Regret Lower Bounds for Learning Linear Quadratic Gaussian Systems

Ingvar Ziemann, Member, IEEE, and Henrik Sandberg, Fellow, IEEE

Abstract—In this article, we establish regret lower bounds for adaptively controlling an unknown linear Gaussian system with quadratic costs. We combine ideas from experiment design, estimation theory, and a perturbation bound of certain information matrices to derive regret lower bounds exhibiting scaling on the order of magnitude $\sqrt{T}$ in the time horizon $T$. Our bounds accurately capture the role of control-theoretic parameters and we are able to show that systems that are hard to control are also hard to learn to control; when instantiated to state feedback systems we recover the dimensional dependency of earlier work but with improved scaling with system-theoretic constants, such as system costs and Gramians. Furthermore, we extend our results to a class of partially observed systems and demonstrate that systems with poor observability structure also are hard to learn to control.

Index Terms—Adaptive control, closed loop identification, fundamental limits, statistical learning.

I. INTRODUCTION

LEARNING algorithms are set to play an increasing role in modern engineering solutions. Early successes include walking robots [1] and playing repeated games, such as [2] and are likely to become increasingly important in modern safety-critical infrastructure, such as smart grids and intelligent transportation. However, their emergence in safety-critical systems is not without problems. Indeed, one of the hallmarks of these early successes is abundant data from a relatively unchanging source, potentially even through simulation access. By contrast, their failure modes as a component of modern engineering systems in dynamic, changing, environments when data are scarce remain poorly understood.

Such an understanding must necessarily be based on two components: 1) the study of fundamental performance limitations, which no algorithm can exceed and 2) the provision of algorithms which match these fundamental limitations. In this work, we focus on the first component and provide an information-theoretic framework for understanding the fundamental limits of adaptive controlling an a priori unknown linear Gaussian system subject to a quadratic cost function. We emphasize the interaction between the hardness of learning and the hardness of control by deriving regret lower bounds that accurately capture various system-theoretic constants: our bounds demonstrate that linear systems with poor controllability (or observability) structure are hard to learn to control.

A. Problem Formulation

We study the fundamental limitations to adaptively learning to control the following parametric system model:

$$X_{t+1} = A(\theta)X_t + B(\theta)U_t + W_t, \quad X_0 \sim N(0, \Gamma_0) \quad (1)$$

with $t = 0, 1, \ldots$ and parameterized by $\theta \in \mathbb{R}^{d_\Theta}$. In other words, the joint law of the $X_t$ is parametrized by $\theta$.

We will also consider the extension to partially observed systems, in which the controller is constrained to rely on past and present observations of the form

$$Y_t = C(\theta)X_t + V_t. \quad (2)$$

The noise processes $W_t$ and $V_T$ are assumed mutually independent iid sequences of mean zero Gaussian noise with fixed covariance matrices $\Sigma_W \succeq 0$ and $\Sigma_V \succeq 0$, respectively. Above $X_t \in \mathbb{R}^{d_X}$ and $Y_t \in \mathbb{R}^{d_Y}$, respectively, denote the observations available to the learner, which are obtained in a sequential fashion after applying the input $U_{t-1} \in \mathbb{R}^{d_U}$ according to (1) and (2) in the case of partially observed systems. The matrices $A(\theta) \in \mathbb{R}^{d_X \times d_X}$, $B(\theta) \in \mathbb{R}^{d_X \times d_U}$, and $C(\theta) \in \mathbb{R}^{d_Y \times d_X}$ are assumed to be known continuously differentiable functions of the unknown parameter, $\theta \in \mathbb{R}^{d_\Theta}$. The unstructured case, to which much attention has been devoted in literature, is recovered by setting $\text{vec}[A(\theta) \quad B(\theta) \quad C(\theta)] = \theta$. We further denote by $\mathcal{Y}_t$ the sigma-field generated by the observations of $Y_1, \ldots, Y_t$ and possible auxiliary randomization, AUX, a random variable with density against the $d_{\text{AUX}}$-dimensional Lebesgue measure. We also point out that the state feedback (SF) case (1) can be recovered from the general partially observable (PO) case (1) and (2) by setting $C(\theta) = I_{d_Y}$ and $\Sigma_V = 0_{d_Y \times d_Y}$. In this case, the nondegeneracy condition above simplifies to $\Sigma_W \succ 0$.

The learner is to design a policy $\pi$, constituted by a sequence of conditional laws of the variables $U_t$ given $\mathcal{Y}_t$. We will say that a policy $\pi$ is admissible if the process $\{U_t\}_{t=0,1,\ldots}$ is adapted to the filtration $\{\mathcal{Y}_t\}_{t=0,1,\ldots}$. Roughly, the learner’s goal is to choose...
such an admissible policy \( \pi \) as to minimize the cumulative cost

\[
V_T^\pi(\theta) \triangleq \mathbb{E}_\theta^\pi \left[ \sum_{t=0}^{T-1} \left( X_t^T Q X_t + U_t^T R U_t \right) + X_T^T Q_T(\theta) X_T \right]
\]

where we have fixed two known positive-definite cost weighting matrices \( Q, R \succ 0, Q \in \mathbb{R}^{d_X \times d_X}, R \in \mathbb{R}^{d_U \times d_U} \) and a terminal cost matrix \( Q_T(\theta) \succeq 0 \). An equivalent formulation is to minimize the cumulative suboptimality due to not knowing the true parameter \( \theta \), namely the regret

\[
R_T^\pi(\theta) \triangleq \mathbb{E}_\theta^\pi \left[ \sum_{t=0}^{T-1} \left( X_t^T Q X_t + U_t^T R U_t \right) + X_T^T Q_T(\theta) X_T \right] - \inf_{\pi} \left[ \sum_{t=0}^{T-1} \left( X_t^T Q X_t + U_t^T R U_t \right) + X_T^T Q_T(\theta) X_T \right].
\]

(3)

Whenever \( \theta = \text{vec} [A \ B \ C] \), we will also allow ourselves the abuse of notation \( R_T^\pi(A,B,C) = R_T^\pi(\theta) \).

In principle, there is no learning involved in the above formulation since the optimal policy \( \pi^*(\theta) \) is admissible. As is standard in modern statistics, we circumvent this issue by requiring that the learner performs well on a neighborhood of instances. Namely, for a fixed tolerance \( \varepsilon > 0 \) we posit that the learner seeks to minimize \( \sup_{\theta \in B(\theta, \varepsilon)} R_T^\pi(\theta) \). It is this latter quantity which we seek to study in terms of fundamental achievable performance limits.

Finally, we make the following standing assumptions about the system (1) and (2).

A1: The pair \((A(\theta), B(\theta))\) is stabilizable and the pair \((A(\theta), B(\theta), C(\theta))\) is detectable.

A2: The terminal cost \( Q_T \) renders the optimal controller stationary; \( Q_T(\theta) = P(\theta) \) where \( P(\theta) \) is the solution to the discrete algebraic Riccati equation [which is reproduced in (5)].

A3: The distribution of \( X_0 \) renders the optimal filter stationary; \( X_0 \sim \mathcal{N}(0, S(\theta)) \), where \( S(\theta) \) is given by (10).

The first assumption A1 guarantees the feasibility of the long-run averaged version of the problem at hand. Assumptions A2 and A3 are made to streamline the exposition; it is possible to derive analogous results in their absence, the only difference being that the quantities related to the Riccati equations (5) and (10) in the regret representation derived in Lemma II.1 become time varying. However, the time-varying versions of these quantities converge at an exponential rate to those used here, so the overall difference is negligible.

B. Contribution

We establish fundamental limits for adaptive control problems by proving local minimax regret lower bounds for linear quadratic regulator (LQR) and linear quadratic Gaussian (LQG) problems. Our lower bounds take the form

\[
\inf_{\pi} \sup_{\theta \in B(\theta, \varepsilon)} R_T^\pi(\theta) \geq c(\theta, \varepsilon) \sqrt{T}
\]

for a constant \( c(\theta, \varepsilon) \) that captures the instance-specific hardness of the problem, such as dependence on the dimension of the unknown parameter \( d_\theta \) and various system quantities depending on the controllability and observability structure of the instance under consideration. We remark the radius \( \varepsilon \) appearing in (4) can be thought of as the strength of an adversary selecting the parameters against which the adaptive algorithm is to compete— we will prove slightly stronger statements in which the strength of the adversary \( \varepsilon \) is vanishing: \( \varepsilon \to 0 \).

We match the dimensional dependencies in prior work [3], but in contrast to it, we obtain much sharper dependencies on system-theoretic constants. This allows us to conclude that systems with poor controllability structure are much harder to learn to control. Moreover, we extend our results to a particular class of partially observed systems and show that poor observability can be analogously detrimental to learning performance.

At a more technical level, we also introduce a new proof approach for establishing regret lower bounds. Our technique relies on linking the nullspace of the Fisher information matrix (60) of the data collected by any policy to the regret incurred by that policy. Informally, for any policy \( \pi \), this idea can be summarized as

\[
\text{Fisher Information}(\pi) = \text{Fisher Information}(\pi^*) + O(\text{Regret})
\]

where \( \pi^* \) is the optimal policy. In other words, policies which generate experiments much more informative than the optimal one must have a significant regret component. The Fisher information thus allows us to make precise the exploration–exploitation tradeoff in adaptive control. We derive our lower bound (4) by combining this insight with a version of van Trees’ inequality (62).

C. Related Work

Regret minimization in the context of linear quadratic systems was first introduced by [4] and [5], following their treatment of regret in the related multiarmed bandit problem [6], see also Guo [7]. The notion of regret captures that there is a tradeoff between performance and information collected. This is what Feldbaum called the dual nature of control [8], [9] or what is now known in the reinforcement learning literature as the exploration–exploitation tradeoff. On the one hand, one wishes for the algorithm to perform well now, but on the other, one also needs the algorithm to be informative about the state of world, as to be able to reject for instance model misspecification or other uncertainty. Regret lower bounds describe the tradeoff between the statistical rate of convergence of an adaptive algorithm and its performance on a nominal instance. Generally, such lower bounds are the consequence of requiring an algorithm to perform well on some class of models, subject to uncertainty about the nominal instance. Such an algorithm must necessarily “explore” to determine which model to optimize for. In turn, such exploration leads to suboptimal performance on the nominal instance. Key algorithmic principles based on this need for exploration date back at least to Simon’s [10] 1956 introduction of certainty equivalence and the notion of dual control [8], [9]. When it comes to linear quadratic systems, an early reference is the self-tuning regulator by [11], in turn inspired by Kalman’s earlier
work [12]. In this context, the primary mathematical issue was first and foremost the convergence of the adaptive controller to the global optimum [13], [14], [15], [16]. These works provide and analyze adaptive algorithms that are asymptotically optimal on average, which one now would perhaps simply call sublinear regret.

Recently, the adaptive linear-quadratic-Gaussian problem has mainly been studied under the assumption of perfect state observability ($C = I_{dx}, V_i = 0$ identically). While the problem has a rich history, it was repopularized by [17] in which the authors provided an algorithm attaining $O(\sqrt{T})$ regret. A number of works following that publication focus on improving and providing more computationally tractable algorithms in this setting [18], [19], [20], [21], [22], [23], [24], [25], [26]. While the emphasis of these works is entirely on providing upper bounds, recently, some effort has been made to understand the complexity of the problem in terms of lower bounds. Notably, Simchowitz and Foster [3] provided nearly matching upper and lower bounds scaling almost correctly with the dimensional dependence given that the entire set of parameters $(A, B)$ are unknown. While Simchowitz and Foster [3] provided lower bounds that scale correctly with the dimension of the problem (1), their bounds are rather loose in terms of system-theoretic quantities. In many situations, these can be exponentially larger than the relevant dimensional factors [27]. With this in mind, our work seeks to understand the hardness of learning to control in terms of such system-theoretic quantities. To this end, we provide refined lower bounds for the SF setting, which further apply to partially observed systems. Later, in our work [28], Th. 8], we leveraged the refined bounds of the present work to show that these “hidden” system-theoretic quantities can have exponential impact on the regret for many reasonable classes of systems. Indeed, there is still a rather large gap to be filled in literature in terms of how control-theoretic parameters (such as Gramians) affect the scaling limits of the smallest possible regret an algorithm can incur. Motivated by this gap, we develop a theory of regret lower bounds based on Cramér–Rao type bounds which are known to yield tight lower bounds in system identification. We see Table I for schematic overview of the abovementioned results.

