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

Linear Systems can be Hard to Learn

Anastasios Tsiamis and George J. Pappas

2021 60th IEEE Conference on Decision and Control (CDC) | 978-1-6654-3659-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/CDC45484.2021.9682778

***Abstract*— In this paper, we investigate when system identification is statistically easy or hard, in the finite sample regime. Statistically easy to learn linear system classes have sample complexity that is polynomial with the system dimension. Most prior research in the finite sample regime falls in this category, focusing on systems that are directly excited by process noise. Statistically hard to learn linear system classes have worst-case sample complexity that is at least exponential with the system dimension, regardless of the identification algorithm. Using tools from minimax theory, we show that classes of linear systems can be hard to learn. Such classes include, for example, under-actuated or under-excited systems with weak coupling among the states. Having classified some systems as easy or hard to learn, a natural question arises as to what system properties fundamentally affect the hardness of system identifiability. Towards this direction, we characterize how the controllability index of linear systems affects the sample complexity of identification. More specifically, we show that the sample complexity of robustly controllable linear systems is upper bounded by an exponential function of the controllability index. This implies that identification is easy for classes of linear systems with small controllability index and potentially hard if the controllability index is large. Our analysis is based on recent statistical tools for finite sample analysis of system identification as well as a novel lower bound that relates controllability index with the least singular value of the controllability Gramian.**

I. Introduction

Linear system identification focuses on using input-output data samples for learning dynamical systems of form:

$$x_{k+1} = Ax_k + Bu_k + Hw_k, \tag{1}$$

where x_k represents the state, u_k represents the control signal, and w_k is the process noise. The statistical analysis of system identification algorithms has a long history [1]. Until recently, the main focus was providing guarantees for the convergence of system identification in the *asymptotic regime* [2]–[4], when the number of collected samples N tends to infinity. Under sufficient persistency of excitation [5], system identification algorithms converge and the asymptotic bounds capture very well how the identification error decays with N qualitatively.

However, our standard asymptotic tools (e.g. the Central Limit Theorem), do not always capture all finite-sample phenomena [6, Ch 2]. Moreover, the identification error depends on various system theoretic constants, like the state space dimension n, which might be hidden under the big-O notation in the asymptotic bounds. As a result, system identification limitations, like the curse of dimensionality, although known to practitioners, are not always reflected in the theoretical asymptotic bounds.

With the advances in high-dimensional statistics [6], there has been a recent shift from asymptotic analysis with infinite data to statistical analysis of system identification with finite samples. Over the past years there have been significant advances in understanding finite sample system identification for both fully-observed systems [7]–[14] as well as partially-observed systems [15]–[23]. A tutorial can be found in [24]. The above approaches offer mainly *data-independent* bounds which reveal how the state dimension n and other system theoretic parameters affect the sample complexity of system identification *qualitatively*. This is different from finite sample data-dependent bounds-see for example bootstrapping [8] or [25], which might be more tight and more suitable for applications but do not necessarily reveal this dependence.

Despite these advances, we still do not fully understand the fundamental limits of when identification is easy or hard. In this paper, we define as statistically easy, classes of systems whose finite-sample complexity is polynomial with the system dimension. Most prior research in the finite-sample analysis of fully observed systems falls in this category by assuming system (1) is fully excited by the process noise w_k. We define as statistically hard, classes of linear systems whose worst-case sample complexity is at least exponential with the system dimension, regardless of the learning algorithm. Using recent tools from minimax theory [13], we show that classes of linear systems which are statistically hard to learn do indeed exist. Such system classes include, for example, under-actuated systems with weak state coupling. The fact that linear systems may contain exponentially hard classes has implications for broader classes of systems, such as nonlinear systems, as well as control algorithms, such as the linear quadratic regulator [26] and reinforcement learning [27].

By examining classes of linear systems that are statistically easy or hard, we quickly arrive at the conclusion that system theoretic properties, such as controllability, fundamentally affect the hardness of identification. In fact, as we show in the paper, structural properties like the controllability index can crucially affect learnability, determining whether a problem is hard or not. In summary, our contributions are the following:

–Learnability of dynamical systems. We define two novel notions of learnability. A class of systems is easy to learn if it exhibits polynomial sample complexity with respect the state dimension n. It is hard to learn if for any possible learning algorithm it has exponential worst-case complexity.

–Exponential sample complexity is possible. We identify non-trivial classes of under-actuated linear systems whose worst-case sample complexity increases exponentially with

The authors are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104. Emails: {atsiamis,pappasg}@seas.upenn.edu. This work is supported by the NSF-Simons grant 2031985, the AFOSR grant FA9550-19-1-0265, and the NSF CPS 2038873 grant.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:56:41 UTC from IEEE Xplore. Restrictions apply.

the state dimension n regardless of learning algorithm.

–Controllability index affects sample complexity. We prove that under the least squares algorithm, the sample complexity is upper-bounded by an exponential function of the system's controllability index. This implies that if the controllability index is small $O(1)$ (with respect to the dimension n), the sample complexity is guaranteed to be polynomial generalizing previous cases.

–New controllability Gramian bound Our sample complexity upper bound is a consequence of a new result that is of independent, system theoretic interest. We prove that for robustly controllable systems, the inverse of the least singular value of the controllability Gramian can grow at most exponentially with the controllability index. Although this has been observed empirically [28], to the best of our knowledge this theoretical bound is new and has implications beyond system identification.

Recently in [29], it was proved that learning to control linear systems online can also be hard. In particular, [29] proved that the transient component of the regret can be exponential with the system dimension. This complexity shows up in the case of unstable unknown linear systems. Here, our setting is different since we focus on the sample complexity of system identification. The source of hardness here is the controllability structure of the dynamical system.

Notation: The transpose operation is denoted by $(\cdot)'$ and the complex conjugate by $*$. By $e_i \in \mathbb{R}^n$ we denote the $i-$th canonical vector. By $\sigma_{\min}$ we denote the least singular value. $\succeq$ denotes comparison in the positive semidefinite cone. The identity matrix of dimension n is denoted by I_n. The spectral norm of a matrix A is denoted by $\|A\|_2$. The notion of controllability and other related concepts are reviewed in the Appendix.

The full proofs of our results can be found in [30].

