Journal of Machine Learning Research 23 (2022) 1-30 Submitted 7/20; Published 2/22

Active Learning for Nonlinear System Identification with Guarantees

Horia Mania HMANIA@BERKELEY.EDU
Michael I. Jordan JORDAN@CS.BERKELEY.EDU
Benjamin Recht BRECHT@BERKELEY.EDU
Department of Electrical Engineering and Computer Science
University of California
Berkeley, CA 94720-1776, USA

Editor: Kenji Fukumizu

Abstract

While the identification of nonlinear dynamical systems is a fundamental building block of model-based reinforcement learning and feedback control, its sample complexity is only understood for systems that either have discrete states and actions or for systems that can be identified from data generated by i.i.d. random inputs. Nonetheless, many interesting dynamical systems have continuous states and actions and can only be identified through a judicious choice of inputs. Motivated by practical settings, we study a class of nonlinear dynamical systems whose state transitions depend linearly on a known feature embedding of state-action pairs. To estimate such systems in finite time identification methods must explore all directions in feature space. We propose an active learning approach that achieves this by repeating three steps: trajectory planning, trajectory tracking, and re-estimation of the system from all available data. We show that our method estimates nonlinear dynamical systems at a parametric rate, similar to the statistical rate of standard linear regression.

Keywords: nonlinear dynamical systems, system identification, least squares, control theory

1. Introduction

The estimation of nonlinear dynamical systems with continuous states and inputs is generally based on data-collection procedures inspired by the study of optimal input design for linear dynamical systems (Schoukens and Ljung, 2019). Unfortunately, these data-collection methods are not sufficient in general to enable the estimation of nonlinear systems. To attempt to circumvent this issue, studies of system identification have either assumed that the available data is informative enough for estimation (Hong et al., 2008; Ljung, 1987; Schoukens and Ljung, 2019) or considered systems for which i.i.d. random inputs produce informative data (Bahmani and Romberg, 2020; Foster et al., 2020; Oymak, 2019; Sattar and Oymak, 2020). However, as we will see, there are many nonlinear dynamical systems that require a more judicious choice of inputs for estimation to be possible.

Inspired by experimental design and active learning, we present a data-collection scheme that is guaranteed to enable system identification in finite time. Our method applies to dynamical systems whose transitions depend linearly on a known feature embedding of state-input pairs. This class of models can capture many types of dynamics and is used widely in system identification (Hong et al., 2008; Ljung, 1987). For example, Ng et al. (2006) used such a model to estimate the dynamics of a helicopter and Brunton et al. (2016) showed that sparse linear regression of polynomial and trigonometric feature embeddings can be used to fit models of the chaotic Lorentz system and of a

fluid shedding behind an obstacle. These models can be parametrized as follows:

$$\mathbf{x}_{t+1} = A_\star \phi(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{w}_t, \tag{1}$$

where $\mathbf{x}_t$ and $\mathbf{u}_t$ are the state and input of the system at time t, and $\mathbf{w}_t$ is stochastic noise. The feature map ϕ is assumed known and the goal is to estimate $A_\star$ from one trajectory by choosing a good sequence of inputs. The input $\mathbf{u}_t$ is allowed to depend on the history of states $\{\mathbf{x}_j\}_{j=0}^t$ and is independent of $\mathbf{w}_t$.

The class of systems (1) contains any linear system, with fully observed states, when the features include the states and inputs of the system. Moreover, any piecewise affine (PWA) system can be expressed using (1) if the support of its pieces is known. First introduced by Sontag (1981) as an approximation of nonlinear systems, PWA systems are a popular model of hybrid systems (Borrelli et al., 2017; Camacho et al., 2010; Heemels et al., 2001) and have been successfully used in a wide range of applications (Borrelli et al., 2006; Geyer et al., 2008; Han and Tedrake, 2017; Marcucci et al., 2017; Sadraddini and Tedrake, 2019; Yordanov et al., 2011).

While linear dynamical systems can be estimated from trajectories induced by i.i.d. random inputs (Simchowitz et al., 2018), the following example shows that this is not possible for PWA systems.

Example 1. Let us consider the feature map $\phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{3d}$ defined by:

$$\phi(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} \mathbf{x} \cdot \mathbb{1}\{\|\mathbf{x}\| \leq \frac{3}{2}\} \\ \mathbf{x} \cdot \mathbb{1}\{\|\mathbf{x}\| > \frac{3}{2}\} \\ \mathbf{u} \cdot \mathbb{1}\{\|\mathbf{u}\| \leq 1\} \end{bmatrix},$$

where $\mathbb{1}\{\cdot\}$ is the indicator function and the multiplication with $\mathbf{x}$ is coordinatewise. We assume there is no process noise and let $A_\star = \begin{bmatrix} \frac{1}{2} I_d & A_2 & I_d \end{bmatrix}$ for some $d \times d$ matrix A_2 and the $d \times d$ identity matrix I_d. Also, we assume $\mathbf{x}_0 = 0$.

Then, since the inputs to the system can have magnitude at most 1, the state of the system can have magnitude larger than $3/2$ only if consecutive inputs point in the same direction. However, the probability that two or more random vectors, uniformly distributed on the unit sphere, point in the same direction is exponentially small in the dimension d. Therefore, if we used random inputs, we would have to wait for a long time in order to reach a state with magnitude larger than $3/2$.

On the other hand, if we chose a sequence of inputs $\mathbf{u}_t = \mathbf{u}$ for a fixed unit vector $\mathbf{u}$, we would be guaranteed to reach a state with norm larger than $3/2$ in a couple of steps. Hence, despite the input constraint, we would be able to reach the region $\|\mathbf{x}\| > 3/2$ with a good choice of inputs. ■

Therefore, in general the estimation of (1) requires a judicious choice of inputs. To address this challenge we propose a method based on trajectory planning, which, at a high level, repeats three steps:

- Given past observations and an estimate $\widehat{A}$, our method plans a reference trajectory from the current state of the system to a high uncertainty region of the feature space.
- Then, our method attempts to track the reference trajectory using $\widehat{A}$.
- Finally, using all data collected so far, our method re-estimates $\widehat{A}$.

The ability to find reference trajectories from a given state to a desired goal set is related to the notion of controllability, a standard notion in control theory. A system is called *controllable* if it is possible to take the system from any state to any other state in a finite number of steps by using an appropriate sequence of inputs. In our case, a system is considered more controllable the bigger we can make the inner product between the system's features and goal directions in feature space. The number of time steps required to obtain a large inner product is called the *planning horizon*.

The controllability of the system and the planning horizon are system-dependent properties that influence our ability to estimate the system. Intuitively, the more controllable a system is, the easier it is to collect the data we need to estimate it. The following informal version of our main result clarifies this relationship.

Theorem 1 (Informal) *Our method chooses actions* $\mathbf{u}_t$ *such that with high probability the ordinary least squares (OLS) estimate* $\widehat{A} \in \arg\min_A \sum_{t=0}^{T-1} \|A\phi(\mathbf{x}_t, \mathbf{u}_t) - \mathbf{x}_{t+1}\|^2$ *satisfies*

$$\|\widehat{A} - A_\star\| \leq \frac{\textit{size of the noise}}{\textit{controllability of the system}} \sqrt{\frac{\textit{dimension} \times \textit{planning horizon}}{\textit{number of data points}}}.$$

This statistical rate is akin to that of standard supervised linear regression, but it has an additional dependence on the controllability of the system and the planning horizon. To better understand why these two terms appear, recall that our method uses $\widehat{A}$, an estimate of $A_\star$, to plan and track reference trajectories. Therefore, the tracking step is not guaranteed to reach the desired region of the feature space. The main insight of our analysis is that when trajectory tracking fails, we are still guaranteed to collect at least one informative data point per reference trajectory. Therefore, in the worst case, the effective size of the data collected by our method is equal to the total number of data points collected over the planning horizon.

In the next section we present our mathematical assumptions and in Section 3 we discuss our method and main result. Section 4 includes a general result derived from prior work concerning linear regression with dependent data. Then, in Section 5 we present in detail the proof of our main result. There is a long line of work studying system identification, which we discuss in Section 6. Finally, Section 7 contains takeaways and open problems.

Notation: The norm $\|\cdot\|$ is the Euclidean norm whenever it is applied to vectors and is the spectral norm whenever it is applied to matrices. We use c_1, c_2, c_3, ... to denote different universal constants. Also, $\mathbb{S}^{p-1}$ is the unit sphere in $\mathbb{R}^p$ and $\mathbb{B}_r^p$ is the ball in $\mathbb{R}^p$ centered at the origin and of radius r. The symbol ■ is used to indicate the end of an example or of a proof. To aid the reader Appendix A lists the definitions of our notation and also discusses the universal constants we use.

2. Assumptions

To guarantee the estimation of (1) we must make several assumptions about the true system we are trying to identify. We denote the dimensions of the states and inputs by d and p respectively. The feature map ϕ maps state-action pairs to feature vectors in $\mathbb{R}^k$.

The main challenge in the estimation of (1) is choosing inputs $\mathbf{u}_t$ so that the minimal singular value of the design matrix is $\Omega(\sqrt{T})$, where T is the length of the trajectory collected from the system. To reliably achieve this we must assume the feature map ϕ has some degree of smoothness. Without a smoothness assumption the noise term $\mathbf{w}_t$ at time t might affect the feature vector $\phi(\mathbf{x}_{t+1}, \mathbf{u}_{t+1})$ at time $t+1$ in arbitrary ways, regardless of the choice of input at time t.

Assumption 1 *The map $\phi\colon \mathbb{R}^d \times \mathbb{B}^p_{r_u} \to \mathbb{R}^k$ is L-Lipschitz.*[1]

In order to use known techniques for the analysis of online linear least squares (Abbasi-Yadkori et al., 2011; Dani et al., 2008; Rusmevichientong and Tsitsiklis, 2010; Simchowitz et al., 2018) we also assume that the feature map ϕ is bounded. For some classes of systems (e.g., certain linear systems) this condition can be removed (Simchowitz et al., 2018).

Assumption 2 *There exists $b_\phi > 0$ such that $\|\phi(\mathbf{x}, \mathbf{u})\| \le b_\phi$ for all $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{u} \in \mathbb{B}_{r_u}$.*

This assumption implies that the states of the system (1) are bounded, a consequence which can be limiting in some applications. To address this issue we could work instead with the system

$$\mathbf{x}_{t+1} = A_\star \phi(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{x}_t + \mathbf{w}_t. \tag{2}$$

In this case, ϕ being bounded implies that the increments $\mathbf{x}_{t+1} - \mathbf{x}_t$ are bounded, allowing the states to grow in magnitude. However, formulation (2) complicates the exposition so we choose to focus on (1).

As mentioned in the introduction, our method relies on trajectory planning and tracking to determine the inputs to the system. Suppose we would like to track a reference trajectory $\{(\mathbf{x}^R_t, \mathbf{u}^R_t)\}_{t\ge 0}$ that satisfies $\mathbf{x}^R_{t+1} = A_\star \phi(\mathbf{x}^R_t, \mathbf{u}^R_t)$. In other words, we wish to choose inputs $\mathbf{u}_t$ to ensure that the tracking error $\|\mathbf{x}_t - \mathbf{x}^R_t\|$ is small. Simply choosing $\mathbf{u}_t = \mathbf{u}^R_t$ does not work even when the initial states $\mathbf{x}_0$ and $\mathbf{x}^R_0$ are equal because the true system (1) experiences process noise.

To ensure that tracking is possible we assume that there always exists an input to the true system that can keep the tracking error small. There are multiple ways to formalize such an assumption. We make the following choice.

