$.

Theorem 4 yields a constant probability suboptimality guarantee based on the analysis technique of Malik et al. (2018); there, the infinite-horizon LQR problem with exact state measurements and no information structure was addressed. Our result extends this analysis as follows. First, we consider distributed control problems, given noisy output information and allow for inclusion of both noise on the initial state, process noise and measurement noise. We achieve this by bounding the variance of the gradient estimate in our Lemma 13, which is reported in Appendix C of our Arxiv report Furieri et al. (2019a). Second, we allow for a success probability $1-\delta$ for any $\delta > 0$ and show how δ affects the sample-complexity. Last, we observe that a *local* gradient dominance constant valid on $\mathcal{G}_{10\delta^{-1}}$ is sufficient for the analysis, whereas Malik et al. (2018) considered a global one. Nonetheless, we note that the finite-horizon framework enjoys a significant simplification because any control policy leads to a finite cost and the feasible region of control policies is always connected. Extension to infinite-horizon requires further work.

Based on Theorem 4, the model-free sample-complexity scales as $\mathcal{O}\left(\frac{d^2}{\epsilon^2\delta^4}\log\frac{1}{\epsilon\delta}\right)$. For the standard centralized LQR problem, model-based methods (e.g. Dean et al. (2017)) can enjoy a better scaling of $\mathcal{O}\left(\frac{d}{\epsilon^2}\log\frac{1}{\delta}\right)$, but extension of these methods to the general $LQ_{\mathcal{K}}$ is non-trivial due to non-existence of a convex reformulation in general. Interestingly, the scaling with respect to the suboptimality gap ϵ is practically unaffected despite using a model-free method.



(a) (b)

Figure 1: In Figure 1(a) we plotted 1) the average number of steps over 10 runs of Algorithm 1 needed to achieve 7 increasingly tight precision levels from $\epsilon = 0.2$ to $\epsilon = 0.02$ and 2) the sample-complexity T predicted by Theorem 4, when η is scaled as $\eta = \mathcal{O}\left(\epsilon r^2\right)$ and r is scaled as $r = \mathcal{O}(\sqrt{\epsilon})$. In Figure 1($b$), we plotted the convergence behaviour, highlighting the maximum and minimum cost achieved at each iteration among the 10 runs.

4.2. Experiments for distributed control

To validate our results, we considered problem $LQ_{\mathcal{K}}$ for $A_t = A$, $B_t = [1 \quad -1 \quad 0]^\mathsf{T}$, $C_t = I$ for $t = 0, 1, 2$, $\mu_0 = 10^{-1} \times [1 \quad -1 \quad 1]^\mathsf{T}$, with $\mathbf{K} \in \mathcal{K} = \text{Sparse}(\mathbf{S})$, where

$$A = \begin{bmatrix} 1 & 0 & -10 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \otimes [1 \quad 0 \quad 0] \,.$$

Furthermore, we consider additive initial state uncertainty uniformly distributed in the interval $[-10^{-2}, 10^{-2}]$, process noise and measurements noise w_t, v_t uniformly distributed in the interval $[-10^{-3}, 10^{-3}]$ for every t. The cost function weights are chosen as $M_t = \frac{1}{4}I$ and $R_t = \frac{1}{4}I$ at each t. It is easy to verify that $\mathbf{KCP}_{12}\mathbf{K} \in \text{Sparse}(\mathbf{S})$ for any $\mathbf{K} \in \text{Sparse}(\mathbf{S})$; hence, $\mathcal{K}$ is QI with respect to $\mathbf{CP}_{12}$ and Theorem 4 holds. Figure 1(a)-1(b) shows that the sample-complexity scales significantly better than the one predicted by Theorem 4 with respect to ϵ, thus validating the corresponding bounds for this example. Additional details and considerations on selecting η and r are reported in Appendix D of our Arxiv report Furieri et al. (2019a).

5. Conclusions

Motivated by the challenges of solving model-based distributed optimal control problems, we studied model-free policy learning subject to subspace constraints. By drawing a novel connection between gradient dominance and QI, we derived sample-complexity bounds on learning the globally optimal distributed controller for a class of problems including QI problems and other instances; for these, the available model-based learning techniques might not converge to a global optimum. One exciting future direction is to extend these results to infinite-horizon, by bridging the gap between dynamical controller synthesis and a gradient-descent landscape. We also envision including safety constraints. Furthermore, significantly sharpening our sample-complexity bounds might be possible with potentially more refined analysis.

References

Yasin Abbasi-Yadkori and Csaba Szepesvári. Regret bounds for the adaptive control of linear quadratic systems. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 1–26, 2011.

Yasin Abbasi-Yadkori, Nevena Lazic, and Csaba Szepesvári. Regret bounds for model-free linear quadratic control. *arXiv preprint arXiv:1804.06021*, 2018.

Marc Abeille and Alessandro Lazaric. Improved regret bounds for thompson sampling in linear quadratic control problems. In *International Conference on Machine Learning*, pages 1–9, 2018.

Krishnakumar Balasubramanian and Saeed Ghadimi. Zeroth-order nonconvex stochastic optimization: Handling constraints, high-dimensionality, and saddle-points. *arXiv preprint arXiv:1809.06474*, 2018.