II. Learnability of System Classes

Consider system (1), where $x_k \in \mathbb{R}^n$ is the state and $u_k \in \mathbb{R}^p$ is the input. By $w_k \in \mathbb{R}^r$ we denote the process noise which is assumed to be Gaussian, i.i.d. with covariance I_r. Without loss of generality the initial state is assumed to be zero $x_0 = 0$.

Assumption 1. *All state parameters are bounded: $\|A\|_2, \|B\|_2, \|H\|_2 \leq M$, for some positive constant $M > 0$. The noise has unknown dimension r and can be degenerate $r \leq n$. All parameters A, B, H, r are considered unknown. Matrices B, H have full column rank $\mathrm{rank}(B) = p \leq n$, $\mathrm{rank}(H) = r \leq n$. We also assume that the system is non-explosive $\rho(A) \leq 1$. Finally, we assume that the control inputs have bounded energy $\mathbb{E}u_t' u_t \leq M$.*

A system identification (SI) algorithm $\mathcal{A}$ receives a finite number N of input-state data $(x_0, u_0), \ldots, (x_N, u_N)$ generated by system (1), and returns an estimate of the unknown system's parameters $\hat{A}_N, \hat{B}_N, \hat{H}_N$. We denote by N the number of collected input-state samples, which are generated during a single roll-out of the system, that is a single trajectory of length N. For simplicity, we focus only on the estimation of matrix A in this paper.

Our goal is to study when the problem of system identification is fundamentally easy or hard. The difficulty is captured by the sample complexity, i.e. how many data N do we need to achieve small identification error with high probability. Formally, let $\epsilon > 0$, $0 < \delta < 1$ be the accuracy and confidence parameters respectively. Then, the sample complexity is the smallest possible number of samples N such that with probability at least $1 - \delta$ we can estimate A with small error $\|A - \hat{A}_N\| \leq \epsilon$. Naturally, the sample complexity increases as the accuracy/confidence parameters ϵ, δ decrease. The sample complexity also increases in general with the state-space dimension n and the bound M on the state space parameters.

Ideally, the sample complexity should grow slowly with $n, M, \epsilon^{-1}, \delta^{-1}$. Inspired by Probably Approximately Correct (PAC) learning [31], [32], we classify an identification problem as easy when the sample complexity depends polynomially on $n, M, \epsilon^{-1}, \delta^{-1}$. For brevity we will use the symbol S to denote the tuple $S = (A, B, H)$. Let $\mathbb{P}_S$ denote the probability distribution of the input-state data when the true parameters of the system are equal to S and we apply a control law $u_t \in \mathcal{F}_t$, where $\mathcal{F}_t \triangleq \sigma(x_0, u_0, \ldots, u_{t-1}, x_t)$ is the sigma algebra generated by the previous outputs and inputs. By $\mathcal{C}_n$ we will denote a class of systems with dimension n.

Definition 1 (poly-learnable classes)**.** *Let $\mathcal{C}_n$ be a class of systems. Consider a trajectory of input-state data $(x_0, u_0), \ldots, (x_N, u_N)$, which are generated by a system S in $\mathcal{C}_n$ under some control law $u_t \in \mathcal{F}_t$, $t \leq N$. We call the class $\mathcal{C}_n$ $\mathrm{poly}(n)-$learnable if there exists an identification algorithm such that the sample complexity is polynomial: for any confidence $0 \leq \delta < 1$ and any tolerance $\epsilon > 0$:*

$$\sup_{S \in \mathcal{C}_n} \mathbb{P}_S(\|A - \hat{A}_N\| \geq \epsilon) \leq \delta, \quad (2)$$
$$\text{for } N \geq \mathrm{poly}(n, 1/\epsilon, \log 1/\delta, M),$$

where $\mathrm{poly}(\cdot)$ is some polynomial function.

Definition 1 provides an intuitive definition for a class $\mathcal{C}_n$ of linear systems whose system identification problem is easy. To prove that a class of systems $\mathcal{C}_n$ is easy, it suffices to provide one algorithm that performs well for any system $S \in \mathcal{C}_n$ in the sense that it requires **at most** a polynomial number of samples. This means that we should obtain sample complexity **upper bounds** across all $S \in \mathcal{C}_n$ which is what the supremum over $S \in \mathcal{C}_n$ achieves in (2). Otherwise, we can construct trivial algorithms that perform well only on one system and fail to identify the other.

In recent work [9], [11], [12], it was shown that under the least squares algorithm, the sample complexity of learning linear systems is polynomial. As we review in Section III, these results hold for classes of linear systems where the noise is isotropic and hence directly exciting all states.

However, if we relax the last assumption it turns out that the sample complexity might degrade dramatically. To raise

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:56:41 UTC from IEEE Xplore. Restrictions apply.

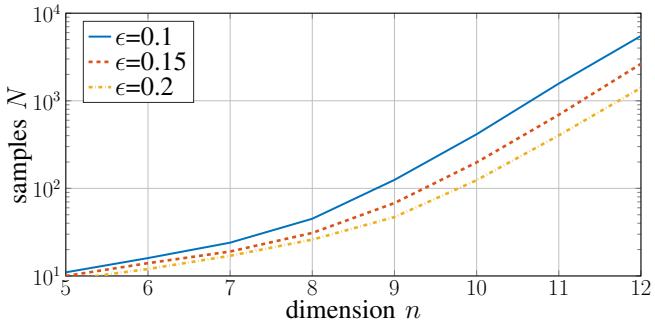


Fig. 1. The minimum number of samples N such that the (empirical) average error $\mathbb{E}\|A - \hat{A}_N\|_2$ is less than ϵ. The sample complexity appears to be increasing exponentially with the dimension n under the least squares algorithm.

this issue, consider the following example. Let $J_n(1)$ be a Jordan block of size n with eigenvalue 1 and let e_n be the $n-$th canonical vector. We simulate the performance of least squares identification for the system

$$x_{k+1} = 0.5J_n(1)x_k + e_n(u_k + w_k) \tag{3}$$

Note that in system (3) the process noise is no longer isotropic. Figure 1 shows the minimum number of samples N required to achieve (empirical) average error $\mathbb{E}\|A - \hat{A}_N\| \leq \epsilon$ (the details of the simulation can be found in Section VI). It seems that the sample complexity increases exponentially rather than polynomially. Are the results in Figure 1 due to the choice of the algorithm or is there a fundamental limitation for all system identification algorithms? We pose the following fundamental problem.