Assumption 3 *There exist positive constants γ and b_u such that for any $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$ and any $\mathbf{u}' \in \mathbb{B}^p_{b_u}$ we have*

$$\min_{\mathbf{u} \in \mathbb{B}^p_{r_u}} \|A_\star \left(\phi(\mathbf{x}, \mathbf{u}) - \phi(\mathbf{x}', \mathbf{u}')\right)\| \le \gamma \|\mathbf{x} - \mathbf{x}'\|. \tag{3}$$

Moreover, if $\|\mathbf{u}'\| \le b_u/2$, there exists $\mathbf{u}$, with $\|\mathbf{u}\| \le b_u$, that satisfies (3)*.*

Suppose we wish to track a trajectory $\{(\mathbf{x}^R_t, \mathbf{u}^R_t)\}_{t\ge 0}$ that satisfies $\mathbf{x}_{t+1} = A_\star \phi(\mathbf{x}_t, \mathbf{u}_t)$. Then, Assumption 3 guarantees the existence of an input $\mathbf{u}_t \in \mathbb{B}^p_{b_u}$ such that

$$\begin{aligned}\|\mathbf{x}_{t+1} - \mathbf{x}^R_{t+1}\| &= \|A_\star \phi(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{w}_t - A_\star \phi(\mathbf{x}^R_t, \mathbf{u}^R_t)\| \\ &\le \gamma \|\mathbf{x}_t - \mathbf{x}^R_t\| + \|\mathbf{w}_t\|.\end{aligned}$$

In other words, Assumption 3 allows us to find an input $\mathbf{u}_t$ such that the tracking error $\|\mathbf{x}_{t+1} - \mathbf{x}^R_{t+1}\|$ is upper bounded in terms of noise $\mathbf{w}_t$ and the tracking error at time t. By induction, Assumption 3 guarantees the existence of inputs to the system such that

$$\|\mathbf{x}_H - \mathbf{x}^R_H\| \le \max_{t=0,\ldots,H-1} \|\mathbf{w}_t\| (1 + \gamma + \ldots + \gamma^{H-1}) + \gamma^H \|\mathbf{x}_0 - \mathbf{x}^R_0\|.$$

1. Since ϕ is continuous and since $\mathbf{u}$ lies in a compact set, we know that any continuous function of $\phi(\mathbf{x}, \mathbf{u})$ achieves its maximum and minimum with respect to $\mathbf{u}$. This is the only reason we assume the inputs to the system are bounded. Alternatively, we could let the inputs be unbounded and work with approximate maximizers and minimizers.

Hence, when $\gamma < 1$ we can choose a sequence of inputs such that the state $\mathbf{x}_H$ at time H is close to $\mathbf{x}_H^R$, as long as the process noise is well behaved.

Note that in Assumption 3 we allow $\gamma \geq 1$. However, we pay a price when γ is large. The larger γ is the more stringent the following assumptions become. Finally, we note that the parameter b_u appearing in Assumption 3 makes it easier for systems to satisfy the assumption than requiring that (3) holds for all $\mathbf{u}'$.

To estimate (1) we must collect measurements of state transitions from feature vectors that point in different directions. To ensure that such data can be collected from the system we must assume that there exist sequences of actions which take the dynamical system from a given state to some desired direction in feature space. This type of assumption can be formulated in terms of controllability. Recall that a linear system $\mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t$ is said to be controllable when the matrix $\begin{bmatrix} B & AB & \dots & A^{d-1}B \end{bmatrix}$ has full row rank. It can be easily checked that for a controllable linear system it is possible to get from any state to any other state in d steps by appropriately choosing a sequence of inputs. Moreover, this notion of controllability can be extended to a class of nonlinear systems, called control affine systems, through the use of Lie brackets (Sastry, 1999; Slotine and Li, 1991). We require, however, a different notion of controllability. In particular, we assume that in the absence of process noise we can take the system (1) from any state to a feature vector that aligns sufficiently with a desired direction in feature space.

Assumption 4 *There exist α and H, a positive real number and a positive integer, such that for any initial state $\mathbf{x}_0$ and goal vector $v \in \mathbb{S}^{k-1}$ there exists a sequence of actions $\mathbf{u}_t$, with $\|\mathbf{u}_t\| \leq b_u/2$, such that $|\langle \phi(\mathbf{x}_t, \mathbf{u}_t), v \rangle| \geq \alpha > 0$ for some $0 \leq t \leq H$, with $\mathbf{x}_{j+1} = A_\star \phi(\mathbf{x}_j, \mathbf{u}_j)$ for all j.*

While we have not seen this notion of controllability anywhere else in the literature, it captures for nonlinear systems an important aspect of the controllability of linear systems. Namely, a linear system is controllable if it is possible to go from any state to any other state in d steps, using an appropriate sequence of actions. Similarly, a nonlinear system satisfies Assumption 4 if it is possible to go from any state to any direction in feature space in H steps. If the assumption is satisfied for some horizon H, it is clear that it is also satisfied for larger horizons. Moreover, one expects that a larger horizon H allows a larger controllability parameter α. As discussed in the introduction, the larger H is, the weaker our guarantee on estimation will be. However, the larger α is, the better our guarantee on estimation will be. Therefore, there is a tension between α and H in our final result. It is also worth noting that a nonlinear system (1) that does not satisfy Assumption 4 would not permit the estimation of $A_\star$ because it would not be possible to explore certain directions in feature space.

Assumptions 1 to 4 impose many constraints. Therefore, it is important to give examples of nonlinear dynamical systems that satisfy these assumptions. We give two simple examples. First we present a synthetic example for which it is easy to check that it satisfies all the assumptions, and then we discuss the simple pendulum.

Example 2. Smoothed Piecewise Linear System When the support sets of the different pieces are known, piecewise affine systems can be easily expressed as (1). However, the feature map ϕ would not be continuous. In this example, we present a smoothed version of a PWA system, which admits a 1-Lipschitz feature map. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 0 & \text{if } x < -1/2, \\ x + 1/2 & \text{if } x \in [-1/2, 1/2], \\ 1 & \text{if } x > 1/2. \end{cases}$$

We also consider the maps $g(\mathbf{x}) = \frac{\mathbf{x}_t}{\|\mathbf{x}_t\|}\min\{\|\mathbf{x}_t\|, b_x\}$ and $h(\mathbf{u}) = \frac{\mathbf{u}}{\|\mathbf{u}\|}\min\{\|\mathbf{u}\|, r_u\}$, for some values b_x and r_u. The maps g and h simply clip the size of the states and inputs. Here both the inputs and the states are d-dimensional. Then, we define the feature map $\phi : \mathbb{R}^{2d} \to \mathbb{R}^{3d}$ as follows:

$$\phi(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} g(\mathbf{x})f(x_1) \\ g(\mathbf{x})(1 - f(x_1)) \\ h(\mathbf{u}) \end{bmatrix},$$

where x_1 denotes the first coordinate of $\mathbf{x}$. Now, let us consider the following dynamical system:

$$\mathbf{x}_{t+1} = \begin{bmatrix} A_1 & A_2 & I_d \end{bmatrix} \phi(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{w}_t, \tag{4}$$

where A_1 and A_2 are two unknown $d \times d$ matrices. Since Assumptions 1-4 have nothing to do with the process noise, in this example we assume $\mathbf{w}_t$ is zero almost surely.

To better understand the system (4) note that when $\|\mathbf{x}_t\| \le b_x$ and $\|\mathbf{u}_t\| \le r_u$ we have

$$\begin{aligned} \mathbf{x}_{t+1} &= A_1\mathbf{x}_t + \mathbf{u}_t \quad \text{if } \; x_{t1} \ge 1/2, \\ \mathbf{x}_{t+1} &= A_2\mathbf{x}_t + \mathbf{u}_t \quad \text{if } \; x_{t1} \le -1/2. \end{aligned}$$

By construction, the feature map of the system is 1-Lipschitz and bounded. Therefore, (4) satisfies Assumptions 1 and 2. We are left to show that we can choose A_1, A_2, b_x, and r_u so that (4) satisfies Assumptions 3 and 4 as well.

Now, let us choose $b_x = 4$. To ensure Assumptions 3 and 4 we choose r_u sufficiently large. More precisely, we choose $r_u = 32(\|A_1\| + \|A_2\|) + 4$ and $b_u = 16(\|A_1\| + \|A_2\|) + 4$, which implies that Assumption 3 is satisfied with $\gamma = 0$ (one can choose u to make the left hand side of (3) zero). Since $r_u \ge 4$, it means that in one time step the state $\mathbf{x}$ can take any value in a ball of radius 4 centered at the origin. This ball intersects with the three bands that determine the function f. Now, let $v = [v_1^\top, v_2^\top, v_3^\top]^\top$ be a $3d$-dimensional unit vector with v_i d-dimensional. Then, the maximum of $\|v_1\|$, $\|v_2\|$, and $\|v_3\|$ is at least $1/\sqrt{3}$. If $\|v_1\| \ge 1/\sqrt{3}$, we can choose $\mathbf{x}$ with $x_{t1} \ge 1/2$, $\|\mathbf{x}\| \le 4$, and $\mathbf{u} = 0$ such that $|\langle v, \phi(\mathbf{x}, \mathbf{u})\rangle| \ge 4 \cdot 1/2 \cdot 1/\sqrt{3} = 2/\sqrt{3}$. The cases $\|v_2\| \ge 1/\sqrt{3}$ and $\|v_3\| \ge 1/\sqrt{3}$ are treated similarly. Therefore, Assumption 4 also holds.

■

Example 3. Simple Pendulum The dynamics of a simple pendulum are described in continuous time by the equation

$$m\ell^2\ddot{\theta}(t) + mg\ell\sin\theta(t) = -b\dot{\theta}(t) + u(t), \tag{5}$$

where $\theta(t)$ is the angle of the pendulum at time t, m is the mass of the pendulum, ℓ is its length, b is a friction coefficient, and g is the gravitational acceleration.

Discretizing (5) according to Euler's method[2] with step size h and assuming stochastic process noise, we obtain the following two-dimensional system:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \begin{bmatrix} a_1 & a_2 \\ h & 0 \end{bmatrix} \begin{bmatrix} x_{t1} \\ \sin(x_{t2}) \end{bmatrix} + \begin{bmatrix} a_3 \\ 0 \end{bmatrix} \mathbf{u}_t + \mathbf{w}_t,$$

2. Using a more refined discretization method, such as a Runge-Kutta method, would be more appropriate. Unfortunately, such discretization methods yield a discrete-time system which cannot be easily put in the form (2).

where x_{t1} and x_{t2} are the coordinates of $\mathbf{x}_t$ and a_1, a_2, and a_3 are unknown real values. The first coordinate of $\mathbf{x}_t$ represents the angular velocity of the pendulum at time t, while the second coordinate represents the angle of the pendulum. Therefore, to put the inverted pendulum in the form of (2) we consider the feature map

$$\phi(\mathbf{x}_t, \mathbf{u}_t) = \begin{bmatrix} x_{t1} \\ \sin(x_{t2}) \\ \mathbf{u}_t \end{bmatrix}.$$

It can be easily checked that this feature map is 1-Lipschitz. While it is not bounded, if the pendulum experiences friction, we can ensure the feature values stay bounded by clipping the inputs $\mathbf{u}_t$; i.e., we replace $\mathbf{u}_t$ with $\operatorname{sgn}(\mathbf{u}_t)\min\{|\mathbf{u}_t|, r_u\}$ for some value r_u.

The simple pendulum satisfies Assumption 4 because we can drive the system in a finite number of steps from any state $\mathbf{x}_t$ to states $\mathbf{x}_{t+H}$ for which the signs of $x_{(t+h)1}$ and $\sin(x_{t1})$ can take any value in $\{-1, 1\}^2$, with their absolute values lower bounded away from zero.

Finally, to check Assumption 3 note that the input $\mathbf{u}$ can impact directly only the first coordinate of the state. Hence, in (3) we can choose $\mathbf{u}$ such that the left hand side depends only on the second coordinates of the states. Since the sine function is 1-Lipschitz, we can choose $\gamma = 1 + h$ to ensure that (3) is satisfied. Assumption 3 is pessimistic because it only considers one step control and the simple pendulum is stabilizable and can track reference trajectories. However, Assumption 3 does not hold with $\gamma < 1$ since the input at time t does not affect the position at time $t + 1$. ■

We now turn to our final two assumptions. We need to make an assumption about the process noise and we also must assume access to an initial $\widehat{A}$ to warm start our method.