Vincent D Blondel and John N Tsitsiklis. A survey of computational complexity results in systems and control. *Automatica*, 36(9):1249–1274, 2000.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. *arXiv preprint arXiv:1710.01688*, 2017.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. Regret bounds for robust adaptive control of the linear quadratic regulator. In *Advances in Neural Information Processing Systems*, pages 4188–4197, 2018.

Sarah Dean, Stephen Tu, Nikolai Matni, and Benjamin Recht. Safely learning to control the constrained linear quadratic regulator. In *2019 American Control Conference (ACC)*, pages 5582–5588. IEEE, 2019.

Salar Fattahi, Nikolai Matni, and Somayeh Sojoudi. Efficient learning of distributed linear-quadratic controllers. *arXiv preprint arXiv:1909.09895*, 2019.

Maryam Fazel, Rong Ge, Sham Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, 2018.

Ghazal Fazelnia, Ramtin Madani, Abdulrahman Kalbat, and Javad Lavaei. Convex relaxation for optimal distributed control problems. *IEEE Trans. on Aut. Contr.*, 62(1):206–221, 2016.

Han Feng and Javad Lavaei. On the exponential number of connected components for the feasible set of optimal decentralized control problems. In *American Control Conference (ACC)*, page 8, 2019.

Claude-Nicolas Fiechter. PAC adaptive control of linear systems. In *Annual Workshop on Computational Learning Theory: Proceedings of the tenth annual conference on Computational learning theory*, volume 6, pages 72–80. Citeseer, 1997.

Luca Furieri and Maryam Kamgarpour. First order methods for globally optimal distributed controllers beyond quadratic invariance. *arXiv preprint arXiv:1909.10999*, 2019a.

Luca Furieri and Maryam Kamgarpour. Unified approach to convex robust distributed control given arbitrary information structures. *IEEE Trans. on Aut. Contr.*, 2019b.

Luca Furieri, Yang Zheng, and Maryam Kamgarpour. Learning the globally optimal distributed LQ regulator. *arXiv preprint arXiv:1912.08774*, 2019a.

Luca Furieri, Yang Zheng, Antonis Papachristodoulou, and Maryam Kamgarpour. Sparsity invariance for convex design of distributed controllers. *arXiv preprint arXiv:1906.06777*, 2019b.

Saeed Ghadimi and Guanghui Lan. Stochastic first-and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013.

Benjamin Gravell, Peyman M. Esfahani, and Tyler Summers. Learning robust control for LQR systems with multiplicative noise via policy gradient. *arXiv preprint arXiv:1905.13547*, 2019.

Sepideh Hassan-Moghaddam, Mihailo R Jovanović, and Sean Meyn. Data-driven proximal algorithms for the design of structured optimal feedback gains. In *American Control Conference (ACC)*, pages 5846–5850. IEEE, 2019.

Zhong-Sheng Hou and Zhuo Wang. From model-based control to data-driven control: Survey, classification and perspective. *Information Sciences*, 235:3–35, 2013.

Hamed Karimi, Julie Nutini, and Mark Schmidt. Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 795–811. Springer, 2016.

Andrew Lamperski and John C Doyle. The $\mathcal{H}_2$ control problem for quadratically invariant systems with delays. *IEEE Trans. on Aut. Contr.*, 60(7):1945–1950, 2015.

Laurent Lessard and Sanjay Lall. Internal quadratic invariance and decentralized control. In *Proceedings of the 2010 American Control Conference*, pages 5596–5601. IEEE, 2010.

Lennart Ljung. Perspectives on system identification. *Annual Reviews in Control*, 34(1):1–12, 2010.

Dhruv Malik, Ashwin Pananjady, Kush Bhatia, Koulik Khamaru, Peter L Bartlett, and Martin J Wainwright. Derivative-free methods for policy optimization: Guarantees for linear quadratic systems. *arXiv preprint arXiv:1812.08305*, 2018.

Horia Mania, Stephen Tu, and Benjamin Recht. Certainty equivalent control of LQR is efficient. *arXiv preprint arXiv:1902.07826*, 2019.

Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017.

Christos H Papadimitriou and John Tsitsiklis. Intractable problems in control theory. *SIAM j. on contr. and opt.*, 24(4):639–654, 1986.

Michael Rotkowitz and Sanjay Lall. A characterization of convex problems in decentralized control. *IEEE Trans. on Aut. Contr.*, 51(2):274–286, 2006.

Hyung Sik Shin and Sanjay Lall. Decentralized control via Groebner bases and variable elimination. *IEEE transactions on automatic control*, 57(4):1030–1035, 2011.

Stephen Tu and Benjamin Recht. The gap between model-based and model-free methods on the linear quadratic regulator: An asymptotic viewpoint. *arXiv preprint arXiv:1812.03565*, 2018.

Hans S Witsenhausen. A counterexample in stochastic optimum control. *SIAM Journal on Control*, 6(1):131–147, 1968.

Kaiqing Zhang, Bin Hu, and Tamer Başar. Policy optimization for $\mathcal{H}_2$ linear control with $\mathcal{H}_\infty$ robustness guarantee: Implicit regularization and global convergence. *arXiv preprint arXiv:1910.09496*, 2019.

Yang Zheng, Luca Furieri, Antonis Papachristodoulou, Na Li, and Maryam Kamgarpour. On the equivalence of youla, system-level and input-output parameterizations. *IEEE Transactions on Automatic Control*, pages 1–8, 2020.