Question 1. *Do there exist classes of linear systems which are hard to learn, meaning not* poly*-learnable by any system identification algorithm? Furthermore, can the sample complexity for a class of linear systems be exponential with state dimension n?*

A class of linear systems $\mathcal{C}_n$ that is not poly-learnable will be viewed as hard. By negating Definition 1, this notion of hardness means that given *any* system identification algorithm, there exist instances $S \in \mathcal{C}_n$ that cannot have polynomial sample complexity. In other words, a system class $\mathcal{C}_n$ is classified as hard when its impossible to find any system identification algorithm that achieves polynomial sample complexity for all $S \in \mathcal{C}_n$. This can be viewed as a fundamental statistical limitation for the chosen class $\mathcal{C}_n$.

Motivated by Figure 1, we define an important subclass of hard problems, namely linear system classes that have worst-case sample complexity that grows exponentially with the dimension n regardless of identification algorithm choice.

Definition 2 (exp-hard classes)**.** *Let $\mathcal{C}_n$ be a class of systems of dimension n. Consider a trajectory of input-output data $(x_0, u_0), \ldots, (x_N, u_N)$, which are generated by a system S in $\mathcal{C}_n$ under some control law $u_t \in \mathcal{F}_t$, $t \leq N$. We call a class $\mathcal{C}_n$ of systems* exp(n)*-hard if the sample complexity is at least exponential with the dimension n: there exist confidence $0 \leq \delta < 1$ and tolerance ϵ parameters such that for any identification algorithm:*

$$\sup_{S \in \mathcal{C}_n} \mathbb{P}_S(\|A - \hat{A}_N\| \geq \epsilon) \leq \delta,$$
$$\textit{only if } N \geq \exp(n),$$

where exp(n) *denotes an exponential function of n.*

System classes $\mathcal{C}_n$ that are exp-hard are an important subset of hard system classes as they are clearly not poly-learnable. However, not all classes that are not poly-learnable are exp-hard.

In order to show that a class of systems $\mathcal{C}_n$ is exp-hard, one must show that for **any** system identification algorithm the worst-case sample complexity is at least exponential in state dimension n. Contrary to poly-learnable problems, for exponential hardness we should establish sample complexity **lower bounds**.

In this paper, we first address Question 1 and show that exp-hard classes of linear systems do indeed exist. While this can be viewed as a fundamental statistical limitation for all system identification algorithms, our results open a new direction of research that classifies when linear systems are easy to learn and when they are hard to learn. This leads to the following important question addressing in this paper.

Question 2. *When is a class of linear systems $\mathcal{C}_n$ guaranteed to be* poly*-learnable?*

Based on prior work, we already have partial answers to Question 2 as we know that linear systems with isotropic noise are poly-learnable. In Section V, we seek to broaden the classes of poly-learnable systems and discover their relation to fundamental system theoretic properties such as controllability.

While Definitions 1, 2 are inspired by PAC learning, they have a different flavor. One of the differences is that the guarantees in Definitions 1, 2 are stated in terms of recovering the state-space parameters, while in PAC learning, they would be stated in terms of the prediction error of the learned model or informally $\sum_{k=0}^{N-1} \mathbb{E}\|x_k - \hat{A}x_{k-1} - \hat{B}u_{k-1}\|^2$.

III. Directly-excited systems are poly-learnable

In this section, we revisit state-of-the-art results in finite-sample complexity for fully-observed linear systems and re-establish that they all lead to polynomial sample complexity. In prior work [9], [11], [12], the class of linear systems considered assumes that the stochastic process noise is isotropic, i.e. $HH' = \sigma_w^2 I_n$. Since all states are directly excited by the process noise, all modes of the system are captured sufficiently in the data. To obtain polynomial complexity, it suffices to use the least squares identification algorithm

$$\begin{bmatrix} \hat{A}_N & \hat{B}_N \end{bmatrix} = \arg\min_{\{F,G\}} \sum_{t=0}^{N-1} \|x_{t+1} - Fx_t - Gu_t\|_2^2 \tag{4}$$

with white noise inputs $u_t \sim \mathcal{N}(0, \sigma_u^2 I)$. Based on the algorithm analysis from [9], let k be a fixed time index

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:56:41 UTC from IEEE Xplore. Restrictions apply.

which is much smaller than the horizon N (see Theorem 2.1 in [9] for details). Let $0 < \delta < 1$ and ϵ be the confidence and accuracy parameters respectively. Then, with probability at least $1-\delta$, the error is $\|A - \hat{A}_N\|_2 \leq \epsilon$ if:

$$N \geq \frac{c\sigma_w^2}{\sigma_{\min}(\Gamma_k)} \frac{1}{\epsilon^2} \left(n \log \frac{n}{\delta} + \log \det(\Gamma_N \Gamma_k^{-1}) \right),$$

where c is a universal constant, and $\Gamma_k = \sigma_u^2 \Gamma_k(A,B) + \sigma_w^2 \Gamma_k(A, I_n)$ is the (combined) controllability Gramian–see Appendix for the definition. Under the isotropic noise assumption, the least singular value of the Gramian Γ_k is bounded away from zero, $\sigma_{\min}(\Gamma_k) \geq \sigma_w^2$.

In a slight departure from [9], [11], [12], we can show that the determinant of the Gramian $\det(\Gamma_N)$ can only increase at most polynomially with the number of samples N and exponentially with state dimension n. This is a direct consequence of the following lemma, which is a new result.

Lemma 1. *Let $A \in \mathbb{R}^{n \times n}$ have all eigenvalues inside or on the unit circle, with $\|A\|_2 \leq M$. Then, the powers of matrix A are bounded by:*

$$\left\|A^k\right\|_2 \leq (ek)^{n-1} \max\{M^n, 1\} \tag{5}$$

Lemma 1 enables us to eliminate the dependence on the condition number of the Jordan form's similarity transformation, which exists in prior bounds and can be arbitrarily large. We avoid this dependence by using the Schur form of A [33]. While this does not alter the already known sample complexity results, it allows us to have sample complexity bounds that are uniform across all systems that satisfy Assumption 1.