Assumption 5 *The random vectors $\mathbf{w}_t$ are independent, zero mean, and $\|\mathbf{w}_t\| \le b_w$ almost surely.*[3] *Also, $\mathbf{w}_t$ is independent of $(\mathbf{x}_t, \mathbf{u}_t)$. Furthermore, we assume*

$$b_w \le \frac{\alpha}{c_1 L(1 + \gamma + \ldots + \gamma^{H-1})}, \tag{6}$$

for some universal constant $c_1 > 4$.

Equation 6 imposes on upper bound on the size of the process noise in terms of system-dependent quantities: the controllability parameter α introduced in Assumption 4, the Lipschitz constant L of the feature map, and the control parameter γ introduced in Assumption 3. An upper bound on b_w is required because when the process noise is too large, it can be difficult or impossible to counteract its effects through feedback. For example, even when the system is linear and stable if the process noise is large, the state would not reach a small neighborhood of the origin, which would prevent algorithm from collecting data in such a neighborhood. Therefore, the size of the process noise needs to be small enough in relation to the system in order to make the task possible.

Finally, we assume access to an initial guess $\widehat{A}$ that is sufficiently close to $A_\star$. Namely, we require an initial estimate $\widehat{A}$ such that $\|\widehat{A} - A_\star\| = \mathcal{O}\left(L^{-1}(1 + \gamma + \ldots + \gamma^{H-1})^{-1}\right)$. To understand the key issue this assumption resolves, suppose we are trying to track a reference trajectory $\{(\mathbf{x}_t^R, \mathbf{u}_t^R)\}_{t\ge 0}$ and $\|(\widehat{A} - A_\star)\phi(\mathbf{x}_t^R, \mathbf{u}_t^R)\|$ is large. Without an assumption on the size of $\|\widehat{A} - A_\star\|$,

3. We can relax this assumption to only require $\mathbf{w}_t$ to be sub-Gaussian. In this case, we would make a truncation argument to obtain an upper bound on all $\mathbf{w}_t$ with high probability.

the magnitude of $(\widehat{A} - A_\star)\phi(\mathbf{x}_t^R, \mathbf{u}_t^R)$ might be large while $\|\phi(\mathbf{x}_t^R, \mathbf{u}_t^R)\|$ is small. Then, making a measurement at a point $(\mathbf{x}_t, \mathbf{u}_t)$ close to $(\mathbf{x}_t^R, \mathbf{u}_t^R)$ might not be helpful for estimation because $\phi(\mathbf{x}_t, \mathbf{u}_t)$ could be zero. Therefore, if $\|\widehat{A} - A_\star\|$ is too large, we might both fail to track a reference trajectory and to collect a useful measurement. For ease of exposition, instead of assuming access to an initial guess $\widehat{A}$, we assume access to a data set.

Assumption 6 *We have access to an initial trajectory $\mathcal{D} = \{(\mathbf{x}_t, \mathbf{u}_t, \mathbf{x}_{t+1})\}_{0 \le t < t_0}$ of transitions from the true system such that*

$$\lambda_{\min}\left(\sum_{t=0}^{t_0-1} \phi(\mathbf{x}_t, \mathbf{u}_t)\phi(\mathbf{x}_t, \mathbf{u}_t)^\top\right) \ge 1 + c_2 b_w^2 L^2 \left(\sum_{i=0}^{H-1} \gamma^i\right)^2 \left(d + k\log(b_\phi^2 T) + \log\left(\frac{\pi^2 T^2}{6\delta}\right)\right),$$

where c_2 is a sufficiently large universal constant (see Section 5.2 for a precise bound on c_2) and T is the number of samples to be collected by our method. In Appendix D we show how to replace T by a fixed quantity $T_\star$.

As shown in Section 4, Assumption 6 guarantees that the OLS estimate $\widehat{A}$ obtained from $\mathcal{D}$ satisfies $\|\widehat{A} - A_\star\| \le \frac{c_3}{\sqrt{c_2}} L^{-1}(1+\gamma+\ldots+\gamma^{H-1})^{-1}$ for some universal constant c_3. Since the features $\phi(\mathbf{x}, \mathbf{u})$ can have magnitude as large as b_ϕ, Assumption 6 only implies $\|(\widehat{A} - A_\star)\phi(\mathbf{x}, \mathbf{u})\| = \mathcal{O}(b_\phi L^{-1}(1+\gamma+\ldots+\gamma^{H-1})^{-1})$. Therefore, Assumption 6 does not imply a stringent upper bound on $\|(\widehat{A} - A_\star)\phi(\mathbf{x}, \mathbf{u})\|$ because b_ϕ can be arbitrarily large relative to L and γ.

3. Main Result

Our method for estimating the parameters of a dynamical system (1) is shown in Algorithm 1. The trajectory planning and tracking routines are discussed in detail in Sections 3.1 and 3.2, respectively. Our method is also presented in one block of pseudo-code in Appendix E. We now state our main result.

Theorem 1 *Suppose $\mathbf{x}_{t+1} = A_\star\phi(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{w}_t$ is a nonlinear dynamical system which satisfies Assumptions 1-5 and suppose $\mathcal{D}$ is an initial trajectory that satisfies Assumption 6. Also, let $\beta = c_4\left(d + k\log(\beta_\phi^2 T) + \log(\pi^2 T^2/(6\delta))\right)^{-1}$ with $c_4 \le \frac{(c_1-4)^2}{144 c_3^2}$ and let:*[4]

$$N_e := \left\lceil \frac{2k\log\left(\frac{2kb_\phi^2}{\log(1+\beta/2)}\right)}{\log(1+\beta/2)} \right\rceil.$$

Then, with probability $1-\delta$, and given parameters T and β, Algorithm 1 outputs $\widehat{A}$ such that

$$\|\widehat{A} - A_\star\| \le c_5 \frac{b_w}{\alpha}\sqrt{\frac{d + k\log(b_\phi^2 T) + \log\left(\frac{\pi^2 T}{6\delta}\right)}{T/H - N_e}},$$

whenever $T \ge \frac{32kb_\phi^2 H}{\alpha^2} + HN_e$.

Algorithm 1 Active learning for nonlinear system identification

Require: Parameters: the feature map ϕ, initial trajectory $\mathcal{D}$, and parameters T, α, and β.

1: Initialize Φ to have rows $\phi(\mathbf{x}_j, \mathbf{u}_j)^\top$ and Y to have rows $(\mathbf{x}_{j+1})^\top$, for $(\mathbf{x}_j, \mathbf{u}_j, \mathbf{x}_{j+1}) \in \mathcal{D}$.
2: Set $\widehat{A} \leftarrow Y^\top \Phi(\Phi^\top \Phi)^{-1}$; i.e., the OLS estimate according to $\mathcal{D}$.
3: Set $t \leftarrow t_0$.
4: **while** $t \leq T + t_0$ **do**
5: Set $\mathbf{x}_0^R \leftarrow \mathbf{x}_t$,
6: Set v to be a minimal eigenvector of $\Phi^\top \Phi$, with $\|v\| = 1$.
7: **Trajectory planning:** find inputs $\mathbf{u}_0^R, \mathbf{u}_1^R, \ldots, \mathbf{u}_r^R$, with $\|\mathbf{u}_j^R\| \leq b_u$ and $r \leq H$, such that

$$\left|\langle \phi(\mathbf{x}_r^R, \mathbf{u}_r^R), v \rangle\right| \geq \frac{\alpha}{2} \text{ or } \phi(\mathbf{x}_r^R, \mathbf{u}_r^R)^\top (\Phi^\top \Phi)^{-1} \phi(\mathbf{x}_r^R, \mathbf{u}_r^R) \geq \beta,$$

where $\mathbf{x}_{j+1}^R = \widehat{A}\phi(\mathbf{x}_j^R, \mathbf{u}_j^R)$ for all $j \in \{0, 1, \ldots, r-1\}$.
8: **Trajectory tracking:** track the reference trajectory $\{(\mathbf{x}_j^R, \mathbf{u}_j^R)\}_{j=0}^r$ and increment t as described in Section 3.2.
9: Set $\Phi^\top \leftarrow [\phi_0, \phi_1, \ldots, \phi_{t-1}]$ and $Y^\top \leftarrow [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t]$, where $(\phi_j, \mathbf{x}_{j+1})$ are all feature-state transitions observed so far.
10: **Re-estimate:** $\widehat{A} \leftarrow Y^\top \Phi(\Phi^\top \Phi)^{-1}$.
11: **end while**
12: Output the last estimate $\widehat{A}$.

There are several aspects of this result worth emphasizing. First, the statistical rate we obtain in Theorem 1 has the same form as the standard statistical rate of linear regression, which is $\mathcal{O}\left(\sigma_w \sqrt{\frac{k}{T}}\right)$, where σ_w is the sub-Gaussian parameter of the noise w. Nonetheless, there are a couple of key distinctions between our result and the standard statistical rate of linear regression. Firstly, while a bound b_w on the norm of the noise implies that the noise is b_w-sub-Gaussian, it is not true that the norm of a σ_w-sub-Gaussian random vector can be bounded with high probability by σ_w. In general, the best available bound would be $\sigma_w\sqrt{k}$, where k is the dimension of the random vector. Therefore, for certain types of process noise our result has a worse dependence on the dimension than the statistical rate of linear regression.

Two other important distinctions are the dependence on the planning horizon H and the controllability term α, both of which are to be expected in our case. Algorithm 1 uses trajectory planning for data collection and the length of the reference trajectories is at most H. Since we can only guarantee one useful measurement per reference trajectory, it is to be expected that we can only guarantee an effective sample size of T/H. The controllability term α is also natural in our result because it quantifies how large the feature vectors can become in different directions. Larger feature vectors imply a larger signal-to-noise ratio, which in turn implies faster estimation.

Theorem 1 shows that our method can estimate $A_\star$ reliably. However, downstream control tasks may be difficult even when a good estimate of $A_\star$ is available. In other words, given a model (1) it may be difficult to find a controller that achieves a desired behavior. Model predictive control (MPC) is a good technique for leveraging a model of the dynamics for control since at each time

4. Recall that c_1 is the universal constant appearing in Assumption 4 and c_3 is the universal constant appearing in the upper bound on the error of the OLS estimate, shown in Section 4.

step it determines the next input to the system by approximately optimizing the cost achieved by trajectories generated using the model (Borrelli et al., 2017). We leave an exploration of the interplay between MPC and system identification to future work.

3.1 Trajectory Planning

The trajectory planning routine shown in Algorithm 1 uses the current estimate $\widehat{A}$ to plan, assuming no process noise, a trajectory from the current state of the system $\mathbf{x}_0^R = \mathbf{x}_t$ to a high-uncertainty region of the feature space, assuming no process noise. More precisely, it finds a sequence of actions $\{\mathbf{u}_j^R\}_{j=0}^r$ which produces a sequence of reference states $\{\mathbf{x}_j^R\}_{j=0}^r$ with the following properties:

- $\mathbf{x}_{j+1}^R = \widehat{A}\phi(\mathbf{x}_j^R, \mathbf{u}_j^R)$.
- The last reference state-action pair $(\mathbf{x}_r^R, \mathbf{u}_r^R)$ is either well aligned with v, the minimum eigenvector of $\Phi^\top\Phi$, or its feature vector is in a high-uncertainty region of the state space. More precisely, $(\mathbf{x}_r^R, \mathbf{u}_r^R)$ must satisfy one of the following two inequalities:

$$\left|\langle \phi(\mathbf{x}_r^R, \mathbf{u}_r^R), v\rangle\right| \geq \frac{\alpha}{2} \text{ or } \phi(\mathbf{x}_r^R, \mathbf{u}_r^R)^\top(\Phi^\top\Phi)^{-1}\phi(\mathbf{x}_r^R, \mathbf{u}_r^R) \geq \beta.$$

It is not immediately obvious that we can always find such a sequence of inputs. In Section 5 we prove that when Assumptions 4 and 6 hold the trajectory planning problem is feasible.