Turning to the more general situation including partial state observability, the key works are [29] and [30]. Simchowitz et al. [29] considered a closely related but more general setting in which the noise model is (possibly) adversarial instead of Gaussian and produce a $O(\sqrt{T})$ regret algorithm. Lale et al. [30] considered a system of the form considered in the present work and give an algorithm attaining $O(T^{2/3})$ regret. However, they also show that when a condition related to the persistency of excitation of the benchmark law holds, their algorithm can attain polylogarithmic regret. However, it is not known when their condition holds for the optimal law. In other words, their notion of regret may differ from ours. We will construct a negative result in this direction that satisfies A1–A3 in Section V; we show that there exists stabilizable and detectable systems with full rank noise on which obtaining logarithmic regret is impossible. At this point, we remark that bounds on the scale $\log T$ have been well known for some time [4], [31], [32], [33]. By contrast, lower bounds on the order $\sqrt{T}$ are a more recent phenomenon starting with [3], [34], [35].

In principle, the recognition that $\sqrt{T}$ regret is often unimprovable stems from the fact that the stochastic adaptive control problem is intimately connected to parameter estimation. Already from the outset, algorithm design has to a large extent been based on certainty equivalence; that is, estimating the parameters and plugging these estimates into an optimality equation, as if they were the ground truth [11], [22]. Our lower bound condition is also closely related to parameter estimation. It is a consequence of the viewpoint that an adaptive controller, to be asymptotically optimal, must generate an experiment asymptotically very similar to one in which the optimal controller has generated the data. If this experiment is “bad” in a certain sense, logarithmic regret becomes impossible. This reasoning is akin to earlier results in experiment design and identification for control. In particular, there is a very interesting result due to [36] which finds that the optimal experiment for minimum variance control is to use the minimum variance controller itself. This is the opposite of the phenomenon which more general adaptive linear-quadratic regulation problems exhibit, as noted for instance by [37] and [38]. Here, application of the optimal feedback law typically yields a singular experiment. However, even under more general circumstances, it still holds true that the optimal experiment design is closed loop [39]. For more on experiment design, see [40].

Our lower bound condition, uninformativeness (Definition III.1), is related to identifiability. Actually, it is inspired by a similar phenomenon in point estimation, which may become

| Paper                           | Setting | Method      | Upper Bound       | Lower Bound       |
|---------------------------------|---------|-------------|-------------------|-------------------|
| Abbasi-Yadkori and Szepesvári [2011] | SF       | Optimism    | $O(\sqrt{T})$    |                   |
| Dean et al. [2018]              | SF       | CE          | $O(T^{2/3})$      |                   |
| Parson et al. [2020]            | SF       | CE          | $O(T^{2/3})$      |                   |
| Mania et al. [2019]             | SF       | CE          | $O(T^{2/3})$      |                   |
| Cohen et al. [2019]             | SF       | Optimism    | $O(T^{2/3})$      |                   |
| Simchowitz and Foster [2020]    | SF       | CE          | $O(\sqrt{T})$    | $\Omega(T^{2/3})$|
| Simchowitz et al. [2020]        | PO       | Gradient    | $O(\sqrt{T})$    |                   |
| Section IV                      | SF       |             |                   | $\Omega(\sqrt{T})$|
| Section V                       | PO       |             |                   |                   |
| Tsianis et al. [2022]           | SF       |             |                   | $\Omega(\sqrt{T})$|

ZIEMANN AND SANDBERG: REGRET LOWER BOUNDS FOR LEARNING LINEAR QUADRATIC GAUSSIAN SYSTEMS 161

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
arbitrarily hard when the Fisher information is singular [41], [42], [43], [44]. Indeed, we will see that un informativeness allows one to draw conclusions analogous to results of [38] about the necessity of identifying the true parameter in adaptive control subject to the data being generated by the optimal controller itself (which one would hope to, at least asymptotically, be close to). In principle, the condition stipulates that there is a lack of curvature (as measured by the loss geometry) in certain directions of parameter space in a neighborhood of the optimal policy. It is precisely this lack of curvature that prevents logarithmic regret, which has been reported in certain special cases of the adaptive LQG problem [4], [7].

The proof approach for our lower bound also relies on methods pioneered in parameter estimation; we use the van Trees’ inequality [45], [46], [47]. This necessarily involves the Fisher information, which, quite naturally, allows for taking problem structure into account by considering different parametrizations of the problem dynamics. We also note that the idea to bound a minimax complexity by a suitable family of Bayesian problems is well known in statistics literature [47]. See also [48], [49], and [50]. We also note that [32] attributes the idea to use Cramér–Rao type bounds to derive performance lower bounds in adaptive control to [31]. Finally, we note in passing that nonsingularity of the Fisher information is strongly related to the size of the smallest singular value of the empirical covariance matrix, which has been the emphasis of some recent advances in linear system identification [51], [52], [53], [54], [55]. In the present work, we ask analogous questions in the regret-minimization setting, in which there is a rather rich interplay between identification and control.

D. Notation

Maxima (respectively, minima) of two numbers $a, b \in \mathbb{R}$ are denoted by $a \lor b = \max(a, b)$ ($a \land b = \min(a, b)$). For two sequences $(a_t)_{t \in \mathbb{Z}}$ and $(b_t)_{t \in \mathbb{Z}}$ we introduce the shorthand $a_t \lesssim b_t$ if there exists a universal constant $C > 0$ and an integer $t_0$ such that $a_t \leq C b_t$ for every $t \geq t_0$. Let $\mathcal{X} \subseteq \mathbb{R}^d$ and let $f, g \in \mathcal{X} \to \mathbb{R}$. We write $f = O(g)$ if $\limsup_{x \to x_0} |f(x)/g(x)| < \infty$, where the limit point $x_0$ is typically understood from the context. We use $\hat{O}$ to hide logarithmic factors and write $f = o(g)$ if $\limsup_{x \to x_0} |f(x)/g(x)| = 0$. We write $f = \Omega(g)$ if $\liminf_{x \to x_0} |f(x)/g(x)| > 0$.

1. Euclidean Spaces: The Euclidean norm on $\mathbb{R}^d$ is denoted $\| \cdot \|_2$, and the unit sphere in $\mathbb{R}^d$ is denoted $S^d-1$. The ball of radius $\varepsilon$ in $\| \cdot \|_2$, centered at $x \in \mathbb{R}^d$ is denoted $B(x, \varepsilon)$. The standard inner product on $\mathbb{R}^d$ is denoted $\langle \cdot, \cdot \rangle$. We embed matrices $M \in \mathbb{R}^{d_1 \times d_2}$ in Euclidean space by vectorization: $\mathbb{E} \in \mathbb{R}^{d_1 \times d_2}$, where vec is the operator that vertically stacks the columns of $M$ (from left to right and from top to bottom). For a matrix $M$ the Euclidean norm is the Frobenius norm, i.e., $\| M \|_F = \| \text{vec } M \|_2$. We similarly define the inner product of two matrices $M, N$ by $(M, N) = (\text{vec } M, \text{vec } N)$. The transpose of a matrix $M$ is denoted by $M^T$ and $\text{tr } M$ denotes its trace. We define (twice) the symmetric component of $M$ by $\text{sym}(M) \triangleq M + M^T$. For a matrix $M \in \mathbb{R}^{d_1 \times d_2}$, we order its singular values $\sigma_1(M), \ldots, \sigma_{d_1}(M)$ in descending order by magnitude. We also write $\| M \|_{\infty}$ for its largest singular value: $\| M \|_{\infty} \triangleq \sigma_1(M)$. To not carry dimensional notation, we will also use $\sigma_{\min}(M)$ for the smallest nonzero singular value. For square matrices $M \in \mathbb{R}^{d \times d}$ with real eigenvalues, we similarly order the eigenvalues of $M$ in descending order as $\lambda_1(M), \ldots, \lambda_d(M)$. In this case, $\lambda_{\min}(M)$ will also be used to denote the minimum (possibly zero) eigenvalue of $M$. For two symmetric matrices $M, N$, we write $M \succ N$ ($M \preceq N$) if $M - N$ is positive (semi)definite.

II. Riccati, Regret, and a Reduction to Bayesian Estimation

We begin by recalling a number of elementary facts regarding the optimal control and filtering of the system (1) and (2), valid in the case the parameter $\theta$ is fixed and $A_1$–$A_3$ hold. For a reference, see for instance [56].

1) Riccati Equations: The expression for the linear system (1) and (2) and the regret (3) are cumbersome to work with directly. However, these can be simplified using Riccati equations. Let us now recall, provided that $\theta = (A, B)$ is stabilizable, that the optimal policy minimizing the long-run average of $V^T(\theta)$, is represented by a stabilizing feedback matrix $K(\theta)$ and can be expressed via $P(\theta)$ which together satisfy

$$P = Q + A^T P A - A^T P B (B^T P B + R)^{-1} B^T P A$$

$$K = -(B^T P B + R)^{-1} (B^T P A)$$

where dependence on $\theta$ has been omitted in (5) and (6) for notational brevity.

Since $X_t$ is not always directly observed in our problem formulation, (5) and (6) do not constitute complete solutions of the LQG problem. Indeed, the asymptotically optimal policy is of the form $U_t = X_t$ where $X_t = E_Q^\pi(X_t | X_t)$, which provided suitable initial conditions (to be specified momentarily), can be expressed recursively

$$\begin{align*}
\dot{X}_{t+1} &= A(\theta) X_t + B(\theta) U_t \\
& \quad + F(\theta)[Y_{t+1} - C(\theta)(A(\theta) X_{t} + B(\theta) U_t)]
\end{align*}$$

(7)

where $F(\theta) \in \mathbb{R}^{d_x \times d_y}$ is given by (11) and is characterized as follows in terms of a Riccati equation (10), dual to (5) and (6). We further denote

$$\nu_t = F(\theta)[Y_{t+1} - C(\theta)(A(\theta) X_t + B(\theta) U_t)]$$

(8)
which plays a role corresponding to that of \( \omega_t \) in (1) for the filtered state (7). The process \( \nu_t \) is iid Gaussian with mean zero and we denote its covariance \( \Sigma_\nu \). It will also be convenient to introduce the one-step ahead prediction \( \zeta_t = E^\theta_0 |X_t|Y_t \) which similarly satisfies a recursion
\[
\zeta_{t+1} = A(\theta)\zeta_t + B(\theta)U_t + F(\theta)|Y_t - C(\theta)\zeta_t|.
\]
(9)
The quantity \( F \) appearing in both the asymptotic Kalman filter recursions (7) and (9) is characterized by the Filter Riccati equation
\[
S = ASA^T - ASC^T(CSC^T + \Sigma_\nu)^{-1}CSA^T + \Sigma_W
\]
(10)
\[
F = SC^T(CSC^T + \Sigma_\nu)^{-1}
\]
where we again omit dependence on \( \theta \) in (10) and (11) for notational brevity.
The quantity \( S(\theta) \) is the (steady state) covariance matrix of \( X_t - \zeta_t \). Similarly, we define \( \Xi(\theta) \) to be the covariance matrix of \( X_t - \bar{X}_t \), which can be expressed in terms of \( S(\theta) \) as
\[
\Xi = S - SC^T(CSC^T + \Sigma_\nu)^{-1}CS.
\]
Finally, we point out that unrolling the filter dynamics (7) allows us to define a sequence of \( \theta \)-parametrized linear maps \( G_t(\theta) : \mathbb{R}^{(d_0 + d_Y)(t+1)} \rightarrow \mathbb{R}^{d_X} \) that act on past inputs and outputs to produce the present state estimate. In other words, \( G_t(\theta) \) is defined by
\[
G_t(\theta)(U_{0:t}, Y_{0:t}) = \bar{X}_t
\]
(12)
where, for a fixed \( \theta \), \( \bar{X}_t \) is given by (7). We now turn to representing the regret (3) in terms of the filtered state (7).