As a result of Lemma 1, we obtain that the system identification problem for linear systems with isotropic noise has polynomial sample complexity. The result can be broadened to the more general case of direct excitation, where the covariance is lower bounded by $HH' + BB' \succeq \sigma_w^2 I_n$, for some $\sigma_w > 0$, as the following theorem states.

Theorem 1 (Directly-excited)**.** *Consider the class $\mathcal{C}_n$ of directly-excited systems $S = (A, B, H) \in \mathbb{R}^{n \times (n+p+r)}$ such that Assumption 1 is satisfied with covariance $HH' + BB' \succeq \sigma_w^2 I_n$, for some $\sigma_w > 0$. The class $\mathcal{C}_n$ is* poly$-$*learnable under the least squares system identification algorithm with white noise input signals $u_k \sim \mathcal{N}(0, I_p)$.*

Directly excited systems includes fully-actuated systems (number of inputs equal to the number of states $p = n$), or systems with isotropic noise as special cases. However, having direct excitation might not always be the case. The combined noise and input matrices might be rank-deficient. For example, we might have actuation noise as in:

$$x_{t+1} = Ax_t + B(u_t + w_t).$$

In general, the noise might be ill-conditioned (zero across certain directions), while it might be physically impossible to actuate every state of the system. We call such systems under-actuated or under-excited. It might still be possible to identify underactuated systems, e.g. if the pair $(A, \begin{bmatrix} H & B \end{bmatrix})$ is controllable. However, as we prove in the next section, the identification difficulty might increase dramatically.

IV. Exp-hard system classes

In this section, we show that there exist common classes of linear systems which are impossible or hard to identify with a finite amount of samples. As we will see, this can happen when systems are under-actuated and under-excited. When only a limited number of system states is directly driven by inputs (or excited by noise) and the remaining states are only indirectly excited, then identification can be inhibited.

A. Controllable systems with infinite sample complexity

For presentation simplicity, let us assume that there are no exogenous inputs $B = 0$. Similar results also hold when $B \neq 0$–see Remark 1. To fully identify the unknown matrix A, it is necessary that the pair (A, H) is controllable. Furthermore, let's assume that the noise is meaningful, that is $\sigma_{\min}(H) \geq \sigma$ for some $\sigma > 0$. However, controllability of (A, H) and $\sigma_{\min}(H) \geq \sigma$ are not sufficient to ensure system identification from a finite number of samples. The following, perhaps unsurprising theorem, shows that for this class of linear systems, the worst-case sample complexity is infinite.

Theorem 2 (Controllability is not sufficient for finite sample complexity)**.** *Consider the class $\mathcal{C}_n$ of systems $S = (A, H) \in \mathbb{R}^{n \times (n+r)}$ such that Assumption 1 is satisfied with (A, H) controllable, and $\sigma_{\min}(H) \geq \sigma$ for some $\sigma > 0$. For any system identification algorithm the sample complexity is infinite: there exist a failure probability $0 \leq \delta < 1$ and a tolerance $\epsilon > 0$ such that we cannot achieve*

$$\sup_{S \in \mathcal{C}_n} \mathbb{P}_S(\|A - \hat{A}_N\| \geq \epsilon) \leq \delta$$

with a finite number of samples N.

Theorem 2 clearly shows that we may need stronger notions of controllability, as done in Section IV-B, in order to find classes of systems whose sample complexity is finite. The proof of Theorem 2 uses tools from minimax theory [13]. Adapting these tools in our setting results in the following.

Lemma 2 (Minimax bounds)**.** *Let $\mathcal{C}_n$ be a class of systems. Consider a confidence $0 < \delta < 1$ and an accuracy parameter $\epsilon > 0$. Denote by $S_1, S_2 \in \mathcal{C}_n$ any pair of two systems with A_1, H_1, A_2, H_2 the respective unknown matrices, such that $\|A_1 - A_2\| \geq 2\epsilon$. Let $\mathrm{KL}(\mathbb{P}_{S_1}, \mathbb{P}_{S_2})$ be the Kullback-Leibler divergence between the probability distributions of the data when generated under S_1, S_2 respectively. Then for any identification algorithm*

$$\sup_{S \in \mathcal{C}_n} \mathbb{P}_S(\|A - \hat{A}_N\| \geq \epsilon) \leq \delta$$

holds only if

$$\mathrm{KL}(\mathbb{P}_{S_1}, \mathbb{P}_{S_2}) \geq \log \frac{1}{3\delta}, \tag{6}$$

for all such pairs $S_1, S_2 \in \mathcal{C}_n$.

Intuitively, to find difficult learning instances we construct systems which are sufficiently separated (2ϵ away). Meanwhile, the systems should be similar enough to generate data with as indistinguishable distributions as possible (small

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:56:41 UTC from IEEE Xplore. Restrictions apply.

KL divergence). If the system is hard to excite, then the distributions of the states will look similar under many different matrices A, leading to smaller KL-divergence. Unless we bound the pair (A, H) away from uncontrollability, it might be impossible to satisfy (6) for all pairs of systems with a finite number of samples. For example consider:

$$A = \begin{bmatrix} 0 & \alpha & 0 \\ 0 & 0 & \beta \\ 0 & 0 & 0 \end{bmatrix}, H = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix},$$

It requires an arbitrarily large number of samples to learn α if the coupling β between $x_{t,2}$ and $x_{t,3}$ is arbitrarily small. The distribution of $x_{t,1}$ remains virtually the same as we perturb α, since the state $x_{t,2}$ is under-excited for small β.

B. Robustly controllable systems can be exp*-hard*

Theorem 2 implies that we need to bound the system away from uncontrollability in order to obtain non-trivial sample complexity bounds. In order to formulate this, we review the notion of distance from uncontrollability, which is the norm of the smallest perturbation that makes (A, H) uncontrollable.

Definition 3 (Distance from uncontrollability [34])**.** *Let $(A, H) \in \mathbb{R}^{n\times(n+r)}$ be controllable. Then, the distance from uncontrollability is given by:*

$$d(A,H) \triangleq \inf \left\{ \| \begin{bmatrix} \Delta A & \Delta H \end{bmatrix} \|_2 : (A+\Delta A, H+\Delta H) \text{ uncontrollable} \right\}, \quad (7)$$

where perturbations $(\Delta A, \Delta H) \in \mathbb{C}^{n\times(n+r)}$ are complex.