From the study of OLS, discussed in Section 4, we know that the matrix $\Phi^\top\Phi$ determines the uncertainty set of OLS. The larger $\lambda_{\min}\left(\Phi^\top\Phi\right)$ is, the smaller the uncertainty set will be. Therefore, to reduce the size of the uncertainty set we want to collect measurements at feature vectors ϕ such that the smallest eigenvalues of $\Phi^\top\Phi+\phi\phi^\top$ are larger than the smallest eigenvalues of $\Phi^\top\Phi$. Ideally, ϕ is a minimal eigenvector of $\Phi^\top\Phi$. However, we cannot always drive the system to such a feature vector, especially in the presence of process noise.

Instead, we settle for feature vectors of the following two types. Firstly, the trajectory planner tries to drive the system to feature vectors ϕ that are well aligned with the minimal eigenvector v of $\Phi^\top\Phi$; i.e., such that $|\langle\phi, v\rangle| \geq \alpha$. Such a data collection scheme is an instance of E-optimal design (Pukelsheim, 1993), which has been shown by Wagenmaker and Jamieson (2020) to produce inputs that allow the estimation of linear dynamics at an optimal rate.

However, if reaching a feature vector that aligns with the minimal eigenvector is not possible, the trajectory planner finds a reference trajectory to a feature vector ϕ such that $\phi^\top(\Phi^\top\Phi)^{-1}\phi \geq \beta$. When this inequality holds our uncertainty about the estimate $\widehat{A}$ in the direction ϕ is large. As shown in Section 5, such feature vectors can be encountered for only a small number of iterations.

Finally, trajectory planning is computationally intractable in general. However, in this work we quantify the data requirements of identifying $A_\star$, leaving computational considerations for future work. We assume access to a computational oracle. This assumption is reasonable since trajectory planning is often solved successfully in practice (Kavraki et al., 1996; LaValle and Kuffner Jr., 2001; Zucker et al., 2013).

3.2 Trajectory Tracking

Now we detail the trajectory tracking component of our method. We saw that the trajectory planner produces a reference trajectory $\{(\mathbf{x}_j^R, \mathbf{u}_j^R)\}_{j=0}^r$, with $r \leq H$. However, the planner assumes no

process noise to generate this reference trajectory. Therefore, if we were to simply plug the sequence of actions $\{\mathbf{u}_j^R\}_{j=0}^r$ into (1), the states of the system would diverge from $\mathbf{x}_j^R$. Instead, after observing each state $\mathbf{x}_t$ of the system (1), our method chooses an input $\mathbf{u}_t$ as follows:

- Given the current state $\mathbf{x}_t$, our method chooses an input $\mathbf{u}_t$ such that
$$\phi(\mathbf{x}_t, \mathbf{u}_t)^\top (\Phi^\top \Phi)^{-1} \phi(\mathbf{x}_t, \mathbf{u}_t) \geq \beta,$$
if there exists such an input. In other words, if there is an opportunity to greedily collect an informative measurement, our method takes it. If this situation is encountered, the trajectory tracker increments t by 1 and then stops tracking and returns.
- If there is no opportunity for greedy exploration, our method chooses an input $\mathbf{u}_t$ that minimizes $\|\widehat{A}(\phi(\mathbf{x}_t, \mathbf{u}_t) - \phi(\mathbf{x}_j^R, \mathbf{u}_j^R))\|$, and then increments t and j by one (t indexes the time steps of the system (1) and j indexes the reference trajectory). Therefore, our method uses closed loop control for data generation since minimizing $\|\widehat{A}(\phi(\mathbf{x}_t, \mathbf{u}_t) - \phi(\mathbf{x}_j^R, \mathbf{u}_j^R))\|$ requires access to the current state $\mathbf{x}_t$. At time t we choose $\mathbf{u}_t$ in this fashion in order to minimize the tracking error $\mathbb{E}\|\mathbf{x}_{t+1} - \mathbf{x}_{t+1}^R\|^2$ at the next time step, where the expectation is taken with respect to $\mathbf{w}_t$.
- Our method repeats these steps until $j = r$; i.e., until it reaches the end of the reference trajectory. When $j = r$ the trajectory tracker sets $\mathbf{u}_t = \mathbf{u}_j^R$, increments t by one, and returns.

4. General Guarantee on Estimation

In this section we provide a general upper bound on the error between an OLS estimate $\widehat{A}$ and the true parameters $A_\star$. The guarantee is based on the work of Simchowitz et al. (2018). We note also that results of this kind have been previously used in the study of online least squares and linear bandits (Abbasi-Yadkori et al., 2011; Dani et al., 2008; Rusmevichientong and Tsitsiklis, 2010). We assume that we are given a sequence of observations $\{(\mathbf{x}_t, \mathbf{u}_t, \mathbf{x}_{t+1})\}_{t\geq 0}$, generated by the system (1), with u_t allowed to depend on $\mathbf{x}_0$, $\mathbf{x}_1$, …, $\mathbf{x}_{t-1}$ and independent of $\mathbf{w}_j$ for all $j \geq t$. In what follows we denote $\phi_t := \phi(\mathbf{x}_t, \mathbf{u}_t)$.

Our method re-estimates the parameters $A_\star$ as more data is being collected. For the purpose of this section let us denote by $\widehat{A}_j$ the OLS estimate obtained using the first j measurements $(\mathbf{x}_t, \mathbf{u}_t, \mathbf{x}_{t+1})$:

$$\widehat{A}_j = \arg\min_A \sum_{t=0}^{j-1} \|A\phi_t - \mathbf{x}_{t+1}\|^2. \tag{7}$$

Proposition 2 *If the system* (1) *satisfies Assumptions 5 and 2 and if* $\lambda_{\min}\left(\sum_{t=0}^{t_0-1} \phi_t \phi_t^\top\right) \geq \underline{\lambda}$*, for some* $\underline{\lambda} > 0$ *and* $t_0 > 0$*, the OLS estimates* (7) *satisfy*

$$\mathbb{P}\left[\exists u \in \mathbb{S}^{k-1} \text{ and } j \geq t_0 \text{ s.t. } \|(\widehat{A}_j - A_\star)u\| \geq \mu_j \sqrt{u^\top \left(\sum_{t=0}^{j-1} \phi_t \phi_t^\top\right)^{-1} u}\right] \leq \delta,$$

where $\mu_j = c_3 b_w \sqrt{d + k \log\left(\frac{b_\phi^2 j}{\underline{\lambda}}\right) + \log\left(\frac{\pi^2 j^2}{6\delta}\right)}$ *for some universal constant* c_3.

Proof By assumption $\lambda_{\min}\left(\sum_{t=0}^{t_0-1}\phi_t\phi_t^\top\right) \geq \underline{\lambda} > 0$. Therefore, $\sum_{t=0}^{j-1}\phi_t\phi_t^\top$ is invertible and

$$\widehat{A}_j - A_\star = W_j^\top \Phi_j (\Phi_j^\top \Phi_j)^{-1},$$

where $W_j^\top = [\mathbf{w}_0, \ldots, \mathbf{w}_{j-1}]$ and $\Phi_j^\top = [\phi_0, \ldots, \phi_{j-1}]$. Now, we fix the index j and we consider the SVD decomposition $\Phi_j = U\Sigma V^\top$. Therefore, $\widehat{A}_j - A_\star = W_j^\top U \Sigma^\dagger V^\top$.

Recall that $\sup_{\mathbf{x},\mathbf{u}} \|\phi(\mathbf{x},\mathbf{u})\|_2 \leq b_\phi$ by assumption. Then, according to the analysis of Simchowitz et al. (2018) we know that $\|W_j^\top U\| \leq \mu_j$ with probability at least $1 - 6\delta/(\pi^2 j^2)$. Note that for all $u \in \mathbb{S}^{k-1}$ we have

$$\begin{aligned}\|(\widehat{A}_j - A_\star)u\| &\leq \|W_j^\top U\|\|\Sigma^\dagger V^\top u\| = \|W_j^\top U\|\sqrt{u^\top V(\Sigma^\dagger)^\top \Sigma^\dagger V^\top u} \\ &= \|W_j^\top U\|\sqrt{u^\top (\Phi_j^\top \Phi_j)^{-1} u}.\end{aligned}$$

Therefore, for a fixed index j, we have

$$\mathbb{P}\left[\exists u \in \mathbb{S}^{k-1} \text{ s.t. } \|(\widehat{A}_j - A_\star)u\|_2 \geq \mu_j \sqrt{u^\top \left(\sum_{t=0}^{j-1}\phi_t\phi_t^\top\right)^{-1} u}\right] \leq \frac{6\delta}{\pi^2 j^2}.$$

A direct application of the union bound yields the desired conclusion. ∎

5. Proof of Theorem 1

First let us observe that when $b_w = 0$ the result is trivial. Because we assume access to an initial trajectory $\mathcal{D}$ which satisfies Assumption 6 we are guaranteed $\widehat{A} = A_\star$ when $b_w = 0$. Therefore, we can assume that $b_w > 0$, which implies that α must be strictly positive according to Assumption 4. Throughout the proof we denote $\phi_t := \phi(\mathbf{x}_t, \mathbf{u}_t)$ and $\phi_j^R := \phi(\mathbf{x}_j^R, \mathbf{u}_j^R)$.

The proof of our result has three parts, which we now outline:

- We show that the trajectory planning step in Algorithm 1 is always feasible.
- We show that during the execution of Algorithm 1 there are at most N_e iterations such that:

$$\max_{\mathbf{u}\in\mathbb{B}_{r_u}} \phi(\mathbf{x}_t,\mathbf{u})^\top (\Phi^\top\Phi)^{-1}\phi(\mathbf{x}_t,\mathbf{u}) \geq \beta \;\text{ or }\; (\phi_j^R)^\top(\Phi^\top\Phi)^{-1}\phi_j^R \geq \beta.$$

- We show that Algorithm 1 collects at least $T/H - N_e$ measurements $(\phi_t, \mathbf{x}_{t+1})$ such that $|\langle \phi_t, v\rangle| \geq \alpha/4$, where v is a minimal eigenvector used to plan the reference trajectories. As a consequence, we show that Algorithm 1 collects measurements $(\phi_t, \mathbf{x}_{t+1})$ such that

$$\lambda_{\min}\left(\sum_{t=1}^{T+t_0}\phi_t\phi_t^\top\right) \geq \mathcal{O}(1)\alpha^2\left(\frac{T}{H} - N_e\right) - \frac{k-1}{2}b_\phi^2. \tag{8}$$

Once we have shown (8) is true, Theorem 1 follows from Proposition 2 and some algebra.

5.1 Part 1 of the Proof of Theorem 1.

We show that the trajectory planning step of Algorithm 1 is always feasible. Let

$$\mu = c_3 b_w \sqrt{d + k \log\left(b_\phi^2 T\right) + \log\left(\frac{\pi^2 T^2}{6\delta}\right)},$$

where c_3 is the universal constant appearing in Proposition 2. Since Assumption 6 guarantees that the minimum eigenvalue of the design matrix is at least 1, we know that

$$\|(\widehat{A} - A_\star)\phi\| \leq \mu \sqrt{\phi^\top \left(\Phi^\top \Phi\right)^{-1} \phi}, \tag{9}$$

for all $\phi \in \mathbb{S}^{k-1}$ and all iterations of Algorithm 1 with probability $1-\delta$.