2) Regret Representation: In terms of the quantities above, it is straightforward to verify that the optimal cost \( V^*_T(\theta) \) can be expressed as (see, e.g., Söderström [56, Th. 11.3]):
\[
V^*_T(\theta) = E^\theta_0 X_0^T P(\theta)X_0 + T \text{tr}(\Sigma_\nu P(\theta))
\]
\[
+ T \text{tr}(QS(\theta)).
\]
(13)
We now provide an alternative representation of the regret (3).

Lemma II.1: Assume A1–A3. Then:
\[
R_T^*(\theta) = \sum_{t=1}^{T-1} E^\theta_0 (U_t \times - K(\theta)E^\theta_0 |X_t|Y_t)|Y_t)
\]
\[
\times (B^T(\theta)P(\theta)B(\theta) + R)(U_t \times - K(\theta)E^\theta_0 |X_t|Y_t)|Y_t).
\]
(14)

Proof: After subtracting constant terms independent of \( \pi \), this is immediate by [56, Lemma 11.2], which gives a general expression for \( V^*_T(\theta) \). ■

Remark II.1: For systems with an observed state described by (1), by embedding them into the description (1) and (2), setting \( C = I_{d_X} \) and \( \Sigma_V = 0 \), (14) becomes
\[
R_T^*(\theta) = \sum_{t=1}^{T} E^\theta_0 (U_t \times - K(\theta)X_t)^T
\]
\[
\times (B(\theta)P(\theta)B(\theta) + R)(U_t \times - K(\theta)X_t).
\]
The problem of regret minimization in LQG can be seen as that of sequentially learning \( \theta \rightarrow K(\theta)E^\theta_0 |X_t|Y_t| \) which is the composition of both the asymptotically optimal SF controller and the Kalman filter.

With this in mind, (14) may be understood as saying that regret minimization is at least as hard as minimizing a cumulative weighted quadratic estimation error for the sequence of estimands \( K(\theta)E^\theta_0 |X_t|Y_t| \). To be clear, our perspective is that we wish to estimate the function value of \( K(\theta)E^\theta_0 |X_t|Y_t| \), where the function to be estimated \( \theta \rightarrow K(\theta)E^\theta_0 |X_t|Y_t| \) is revealed at time \( t \). By relaxing the local minimax problem to a Bayesian setting, the entire trajectory \( Y_{0:t} \) is then interpreted as a noisy observation of the underlying parameter \( \theta \). A natural approach for variance lower bounds is to rely on Fisher information and use (Bayesian) Cramér–Rao bounds. The discussion above is summarized by the following observation.

Lemma II.2: Fix \( \varepsilon > 0 \) and let \( M \) be a smooth and compactly supported prior on \( B(\theta, \varepsilon) \). Fix also a function \( \tau : \mathbb{R} \rightarrow \mathbb{R} \) satisfying \( \tau(t) \geq t, t \in \mathbb{N} \). As long as A1–A3 hold, it holds for any admissible policy \( \pi \) that
\[
\sup_{\theta \in B(\theta, \varepsilon)} R_T^*(\theta) \geq \inf_{\theta \in B(\theta, \varepsilon)} \sum_{t=0}^{T-1} E^\theta_0 (\bar{U}_t - K(\theta)E^\theta_0 |X_t|Y_t)|Y_t|^T
\]
\[
\times (X_tX_t^T)(E[\theta]|Y_{\tau(t)}| - K(\theta))^T.
\]
(15)

III. EXPERIMENT DESIGN AND REGRET
Let us reiterate the point made above following Lemma II.1 that we intend to lower bound the regret by estimation-theoretic
methods. From this perspective, the adaptive policy—the decision variable of the learner—generates an experiment, namely a sequence of input–output pairs \((Y_{0:T-1}, U_{0:T-1})\).

To make precise the notion of an experiment—and the information contained in such an experiment—let us introduce the concept of Fisher information. This quantity can be thought of as a form of signal-to-noise ratio of the observed measurements with respect to the unknown parameter \(\theta\). We are interested in the Fisher information pertaining to the information available to the learner in the setting (1), where the measurement is the input–output pair \((Y_{0:T-1}, U_{0:T-1})\). Denote by \(p_{\pi,\theta,T}\) the joint density of the random variable \((\pi X, Y_0, \ldots, Y_{T-1})\) under policy \(\pi\) (conditionally on the the parameter \(\theta\)). The following information quantity serves as the basis for our analysis and is a policy-dependent measure of information available to the learner about the uncertain parameter \(\theta\)

\[
 I(\theta; \pi, T) \triangleq I_p(\theta) \tag{17}
\]

with \(I_p(\theta)\), as in (60).

A. Optimal Policies and Degenerate Experiments

Naively, the perspective discussed in following Lemma II.1 viewing (14) as a cumulative estimation error suggests a lower bound on the scale \(\log T\), since one might think that (17) should scale linearly in time \(T\). In this case, the associated parameter estimation errors variances should decay as \(1/T\). Indeed, this is the case of certain instances, see, e.g., [4].

However, when the Fisher information corresponding to the experiment of running the optimal policy is singular, this reasoning fails. We will see that the Fisher information corresponding to low regret algorithms have nearly singular information. Namely, if \(I(\theta; \pi_*, T)\) given by (17) is singular, and this singularity is relevant for identifying \(K(\theta)\), we expect there to be a nontrivial tradeoff between exploration and exploitation. For instance, we will see that when the experiment corresponding to the optimal policy is degenerate, any algorithm with \(O(\sqrt{T})\)-regret necessarily generates an experiment in which the smallest (relevant) singular value only scales as \(\sqrt{T}\); (17) scales sublinearly in certain relevant directions. Put yet differently: if the optimal policy \(\pi_*\) yields an experiment in which \(K(\theta)\) is not locally identifiable, we expect the regret to be \(\Omega(\sqrt{T})\).

We now make precise the above reasoning by imposing two conditions, which together rule out the possibility of logarithmic regret. The first condition states that the optimal policy \(\pi_*\) does not persistently excite the parameters for local identifiability in terms of Fisher information.

**Definition III.1:** Fix \(\varepsilon > 0\) and a subspace \(U\) of \(\mathbb{R}^{d_\Theta}\). The instance \((\theta, A(\cdot), B(\cdot), C(\cdot), Q, R, \Sigma W, \Sigma V)\) is \((U, \varepsilon)\)-locally uninformative if there exists a neighborhood \(B(\theta, \varepsilon)\) such that for all \(\theta \in B(\theta, \varepsilon)\) and \(v \in U\):

1) \(I(\theta; \pi_* \theta), T) v = 0\) for all \(T\); and
2) \(D_{\theta} \mathbf{vec} K(\theta) | v \neq 0\) if \(v \neq 0\).

Any subspace \(U \subset \mathbb{R}^{d_\Theta}\), with all nonzero \(v \in U\) satisfying the above condition and of maximal dimension (i.e., largest possible satisfying the constraints), is called a (control) information singular subspace. The condition requires that the optimal policy pertaining to the instance \(\theta\) does not persistently excite any instance in a small neighborhood around \(\theta\) in the relative interior of \(U\). By this construction, the dimension of \(U\) captures the number of directions the learner needs to explore beyond those directions which the optimal policy does not explore. The second part of the condition, that \(D_{\theta} \mathbf{vec} K(\theta) | \theta \neq 0\), pertains to the change of variables \(\theta \mapsto K(\theta)\) and relates to the fact that the learner must not necessarily be able to identify \(\theta\) from an optimally regulated trajectory, but rather \(K(\theta)\).

The second condition, presented as follows, is that having bounded regret growth, say on the order \(\sqrt{T}\), effectively constrains the experiments available to the learner on the subspace the optimal policy does not explore. In other words, the condition formalizes the exploration–exploitation tradeoff in LQG in terms of a regret constraint on Fisher information. This is reflected in the proof strategy we pursue in the sequel. Namely, we restrict attention to those policies which attain low regret \(O(\sqrt{T})\). However, these policies necessarily generate experiments with relatively little information content, which in turn implies that the regret of these policies cannot be too small, that is at least \(\Omega(\sqrt{T})\).

**Definition III.2:** Fix an \((U, \varepsilon)\)-locally uninformative instance \((\theta, A(\cdot), B(\cdot), C(\cdot), Q, R, \Sigma W, \Sigma V)\) and a constant \(L > 0\). We say that the instance is \((U, L)\)-information-regret-bounded if for any policy \(\pi\), for all \(T \in \mathbb{N}\), for all \(\theta \in B(\theta, \varepsilon)\) and any matrix \(V_0\) with orthonormal columns spanning \(U\)

\[
 \text{tr} V_0^T I(\pi; \pi, T) V_0 \leq LR_T^2(\theta). \tag{18}
\]

Roughly speaking, **Definition III.2** asks that \(\dim U\)-many eigenvalues of the information matrix, pertaining to a particular policy \(\pi\), satisfy a perturbation bound with respect to the regret of that same policy, \(\pi\). In particular, if the condition holds, any policy with \(O(\sqrt{T})\) regret will yield an information matrix of which the smallest eigenvalue is also \(O(\sqrt{T})\). This should be contrasted with the typical parametric iid design scenario, in which the information matrix scales linearly with the samples.

The conditions given in Definitions III.1 and III.2 reveal the key elements needed to prove a regret lower bound on the order of magnitude \(\sqrt{T}\), as is done in Theorem IV.1. However, the question remains as to which systems these conditions actually apply. To this end, we spend the remainder of this section demonstrating that the conditions given in Definitions III.1 and III.2 are far from vacuous. We prove in Section III-B that a large class of SF systems satisfy both Definition III.1 and III.2. For instance Lemma III.2 together with Proposition III.2 proves that almost any SF system with both \(A\) and \(B\) completely unknown satisfies these conditions. The corresponding nullspace is rather more difficult to characterize for partially observed systems, and we postpone our discussion of these to Section V. However, at a high level the proof approach detailed as follows is still valid: one needs to characterize the parameter directions that the optimal policy does not persistently excite and then show that these unexplored directions are necessary for correctly identifying the optimal policy.\(^2\)

\(^2\)The notion of under-explored parameter direction becomes more subtle for partially observed systems for two reasons: nonuniqueness of realization and the fact that the policy itself has an internal state.
B. Low Regret Experiments

We now initiate our study of what we informally refer to as low regret experiments. As mentioned above, the main idea is that if the optimal policy $K(\theta)$ does not provide sufficient exploration, its application to the system yields a degenerate information matrix. The next step is to note that any controller with bounded regret—which can be thought of as the norm of a particular controller to the optimal controller—cannot yield a particularly good experiment either. While such an experiment is not necessarily singular, it should at least be ill-conditioned. It will be convenient to first calculate the Fisher information of our “experiments.”

Lemma III.1: Fix $T \in \mathbb{N}$. Suppose $C = I_{d_{x}}$ and $V_{t} = 0$ for all $t$. The Fisher information under any policy $\pi$ is given by

$$I(\pi; \theta, T) = \mathbb{E}_{\theta}^{T-1} \sum_{t=0}^{T-1} \mathbb{E}_{\theta}^{T-1} \sum_{t=0}^{T-1} D_{\theta} \mathbb{E}_{\theta}^{T-1} \sum_{t=0}^{T-1} D_{\theta} [A(\theta)X_{t} + B(\theta)U_{t}]^T \times \Sigma_{W}^{-1} D_{\theta}[A(\theta)X_{t} + B(\theta)U_{t}], \quad (19)$$

Proof: Follows immediately by the chain rule for Fisher information and the conditional dependence structure $X_{t+1}|(\text{AUX}, X_{0}, \ldots, X_{t}) \sim N(A(\theta)X_{t} + B(\theta)U_{t}, \Sigma_{W})$.