Let us now consider linear systems that are robustly controllable. That is, classes of controllable linear systems whose distance from uncontrollability is lower bounded. The lower bound is allowed to degrade gracefully (polynomially) with the system dimension n.

Assumption 2 (Robust Controllability)**.** *Assume class $\mathcal{C}_n$ of linear systems that is robustly controllable, that is $(A, H) \in \mathbb{R}^{n\times(n+m)}$ is μ-away from uncontrollability:*

$$d(A, H) \geq \mu, \quad (8)$$

for some positive $\mu \geq 0$, with $\mu^{-1} \leq \text{poly}(n)$.

Assumption 2 is not restrictive as long as we allow the bound to degrade with the dimension. Common systems like the $n-$th order integrator have distance that degrades linearly with n–see Lemma 3 in the Appendix. However, even for system classes that satisfy Assumption 2, the next theorem shows that system identification can be exp-hard.

Theorem 3 (Exp(n)-hard classes)**.** *Consider the set $\mathcal{C}_n$ of systems $S = (A, H)$ such that Assumptions 1, 2 are satisfied with $d(A, H) \geq \mu = (8(n+1))^{-1}$. Then, for any system identification algorithm $\mathcal{A}$ the sample complexity is exponential in the state dimension n. There exist a confidence $0 \leq \delta < 1$ and a tolerance $\epsilon > 0$ such that*

$$\sup_{S\in\mathcal{C}_n} \mathbb{P}_S(\|A - \hat{A}_N\| \geq \epsilon) \leq \delta$$

is satisfied only if

$$N \geq \frac{4^{n-3}}{3\epsilon^2} \log \frac{1}{\delta}.$$

Theorem 3 shows that even for robustly controllable classes of linear systems satisfying Assumptions 1, 2, any system identification algorithm will have worst-case sample complexity that depends exponentially on the system dimension n. The proof of Theorem 3 is based once more on minimax theory used in Lemma 2.

The reason for this learning difficulty is due to the need for indirect excitation. Consider, for example, chained systems, where every state indirectly excites the next one. If the states are weakly-coupled, then the exploratory signal (noise or input) attenuates exponentially fast along the chain. As a concrete example, consider the following system for $\rho < 0.5$:

$$A = \begin{bmatrix} \rho & \rho & 0 & \cdots & 0 & 0 \\ 0 & \rho & \rho & \cdots & 0 & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & \cdots & \rho & \rho \\ 0 & 0 & 0 & \cdots & 0 & \rho \end{bmatrix}, H = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 0 & \rho \end{bmatrix} \quad (9)$$

which satisfies Assumptions 1, 2. Matrix A has a chained structure with weak coupling between the states. Noise can only excite states $x_{t,1}, x_{t,n}$ directly. Until the exploratory noise signal reaches $x_{t,2}$ it decreases exponentially fast with the dimension n. As a result, it is difficult to learn A_{12} due to lack of excitation. In terms of Lemma 2, the distribution of $x_{t,1}$ will remain virtually the same if we perturb A_{12} since $x_{t,2}$ is under-excited.

Remark 1 (Exogenous inputs)**.** *When $B \neq 0$ similar results hold but with an additional interpretation. Consider system* (9) *but with $H = e_1$, $B = \rho e_n$. Then, if we apply white-noise input signals we have two possibilities: i) the control inputs have bounded energy per Assumption 1 but we suffer from exponential sample complexity or ii) we obtain polynomial sample complexity but we allow the energy of the inputs to increase exponentially with the dimension. From this alternative viewpoint a system is hard to learn if it requires exponentially large control inputs.*

V. Controllability index affects learnability

Structural system properties of an underactuated system, such as the chained structure in the dynamics, can be critical in making system identification easy or hard. This poses novel questions about understanding how system theoretic properties affect system learnability as defined in Definitions 1 and 2. We begin a new line of inquiry by characterizing how the controllability index κ, a critical structural system property, affects the statistical properties of system identification. A brief review of the concept of controllability index can be found in the Appendix. It can be viewed as a structural measure of whether a system is directly actuated or underactuated resulting in long chains. The following theorem, is the first result connecting the controllability index with sample complexity bounds.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:56:41 UTC from IEEE Xplore. Restrictions apply.

Theorem 4 (Controllability index-dependent upper bounds)**.** *Consider the set $\mathcal{C}_n$ of systems $S = (A, H)$ such that Assumption 1 is satisfied. Let Assumption 2 be satisfied for the pair $(A, \begin{bmatrix} H & B \end{bmatrix})$. Furthermore assume that the controllability index of all systems in the class is upper bounded by κ. Then, under the least squares system identification algorithm and white noise inputs $u_k \sim \mathcal{N}(0, I_p)$, we obtain that*

$$\sup_{S \in \mathcal{C}_n} \mathbb{P}_S(\|A - \hat{A}_N\| \geq \epsilon) \leq \delta$$

is satisfied for

$$N \geq \text{poly}^{\kappa}(n, M)\text{poly}(\epsilon^{-1}, \log 1/\delta).$$

Theorem 4 formalizes our intuition since the controllability index is the length of the chain from input excitation towards the most distant state in the chain. Hence, systems with a large number of inputs (or noise) and small controllability index ($\kappa << n$) are easy to identify. The directly excited case with isotropic noise, presented in Theorem 1, is a special case corresponding to a controllability index $\kappa = 1$, recovering prior polynomial bounds.

The implications of Theorems 3, 4 illustrate the impact controllability properties have on system learnability–see Figure 4. Classes of systems with small controllability index $O(1)$ have polynomial sample complexity. Classes where the index grows linearly $\Omega(n)$ can be exponentially hard in the worst case in general. There might still be subclasses of systems with large controllability indexes which nonetheless can be identified with a polynomial number of samples. However, we cannot provide any guarantees without further assumptions.

The proof of Theorem 4 is based on the results of [9] and crucially depends on the following system theoretic result that bounds the least singular value of the controllability Gramian (a quantitative measure of controllability) with the controllability index (a structural measure of controllability).