Now, let $\beta = c_4 \left(d + k \log(\beta_\phi^2 T) + \log(\pi^2 T^2/(6\delta))\right)^{-1}$ with $c_4 \leq c_1^2/(4c_3^2)$. Then, since $\alpha \geq c_1 L b_w (1 + \gamma + \ldots + \gamma^{H-1})$, we have

$$\beta \leq \left(\frac{\alpha}{2L(1+\gamma+\ldots+\gamma^{H-1})\mu}\right)^2. \tag{10}$$

Let us $\tilde{\mathbf{x}}_0$ be equal to the initial state $\mathbf{x}_0^R$ of the trajectory planning and let $v \in \mathbb{R}^k$ be the desired goal direction. By Assumption 4 we know that there must exist a sequence of inputs $\tilde{\mathbf{u}}_0, \tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_r$, with $r \leq H$ and $\|\tilde{\mathbf{u}}_j\| \leq b_u/2$, such that $|\langle \phi(\tilde{\mathbf{x}}_r, \tilde{\mathbf{u}}_r), v\rangle| \geq \alpha$, where $\tilde{\mathbf{x}}_{j+1} = A_\star \phi(\tilde{\mathbf{x}}_j, \tilde{\mathbf{u}}_j)$. Now, let $\mathbf{x}_{j+1}^R = \widehat{A}\phi(\mathbf{x}_j^R, \mathbf{u}_j^R)$, where $\mathbf{u}_j^R$ is any input vector with $\|\mathbf{u}_j^R\| \leq b_u$ such that

$$\|A_\star[\phi(\mathbf{x}_j^R, \mathbf{u}_j^R) - \phi(\tilde{\mathbf{x}}_j, \tilde{\mathbf{u}}_j)]\| \leq \gamma \|\mathbf{x}_j^R - \tilde{\mathbf{x}}_j\|, \tag{11}$$

for $j < r$. Assumption 4 guarantees the existence of $\mathbf{u}_j^R$. We set $\mathbf{u}_r^R = \tilde{\mathbf{u}}_r$ and denote $\tilde{\phi}_j = \phi(\tilde{\mathbf{x}}_j, \tilde{\mathbf{u}}_j)$ and $\phi_j^R = \phi(\mathbf{x}_j^R, \mathbf{u}_j^R)$.

Case 1. There exists $j \in \{0, 1, 2, \ldots, r\}$ such that $(\phi_j^R)^\top \left(\Phi^\top \Phi\right)^{-1} \phi_j^R \geq \beta$. If this is the case, we are done because we found a feasible sequence of inputs $\mathbf{u}_0^R, \mathbf{u}_1^R, \ldots, \mathbf{u}_j^R$.

Case 2. We have $(\phi_j^R)^\top \left(\Phi^\top \Phi\right)^{-1} \phi_j^R \leq \beta$ for all $j \in \{0, 1, 2, \ldots, r\}$. In this case, we have

$$\tilde{\mathbf{x}}_{j+1} - \mathbf{x}_{j+1}^R = A_\star \tilde{\phi}_j - \widehat{A}\phi_j^R = A_\star(\tilde{\phi}_j - \phi_j^R) + (A_\star - \widehat{A})\phi_j^R.$$

Therefore, using (9), (10), and (11) we find

$$\begin{aligned}\|\tilde{\mathbf{x}}_{j+1} - \mathbf{x}_{j+1}^R\| &\leq \|A_\star(\tilde{\phi}_j - \phi_j^R)\| + \|(A_\star - \widehat{A})\phi_j^R\| \\ &\leq \gamma \|\tilde{\mathbf{x}}_j - \mathbf{x}_j^R\| + \frac{\alpha}{2L(1+\gamma+\ldots+\gamma^{H-1})}.\end{aligned}$$

Applying this inequality recursively, we find $\|\tilde{\mathbf{x}}_r - \mathbf{x}_r^R\| \leq \frac{\alpha}{2L}$, which implies $|\langle \phi_r^R, v\rangle| \geq \alpha/2$ because $\|\phi_r^R - \tilde{\phi}_r\| \leq L\|\tilde{\mathbf{x}}_r - \mathbf{x}_r^R\|$ by Assumption 1 and $|\langle \tilde{\phi}_r, v\rangle| \geq \alpha$ by construction. Hence, we constructed a feasible sequence of inputs $\{\mathbf{u}_j\}_{j=0}^r$ and Part 1 of the proof is complete.

5.2 Part 2 of the Proof of Theorem 1.

Now, we show that the number of iterations for which Algorithm 1 satisfies

$$\max_{\mathbf{u}\in\mathbb{B}_{r_u}} \phi(\mathbf{x}_t,\mathbf{u})^\top(\Phi^\top\Phi)^{-1}\phi(\mathbf{x}_t,\mathbf{u}) \geq \beta \;\text{ or }\; (\phi_j^R)^\top(\Phi^\top\Phi)^{-1}\phi_j^R \geq \beta \tag{12}$$

is upper bounded by

$$N_e := \left\lceil \frac{2k\log\left(\frac{2kb_\phi^2}{\log(1+\beta/2)}\right)}{\log(1+\beta/2)} \right\rceil. \tag{13}$$

We rely on the following proposition whose proof is deferred to Appendix B.

Proposition 3 *Let M_0 be a positive definite matrix and let us consider a sequence of vectors $\{v_t\}_{t\geq 1}$ in $\mathbb{R}^k$ with $\max_{t\geq 1}\|v_t\| \leq b$. Then, the number of vectors v_{t+1} such that*

$$v_{t+1}^\top\left(M_0+\sum_{i=1}^t v_i v_i^\top\right)^{-1} v_{t+1} \geq \beta,$$

is upper bounded by

$$\left\lceil \frac{2k\log\left(\frac{2kb^2}{\lambda_k(M_0)\log(1+\beta)}\right)}{\log(1+\beta)} \right\rceil. \tag{14}$$

Given Proposition 3, to prove (13) it suffices to show that during each iteration of Algorithm 1 when (12) occurs our method collects a measurement $(\phi_t, \mathbf{x}_{t+1})$ such that $\phi_t^\top(\Phi^\top\Phi)^{-1}\phi_t \geq \beta/2$.

By the definition of our trajectory tracker, whenever $\sup_{\mathbf{u}} \phi(\mathbf{x}_t,\mathbf{u})^\top(\Phi^\top\Phi)^{-1}\phi(\mathbf{x}_t,\mathbf{u}) \geq \beta$ we collect a measurement $(\phi_t, \mathbf{x}_{t+1})$ such that $\phi_t^\top(\Phi^\top\Phi)^{-1}\phi_t \geq \beta$.

Next, we show that when $(\phi_j^R)^\top(\Phi^\top\Phi)^{-1}\phi_j^R \geq \beta$, for some $j \leq r$, Algorithm 1 is guaranteed to collect a measurement $(\phi_t, \mathbf{x}_{t+1})$ such that $\phi_t^\top(\Phi^\top\Phi)^{-1}\phi_t \geq \beta/2$. Let s be the smallest index in the reference trajectory such that $(\phi_s^R)^\top(\Phi^\top\Phi)^{-1}\phi_s^R \geq \beta$.

For the remainder of this section we re-index the trajectory $\{(\mathbf{x}_t,\mathbf{u}_t)\}_{t\geq 0}$ collected by Algorithm 1 so that $\mathbf{x}_j^R = \mathbf{x}_j$ for all $j \in \{0,1,\ldots,s\}$. Then, we show that $(\phi_s^R)^\top(\Phi^\top\Phi)^{-1}\phi_s^R \geq \beta$ implies the existence of $j \in \{0,1,\ldots,s\}$ such that $\phi_j^\top(\Phi^\top\Phi)^{-1}\phi_j \geq \beta/2$.

Let $\Delta = \phi_s^R - \phi_s$. The Cauchy-Schwarz inequality implies

$$\begin{aligned}\phi_s(\Phi^\top\Phi)^{-1}\phi_s &= \phi_s^R(\Phi^\top\Phi)^{-1}\phi_s^R + \Delta^\top(\Phi^\top\Phi)^{-1}\Delta + 2\Delta^T(\Phi^\top\Phi)^{-1}\phi_s^R \\ &\geq \left(\sqrt{\phi_s^R(\Phi^\top\Phi)^{-1}\phi_s^R} - \sqrt{\Delta^\top(\Phi^\top\Phi)^{-1}\Delta}\right)^2.\end{aligned}$$

Then, as long as $\Delta^\top(\Phi^\top\Phi)^{-1}\Delta \leq \frac{\beta}{2}(3-2\sqrt{2})$, we are guaranteed to have $\phi_s^\top(\Phi^\top\Phi)^{-1}\phi_s \geq \beta/2$.

Now, since s is the smallest index such that $(\phi_s^R)^\top(\Phi^\top\Phi)^{-1}\phi_s^R \geq \beta$, we know that for all $j \in \{0,1,\ldots,s-1\}$ we have $(\phi_j^R)^\top(\Phi^\top\Phi)^{-1}\phi_j^R \leq \beta$. Also, we can assume that during reference tracking we do not encounter a state $\mathbf{x}_j$, with $j \in \{0,1,\ldots,s-1\}$, such that

$$\max_{\mathbf{u}\in\mathbb{B}_{r_u}} \phi(\mathbf{x}_j,\mathbf{u})^\top(\Phi^\top\Phi)^{-1}\phi(\mathbf{x}_t,\mathbf{u}) \geq \beta,$$

because we already treated this case. Now, let us consider the difference

$$\mathbf{x}_{j+1} - \mathbf{x}_{j+1}^R = A_\star \phi_j + \mathbf{w}_j - \widehat{A}\phi_j^R = (A_\star - \widehat{A})\phi_j + \mathbf{w}_j - \widehat{A}[\phi_j^R - \phi_j].$$

We obtain

$$\|\mathbf{x}_{j+1} - \mathbf{x}_{j+1}^R\| \leq \|(A_\star - \widehat{A})\phi_j\| + b_w + \|\widehat{A}[\phi_j^R - \phi_j]\|.$$

Let us denote $\delta_j(\mathbf{u}) = \phi(\mathbf{x}_j, \mathbf{u}) - \phi_j^R$. Hence, $\delta_j(\mathbf{u}_j) = \phi_j - \phi_j^R$. Now, let $\mathbf{u}_\star \in \mathbb{B}_{r_u}$ an input such that $\|A_\star \delta_t(\mathbf{u}_\star)\| \leq \gamma \|\mathbf{x}_j - \mathbf{x}_j^R\|$, which we know exists by Assumption 3 (note that $\mathbf{u}_\star$ depends on the index j, but we dropped this dependency from the notation for simplicity). Since our method attempts trajectory tracking by choosing $\mathbf{u}_j \in \arg\min_{\mathbf{u}\in\mathbb{B}_{r_u}} \|\widehat{A}(\phi(\mathbf{x}_t, \mathbf{u}) - \phi(\mathbf{x}_j^R, \mathbf{u}_j^R))\|$ we have

$$\begin{aligned}
\|\widehat{A}\delta_j(\mathbf{u}_j)\| &\leq \|\widehat{A}\delta_j(\mathbf{u}_\star)\| \leq \|A_\star \delta_j(\mathbf{u}_\star)\| + \|(A_\star - \widehat{A})\delta_j(\mathbf{u}_\star)\| \\
&\leq \gamma\|\mathbf{x}_j - \mathbf{x}_j^R\| + \|(A_\star - \widehat{A})\delta_j(\mathbf{u}_\star)\| \\
&\leq \gamma\|\mathbf{x}_j - \mathbf{x}_j^R\| + \|(A_\star - \widehat{A})\phi(\mathbf{x}_j, \mathbf{u}_\star)\| + \|(A_\star - \widehat{A})\phi_j^R\|.
\end{aligned}$$

As mentioned above, we can assume $\phi(\mathbf{x}_j, \mathbf{u}_\star)^\top (\Phi^\top \Phi)^{-1} \phi(\mathbf{x}_j, \mathbf{u}_\star) < \beta$. Also, recall that

$$(\phi_j^R)^\top (\Phi^\top \Phi)^{-1} \phi_j^R \leq \beta,$$

since $j < s$ and s is the smallest index so that this inequality does not hold. Hence, Proposition 2 implies that $\|(A_\star - \widehat{A})\phi(\mathbf{x}_t, \mathbf{u}_\star)\| \leq \mu\sqrt{\beta}$ and $\|(A_\star - \widehat{A})\phi_r^R\| \leq \mu\sqrt{\beta}$. Putting everything together we find

$$\|\mathbf{x}_{j+1} - \mathbf{x}_{j+1}^R\| \leq \gamma\|\mathbf{x}_j - \mathbf{x}_j^R\| + 3\mu\sqrt{\beta} + b_w.$$

Then, since the reference trajectory is initialized with the state $\mathbf{x}_0^R = \mathbf{x}_0$, we find

$$\begin{aligned}
\|\Delta\| = \|\phi_s - \phi_s^R\| &\leq L(b_w + 3\mu\sqrt{\beta})(1 + \gamma + \ldots + \gamma^{s-1}) \\
&= (3c_3\sqrt{c_4} + 1)Lb_w(1 + \gamma + \ldots + \gamma^{s-1}),
\end{aligned}$$

where the last identity follows because $\mu\sqrt{\beta} = c_3\sqrt{c_4}b_w$.