For more details, see the arxiv version of this manuscript. $\blacksquare$

To make the dependence on $[A(\theta) B(\theta)]$ more explicit, we may rewrite (19) by vectorizing

$$I(\theta; \pi, T) = \mathbb{E}_{\theta}^{T-1} \sum_{t=0}^{T-1} D_{\theta} \mathbb{E}_{\theta}^{T-1} \sum_{t=0}^{T-1} D_{\theta} [A(\theta)B(\theta)]^T \times \Sigma_{W}^{-1} \[Z_{t}Z_{t}^T \otimes \Sigma_{W}^{-1}]$$

where $Z_{t}^T = [X_{t}^T \ U_{t}^T]$.

Note that for the simple parametrization $\text{vec}[A(\theta) B(\theta)] = \theta$, the Jacobian $D_{\theta} \mathbb{E}_{\theta}^{T-1} \sum_{t=0}^{T-1} D_{\theta} [A(\theta) B(\theta)]$ is equal to the identity matrix $I_{d_{x}}$. In this case, (20) is proportional to the covariance matrix used in the “denominator” of the least squares estimator. This is satisfying, as this means that our results imply that if the least squares estimator becomes ill-conditioned, there is little else that can be done. A more direct consequence of the representation (20) is that an algebraic condition for uninformativeness is straightforward to derive.

Proposition III.1: The instance $(\theta, A(\cdot), B(\cdot), Q, R, \Sigma_{W})$ is $(\mathcal{U}, \varepsilon)$-locally uninformative if and only if for every $v \in \mathcal{U} \setminus \{0\}$ and every $\tilde{\theta} \in B(\theta, \varepsilon)$

$$v \in \ker \left[ D_{\theta} \mathbb{E}_{\theta}^{T-1} \sum_{t=0}^{T-1} D_{\theta} [A(\tilde{\theta}) B(\tilde{\theta})] \right]^T \left[ H(\theta)H^T(\theta) \otimes \Sigma_{W}^{-1} \right]$$

$$D_{\theta} \mathbb{E}_{\theta}^{T-1} \sum_{t=0}^{T-1} D_{\theta} [A(\tilde{\theta}) B(\tilde{\theta})]$$

$$v \notin \ker D_{\theta} \mathbb{E}_{\theta}^{T-1} \sum_{t=0}^{T-1} D_{\theta} [A(\tilde{\theta}) B(\tilde{\theta})] \quad (21)$$

where

$$H(\theta) = \begin{bmatrix} I_{d_{x}} \colon K(\theta) \end{bmatrix}.$$ 

Proof: Write, for each $t$ and $\tilde{\theta} \in B(\theta, \varepsilon) \cap \mathcal{U},$

$$D_{\theta} [A(\tilde{\theta}) B(\tilde{\theta})] Z_{t} = \left[D_{\theta} \mathbb{E}_{\theta}^{T-1} \sum_{t=0}^{T-1} D_{\theta} [A(\tilde{\theta}) B(\tilde{\theta})] Z_{t}\right]^T \Sigma_{W}^{-1} D_{\theta} [A(\tilde{\theta}) B(\tilde{\theta})] Z_{t}$$

and the result is established. $\blacksquare$

In Proposition III.2, we specialize the above result to the parametrization in which $\text{vec}[A(\theta) B(\theta)] = \theta$.

1) Information Comparison: We next turn our attention to establishing that uninformative SF systems satisfy the information-regret-boundedness property.

Lemma III.2: Suppose $C = I_{d_{x}}$ and $V_{t} = 0$ for all $t$ and fix a $(\mathcal{U}, \varepsilon)$-locally uninformative instance $(\theta, A(\cdot), B(\cdot), Q, R, \Sigma_{W})$. For every $\theta' \in B(\theta, \varepsilon)$ and every $T \in \mathbb{N}$

$$\mathbb{E}_{\theta}^{T-1} \sum_{t=0}^{T-1} D_{\theta} \mathbb{E}_{\theta}^{T-1} \sum_{t=0}^{T-1} D_{\theta} [A(\tilde{\theta}) B(\tilde{\theta})]$$

$$\leq \mathbb{E}_{\theta}^{T-1} \sum_{t=0}^{T-1} D_{\theta} \mathbb{E}_{\theta}^{T-1} \sum_{t=0}^{T-1} D_{\theta} [A(\tilde{\theta}) B(\tilde{\theta})]$$

$$\leq \left(\sup_{\theta \in B(\theta, \varepsilon)} \left\| D_{\theta} [A(\tilde{\theta}) B(\tilde{\theta})] \right\|_{\infty} \right) \times \left(\left\| (B^T P(\theta) B + R)^{-1} \right\|_{\infty} \right)$$

where as before, the columns of $V_{0}$ span $\mathcal{U}$.

In other words, regret bounds information and uninformative SF systems are $L$-information-regret-bounded with $L = \sup_{\theta \in B(\theta, \varepsilon)} \left\| D_{\theta} [A(\tilde{\theta}) B(\tilde{\theta})] \right\|_{\infty} \times \left(\left\| (B^T P(\theta) B + R)^{-1} \right\|_{\infty} \right)$.

While the full proof of Lemma III.2 can be found in Appendix A3, the intuition for the following proof is as follows. Both the regret $R_{\pi}^{T}(\theta)$ and the Fisher information $I(\pi; \theta, T)$ are expectations of quadratic forms in the variables $X_{t}, U_{t}, t = 0, \ldots, T-1$. Knowing that the optimal policy $\pi_{*}$ renders the $I(\pi_{*}; \theta, T)$ singular, we can control the small eigenvalues of $I(\pi; \theta, T)$ in terms of the regret of that policy by a simple Taylor approximation. In other words, we regard the regret of a policy simply as a measure of its deviation from $\pi_{*}$ and incorporate this as a constraint on experiment design.

2) Unstructured Uncertainty: In literature, much attention has been given to the case in which both $A$ and $B$ are completely unknown. This corresponds to the parametrization $\text{vec}[A(\theta) B(\theta)] = \theta$. Our characterization of the
nullspace of the Fisher information takes a particularly simple form for this parametrization.

**Proposition III.2:** Suppose that vec \([A(\theta) \ B(\theta)] = \theta\). Suppose further that \(\det(A + BK) \neq 0\). Then, the information singular subspace \(\mathcal{U}\) is unique, is equal to \(\ker HH^T \otimes \Sigma_W^{-1}\), and has dimension

\[
\dim \mathcal{U} = d_X d_U
\]

where \(H\) is as in (22).

**Remark III.1:** If the system realization \((A, B, \sqrt{\mathcal{U}})\) is minimal and \(B\) has full column rank then by [38, Lemma 3.4] \(\det(A + BK) \neq 0\) is equivalent to \(\det A \neq 0\).

**Proof:** Since \(HH^T \in \mathbb{R}^{(d_X + d_A) \times (d_X + d_A)}\) is the outer product of two tall matrices, with an identity of size \(d_X\) in the first, top-left, block, we have \(\dim \ker HH^T = d_U\). Moreover,

\[\ker HH^T = \{(x, u) \in \mathbb{R}^{d_X + d_A} : x = -K^T u\} \]

so that for any \(w \in \mathbb{R}^{d_A}\), any vector, \(\theta\) of the form \((u \in \mathbb{R}^{d_U})\)

\[
\theta = \begin{bmatrix} -K^T u \\ u \end{bmatrix} \otimes w = \begin{bmatrix} (K^T u) \otimes w \\ u \otimes w \end{bmatrix}
\]

satisfies \(\theta \in \ker HH^T \otimes \Sigma_W^{-1}\). We note that the dimension of the span of such \(\theta\) is \(d_X d_A\), since there are no constraints on the choice of \(u\) and \(w\). Moreover, all such \(\theta\) satisfying (24) can be obtained as the image of the composition of the vectorization operator with \(\Delta \mapsto [-\Delta K \ \Delta] \in \mathbb{R}^{d_X \times (d_X + d_A)}\). To see this, write

\[
\text{vec} \begin{bmatrix} -\Delta K \\ \Delta \end{bmatrix} = \begin{bmatrix} \text{vec} \Delta K \\ \text{vec} \Delta \end{bmatrix} = \begin{bmatrix} (K^T \otimes I) \text{vec} \Delta \\ \text{vec} \Delta \end{bmatrix}
\]

(25)

so that identification follows by setting \(\text{vec} \Delta = u \otimes w\) in (24).

We recall [3, Lemma 2.1] (see also [20]) which establishes that

\[
\frac{d}{da} K(A - a\Delta K(A, B), B + \Delta) \bigg|_{a=0} = -(R + B^T PB)^{-1} D^T P(A + BK)
\]

(26)

where we have allowed ourselves some abuse of notation in the obvious identification of \(K(A, B) = K(\theta)\) to ease the translation from [3]. Now, what is important is that, provided that \(A + BK\) is nonsingular, (26) is nonzero for all nonzero \(\Delta \in \mathbb{R}^{d_X \times d_A}\), and \(D_\theta \text{vec} K(\cdot)\) has nonzero action on \(\theta\) as in (24). Hence combining (24) and (25) with (26) implies that \(\dim \mathcal{U} \geq d_X d_A\). However, this is maximal since the rank of \(HH^T \otimes \Sigma_W^{-1}\) is \(d_X^2\). Hence \(\dim \mathcal{U} = d_X d_A\) and the subspace \(\mathcal{U}\) is in fact unique (it consists of the entire kernel of the Fisher information).

In other words, what we have shown is the orthogonality of the two nullspaces defined in Proposition III.1 when specialized to the LQR setup with unknown and unstructured \(A\) and \(B\) matrix. It is interesting to note that we arrive at the variation of parameters \([A - \Delta K \ B + \Delta]\) after the change of coordinates (24) and (25) as a consequence of checking parameter variations that lead to a degenerate information matrix, whereas the authors in [3] arrived at the same variation by directly considering variations which generate indistinguishable trajectories. Of course, these perspectives are nearly equivalent, as the singularity of Fisher information implies the (local) indistinguishability of the distributions of the trajectories.

Moreover, Proposition III.2 is also related to a much earlier observation of Polderman [38]. He established the necessity of identifying the true parameter \(\theta\) to identify \(K(\theta)\) which is mirrored in our result. We show that no elements in the nullspace of Fisher information at \(\theta\) are in the nullspace of the derivative of \(K(\theta)\). In other words, small variations in the parameter space with singular information under the optimal policy yield small variations in optimal policy. Proposition III.2 can thus be seen as the local analogue of Polderman’s identifiability result.

**IV. REGRET LOWER BOUNDS FOR STATE FEEDBACK SYSTEMS**

We now state our main lower bound for SF systems in asymptotic form. The proof actually yields a nonasymptotic lower bound which can be found in the arxiv version of this article.

**Theorem IV.1:** Assume that \(A_1\) and \(A_2\) apply to the system \((\theta, A(\cdot), B(\cdot), Q, R, \Sigma_W)\). We further assume that this instance is \(\epsilon\)-locally uninformative for some \(\epsilon > 0\) and \((\mathcal{U}, L)\)-information-regret-bounded. Let \(d \in \{1, \ldots, \dim \mathcal{U}\}\). There exists a matrix \(W_0\) with \(d\) orthonormal columns which all lie in \(\mathcal{U}\) such that for any admissible policy \(\pi\) and any \(\alpha \in (0, 1/4)\)

\[
\liminf_{T \to \infty} \sup_{\theta \in B(\theta, T^{-\alpha})} \mathbb{R}^{\mathcal{U}}(\theta') \geq \frac{1 + \dim \mathcal{U} - d}{8L} \times \left( \text{tr} \left( (\Sigma_X^-)(\theta) \otimes (B^T P(\theta) B(\theta) + R) \right) \right)^{1/2}
\]

(28)

where

\[
\Sigma_X^- = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}^{\pi, (\theta)}_{t} [X^T_t X_t^T].
\]

We now comment on the hardness result (28).

1) First and foremost, logarithmic regret is impossible under the hypotheses: the conjunction of uninformativeness and information-regret-boundedness implies that there is insufficient curvature near the optimal instance making significant exploration necessary. This leads to regret on the order of magnitude \(\sqrt{T}\).