Theorem 5 (Controllability gramian bound)**.** *Consider a system (A, H) that satisfies Assumptions 1, 2. Let κ be its controllability index. Then, the least singular value of the gramian Γ_κ is lower bounded by:*

$$\sigma_{\min}^{-1}(\Gamma_\kappa) \leq \text{poly}^{\kappa}(M/\mu).$$

The above theorem is of independent interest, since it states that the controllability index rather than the dimension n controls how fast the controllability Gramian degrades. While the above bound may be loose in general, it gives us qualitative insights about how system structure affects the hardness of input excitation and system identification. Our proof exploits the so-called "staircase" (or Hessenberg) canonical representation (12) of state space systems [35]–see Appendix. The main idea is that if a system is robustly controllable then the coupling between the states is bounded away from zero. Hence, we can avoid the essentially uncontrollable systems of Theorem 2 which lead to infinite sample complexity.

VI. Simulations

We study three simulation scenarios to illustrate the qualitative implications of our results. In the first two cases, we verify that the sample complexity of the least squares algorithm can indeed grow exponentially with the dimension. In the third case, we investigate how the controllability index affects the sample complexity. In all cases, we perform Monte Carlo simulations to compute the empirical mean error $\|A - \hat{A}_N\|_2$ and we count the number of samples required to have error less than ϵ, for some $\epsilon > 0$. For numerical stability in the least squares estimator (4) we used a regularization term (ridge regression) with coefficient 0.001.

In the first example in Section II, Figure 1, we used 1000 Monte Carlo iterations to approximate the empirical average. We modeled the noise as gaussian with $w_k \sim \mathcal{N}(0, 0.5)$ and used white noise inputs $u_k \sim \mathcal{N}(0, 10)$. The sample complexity of the least squares algorithm seems to be exponential with the dimension. In Section IV, we showed that such systems exhibit exponential sample complexity due to the weak coupling between the states.

In the second example, we study the behavior of Jordan blocks actuated from the last state. Let $J_n(\lambda)$ be a Jordan block of dimension n and eigenvalues all λ. We consider the system $A = J_n(\lambda)$, $H = 0.1e_n$, $B = 5e_n$, which means we excite directly only state $x_{t,n}$. We repeat the same experiment as before for 1000 Monte Carlo simulations with $w_k, u_k \sim \mathcal{N}(0, 1)$ and for $\epsilon = 0.005$. In Figure 2, it seems that the complexity of the least squares algorithm is also exponential when $0 < \lambda < 1$. In this case the coupling between the states is not weak. However, certain subspaces might still be hard to excite. As λ approaches the unit circle eigenvalue 1 the complexity improves. For $\lambda = 1$, after $n = 9$ Matlab returned inaccurate results as the condition number of the data becomes very large. Hence, we do not report any results beyond $n = 9$. However, based on simulations for small n it might be possible that the system can be learned by only a polynomial number of samples. The intuition might be that in this case instability helps with excitation [9]. It is an open problem to prove or disprove exponential lower bounds for the Jordan block when $0 < \lambda < 1$. Similarly, we leave it as an open problem to prove or disprove polynomial upper bounds for the Jordan block when $\lambda = 1$.

In the third example, we consider the Jordan block $A = J_n(0.5)$ with noise $H = 0.1e_n$. We start from $B = 5e_n$ and we gradually add more exogenous inputs to decrease the controllability index: we try $B = 5\begin{bmatrix} e_n & e_{\lceil n/2 \rceil} \end{bmatrix}$ and $B = 5\begin{bmatrix} e_n & e_{n-2} & \dots \end{bmatrix}$ which correspond to indices $\kappa = \lceil n/2 \rceil$ and $\kappa = 2$ respectively. We repeat the same experiment as before for 1000 Monte Carlo simulations with $w_k, u_k \sim \mathcal{N}(0, 1)$ and for $\epsilon = 0.005$. In Figure 3, it seems that the sample complexity remains exponential when $\kappa = \lceil n/2 \rceil$. However, when $\kappa = 2$ there is a phase transition and the sample complexity becomes polynomial with the dimension.

VII. Conclusion

The results of this paper paint a broader and more diverse landscape about the statistical complexity of learning

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:56:41 UTC from IEEE Xplore. Restrictions apply.

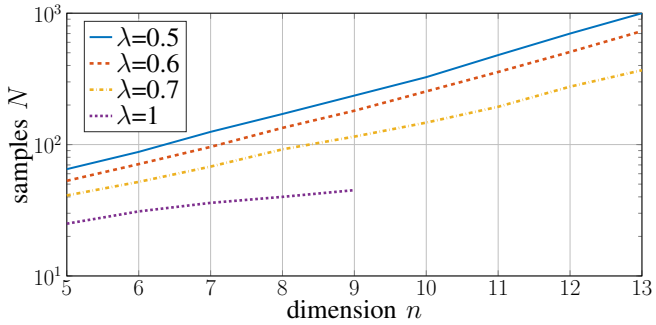


Fig. 2. Sample complexity of identifying the Jordan block of size n and eigenvalues all λ, actuated from the last state. The figure shows the minimum number of samples N such that the (empirical) average error $\mathbb{E}\|A - \hat{A}_N\|_2$ is less than 0.005. The sample complexity appears to be increasing exponentially with the dimension n for $\lambda < 1$. For $\lambda = 1$, Matlab returns inaccurate results for $n \geq 10$ since the condition number of the data is very large. However, in the regime $5 \leq n \leq 9$, the complexity seems to be polynomial, increasing in 5 sample increments.