Then, as long as $c_2 \geq \frac{2(3c_3\sqrt{c_4}+1)^2}{(3-2\sqrt{2})c_4}$, Assumption 6 offers a lower bound on $\lambda_{\min}(\Phi^\top\Phi)$ which ensures that $\Delta^\top(\Phi^\top\Phi)^{-1}\Delta \leq \frac{\beta}{2}(3 - 2\sqrt{2})$, implying $\phi_s^\top(\Phi^\top\Phi)^{-1}\phi_s \geq \beta/2$.

To summarize, we have shown whenever Algorithm 1 encounters a situation in which either

$$\sup_{\mathbf{u}} \phi(\mathbf{x}_t, \mathbf{u})^\top (\Phi^\top \Phi)^{-1} \phi(\mathbf{x}_t, \mathbf{u}) \geq \beta \;\text{ or }\; (\phi_j^R)^\top (\Phi^\top \Phi)^{-1} \phi_j^R \geq \beta, \tag{15}$$

it collects a measurement $(\phi_t, \mathbf{x}_{t+1})$ such that $\phi_t(\Phi^\top\Phi)^{-1}\phi_t \geq \beta/2$. Hence, according to Proposition 3, the event (15) can occur at most N_e times (the value N_e was defined in (13)).

5.3 Part 3 of the Proof of Theorem 1.

In this final part of the proof we analyze what happens when the trajectory planning problem returns a reference trajectory $(\mathbf{x}_j^R, \mathbf{u}_j^R)$ for which $|\langle \phi_r^R, v\rangle| \geq \alpha/2$, where v is a minimal eigenvector with unit norm of $\Phi^\top \Phi$.

During its execution the algorithm produces T/H reference trajectories. Part 2 of the proof implies that at least $T/H - N_e$ of the reference trajectories satisfy $|\langle \phi_r^R, v\rangle| \geq \alpha/2$, with all states $\mathbf{x}_t$ encountered during tracking satisfying $\sup_{\mathbf{u}} \phi(\mathbf{x}_t, \mathbf{u})^\top \left(\Phi^\top \Phi\right)^{-1} \phi(\mathbf{x}_t, \mathbf{u}) \leq \beta$ and all reference features ϕ_j^R satisfying $(\phi_j^R)^\top (\Phi^\top \Phi)^{-1} \phi_j^R \leq \beta$.

Following the same argument as in Part 2 of the proof we know that tracking the reference trajectory in this case takes the system to a state $\mathbf{x}_t$ such that

$$\|\mathbf{x}_t - \mathbf{x}_r^R\| \leq (3c_3\sqrt{c_4} + 1) b_w (1 + \gamma + \ldots + \gamma^{r-1}),$$

which implies by Assumption 1 that

$$\|\phi_t - \phi_r^R\| \leq (3c_3\sqrt{c_4} + 1) L b_w (1 + \gamma + \ldots + \gamma^{r-1}).$$

This last inequality implies that $|\langle \phi_t, v\rangle| \geq \alpha/4$ if $3c_3\sqrt{c_4}+1 \leq c_1/4$. Recall that the only condition we imposed so far on c_4 is $c_4 \leq c_1^2/(4c_3^2)$ in Part 1 of the proof. Hence, since $c_1 > 4$, we can choose c_4 such that $c_4 \leq \frac{(c_1-4)^2}{144 c_3^2}$ and $c_4 \leq c_1^2/(4c_3^2)$. Then, $3c_3\sqrt{c_4} + 1 < c_1/4$. Now, to finish the proof of Theorem 1 we rely on the following result, whose proof is deferred to Appendix C.

Proposition 4 *Let $\mathcal{V} \subset \mathbb{R}^k$ be a bounded set, with $\sup_{v\in\mathcal{V}} \|v\| \leq b$, such that for any $u \in \mathbb{S}^{k-1}$ there exists $v \in \mathcal{V}$ with $|\langle u, v\rangle| \geq \alpha$. Then, for all $T \geq 0$, given any sequence of vectors $\{v_t\}_{t\geq 0}$ in $\mathbb{R}^k$ we have*

$$\lambda_{\min}\left(\sum_{i=1}^{T} v_i v_i^\top\right) \geq \frac{\alpha^2 K(T)}{2k} - \frac{k-1}{2}\left(b^2 - \frac{\alpha^2}{2}\right),$$

where $K(T)$ is the number of times v_{t+1} belongs to $\mathcal{V}$ and $|\langle v_{t+1}, \tilde{v}_{t+1}\rangle| \geq \alpha$, with $\tilde{v}_{t+1} \in \arg\min_{\|v\|=1} v^\top \left(\sum_{i=1}^t v_i v_i^\top\right) v$ and $t < T$.

We have shown that at least $T/H - N_e$ times the algorithm collects a state transition $(\mathbf{x}_t, \mathbf{u}_t, \mathbf{x}_{t+1})$ for which ϕ_t is at least $\alpha/4$ aligned with the minimal eigenvector of $\Phi^\top \Phi$, where Φ is the matrix of all ϕ_j observed prior to the last trajectory planning episode. Therefore, Proposition 4 implies that Algorithm 1 collects a sequence of measurements $(\phi_t, \mathbf{x}_{t+1})$ such that

$$\lambda_{\min}\left(\sum_{t=1}^{T+t_0} \phi_t \phi_t^\top\right) \geq \frac{\alpha^2}{32}\left(\frac{T}{H} - N_e\right) - \frac{k-1}{2} b_\phi^2,$$

where b_ϕ plays the role of b in Proposition 4. Putting this result together with Proposition 2 yields the desired conclusion.

6. Related Work

System identification, being one of the cornerstones of control theory, has a rich history, which we cannot hope to summarize here. For an in-depth presentation of the field we direct the interested reader to the book by Ljung (1987) and the review articles by Åström and Eykhoff (1971), Bombois et al. (2011), Chiuso and Pillonetto (2019), Hong et al. (2008), Juditsky et al. (1995), Ljung et al. (2020), Schoukens and Ljung (2019), and Sjöberg et al. (1995). Instead, we discuss recent studies of system identification that develop finite-time statistical guarantees.

Most recent theoretical guarantees of system identification apply to linear systems under various sets of assumptions (Campi and Weyer, 2002; Dahleh et al., 1993; Faradonbeh et al., 2018; Fattahi et al., 2019; Hardt et al., 2018; Hazan et al., 2017, 2018; Oymak and Ozay, 2019; Sarkar and Rakhlin, 2019; Sarkar et al., 2021, 2019; Simchowitz et al., 2018, 2019; Sun et al., 2020; Tsiamis and Pappas, 2019; Tsiamis et al., 2019; Wagenmaker and Jamieson, 2020). Notably, Simchowitz et al. (2018) derived sharp rates for the non-adaptive estimation of marginally stable systems. Then, Sarkar and Rakhlin (2019) developed a more general analysis that also applies to a certain class of unstable linear systems. Both of these studies assumed that the estimation method can directly observe the state of the system. We make the same assumption in our work. However, in many applications full state observation is not possible. Recently, Simchowitz et al. (2019) proved that marginally stable linear systems can be estimated from partial observations by using a prefiltered least squares method. From the study of linear dynamics, the work of Wagenmaker and Jamieson (2020) is the closest to our own. Inspired by E-optimal design (Pukelsheim, 1993), the authors propose and analyze an adaptive data collection method for linear system identification which maximizes the minimal eigenvalue $\lambda_{\min}(\sum_{t=0}^{T-1} \mathbf{x}_t \mathbf{x}_t^\top)$ under power constraints on the inputs. Wagenmaker and Jamieson (2020) prove matching upper and lower bounds for their method.

There is comparatively little known about the sample complexity of nonlinear system identification. Oymak (2019) and Bahmani and Romberg (2020) studied the estimation of the parameters A and B of a dynamical system of the form $\mathbf{x}_{t+1} = \phi(A\mathbf{x}_t + B\mathbf{u}_t)$, where ϕ is a known activation function and the inputs u_t are i.i.d. standard Gaussian vectors. Importantly, in this model both $\mathbf{x}_t$ and $\mathbf{u}_t$ are observed and there is no unobserved noise, which makes estimation easy when the map ϕ is invertible. In follow-up work, Sattar and Oymak (2020) and Foster et al. (2020) generalized these results. In particular, Foster et al. (2020) took inspiration from the study of generalized linear models and showed that a method developed for the standard i.i.d. setting can estimate dynamical systems of the form $\mathbf{x}_{t+1} = \phi(A\mathbf{x}_t) + \mathbf{w}_t$ at an an optimal rate, where $\mathbf{w}_t$ is unobserved i.i.d. noise. All these works share a common characteristic, they study systems for which identification is possible through the use of non-adaptive inputs. We take the first step towards understanding systems that require adaptive methods for successful identification.

In a different line of work, Singh et al. (2021) proposed a learning framework for trajectory planning from learned dynamics. They propose a regularizer of dynamics that promotes stabilizability of the learned model, which allows the tracking of reference trajectories based on estimated dynamics. Also, Khosravi and Smith (2020a) and Khosravi and Smith (2020b) developed learning methods that exploit other control-theoretic priors. Nonetheless, none of these works characterize the sample complexity of the problem.

While most work that studies sample-complexity questions in the setting of tabular MDPs focuses on finding optimal policies, Jin et al. (2020) and Wolfer and Kontorovich (2019) recently analyzed data collection for system identification. More precisely, Jin et al. (2020) developed an ef-

ficient algorithm for the exploration of tabular MDPs that enables near-optimal policy synthesis for an arbitrary number of reward functions, which are unknown during data collection, while Wolfer and Kontorovich (2019) derived minimax sample complexity guarantees for the estimation of ergodic Markov chains. Finally, we note that Abbeel and Ng (2005) quantified the sample complexity of learning policies from demonstrations for tabular MDPs and for a simpler version of the model class (1).

Finally, Kakade et al. (2020) already followed up on our work, showing that when (1) has an associated cost function one can solve online the control problem by relying on optimism in the face of uncertainty. Their proposed method achieves achieves $\mathcal{O}(\sqrt{T})$ regret and, as in our case, it is not computationally tractable.

7. Discussion and Open Problems

The field of system identification has broadened the range of real-world applications of control theory, and promises to further expand the field in the future, particularly as data collection becomes more pervasive. In this work we proposed and analyzed a method that estimates a class of nonlinear dynamical systems in finite time by adaptively collecting data that is informative enough. While this results takes us closer to understanding the fundamental limits of data-driven control, there are many limitations to our model and approach. We end with a list of open questions:

- To solve trajectory planning problems we assumed access to a computational oracle. Is it possible to develop a method that has good statistical guarantees and is also computationally tractable? In practice, successful nonlinear control is often based on linearizations of the dynamics. Is it possible to quantify the sample complexity of system identification when trajectory planning is implemented using linearizations?

- Our method relies on full state observations. However, in many applications full state observations are impossible. Is it possible to obtain finite-time statistical guarantees for nonlinear system identification from partial observations?

- Our guarantee holds only when the true system being identified lies in the model class (1). When the true system is not part of the model class, how much data is needed to find the best model in class? Ross and Bagnell (2012) studied this problem under a generative model.

- Only fully actuated systems can satisfy Assumption 3 with $\gamma < 1$. Is it possible to extend our result to systems that require multiple time steps to recover from disturbances?

- Assumption 4 allows only systems whose feature vectors can align with any direction. What if the feature vectors can align only with vectors in a subspace? In this case, it is not possible to recover $A_\star$ fully. However, in this case, it would not be necessary to know $A_\star$ fully in order to predict or control. Is it possible to estimate $A_\star$ only in the relevant directions?

- What if we consider infinite-dimensional feature maps ϕ? For example, can we develop a statistical theory of learning reproducing kernel Hilbert space models of dynamical systems?