2) By construction, the columns of \(W_0\) are orthogonal to the nullspace of \(D_\theta \text{vec}(K(\theta))\). Hence, the trace appearing in the square root in (28) is nonzero (with dimensional dependence \(d\)).

3) There is considerable flexibility in the parametrization. This allows us to “concentrate” the lower bound on particular system parameters which we might expect to.
be particularly hard to learn. We will pursue this theme further in the next chapter.

4) The lower bound is proportional to \( \sqrt{\frac{\dim U}{L}} \). This factor captures the tension between exploration and exploitation. As noted before \( \dim U \) captures the number of directions not excited by the optimal policy while the term \( L \) captures the sensitivity of the cost to exploration in these directions.

5) The factors \( \Sigma_X^* \) and \( P \) can be thought of as to capture the control-theoretic hardness of the particular instance under consideration. These will be large if the optimal policy operates near marginal stability.

6) The term \( D_\theta \vec{\text{vec}}(K(\theta)) \) does not have any intrinsic interpretation. It is simply a Jacobian term arising in the implicit change of variables appearing as a consequence of our choice of parameter geometry before applying Theorem A.1. Put differently, this term arises since we seek to estimate \( K(\theta) \) and not \( A(\theta) \) or \( B(\theta) \) which are the matrices that we have parametrized.

**Improving the Lower bound of [3]:** In the setting in which both \( A \) and \( B \) are completely unknown \( \vec{\text{vec}}[A(\theta) \ B(\theta)] = \theta \), we have the following result, recovering an earlier result of [3] but with improved constants.

**Corollary IV.1:** Assume that \( A1 \) and \( A3 \) hold for the fixed tuple \( (A, B) \) with corresponding optimal policy \( K \), Riccati matrix \( P \), and state covariance matrix \( \Sigma_X \) [recall (29)]. Then, for every \( \alpha \in (0, 1/4) \) and any admissible policy \( \pi \)

\[
\lim_{T \to \infty} \inf_{\|A' - A\|_2, B'} \sup \frac{R_T^\pi(A', B')}{\sqrt{T}} \geq \frac{d_U \lambda_{\min}(\Sigma_W) \text{tr} \left( (P[\Sigma_X^* - I_{d_x}]P) \right)}{d_X (1 + \|K K^T\|_\text{op})}
\]

for some universal positive constant \( c > 0 \). Under the same hypotheses we also have that any admissible policy \( \pi \) satisfies

\[
\lim_{T \to \infty} \inf_{\|A' - A\|_2, B'} \sup \frac{R_T^\pi(A', B')}{\sqrt{T}} \geq c' \sqrt{d_X d_U^2}
\]

with \( c' > 0 \). The dimensional dependency in (31) is optimal [3]. However, comparing with [3] our lower bound exhibits improved scaling in system-theoretic constants. This is especially true of (30). Their bound scales as \( 1/\|P^2(\theta)\|_\text{op} \) which becomes small if the optimal controller is operates near marginal stability. By contrast, our bound captures the intuition that systems which are hard to optimally regulate are also hard to learn to optimally regulate. To appreciate this contrast, consider for instance an open-loop unstable scalar system

\[
X_{t+1} = aX_t + bU_t + W_t
\]

with \( |a| \geq 1 \). If we instantiate our lower bound and let \( |b| \to 0 \), the right-hand side of (30) tends to positive infinity (to see this, use \( K^T R K \preceq P \)). By contrast, the corresponding lower bound in [3, Th. 1]—which is proportional to \( 1/\|P\|_{\text{op}}^2 \)—tends to 0—see also Fig. 1.

**V. Extension to Partially Observed Systems**

In this section, we seek to understand the hardness of adaptive LQR in the partially observed setting (1) and (2). We will see that there are two new failure modes that arise due to poor observability of the inputs and due to poor observability of the state. Derivations for this part are analogous to the SF setting and have been relegated to the arxiv version of this manuscript.

**A. Easy-to-Analyze Family of Systems**

It is more difficult to analyze the reduction of Lemma II.2 when the optimal filter is nontrivial. Here, we will circumvent this issue by considering a particular class of systems in which the effect of filtering can be separated from that of control in the context of regret minimization. Let \( d_X \) be divisible by 3 and consider

\[
A(\theta) = \begin{bmatrix} A_{11} & 0 & 0 \\ I_{d_X/3} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B(\theta) = \begin{bmatrix} 0_{d_{x/3} \times d_{x/3}} \\ B_2(\theta) \end{bmatrix}
\]

\[
C(\theta) = \begin{bmatrix} C_{11} & 0 & 0 \\ 0 & 0 & C_{23} \end{bmatrix}
\]

where \( A_{11} \in \mathbb{R}^{d_{x/3} \times d_{x/3}} \) is a fixed stable matrix, \( C_{11}, C_{23} \in \mathbb{R}^{d_{x/3} \times d_{x/3}} \) are also fixed and \( 0 \in \mathbb{R}^{d_{x/3} \times d_{x/3}} \), such that the only uncertainty in the parameters is involved in actuation (via \( B_2(\theta) \in \mathbb{R}^{d_{x/3} \times d_{x/3}} \)). For simplicity fix \( \lambda > 0 \) and set \( Q = I_{d_X} \) and \( R = \lambda I_{d_U} \). To simplify the exposition, we will also set \( \theta = \)
vec $B_2(\theta)$. Let us also assume that $\Sigma_W > 0$ and $\Sigma_V > 0$ have the following block structure:

$$\Sigma_W = \begin{bmatrix}
\Sigma_{W_1} & 0 & 0 \\
0 & \Sigma_{W_2} & 0 \\
0 & 0 & \Sigma_{W_3}
\end{bmatrix}$$

$$\Sigma_V = \begin{bmatrix}
\Sigma_{V_1} & 0 \\
0 & \Sigma_{V_2}
\end{bmatrix}$$

and that as before the sequence $\{W_i\}_{i \in \mathbb{N}}$ and $\{V_i\}_{i \in \mathbb{N}}$ are mutually independent and iid Gaussian.

This gives rise to the following family of linear dynamical systems:

$$
\begin{bmatrix}
X_{t+1}^1 \\
X_{t+1}^2 \\
X_{t+1}^3
\end{bmatrix} = \begin{bmatrix}
A_{11} & 0 & 0 \\
I_{d_x/3} & 0 & 0 \\
0 & 0 & C_{23}
\end{bmatrix} \begin{bmatrix}
X_t^1 \\
X_t^2 \\
X_t^3
\end{bmatrix} + \begin{bmatrix}
\theta \\
B(\theta) \\
W_t^3
\end{bmatrix} U_t + \begin{bmatrix}
W_t^1 \\
W_t^2 \\
W_t^3
\end{bmatrix}
$$

$$
\begin{bmatrix}
Y_t^1 \\
Y_t^2
\end{bmatrix} = \begin{bmatrix}
C_{11} & 0 & 0 \\
0 & 0 & C_{23}
\end{bmatrix} \begin{bmatrix}
X_t^1 \\
X_t^2 \\
X_t^3
\end{bmatrix} + \begin{bmatrix}
V_t^1 \\
V_t^2
\end{bmatrix}.
$$

(34)

Notice that the first (block-)coordinate of $Y$ is a noisy observation of the first coordinate of the state: $Y_t^1 = C_{11}X_t^1 + V_t^1$. The second coordinate of $Y$ is simply a noisy observation of the input action $Y_t^2 = C_{23}B(\theta)U_{t-1} + C_{23}W_{t-1}^2 + V_t^2$. Let us also note that for A1 to hold—i.e., for $(A(\theta), C(\theta))$ to be detectable—it is sufficient that $(A_{11}, C_{11})$ is detectable. Since $A_{11}$ is stable by assumption, A1 is immediate. However, (34) is not observable.

The structure (34) means that the regret has the following simple representation.

**Lemma V.1:** Assume that A1–A3 holds. For every instance of the form (34)

$$R^2_T(\theta) = \sum_{t=0}^{T-1} \left( E \left[ (U_t - K_1(\theta)E^0_t[X_t^1|Y_t])^T \right. \right.$$

$$\left. \times (2B^T_t(\theta)B(\theta) + \lambda I_{d_\theta}) (U_t - K_1(\theta)E^0_t[X_t^1|Y_t]) \right]\right)^T$$

where

$$K_1(\theta) = -(2B^T_1(\theta)B(\theta) + \lambda I_{d_\theta})^{-1} B^T_1(\theta).$$

(36)

The next step is to notice that $E^0_t[X_t^1|Y_t]$ is independent of both $\theta$ and $\pi$. To see this, notice that as long as A1 and A3 hold, the evolution of $E^0_t[X_t^1|Y_t]$ is given by the following filtering equation:

$$\hat{X}_{t+1}^1 = A_{11}\hat{X}_t + F_t[Y_{t+1} - C_{11}A_{11}\hat{X}_t]$$

(37)

where we recall that filter gain $F_t$ is characterized by

$$S_1 = A_{11}S_{11}A_{11}^T$$

$$- A_{11}S_{11}C_{11}^T(C_{11}S_{11}C_{11}^T + \Sigma_{V_t})^{-1}C_{11}S_{11}A_{11}^T + \Sigma_{W_t}$$

so that

$$F_t = S_{11}C_{11}^T(C_{11}S_{11}C_{11}^T + \Sigma_{V_t})^{-1}. \quad (38)$$

All the quantities appearing in (37)–(39) are independent of both $\theta$ and $\pi$. Invoking Lemma II.2 we obtain the following relaxation.

**Lemma V.2:** Fix $\theta \in \mathbb{R}^{d_\theta}$ and $\epsilon > 0$. Assume that A1–A3 hold for every $\theta \in B(\theta, \epsilon)$. Fix a function $\tau : \mathbb{Z} \to \mathbb{Z}$ with $\tau(t) \geq t$ and a smoothly compactly supported prior $\mu \in C_c^\infty(B(\theta, \epsilon))$. Then, for every instance of form (34)

$$\sup_{\theta \in B(\theta, \epsilon)} R^2_T(\theta)' \geq \sum_{t=0}^{T-1} \mathbb{E}_{\theta_{t+1}} E_{\theta_{t+1}} \mathbb{E}_{\theta_{t+1}}$$

$$\mathbb{E} \left[ \left( N_\pi [K_1(\theta)|Y_{t+1} - K_1(\theta)] \right) \left( \hat{X}_{t+1}^1(X_{t+1}^1)^T \right) \right]$$

(40)

for any matrix $N_\pi$ satisfying $N_\pi \leq 2B^T_1(\theta_2)B_2(\theta_2) + \lambda I_d_\theta$ for every $\theta_2 \in B(\theta, \epsilon)$, where $K_1$ is as in (36) and where $X_{t+1}^1$ is given by (37).

### B. Fisher Information Bounds

The following program parallels the development for the SF setting. That is, we seek to characterize the information matrix for policies with low regret $(O(\sqrt{T}))$. We begin by providing an expression analogous to (19) for the system (34).

**Lemma V.3:** Fix $T \in \mathbb{N}$ and consider the system (34). The Fisher information under any policy $\pi$ is given by

$$I(\pi; \theta, T) = E^0_T \sum_{t=0}^{T-1} \left( D_\theta [C_{23}B(\theta)U_t] \right)^T$$

$$\left(C_{23} \Sigma_{W_3} C_{23}^T + \Sigma_{V_2} \right) \left( D_\theta [C_{23}B(\theta)U_t] \right).$$

(41)

Equipped with (41), uninformativeness (recall Definition III.1) is readily characterized as follows.

**Proposition V.1:** For any $\theta$, the instance (34) is $(U, \epsilon)$-locally uninformative if for every $v \in U \setminus \{0\}$

$$v \notin \ker \left( K_1(\theta)K_1^T(\theta) \otimes C_{23} C_{23}^T \right) \setminus \ker D_\theta \text{vec } K_1(\theta).$$

(42)

The proof of Proposition V.1 is almost identical to that of Proposition III.1 and is thus omitted.

Let us now investigate the second part of condition (42).