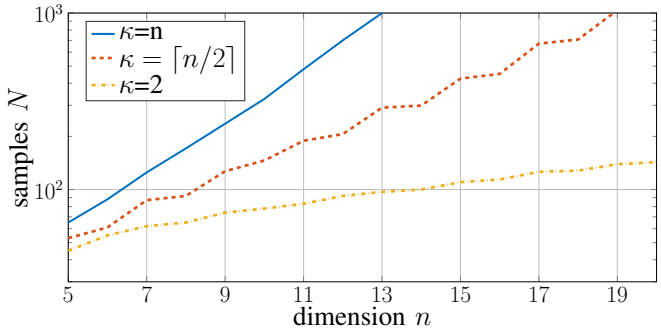


Fig. 3. Sample complexity of identifying the Jordan block $J_n(0.5)$ of size n and eigenvalues all 0.5, for different values of the controllability index. The figure shows the minimum number of samples N such that the (empirical) average error $\mathbb{E}\|A - \hat{A}_N\|_2$ is less than 0.005. The sample complexity appears to be increasing exponentially with the dimension n for $\kappa = \Theta(n)$. For $\kappa = 2$, the sample complexity is much smaller and increases polynomially.

linear systems, summarized in Figure 4 according to the controllability index κ of the considered system class. While statistically easy cases that were previously known are captured by Theorem 1, we also showed that hard system classes exist (Theorem 3). By exploiting structural system theoretic properties, such as the controllability index, we broadened the class of easy to learn linear systems (Theorem 4).

Our results pose numerous future questions for exploiting other system properties (e.g. observability) for efficiently learning classes of partially-observed linear systems or nonlinear systems. It remains an open problem to prove whether or not the $n-$th order integrator is poly-learnable as discussed in Section VI. In this paper, we studied when system identification is hard. Another important problem is when learning-based control is hard. The main question is whether/when hardness of system identification implies hardness of learning-based control. It might be possible that for some control tasks we might not need to identify the whole system in order to control it.

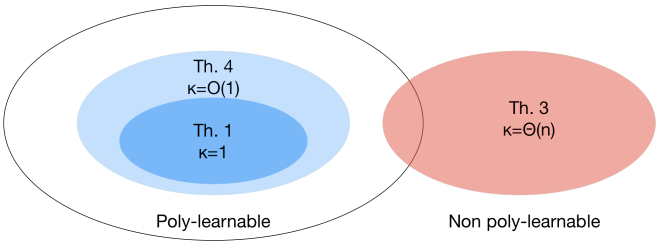


Fig. 4. Sample complexity classes for linear systems according to their controllability index.

REFERENCES

[1] L. Ljung, *System Identification: Theory for the User*. Prentice Hall, 1999.

[2] M. Deistler, K. Peternell, and W. Scherrer, "Consistency and relative efficiency of subspace methods," *Automatica*, vol. 31, no. 12, pp. 1865–1875, 1995.

[3] D. Bauer, M. Deistler, and W. Scherrer, "Consistency and asymptotic normality of some subspace algorithms for systems without observed inputs," *Automatica*, vol. 35, no. 7, pp. 1243–1254, 1999.

[4] A. Chiuso and G. Picci, "The asymptotic variance of subspace estimates," *Journal of Econometrics*, vol. 118, no. 1-2, pp. 257–291, 2004.

[5] E.-W. Bai and S. S. Sastry, "Persistency of excitation, sufficient richness and parameter convergence in discrete time adaptive control," *Systems & control letters*, vol. 6, no. 3, pp. 153–163, 1985.

[6] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. Cambridge University Press, 2018, vol. 47.

[7] M. C. Campi and E. Weyer, "Finite sample properties of system identification methods," *IEEE Transactions on Automatic Control*, vol. 47, no. 8, pp. 1329–1334, 2002.

[8] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "On the sample complexity of the linear quadratic regulator," *arXiv preprint arXiv:1710.01688*, 2017.

[9] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht, "Learning Without Mixing: Towards A Sharp Analysis of Linear System Identification," *arXiv preprint arXiv:1802.08334*, 2018.

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

[11] T. Sarkar and A. Rakhlin, "Near optimal finite time identification of arbitrary linear dynamical systems," *arXiv preprint arXiv:1812.01251*, 2018.

[12] S. Fattahi, N. Matni, and S. Sojoudi, "Learning sparse dynamical systems from a single sample trajectory," *arXiv preprint arXiv:1904.09396*, 2019.

[13] Y. Jedra and A. Proutiere, "Sample complexity lower bounds for linear system identification," in *IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 2676–2681.

[14] A. Wagenmaker and K. Jamieson, "Active learning for identification of linear dynamical systems," in *Conference on Learning Theory*. PMLR, 2020, pp. 3487–3582.

[15] S. Oymak and N. Ozay, "Non-asymptotic Identification of LTI Systems from a Single Trajectory," *arXiv preprint arXiv:1806.05722*, 2018.

[16] T. Sarkar, A. Rakhlin, and M. A. Dahleh, "Finite-Time System Identification for Partially Observed LTI Systems of Unknown Order," *arXiv preprint arXiv:1902.01848*, 2019.

[17] M. Simchowitz, R. Boczar, and B. Recht, "Learning Linear Dynamical Systems with Semi-Parametric Least Squares," *arXiv preprint arXiv:1902.00768*, 2019.

[18] A. Tsiamis and G. J. Pappas, "Finite Sample Analysis of Stochastic System Identification," in *IEEE 58th Conference on Decision and Control (CDC)*, 2019.

[19] B. Lee and A. Lamperski, "Non-asymptotic Closed-Loop System Identification using Autoregressive Processes and Hankel Model Reduction," *arXiv preprint arXiv:1909.02192*, 2019.

[20] Y. Zheng and N. Li, "Non-asymptotic identification of linear dynamical systems using multiple trajectories," *IEEE Control Systems Letters*, vol. 5, no. 5, pp. 1693–1698, 2020.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:56:41 UTC from IEEE Xplore. Restrictions apply.