Acknowledgments

This research is generously supported in part by ONR awards N00014-17-1-2191, N00014-17-1-2401, and N00014-18-1-2833, NSF CPS award 1931853, and the DARPA Assured Autonomy program (FA8750-18-C-0101). We would like to thank Alekh Agarwal, Francesco Borrelli, Debadeepta Dey, Munther Dahleh, Ali Jadbabaie, Sham Kakade, Akshay Krishnamurthy, John Langford, and Koushil Sreenath for useful comments and for pointing out relevant related work. We would also like to thank Michael Muehlebach for detailed and valuable feedback on our manuscript.

Appendix A. Notation

We list the meanings of some of our notation. We also discuss our requirements on the universal constants we use.

Symbol	Definition
d	dimension of the states
p	dimension of the inputs
k	dimension of the features $\phi(\mathbf{x}, \mathbf{u})$
r_u	upper bound on the Euclidean norm of the inputs
L	Lipschitz constant of the feature map ϕ
b_ϕ	upper bound on the Euclidean norm of the feature map ϕ
γ	multiplicative factor that upper bounds how much a tracking error can be shrunk in one time step, appearing in Assumption 3
b_u	upper bound on the Euclidean norm of the inputs needed in Assumption 3 to shrink the tracking errors
α	lower bound on well the state-input features of the system can align with given directions, appearing in Assumption 4
H	a positive integer that upper bounds the horizon needed to align the features of the system with a given direction, appearing in Assumption 4
$\mathcal{D}$	the initial data set
β	hyperparameter in Algorithm 1
$c_1, c_2, c_3 \ldots$	different universal constants
c_1	universal constant appearing in Assumption 5 regarding the size of the noise.
c_2	universal constant appearing Assumption 6 regarding the initialization of the algorithm.
c_3	universal constant appearing in the bound on the performance of OLS (Proposition 2)
c_4	universal constant appearing in our choice of the hyperparameter β (Theorem 1)
c_5	universal constant appearing in our method's final statistical rate (Theorem 1)

Now, we discuss the universal constants. The constant c_1 can be chosen to be any number greater than 4. The constant c_3 can be any positive value that makes Proposition 2 true. Once c_1 and c_3 are chosen, the constant c_4 can be chosen to be any positive number such that $c_4 \leq \frac{(c_1-4)^2}{144c_3^2}$. Finally, once c_1, c_2, and c_3 are set, the constant c_2 can be any value such that $c_2 \geq \frac{2(3c_3\sqrt{c_4}+1)^2}{(3-2\sqrt{2})c_4}$.

Appendix B. Proof of Proposition 3

To make this section self-contained we restate Proposition 3 here. This proposition shows that for any sequence $\{v_t\}_{t\geq 1}$ the quantity $v_{t+1}^\top \left(M_0 + \sum_{i=1}^t v_i v_i^\top\right)^{-1} v_{t+1}$ is large only a small number of times. For linear regression, this fact implies that whenever we collect a measurement with covariates v_{t+1}, the estimation error shrinks by a factor greater than one.

Proposition 5 *Let M_0 be a positive definite matrix and let us consider a sequence of vectors $\{v_t\}_{t\geq 1}$ in $\mathbb{R}^k$ with $\max_{t\geq 1}\|v_t\| \leq b$. Then, the number of vectors v_{t+1} such that*

$$v_{t+1}^\top \left(M_0 + \sum_{i=1}^t v_i v_i^\top\right)^{-1} v_{t+1} \geq \beta,$$

is upper bounded by

$$\left\lceil \frac{2k \log\left(\frac{2kb^2}{\lambda_k(M_0)\log(1+\beta)}\right)}{\log(1+\beta)} \right\rceil. \tag{14}$$

The next lemma relates scaling ellipsoids in one direction with the scaling of their volumes. A proof of this result can be found in the work by Abbasi-Yadkori et al. (2011).

Lemma 6 *Suppose M and N are two positive definite matrices with $M \succ N \succ 0$. Then,*

$$\sup_{v\neq 0} \frac{v^\top M v}{v^\top N v} \leq \frac{\det(M)}{\det(N)}.$$

Now, we are ready to prove Proposition 3. We denote by $N_t = N_0 + \sum_{i=1}^t v_i v_i^\top$. First, we prove that $\det(N_{t+1}^{-1}) \leq \det(N_t^{-1})/(1+\beta)$ whenever $v_{t+1}^\top N_t^{-1} v_{t+1} \geq \beta$.

By definition we have $N_{t+1} \succeq N_t \succ 0$. Therefore, $N_{t+1}^{-1} \preceq N_t^{-1}$. Now, we apply the Sherman-Morrison rank-one update formula to find

$$\begin{aligned} v_{t+1}^\top N_{t+1}^{-1} v_{t+1} &= v_{t+1}^\top N_t^{-1} v_{t+1} - \frac{\left(v_{t+1}^\top N_t^{-1} v_{t+1}\right)^2}{1 + v_{t+1}^\top N_t^{-1} v_{t+1}} \\ &= \left(1 - \frac{v_{t+1}^\top N_t^{-1} v_{t+1}}{1 + v_{t+1}^\top N_t^{-1} v_{t+1}}\right) v_{t+1}^\top N_t^{-1} v_{t+1}. \end{aligned}$$

Since the function $x \mapsto \frac{x}{1+x}$ is increasing for $x > -1$, we find

$$v_{t+1}^\top N_{t+1}^{-1} v_{t+1} \leq \frac{v_{t+1}^\top N_t^{-1} v_{t+1}}{1+\beta},$$

whenever $v_{t+1}^\top N_t^{-1} v_{t+1} \geq \beta$. Then, Lemma 6 implies that $\det(N_{t+1}^{-1}) \leq \det(N_t^{-1})/(1+\beta)$ whenever $v_{t+1}^\top N_t^{-1} v_{t+1} \geq \beta$, which in turn implies $\det(N_{t+1}) \geq (1+\beta)\det(N_t)$ whenever $v_{t+1}^\top N_t^{-1} v_{t+1} \geq \beta$.

Intuitively, $\det(N_{t+1}) \geq (1+\beta)\det(N_t)$ cannot happen too often because the vectors v_t are bounded. Therefore, the number of times $v_{t+1}^\top N_t^{-1} v_{t+1} \geq \beta$ must be small. To prove this fact

we study the growth of the eigenvalues of N_t. We have shown that $\det(N_{t+1}) \geq (1+\beta)\det(N_t)$ whenever $v_{t+1}^\top N_t^{-1} v_{t+1} \geq \beta$, which means that the eigenvalues of N_t grow exponentially with the number of times $v_{t+1}^\top N_t^{-1} v_{t+1} \geq \beta$. We offer an exponential lower bound on the growth of N_t's eigenvalues and we also present an upper bound on the size of the eigenvalues, using the fact that the vectors v_t are bounded. Putting these two bounds together shows that $v_{t+1}^\top N_t^{-1} v_{t+1} \geq \beta$ cannot happen too often. Now, we make this argument precise.

Let us denote by $\lambda_1(t), \lambda_2(t), \ldots, \lambda_k(t)$ the eigenvalues of N_t sorted in decreasing order. Recall that $\lambda_i(t)$ is a non-decreasing function of t. Now, let $\varepsilon_{i,t} = \log_{1+\beta}(\lambda_i(t)/\lambda_i(t-1))$. Therefore, we have $\lambda_i(t) = (1+\beta)^{\varepsilon_{i,t}}\lambda_i(t-1)$. We know $\varepsilon_{i,t} \geq 0$ for all i and t and we know that $\sum_{i=1}^k \varepsilon_{i,t} \geq 1$ when $v_t^\top N_{t-1}^{-1} v_t \geq \beta$ because $\det(N_t) \geq (1+\beta)\det(N_{t-1})$.

We first prove an upper bound on the growth of $\lambda_i(t)$ as a function of t. By definition, we have $\lambda_i(t) = (1+\beta)^{\sum_{j=1}^t \varepsilon_{i,j}}\lambda_i(0) \geq (1+\beta)^{\sum_{j=1}^t \varepsilon_{i,j}}\lambda_k(0)$. Since $\max_j \|v_j\| \leq b$, we know that $\lambda_i(t+1) \leq \lambda_i(t) + b^2$. Therefore,

$$(1+\beta)^{\varepsilon_{i,t+1}} = \frac{\lambda_i(t+1)}{\lambda_i(t)} \leq 1 + \frac{b^2}{\lambda_i(t)} \leq 1 + \frac{b^2}{(1+\beta)^{\sum_{j=1}^t \varepsilon_{i,j}}\lambda_k(0)}.$$

In other words, we have

$$\varepsilon_{i,t+1} \leq \frac{\log\left(1 + \frac{b^2}{(1+\beta)^{\sum_{j=1}^t \varepsilon_{i,j}}\lambda_k(0)}\right)}{\log(1+\beta)} \leq \frac{b^2}{(1+\beta)^{\sum_{j=1}^t \varepsilon_{i,j}}\lambda_k(0)\log(1+\beta)}.$$

We denote $\rho = \log\left(\frac{2kb^2}{\lambda_k(0)\log(1+\beta)}\right) / \log(1+\beta)$. Therefore, when $\sum_{j=1}^t \varepsilon_{i,j} > \rho$, we have $\varepsilon_{i,t+1} \leq 1/(2k)$, which shows that when $\lambda_i(t)$ grew past a certain point as a function of t it can grow by at most a factor of $(1+\beta)^{1/2k}$ at the next time step.

Now, we prove a lower bound on the growth of $\lambda_i(t)$ as a function of t. Suppose there are n vectors v_j such that $v_j^\top N_{j-1}^{-1} v_j \geq \beta$ with $j \leq t$. Since $\sum_{i=1}^k \varepsilon_{i,j} \geq 1$ whenever $v_j^\top N_{j-1}^{-1} v_j \geq \beta$, we have $\sum_{j=1}^t \sum_{i=1}^k \varepsilon_{i,j} \geq n$. Moreover, each time j such that $v_j^\top N_{j-1}^{-1} v_j \geq \beta$ we know that

$$\begin{aligned}
\varepsilon_{i,j} &\geq 1 - \sum_{i' \neq i} \varepsilon_{i',j} \geq 1 - \sum_{i' : \sum_{s=1}^{j-1} \varepsilon_{i',s} < \rho} \varepsilon_{i',j} - \sum_{i' : \sum_{s=1}^{j-1} \varepsilon_{i',s} \geq \rho} \varepsilon_{i',j} \\
&\geq 1 - \sum_{i' : \sum_{s=1}^{j-1} \varepsilon_{i',s} < \rho} \varepsilon_{i',j} - \sum_{i' : \sum_{s=1}^{j-1} \varepsilon_{i',s} \geq \rho} \frac{1}{2k} \geq \frac{1}{2} - \sum_{i' : \sum_{s=1}^{j-1} \varepsilon_{i',s} < \rho} \varepsilon_{i',j}. \qquad (16)
\end{aligned}$$

This last bound holds for all j such that $v_j^\top N_{j-1}^{-1} v_j \geq \beta$ and we assumed there are n such j with $j \leq t$. Also, note that

$$\sum_{j=1}^t \sum_{i' : \sum_{s=1}^{j-1} \varepsilon_{i',s} < \rho} \varepsilon_{i',j} \leq \sum_{i' \neq i} \min\left\{\rho, \sum_{j=1}^t \epsilon_{i',j}\right\}.$$

Hence, summing (16) over j, we obtain a lower bound on the growth of $\lambda_i(t)$ up to time t

$$\log_{1+\beta}\left(\frac{\lambda_i(t)}{\lambda_k(0)}\right) \geq \sum_{j=1}^{t} \varepsilon_{i,j} \geq \frac{n}{2} - \sum_{i' \neq i} \min\left\{\rho, \sum_{j=1}^{t} \varepsilon_{i',j}\right\} \geq \frac{n}{2} - (k-1)\rho.$$

Then, once $n \geq 2k \log\left(\frac{2kb^2}{\lambda_k(0)\log(1+\beta)}\right) / \log(1+\beta)$, we obtain

$$\sum_{j=1}^{t} \varepsilon_{i,j} \geq \log\left(\frac{2kb^2}{\lambda_k(0)\log(1+\beta)}\right) / \log(1+\beta) = \rho,$$

which implies $\varepsilon_{i,j} < \frac{1}{2k}$ for all $j > t$. Since i was chosen arbitrarily, we see that whenever $n \geq 2k \log\left(\frac{2kb^2}{\lambda_k(0)\log(1+\beta)}\right)$ and $j > t$ we get $\sum_{i=1}^{k} \varepsilon_{ij} < k\frac{1}{2k} < 1$. Hence, n must be smaller or equal than $\left\lceil 2k \log\left(\frac{2kb^2}{\lambda_k(0)\log(1+\beta)}\right)\right\rceil$.