**Lemma V.4:** The nullspace $\ker D_\theta \text{vec } K_1(\theta)$ is equal to the set of solutions $v$ of

$$v = 2\Pi^{-1} \left( [K_1^T B_2 \otimes I_{d_\theta}] + [K_1^T \otimes B_2^T] \right) v.$$ 

(43)

where $K_1$ and $B_2$ are evaluated at $\theta$ and where the permutation matrix $\Pi$ maps vec $M$ to vec $M^T$.

The next proposition exploits the geometric description of $\ker D_\theta \text{vec } K_1(\theta)$ provided in Lemma V.4 to give a more direct characterization of uninformativeness.

**Proposition V.2:** Fix $\theta \in \mathbb{R}^{d_\theta}$, $\epsilon > 0$ and suppose that $K_1(\theta)K_1^T(\theta)$ is singular. Set

$$U = \{u \otimes w \in \mathbb{R}^{d_\theta} : u \in \ker K_1(\theta)K_1^T(\theta), w \in \mathbb{R}^{d_x/3} \}. \quad (44)$$

The instance (34) is \((U, \varepsilon)\)-uninformative.

Note that \(\ker K_1(\theta)K_1^T(\theta)\) is nontrivial as soon as \(d_U > d_x/3\).

\textbf{a) Information Comparison:} The final preliminary lemma required will be used in Lemma III.2. We again show that regret bounds the small singular values of the Fisher information.

\textbf{Lemma V.5:} Fix \(\theta \in \mathbb{R}^{d_{\exists}}, \varepsilon > 0\), consider the instance (34) and suppose that \(K_1(\theta)K_1^T(\theta)\) is singular. Set

\[
U = \{ u \otimes w \in \mathbb{R}^{d_{\exists}} : u \in \ker K_1(\theta)K_1^T(\theta), w \in \mathbb{R}^{d_x/3}\}.
\]

(45)

For every \(\theta' \in B(\theta, \varepsilon)\) and every \(T \in \mathbb{N}\)

\[
\text{tr} V_0^T \mathbb{I}(\pi; \theta', T)V_0 \leq \text{tr}(C^T_{23}(C_{23}\Sigma_{W_1}C^T_{23} + \Sigma_{V_2})^{-1}C_{23}) \times \| (2B^T_2B_2 + \lambda I_{d_0})^{-1} \|_\text{op} R_T^2(\theta)
\]

where as before, the columns of \(V_0\) span \(U\).

\textbf{C. Hardness Result for LQG:} Equipped with the characterization of the (restricted) nullspace of Propositions V.1 and V.2 and the information-regret bound of Lemma V.5 we are ready to establish the following analogue of Theorem IV.1. For brevity, we only state the asymptotic version.

\textbf{Corollary V.1:} Consider the system (34). Suppose A1 and A2 apply for every \(\theta' \in B(\theta, \varepsilon)\). Let \(d \in \{1, \ldots, \dim U\}\), where \(U\) is given by (44). Let \(L \triangleq \text{tr}(C^T_{23}(C_{23}\Sigma_{W_1}C^T_{23} + \Sigma_{V_2})^{-1}C_{23})\| (2B^T_2B_2 + \lambda I_{d_0})^{-1} \|_\text{op} \).

There exists a matrix \(W_0\) with \(d\) orthonormal columns which all lie in \(U\) such that for any admissible policy \(\pi\) and any \(\alpha \in (0, 1/4)\)

\[
\liminf_{T \to \infty} \sup_{\theta' \in B(\theta, T^{-\alpha})} \frac{R_T^2(\theta')}{\sqrt{T}} \geq \frac{1 + \dim U - d}{8L} \times \text{tr} \left((\Sigma_X^1 \otimes (2B^T_2(\theta)B_2(\theta) + R)) \right)
\]

\[
\times (D_0 \text{vec}(K_1(\theta))W_0W_0^T(D_0 \text{vec}(K_1(\theta))^T)^{1/2}
\]

(46)

where \(\Sigma_X^1 \triangleq \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} E X_t^1 (X_t^1)^T\).

In light of Lemmas V.2 and V.5, the proof is almost identical to that of Theorem IV.1 and is thus omitted. The result can be significantly simplified if stated asymptotically.

While the lower bound (46) is interpreted much like its fully observed analogue (28) (in particular, see the ensuing discussion), there are two new failure modes that arise. First, as \(\| C_{23} \|_\text{op}\) tends to zero, the lower bound (46) diverges. In this case, the learner faces vanishing information available to them about \(B_2\), which is needed to regulate the system. Second, it is no longer the “state” covariance that enters the bound (46), but the \textit{filtered} state covariance \(\Sigma_X^1\), which depends both on the stability and detectability of the first mode. In particular, this term does not diverge as stability is lost \(\rho(A_{11}) \to 1\) but rather if \(\rho(A_{11}) \to 1\) and observability is lost (e.g., \(\| C_{11} \|_\text{op} \to 0\)). In other words, regret minimization becomes hard as observability of unstable modes is lost.

It is also interesting to note that it has been proven by [30] that logarithmic regret against the best possible, in hindsight, persistently exciting controller\(^3\) whenever the covariance of the measurement noise is positive definite, i.e., \(\Sigma_V \succ 0\). Unfortunately, it is not clear whether the optimal LQG controller is persistently exciting under these hypotheses and so the notion of regret in [30] may differ from the standard one. With this in mind, we thus also prove a negative result, showing that \(\Sigma_V \succ 0\) is not sufficient for logarithmic regret without further assumption. Finally, we point out that the system used in our construction is not minimal, in the sense that (34) is not observable. By contrast, known upper bounds apply to controllable and observable systems [29], [30]. Thus, while our bounds show that logarithmic regret is \textit{not} always possible, even with full rank output noise, it does not rule out this possibility in the exact setting of [30].

\section{Appendix}

\textbf{Proofs for the State Feedback Setting}

\textbf{A. Proof of Theorem IV.1}

Throughout the proof we use \(\varepsilon = T^{-\alpha}\) and set \(T\) sufficiently large such that the quantities \(P(\cdot)\) and \(K(\cdot)\) are continuous over \(B(\theta, \varepsilon)\), which is guaranteed as soon as \(T\) is sufficiently large by [22, Proposition 1].\(^2\) Moreover, we may assume that \(\text{sup}_{\theta' \in B(\theta, \varepsilon)} R_T^2(\theta') \leq C\sqrt{T}\) for a constant \(C > 0\) to be determined later.

Before we proceed with the main argument of the proof, we also choose the matrix \(W_0 \in \mathbb{R}^{d_{x} \times d}\) as \(W_0 = V_0I_0\), where \(I_0 \in \mathbb{R}^{\dim U \times \dim U}\) is a free variable and as before the columns of \(V_0\) span \(U\). Note that by construction the columns of \(W_0\) are linear combinations of the columns of \(V_0\) and hence elements of \(U\). We also introduce a smooth, compactly supported prior \(\mu\) on \(\{ \theta + W_0v : \| v \| \leq \varepsilon \}\). In the sequel, we write \(D_0\) for Jacobian in \(\theta\)-space and \(D_\theta\) for Jacobian in \(v\)-space (e.g., of the composite function \(\text{vec } K(\theta + W_0v)\)). By restricting \(I_0\) to be norm-preserving and then optimizing over \(I_0\), the variational characterization of eigenvalues yields that

\[
\lambda_1 (W_0^T \text{vec}(\Theta, \pi, T)W_0 + \lambda_{\max} (J(\mu)) 
\]

\[
\leq \lambda_1 (W_0^T \text{vec}(\Theta, \pi, T)W_0 + \lambda_{\max} (J(\mu)) 
\]

(47)

The proof now proceeds by invoking Lemma II.2 and (16) to relax the suprenum in the theorem statement to an expectation,
and in particular to find that
\[
\inf_{\pi} \sup_{\theta' \in B(\theta, \epsilon)} R^\pi_T(\theta) \geq \sum_{t=0}^{T-1} E_{\theta' \sim \mu} E_0 \text{tr} \left[ N_s (E[K(\Theta)]) Y_{t-1} - K(\Theta) \right] \\
\times \left( X_t X_t^T \right) (E[K(\Theta)]) Y_{t-1} - K(\Theta))^T \right].
\]

Introduce further a positive-semidefinite matrix \( \Psi \) and the event
\[
\mathcal{E} \equiv \left\{ \sum_{t=0}^{T-1} X_t X_t^T \geq \Psi T \right\}.
\]

By definition of \( \mathcal{E} \)
\[
\inf_{\pi} \sup_{\theta' \in B(\theta, \epsilon)} R^\pi_T(\theta) \geq T E \text{tr} \left( \left( E[\sqrt{N_s} K(\Theta) \sqrt{\Psi}] Y_{t-1} - \sqrt{N_s} K(\Theta) \sqrt{\Psi} \right) 1_{\mathcal{E}} \right)^2.
\]

We now continue by writing the norm (inner product) in (48) as the trace of an outer product
\[
(48) = T E \text{tr} \left( \left( E[\sqrt{N_s} K(\Theta) \sqrt{\Psi}] Y_{t-1} - \sqrt{N_s} K(\Theta) \sqrt{\Psi} \right) 1_{\mathcal{E}} \right)^2.
\]

Now, since \( E[K(\Theta)] Y_{t-1} \) is an estimator of \( K(\Theta) \), we may apply Theorem A.1
\[
(49) \geq T E \text{tr} \left( \left( \Psi \otimes N_s \right) E \left[ (D_0 K(\Theta) W_0 1_{\mathcal{E}}) \right] \right) \times \left( W_0^T E \left[ (D_0 K(\Theta) 1_{\mathcal{E}})^T \right] \right).
\]

Above, (50) uses the fact that \( \Theta \) is a \( W_0 \)-affine translate of \( \theta \), hence the Jacobian transforms as \( D_0 \text{vec} K(V) = D_0 \text{vec} K(\Theta) W_0 \), where \( V \) is any random variable such that \( \Theta = \theta + W_0 V \). Similar reasoning yields that the information matrix in \( \theta \)-space can be written as \( W_0^T E \left[ (D_0 K(\Theta) 1_{\mathcal{E}})^T \right] \).

Hence, if we combine (50) with (47) we find that
\[
\inf_{\pi} \sup_{\theta' \in B(\theta, \epsilon)} R^\pi_T(\theta) \geq \frac{T(1 + \dim U - d)}{LC \sqrt{T} + \text{tr}(J(\mu))} \times \mathbf{P}^2(\mathcal{E}) \inf_{\theta', \delta' \in B(\theta, \epsilon)} \text{tr} \left( (\Psi \otimes N_s)(D_0 \text{vec}(K(\theta'))) \right) \times W_0 W_0^T (D_0 \text{vec}(K(\theta')))^T.
\]

To finish the proof, we need to control the event \( \mathcal{E} \). We will show that if for some (universal) polynomial function poly
\[
T \geq \sup_{\theta' \in B(\theta, \epsilon)} \text{poly} \left( (g^{d_k}, \|B\|_{\text{op}}, \|A\|_{\text{op}}, \ldots) \right)
\]

where we have omitted dependency on \( \theta' \) in the arguments of poly in (52). We show that \( \mathbf{P}(\mathcal{E}) \geq 1/2 \)—in fact we establish the stronger statement that \( \inf_{\theta} \mathbf{P}_\theta(\mathcal{E}) \geq 1/2 \)—with any \( \Psi \) satisfying
\[
(\Psi \otimes N_s)(D_0 \text{vec}(K(\theta'))) \leq \sum_{j=1}^{T^{1/16}} \left( (A(\theta') + B(\theta') K(\theta'))^j \right) \times \left( \Sigma W - I_{d_s}(\|\Sigma W\|_{\text{op}} T^{-1/4} + T^{-1/4} + T^{-1/8}) \right) \times ((A(\theta') + B(\theta') K(\theta'))^T)^j).
\]

for all \( \theta' \in B(\theta, \epsilon) \) in Section A.11. Notice that
\[
\sum_{j=0}^{T^{1/16}} \left( (A(\theta') + B(\theta') K(\theta'))^j \right) \Sigma W ((A(\theta') + B(\theta') K(\theta'))^T)^j
\]

converges uniformly to \( \Sigma^*_{\chi}(\theta') \) by unrolling the dynamics and using the fact that the spectral radius of \( A(\theta') + B(\theta') K(\theta') \) is strictly separated from 1 for \( \theta' \) in some neighborhood of \( \theta \).