[21] H. Lee, "Improved rates for identification of partially observed linear dynamical systems," *arXiv preprint arXiv:2011.10006*, 2020.
[22] S. Lale, K. Azizzadenesheli, B. Hassibi, and A. Anandkumar, "Logarithmic regret bound in partially observable linear dynamical systems," *arXiv preprint arXiv:2003.11227*, 2020.
[23] P. Rashidinejad, J. Jiao, and S. Russell, "Slip: Learning to predict in unknown dynamical systems with long-term memory," *arXiv preprint arXiv:2010.05899*, 2020.
[24] N. Matni and S. Tu, "A tutorial on concentration bounds for system identification," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 3741–3749.
[25] A. Carè, B. C. Csáji, M. C. Campi, and E. Weyer, "Finite-sample system identification: An overview and a new correlation method," *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 61–66, 2018.
[26] B. Recht, "A tour of reinforcement learning: The view from continuous control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 2, no. 1, pp. 253–279, 2019.
[27] S. S. Du, S. M. Kakade, R. Wang, and L. F. Yang, "Is a good representation sufficient for sample efficient reinforcement learning?" *arXiv preprint arXiv:1910.03016*, 2019.
[28] G. Baggio, V. Katewa, and F. Pasqualetti, "Data-driven minimum-energy controls for linear systems," *IEEE Control Systems Letters*, vol. 3, no. 3, pp. 589–594, 2019.
[29] X. Chen and E. Hazan, "Black-box control for linear dynamical systems," in *Conference on Learning Theory*. PMLR, 2021, pp. 1114–1143.
[30] A. Tsiamis and G. J. Pappas, "Linear systems can be hard to learn," *arXiv preprint arXiv:2104.01120*, 2021.
[31] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.
[32] C. Dann, T. Lattimore, and E. Brunskill, "Unifying PAC and regret: Uniform PAC bounds for episodic reinforcement learning," *arXiv preprint arXiv:1703.07710*, 2017.
[33] R. A. Horn and C. R. Johnson, *Matrix analysis*, 2nd ed. Cambridge University Press, 2012.
[34] R. Eising, "Between controllable and uncontrollable," *Systems & Control Letters*, vol. 4, no. 5, pp. 263–264, 1984.
[35] P. M. V. Dooren, "Numerical linear algebra for signals systems and control," *Draft notes prepared for the Graduate School in Systems and Control*, 2003.
[36] D. Kulkarni, D. Schmidt, and S.-K. Tsui, "Eigenvalues of tridiagonal pseudo-Toeplitz matrices," *Linear Algebra and its Applications*, vol. 297, no. 1, pp. 63–80, 1999.

APPENDIX

We briefly review the concept of controllability and other related concepts. We consider the pair (A, H), but the same definitions hold also for (A, B). The controllability matrix of (A, H) is defined as

$$\mathcal{C}_k(A,H) \triangleq \begin{bmatrix} H & AH & \cdots & A^{k-1}H \end{bmatrix}, k \geq 1.$$

The pair (A, H) is *controllable* when the controllability matrix $\mathcal{C}_n(A, H)$ has full column rank n. The *controllability Gramian* at time k is defined as :

$$\Gamma_k(A,H) \triangleq \mathcal{C}_k(A,H)\mathcal{C}_k'(A,H) = \sum_{i=0}^{k-1} A^i H H' (A')^i.$$

Common systems like the integrator are not only controllable but also robustly controllable. Their distance to uncontrollability degrades linearly with the dimension n.

Lemma 3. *System* (9) *is μ-bounded away from uncontrollability with $\mu^{-1} \leq \rho^{-1}(n+1)$.*

The proof follows from the fact that the distance form uncontrollability is equivalently given by the formula [34]:

$$d(A,H) = \inf_{s \in \mathbb{C}} \sigma_{\min}(\begin{bmatrix} A - sI & H \end{bmatrix}), \tag{10}$$

and results about the eigenvalues of Toeplitz matrices [36].

If H is not a column matrix, the full column rank condition might be satisfied earlier for some $k \leq n$. The minimum time that we achieve controllability is the *controllability index*:

$$\kappa(A,H) \triangleq \min\{k \geq 1 : \operatorname{rank}(\mathcal{C}_k(A,H)) = n\}. \tag{11}$$

It is the lag between the time the disturbance w_t is applied and the time $t+\kappa$ by which we see the effect of that disturbance in all states. This lag is non-trivial if the number of disturbances $r < n$ is smaller than the number of states; in this case we call the system underactuated.

Based on the fact that the rank of the controllability matrix at time κ is n, we can show that the pair (A, H) admits the following canonical representation, under a unitary similarity transformation [35].

Proposition 1 (Staircase form)**.** *Consider a controllable pair (A, H) with controllability index κ and controllability matrix $\mathcal{C}_k$, $k \geq 0$. There exists a unitary similarity transformation U such that $U'U = UU' = I$ and:*

$$U'H = \begin{bmatrix} H_1' & 0 & \cdots & 0 \end{bmatrix}'$$
$$U'AU = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,\kappa-1} & A_{1,\kappa} \\ A_{2,1} & A_{2,2} & \cdots & A_{3,\kappa-1} & A_{2,\kappa} \\ 0 & A_{3,2} & \cdots & A_{3,\kappa-1} & A_{3,\kappa} \\ 0 & 0 & \cdots & A_{4,\kappa-1} & A_{4,\kappa} \\ \vdots & & & \vdots & \\ 0 & 0 & \cdots & A_{\kappa,\kappa-1} & A_{\kappa,\kappa} \end{bmatrix}, \tag{12}$$

where $A_{i,j} \in \mathbb{R}^{r_i \times r_j}$ are block matrices, with $r_i = \operatorname{rank}(\mathcal{C}_i) - \operatorname{rank}(\mathcal{C}_{i-1})$, $r_1 = r$, $H_1 \in \mathbb{R}^{r \times r}$. Moreover, the matrices $A_{i+1,i}$ have full row rank $\operatorname{rank}(A_{i+1,i}) = r_{i+1}$ and the sequence r_i is decreasing.

The above representation is useful as it captures the coupling between the several sub-states via the matrices $A_{i+1,i}$. The proof of Theorem 5 exploits this idea. If a system is robustly controllable then these matrices are bounded away from being row-rank deficient.

Lemma 4 (Staircase form lower bound)**.** *Let $(A, H) \in \mathbb{R}^{n \times (n+r)}$ be controllable and let Assumption 2 hold. Consider the staircase form of (A, H), with $A_{i+1,i}$ the sub-diagonal matrices, for $i = 1, \ldots, \kappa - 1$, where κ is the controllability index. Then, we have $A_{i+1,i}A_{i+1,i}' \succeq \mu^2 I_{r_{i+1}}$ for all $i = 1, \ldots, \kappa - 1$.*

The proof of Theorem 5 exploits the above result along with the triangular structure of (12), the fact that the similarity transformation is unitary, and the shift structure of the controllablity matrix. The details can be found in the arXiv version of the paper.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:56:41 UTC from IEEE Xplore. Restrictions apply.