Taking this for granted for now, we now choose
\[
C = \sqrt{(1 + \dim U - d)} \inf_{\theta', \delta' \in B(\theta, \epsilon)} \left( \text{tr} \left( (\Psi \otimes N_s)(D_0 \text{vec}(K(\theta'))) \right) \right)^{1/2}.
\]

Hence with this choice of \( C \) and as long as \( \sqrt{T} \geq \frac{\text{tr}(J(\mu))}{LC} \) we are guaranteed the claimed lower bound. This finishes the proof, since if \( \sup_{\theta' \in B(\theta, \epsilon)} R^\pi_T(\theta) \leq C \sqrt{T} \) does not hold for our choice of \( C \), the statement holds trivially.

1) **Bounding \( \mathbf{P}(\mathcal{E}) \):** Fix any \( \theta' \in B(\theta, \epsilon) \) and denote \( A = A(\theta') \), \( B = B(\theta') \), \( P = P(\theta') \), \( R^\pi_T = R^\pi_T(\theta') \) and \( K = K(\theta') \) to avoid cumbersome notation. 3 The evolution of the sum over \( X_t X_t^T \) under \( \theta' \) is given by
\[
\sum_{t=0}^{T-1} X_t X_t^T = \sum_{j=0}^{m} (A + B K)^j \left[ \sum_{t=0}^{T-1} W_{t-j} W_{t-j}^T \right] ((A + B K)^T)^j + \sum_{j=0}^{m} (A + B K)^j \left[ \sum_{t=0}^{T-1} \text{sym} \left[ B(U_{t-j} + AX_{t-j}) \right] (W_{t-j})^T \right] ((A + B K)^T)^j.
\]

3This convention applies to Appendix A.11 alone.
we know that the instance is 

\[
\sum_{j=0}^{m} (A + BK)^j \left[ \sum_{t=0}^{T-1} \text{sym} \right] [(A + BK)X_{t-j}(B(U_{t-j} - K X_{t-j}))^T] ((A + BK)^T)^j
\]

(54)

where quantities with negative indexes are taken to be zero.

The strategy is now as follows. We wish to show that the first term in (54) is dominant and that it approximates the covariance of the optimal policy. The second term will be bounded by a martingale argument and the third term will be bounded using the hypothesis that \( R_T^2 \leq CV_T^2 \). To this end, we now introduce the following three events:

\[
\begin{align*}
\mathcal{E}_1 & \triangleq \left\{ \exists j : \sum_{t=0}^{T-1} W_{t-j}W_{t-j}^T \not\in \Sigma_W - I_{d_X} \| \Sigma_W \|_{\text{op}} \geq T^{3/4} \right\} \\
\mathcal{E}_2 & \triangleq \left\{ \exists j : \left\| \sum_{t=0}^{T-1} \text{sym} [B(U_{t-j} + AX_{t-j})] (W_{t-j})^T \right\|_{\text{op}} \geq T^{3/4} \right\} \\
\mathcal{E}_3 & \triangleq \left\{ \exists j : \left\| \sum_{t=0}^{T-1} \text{sym}(A + BK)X_{t-1}(B(U_{t-j} - K X_{t-j}))^T \right\|_{\text{op}} \geq T^{7/8} \right\}.
\end{align*}
\]

The idea is to derive bounds on each of these failure-events, which we combine to arrive at a high probability lower bound on (54). Together these yield the desired control of the state covariance. Details are standard and have been relegated to the supplementary (arxiv) version of this article.

## B. Proof of Corollary IV.1

We consider the coordinates

\[
\begin{align*}
\text{vec} A(\theta) & = \text{vec} A - \text{vec}[(\text{vec}^{-1} \theta) K] \\
\text{vec} B(\theta) & = \text{vec} B + \theta.
\end{align*}
\]

By virtue of Proposition III.2 we know that the instance is \( \varepsilon \)-locally uninformative (for every \( \varepsilon > 0 \)) with \( d_0 = \dim U = d_X d_U \). Hence for every \( d \leq d_X d_U \), we have by Theorem IV.1 that there exists \( W_0 \) with \( d \)-many independent elements that satisfy \( \text{span} \ W_0 \subset U \). Note also that for \( d = d_X d_U \) we have \( W_0 W_0^T = I_{d_X} \).

We need to analyze the term

\[
\text{tr} \left( \left[ \Sigma_X(\theta) \right] \otimes (B^T(\theta) P(\theta) B(\theta) + R) \right)
\]

\[
(D_\theta \text{ vec} K(\theta)) W_0 W_0^T (D_\theta \text{ vec} K(\theta))^T.
\]

For this parameterization, appealing to Proposition III.2 and by virtue of (26) (see also Simchowitz and Foster [3]), we have an explicit expression for \( D_\theta \text{ vec} K(\theta) \).

\[
D_\theta \text{ vec} K(\theta) = -((A(\theta) + B(\theta) K(\theta))^T P(\theta)) \otimes (B^T(\theta) P(\theta) B(\theta) + R)^{-1}.
\]

Straightforward calculations involving the trace cyclic property and the Lyapunov identity \( (A + BK)^{\Sigma_X}(A + BK)^T = \Sigma_X - I_{d_X} \) now yield

\[
\text{tr} \left( \left[ \Sigma_X(\theta) \right] \otimes (B^T(\theta) P(\theta) B(\theta) + R) \right)
\]

\[
\times (D_\theta \text{ vec} K(\theta)) W_0 W_0^T (D_\theta \text{ vec} K(\theta))^T)
\]

\[
= \text{tr} \left( \left[ (P(\theta) \Sigma_X(\theta) - I_{d_X}) P(\theta) \right]\otimes (B^T(\theta) P(\theta) B(\theta) + R)^{-1} \right) W_0 W_0^T.
\]

(57)

Using Lemma III.2 we may choose

\[
L = \text{tr}(\Sigma_w^{-1})
\]

\[
\times \left( \sup_{\theta \in B(\theta, \varepsilon)} \| D_\theta [A(\theta) B(\theta)] \|_2 \right) \| (B^T P(\theta) B + R)^{-1} \|_{\text{op}}
\]

with \( \varepsilon \)-arbitrary. Using (56), we have \( \| D_\theta [A(\theta) B(\theta)] \|_2 \leq 1 \lor \| K(\theta) K^T(\theta) \|_{\text{op}} \). Next, note that

\[
\| (B^T P(\theta) B + R)^{-1} \|_{\text{op}} = \frac{1}{\sigma_{\min}(B^T P(\theta) B(\theta) + R)}
\]

that \( \text{tr} \Sigma_w^{-1} \leq \sigma_{\max}(\Sigma_w^{-1}) d_X = \frac{d_X}{\sigma_{\min}(\Sigma_w)} \). Hence, we may take

\[
L \leq \frac{d_X(1 \lor \| K(\theta) K^T(\theta) \|_{\text{op}})}{\sigma_{\min}(\Sigma_w) \times \sigma_{\min}(B^T P(\theta) B(\theta) + R)}.
\]

(58)

By combining (57) with (58) and invoking Theorem IV.1 the result follows.

### C. Proof of Lemma III.2

Fix \( \theta' \in B(\theta, \varepsilon) \) and let the columns of \( V_0 \) span the information singular subspace \( U \). We now introduce a quadratic potential, \( f : \mathbb{R}^{T(d_X d_U + d_0)} \rightarrow \mathbb{R} \), in terms of the dummy variables \( \eta_j \in \mathbb{R} d_X d_U \)

\[
f(\eta_j; T-1) = \text{tr} \left( V_0^T \left( \sum_{j=1}^{T-1} (D_\theta [A(\theta') B(\theta')]) \right)^T \right.
\]

\[
\left. \times (\eta_j - 1 \Sigma^{-1}(\theta)) (D_\theta [A(\theta') B(\theta')]) \right) V_0.
\]

(59)

Observe that \( E_0^T f(Z_0; T-1) = \text{tr} V_0^T \mathbb{I}(\pi; \theta', T) V_0 \). Our next observation is that the restriction of \( f \) to the subspace \( \eta_j = \)
where we have used the fact that \( f \) is convex quadratic function with minimum \( 0 \), attained at all points in \( F_0 \), so that its Taylor-expansion around such a point is just a quadratic form.

By introducing a factor \( I_{du} = (B^T P(\theta) B + R)^{-1}(B^T P(\theta) B + R) \) we obtain

\[
f(\eta_{0:T-1}) \leq \frac{1}{2} \| B^T P(\theta) B + R \|_F \| \nabla^2 \psi_{\eta_{0:T-1}} f \|_F \\
+ \sum_{t=0}^{T-1} \left(\psi_{\eta_{t+1}} - K(\theta) \psi_{\eta_{t0}}\right)^T (B^T P(\theta) B + R) \left(\psi_{\eta_{t+1}} - K(\theta) \psi_{\eta_{t0}}\right).
\]

If we combine the above with the observation that

\[
\| \nabla^2 \psi_{\eta_{0:T-1}} f \|_F \leq \| D_\theta[A(\theta') B(\theta')] \|_F^2 \text{tr}(\Sigma W^{-1})
\]

the proof is concluded. \( \blacksquare \)

D. Fisher Information and van Trees’ Inequality

Fix an observation \( Y \), a random vector taking values in \( \mathbb{R}^n \) with (conditional) density \( p_0(\cdot) = p(\cdot; \theta) \). That is, instead of fixing the parameter \( \theta \), we let \( \Theta \) be a random vector taking values in \( \mathbb{R}^{d_\theta} \) and suppose that it has density \( \mu \). The following two quantities are key to measuring estimation performance of \( \Theta \) from the sample \( Y \) with respect to square loss

\[
I_p(\theta) = \int \left( \begin{array}{c} \nabla \theta p(y|\theta) \\ p(y|\theta) \end{array} \right) \left( \begin{array}{c} \nabla \theta p(y|\theta) \\ p(y|\theta) \end{array} \right)^T p(y|\theta) dy, \quad \text{and}
\]

\[
J(\mu) = \int \left( \begin{array}{c} \nabla \theta \mu(\theta) \\ \mu(\theta) \end{array} \right) \left( \begin{array}{c} \nabla \theta \mu(\theta) \\ \mu(\theta) \end{array} \right)^T \mu(\theta) d\theta.
\]

See [50] for further details about these integrals and their existence.

Let us also introduce a function \( \psi : \mathbb{R}^{d_\theta} \rightarrow \mathbb{R}^n \). The purpose is to relax the setup above slightly, in that we seek to establish lower bounds for estimating \( \psi(\theta) \) instead of just \( \theta \). We impose the following regularity conditions.

B1: \( \mu \in C_c(\mathbb{R}^{d_\theta}); \) the prior is smooth with compact support.

B2: \( p(y|\cdot) \) is continuously differentiable on the domain of \( \mu \) for almost every \( y \).

B3: The score has mean zero; \( \int \frac{\nabla \theta p(y|\theta)}{p(y|\theta)} p(y|\theta) dy = 0 \).

B4: \( J(\mu) \) is finite and \( I_p(\theta) \) is a continuous function of \( \theta \) on the domain of \( \mu \).

B5: \( \psi \) is differentiable on the domain of \( \mu \).

The following theorem is a less general adaption from [46] which suffices for our needs.

Theorem A.1: Fix two random variables \( (Y, \Theta) \sim p(\cdot|\cdot; \mu(\cdot)) \) and suppose that B1–B5 hold. Let further \( \mathcal{E} \) be a \( \sigma(\Theta) \)-measurable event. Then, for any \( \sigma(\Theta) \)-measurable \( \psi \)

\[
\mathbb{E} \left[ (\psi - \psi(\Theta))(\psi - \psi(\Theta))^T 1_{\mathcal{E}} \right] \\
\geq \mathbb{E}[(\psi - \psi(\Theta)) 1_{\mathcal{E}}]^T [\mathcal{E} I_{\mu}(\theta) + J(\mu)]^{-1} \mathbb{E}[(\psi - \psi(\Theta)) 1_{\mathcal{E}}].
\]

Acknowledgment

The authors wish to express their gratitude to Bruce Lee, Anastasios Tsiamis, and Yishao Zhou for helpful suggestions and feedback. This work was completed while Ingvar Ziemann was still at KTH.

References

[1] Y. Yang, K. Caluwaerts, A. Iscen, T. Zhang, J. Tan, and V. Sindhwani, “Data efficient reinforcement learning for legged robots,” in Proc. Conf. Robot Learning, 2020, pp. 1–10.

[2] D. Silver et al., “Mastering the game of go without human knowledge,” Nature, vol. 550, no. 7676, pp. 354–359, 2017.

[3] M. Simchowitz and D. Foster, “Naive exploration is optimal for online LQR,” in Proc. Int. Conf. Mach. Learn., 2020, pp. 8937–8948.

[4] T. L. Lai, “Asymptotically efficient adaptive control in stochastic regression models,” Adv. Appl. Math., vol. 7, no. 1, pp. 23–45, 1986.

[5] Tze Leung Lai and Ching-Zong Wei, “Extended least squares and their applications to adaptive control and prediction in linear systems,” IEEE Trans. Autom. Control, vol. 31, no. 10, pp. 989–906, Oct. 1986.

[6] T. L. Lai and H. Robbins, “Asymptotically efficient adaptive allocation rules,” Adv. Appl. Math., vol. 6, no. 1, pp. 4–22, 1985.

[7] L. Guo, “Convergence and logarithm laws of self-tuning regulators,” Automatica, vol. 31, no. 3, pp. 435–450, 1995.

[8] AA Feldbaum, “Dual control theory. I,” Avtomatika i Telemekhanika, vol. 21, no. 9, pp. 1240–1249, 1960.

[9] AA Feldbaum, “Dual control theory. II,” Avtomatika i Telemekhanika, vol. 21, no. 11, pp. 1453–1464, 1960.

[10] H. A. Simon, “Dynamic programming under uncertainty with a quadratic criterion function,” Econometrica J. Econometric Soc., pp. 74–81, 1956.

[11] K. J. Aström and B. Wittenmark, “On self tuning regulators,” Automatica, vol. 9, no. 2, pp. 185–199, 1973.

[12] R. Kalman, “Design of a self-optimizing control system,” Trans. Amer. Soc. Mech. Engineers, vol. 80, pp. 468–478, 1958.

[13] G. C. Goodwin, P. J. Ramadge, and P. E. Caines, ”Discrete time stochastic adaptive control,” SIAM J. Control Optim., vol. 19, no. 6, pp. 829–853, 1981.

[14] Tze Leung Lai and Ching Zong Wei, “Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems,” Ann. Statist., vol. 10, no. 1, pp. 154–166, 1982.

[15] A. Becker, P. Kumar, and Ching-Zong Wei, “Adaptive control with the stochastic approximation algorithm: Geometry and convergence,” IEEE Trans. Autom. Control, vol. 30, no. 4, pp. 330–338, Apr. 1985.

[16] M. Campi and P. R. Kumar, “Adaptive linear quadratic Gaussian control: The cost-biased approach revisited,” SIAM J. Control Optim., vol. 36, no. 6, pp. 1890–1907, 1998.

[17] Y. Abbasi-Yadkori and C. Szepesvári, “Regret bounds for the adaptive control of linear quadratic systems,” in Proc. 24th Annu. Conf. Learn. Theory, 2011, pp. 1–26.

[18] Y. Ouyang, M. Gagrani, and R. Jain, “Posterior sampling-based reinforcement learning for control of unknown linear systems,” in Proc. 22nd Int. Conf. Artif. Intell. Statist., 2019, pp. 3108–3117.
H. Mania, S. Tu, and B. Recht, “Certainty equivalence is efficient for linear quadratic control,” in *Adv. Neural Inf. Process. Syst.*, vol. 32, pp. 10154–10164, 2019.

A. Cohen, T. Koren, and Y. Mansour, “Learning linear-quadratic regulators efficiently with only $\sqrt{T}$ regret,” in *Proc. Int. Conf. Mach. Learn.*, 2019.

Mohammad Kazem Shirani Farahani, A. Tewari, and G. Michailidis, “Input perturbations for adaptive control and learning,” *Automatica*, vol. 117, 2020, Art. no. 1018950.

M. Abeille and A. Lazaric, “Optimistic exploration in linear-quadratic regulators via Lagrangian relaxation,” in *Proc. Int. Conf. Mach. Learn.*, 2020.

Y. Jedra and A. Proutiere, “Minimal expected regret in linear quadratic control,” in *Proc. Int. Conf. Artif. Intell.*, 2022.

A. Tsiamis and George J. Pappas, “Linear systems can be hard to learn,” in *Proc. 60th IEEE Conf. Decis. Control (CDC)*, 2021.

A. Tsiamis, I. Ziemann, M. Morari, N. Matni, and G. J. Pappas, “Learning to control linear systems can be hard,” in *Proc. Conf. Learn. Theory*, 2022, pp. 3820–3857.

M. Simchowitz, K. Singh, and E. Hazan, “Improper learning for non-stochastic control,” in *Proc. Conf. Learn. Theory*, 2020, pp. 3320–3346.

S. Lale et al., “Logarithmic regret bound in partially observable linear dynamical systems,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 20876–20888.

A. Nemirovski and Y. Z. Tsypkin, “On optimal algorithms of adaptive control,” *Automatica and Telemekhanika*, vol. 12, pp. 64–77, 1984.

M. Raginsky, “Divergence-based characterization of fundamental limitations of adaptive dynamical systems,” in *Proc. 48th Ann. Allerton Conf. Commun., Control, Comput.*, 2010, pp. 107–114.

A. Rantzer, “Concentration bounds for single parameter adaptive control,” in *Proc. Annu. Amer. Control Conf.*, 2018, pp. 1862–1866.

A. Cassel, A. Cohen, and T. Koren, “Logarithmic regret for learning linear quadratic regulators efficiently,” in *Proc. Int. Conf. Mach. Learn.*, 2020.

I. Ziemann and H. Sandberg, “On a phase transition of regret in linear quadratic control: The memoryless case,” *IEEE Control Syst. Lett.*, vol. 5, no. 2, pp. 695–700, Apr. 2021.

M. Gevers and L. Ljung, “Optimal experiment designs with respect to the intended model application,” *Automatica*, vol. 22, no. 5, pp. 543–554, 1986.

W. Lin, P. R. Kumar, and T. I. Seidman, “Will the self-tuning approach work for general cost criteria?,” *Syst. Control Lett.*, vol. 6, no. 2, pp. 77–85, 1985.

J. W. Polderman, “On the necessity of identifying the true parameter in adaptive LQ control,” *Syst. Control Lett.*, vol. 8, no. 2, pp. 87–91, 1986.

H. Hjalmarsson, M. Gevers, and F. D. Bruyne, “For model-based control design, closed-loop identification gives better performance,” *Automatica*, vol. 32, no. 12, pp. 1659–1673, 1996.

F. Pukelsheim, *Optimal Design of Experiments*. Philadelphia, PA, USA: SIAM, 2006.

Thomas I. Rothenberg, “Identification in parametric models,” *Econometrica: J. Econometric Soc.*, vol. 39, no. 3, pp. 577–591, 1971.

R. Goodrich and P. Caines, “Necessary and sufficient conditions for local second-order identifiability,” *IEEE Trans. Autom. Control*, vol. 24, no. 1, pp. 125–127, Feb. 1979.

P. Stoica and T. Söderström, “On non-singular information matrices and local identifiability,” *Int. J. Control*, vol. 36, no. 2, pp. 323–329, 1982.

P. Stoica and T. L. Marzetta, “Parameter estimation problems with singular information matrices,” *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 87–90, Jan. 2001.

L. H. van Trees, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. Hoboken, NJ, USA: Wiley, 2004.

B.-Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai, “Some classes of global Cramér–Rao bounds,” *Ann. Statist.*, vol. 15, no. 4, pp. 1421–1438, 1987.

R. D. Gill and B. Y. Levit, “Applications of the van Trees inequality: A Bayesian Cramér–Rao bound,” *Bernoulli*, vol. 1, no. 1/2, pp. 59–70, 1995.

A. W. Van der Vaart, *Asymptotic Statistics*. New York, NY, USA: Cambridge Univ. Press, 2000.

A. B. Tsybakov, *Introduction to Nonparametric Estimation*. New York, NY, USA: Springer, 2008.

I. A. Bragimrov and R. Z. Has’minskii, *Statistical Estimation: Asymptotic Theory*, vol. 16, New York, NY, USA: Springer, 2013.

M. K. S. Faradonbeh, A. Tewari, and G. Michailidis, “Finite time identification in unstable linear systems,” *Automatica*, vol. 96, pp. 342–353, 2018.

M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht, “Learning without mixing: Towards a sharp analysis of linear system identification,” in *Proc. Mach. Learn. Res.*, vol. 75, S. Bubeck, V. Perchet, and P. Rigollet, Eds., Jul. 2018, pp. 439–473.

T. Sarkar and A. Rakhlin, “Near optimal finite time identification of arbitrary linear dynamical systems,” in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 5610–5618.

Y. Jedra and A. Proutiere, “Finite-time identification of stable linear systems optimality of the least-squares estimator,” in *Proc. 59th IEEE Conf. Decis. Control*, 2020, pp. 996–1001.

A. J. Wagenmaker, M. Simchowitz, and K. Jamieson, “Task-optimal exploration in linear dynamical systems,” in *Proc. Int. Conf. Mach. Learn.*, 2021.

T. Söderström, *Discrete-Time Stochastic Systems: Estimation and Control*. London, U.K.: Springer, 2002.

Ingvar Ziemann (Member, IEEE) received the bachelor’s degree in economics and finance from Stockholm School of Economics, Stockholm, Sweden, in 2014, and second the bachelor’s degree in mathematics from Stockholm University/Royal Institute of Technology (KTH), Stockholm, Sweden, in 2015, and the master’s degree in economics and finance from Stockholm School of Economics, Stockholm, Sweden, in 2017, and the second master’s degree in mathematics from Stockholm University/Royal Institute of Technology (KTH), Stockholm, Sweden, in 2018, and the Ph.D. degree from the Division of Decision and Control Systems, KTH, under the supervision of Henrik Sandberg, in 2022.

He is currently a Postdoctoral Researcher with the University of Pennsylvania, Philadelphia, PA, USA. His research interests include statistical and information theoretic tools to study learning-enabled control methods, with a current interest in studying how learning algorithms generalize in the context of dynamical systems.

Dr. Ziemann was the recipient of the Swedish Research Council International Postdoc Grant, the 2022 IEEE Conference on Decision and Control Best Student Paper Award, and the 2017 Stockholm Mathematics Center Excellent Master Thesis Award.

Henrik Sandberg (Fellow, IEEE) received the M.Sc. degree in engineering physics and the Ph.D. degree in automatic control from Lund University, Lund, Sweden, in 1999 and 2004, respectively.

From 2005 to 2007, he was a Postdoctoral Scholar with the California Institute of Technology, Pasadena, CA, USA. In 2013, he was a Visiting Scholar with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA. He has also held visiting appointments with the Australian National University, Canberra, ACT, Australia, and the University of Melbourne, Parkville, VIC, Australia. He is currently a Professor with the Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm, Sweden. His research interests include security of cyber-physical systems, power systems, model reduction, and fundamental limitations in control.

Dr. Sandberg was the recipient of the Best Student Paper Award from the IEEE Conference on Decision and Control in 2004, an Ingvar Carlsson Award from the Swedish Foundation for Strategic Research in 2007, and a Consolidator Grant from the Swedish Research Council in 2016. He has served on the editorial boards of IEEE TRANSACTIONS ON AUTOMATIC CONTROL and the IFAC Journal *Automatica*. 