Appendix C. Proof of Proposition 4

To make this section self-contained we restate Proposition 4 here. Intuitively, this proposition shows that when there are many vectors v_t that align well with the minimum eigenvectors of $\sum_{i=1}^{t-1} v_i v_i^\top$, the minimum eigenvalue of $\sum_{i=1}^{T} v_i v_i^\top$ must be large.

Proposition 7 *Let $\mathcal{V} \subset \mathbb{R}^k$ be a bounded set, with $\sup_{v\in\mathcal{V}}\|v\| \leq b$, such that for any $u \in \mathbb{S}^{k-1}$ there exists $v \in \mathcal{V}$ with $|\langle u, v\rangle| \geq \alpha$. Then, for all $T \geq 0$, given any sequence of vectors $\{v_t\}_{t\geq 0}$ in $\mathbb{R}^k$ we have*

$$\lambda_{\min}\left(\sum_{i=1}^{T} v_i v_i^\top\right) \geq \frac{\alpha^2 K(T)}{2k} - \frac{k-1}{2}\left(b^2 - \frac{\alpha^2}{2}\right),$$

where $K(T)$ is the number of times v_{t+1} belongs to $\mathcal{V}$ and $|\langle v_{t+1}, \tilde{v}_{t+1}\rangle| \geq \alpha$, with $\tilde{v}_{t+1} \in \arg\min_{\|v\|=1} v^\top \left(\sum_{i=1}^{t} v_i v_i^\top\right) v$ and $t < T$.

Remark: If we choose v_i to cycle over standard basis vectors, we see that $\lambda_{\min}\left(\sum_{i=1}^{T} v_i v_i^\top\right) = \frac{\alpha^2 T}{k}$ when T is a multiple of k. Therefore, we cannot hope to have a lower bound in Proposition 4 better than $\frac{\alpha^2 T}{k}$. Our proof can be refined to show

$$\lambda_{\min}\left(\sum_{i=1}^{T} v_i v_i^\top\right) \geq \frac{\alpha^2 T}{k} - \mathcal{O}(\sqrt{T}),$$

where the $\mathcal{O}(\cdot)$ notation hides dependencies on α, b, and k.

Clearly, the sum of the eigenvalues of $\sum_{i=1}^{t} v_i v_i^\top$ grows linearly with t. However, it is not clear whether just some of the eigenvalues grow or whether all of them grow with t. To prove Proposition 4 we need the following lemma, which intuitively shows that the sum of the smallest eigenvalues cannot lag behind the larger eigenvalues by too much.

Lemma 8 *Let $M \in \mathbb{R}^{k\times k}$ be a positive semi-definite matrix. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ be the eigenvalues of M and let $u_1, u_2, \ldots, u_k$ be the corresponding eigenvectors of unit norm. Suppose v is a vector in $\mathbb{R}^k$ such that $\|v\| \leq b$ and $|\langle v, u_k\rangle| \geq \alpha$. Let $\nu_1 \geq \nu_2 \geq \ldots \geq \nu_k$ be the eigenvalues of $M + vv^\top$. Then, for any $s \in \{2, \ldots k\}$ such that $\lambda_{s-1} \geq \lambda_s + b^2 - \alpha^2/2$ we have*

$$\sum_{i=s}^{k} \nu_i \geq \sum_{i=s}^{k} \lambda_i + \frac{\alpha^2}{2}. \tag{17}$$

Proof First, we express v in M's eigenbasis: $v = \sum_{i=1}^k z_i u_i$. Then, by assumption we know that $\|v\|^2 = \sum_{i=1}^k z_i^2 \leq b^2$ and $z_k^2 \geq \alpha^2$. Using a result by Bunch et al. (1978) we know that $\nu_1 \geq \lambda_1$ and $\nu_i \in [\lambda_i, \lambda_{i-1}]$ for every $i \in \{2, \ldots k\}$ and that the k eigenvalues ν_i are the k solutions of the secular equation:

$$f(\nu) := 1 + \sum_{i=1}^{k} \frac{z_i^2}{\lambda_i - \nu} = 0 \tag{18}$$

if $z_i \neq 0$ for all i. If $z_i = 0$, there is an eigenvalue ν_j such that $\nu_j = \lambda_i$. We assume $z_i \neq 0$ for all i.

If $\nu_s \geq \lambda_s + \frac{\alpha^2}{2}$, there is nothing to prove. Let us assume $\nu_s < \lambda_s + \frac{\alpha^2}{2}$. Hence, the eigenvalues $\nu_s, \nu_{s+1}, \ldots, \nu_k$ lie in the interval $[\lambda_k, \lambda_s + \alpha^2/2)$. For any $\nu \in [\lambda_k, \lambda_s + \alpha^2/2)$ we have

$$0 \leq \zeta(\nu) := \sum_{i=1}^{s-1} \frac{z_i^2}{\lambda_i - \nu} \leq \frac{\sum_{i=1}^{s-1} z_i^2}{\lambda_{s-1} - \nu} \tag{19}$$

$$\leq \frac{b^2 - \alpha^2}{\lambda_{s-1} - \nu} \leq \frac{b^2 - \alpha^2}{b^2 - \alpha^2} = 1. \tag{20}$$

By rewriting the equation $f(\nu) = 0$, for any solution $\nu_\star$ which lies in $[\lambda_k, \lambda_s + \alpha^2/2)$ we obtain

$$0 = 1 + \sum_{i=s}^{k} \frac{z_i^2}{(\lambda_i - \nu_\star)(1 + \zeta(\nu_\star))} \leq 1 + \sum_{i=s}^{k} \frac{z_i^2}{2(\lambda_i - \nu_\star)}$$

because $1 < 1 + \zeta(\nu_\star) \leq 2$ and $\sum_{i=s}^k \frac{z_i^2}{(\lambda_i - \nu_\star)} < 0$. Now, let ν_j be the unique solution $f(\nu_j) = 0$ in the interval $[\lambda_j, \lambda_{j-1}]$ for $j \in \{s+1, \ldots, k\}$ or in the interval $[\lambda_s, \lambda_s + \alpha^2/2)$ for $j = s$.

Since the function $g(\nu) = 1 + \sum_{i=s}^k \frac{z_i^2}{2(\lambda_i - \nu)}$ is increasing on the interval $[\lambda_j, \lambda_{j-1}]$ (if $j = s$, the interval is $[\lambda_s, \infty)$) we know that the unique solution $\nu'_j \in [\lambda_j, \lambda_{j-1}]$ of the equation $g(\nu) = 0$ satisfies $\nu'_j \leq \nu_j$ for all $j \in \{s, \ldots, k\}$.

Therefore, we have shown that $\sum_{j=s}^k \nu_j \geq \sum_{j=s}^k \nu'_j$, where ν'_j are the solutions to the equation

$$1 + \sum_{i=s}^{k} \frac{z_i^2}{2(\lambda_i - \nu)} = 0.$$

However, the solutions of this equation are the eigenvalues of $Q = \operatorname{diag}(\lambda_s, \lambda_{s+1}, \ldots, \lambda_k) + \frac{1}{2} z z^\top$, where $z = [z_s, z_{s+1}, \ldots, z_k]^\top$. Hence,

$$\sum_{j=s}^{k} \nu_j \geq \sum_{j=s}^{k} \nu'_j = \mathbf{tr}(Q) = \sum_{j=s}^{k} \lambda_j + \frac{1}{2} \sum_{j=s}^{k} z_j^2 \geq \sum_{j=s}^{k} \lambda_j + \frac{\alpha^2}{2}.$$

■

Now we can turn back to the proof of Proposition 4. Let $\lambda_i(t)$ be the i-th largest eigenvalue of $\sum_{j=1}^{t} v_j v_j^\top$ and let $K(t)$ be the number of times

$$v_{j+1} \in \{v | v \in \mathcal{V} \text{ and } |\langle \tilde{v}_{j+1}, v\rangle| \geq \alpha\}$$

with $\tilde{v}_{j+1} \in \arg\min_{\|v\|=1} v^\top \left(\sum_{i=1}^{j} v_i v_i^\top\right) v$ and $j < t$.

We prove that $\sum_{i=j}^{k} \lambda_i(t)$ grows linearly with $K(t)$ by induction on j. The base case $j = 1$ is clearly satisfied since it corresponds to the sum of all eigenvalues. Suppose we know that $\sum_{i=j-1}^{k} \lambda_i(t) \geq c_{j-1}\alpha^2 K(t) - d_{j-1}$ for all $t \geq 1$, where $c_{j-1} > 0$ and $d_{j-1} \geq 0$ are some real values. Since $\|v_j\| \geq \alpha$ for all j, we can choose $c_1 = 1$ and $d_1 = 0$. Now, we lower bound $\sum_{i=j}^{k} \lambda_i(t)$ as a function of t. To this end, we define t_j to be the maximum time in $\{1, 2, \ldots, t\}$ such that $\lambda_{s-1}(t_j) - \lambda_s(t_j) < b^2 - \frac{\alpha^2}{2}$ for all $s \in \{j, j+1, \ldots, k\}$.

Then, Lemma 8 and our induction hypothesis guarantee that

$$\begin{aligned}\sum_{i=j}^{k} \lambda_i(t) &\geq \sum_{i=j}^{k} \lambda_i(t_j) + \frac{\alpha^2(K(t) - K(t_j))}{2} \\ &\geq \frac{\alpha^2(K(t) - K(t_j))}{2} + c_{j-1}\alpha^2 K(t_j) - d_{j-1} - \lambda_{j-1}(t_j).\end{aligned}$$

By the definition of t_j we know that $\lambda_i(t_j) \geq \lambda_{j-1}(t_j) - (i - j + 1)(b^2 - \alpha^2)$ for all $i \geq j$. Therefore, we have the lower bound:

$$\sum_{i=j}^{k} \lambda_i(t) \geq \sum_{i=j}^{k} \lambda_i(t_j) \geq (k - j + 1)\lambda_{j-1}(t_j) - \frac{(k - j + 1)(k - j + 2)}{2}\left(b^2 - \frac{\alpha^2}{2}\right).$$

We minimize the maximum of the previous two lower bounds with respect to $\lambda_{j-1}(t_j)$, which can be done by finding the value of $\lambda_{j-1}(t_j)$ which makes the two lower bounds equal. Then, we find

$$\sum_{i=j}^{k} \lambda_i(t) \geq \frac{\alpha^2}{2}\frac{k - j + 1}{k - j + 2}\left((2c_{j-1} - 1)K(t_j) + K(t)\right) - \frac{k - j + 1}{k - j + 2}d_{j-1} - \frac{k - j + 1}{2}\left(b^2 - \frac{\alpha^2}{2}\right).$$

Case 1: $2c_{j-1} \geq 1$. Then, since $K(t_j) \geq 0$, we obtain

$$\sum_{i=j}^{k} \lambda_i(t) \geq \frac{\alpha^2}{2}\frac{k - j + 1}{k - j + 2}K(t) - \frac{k - j + 1}{k - j + 2}d_{j-1} - \frac{k - j + 1}{2}\left(b^2 - \frac{\alpha^2}{2}\right).$$

Case 2: $2c_{j-1} < 1$. Then, since $K(t_j) \leq K(t)$, we obtain

$$\sum_{i=j}^{k} \lambda_i(t) \geq \alpha^2\frac{k - j + 1}{k - j + 2}c_{j-1}K(t) - \frac{k - j + 1}{k - j + 2}d_{j-1} - \frac{k - j + 1}{2}\left(b^2 - \frac{\alpha^2}{2}\right).$$

Now, we can define recursively

$$c_j = \frac{k-j+1}{k-j+2} c_{j-1}$$
$$d_j = \frac{k-j+1}{k-j+2} d_{j-1} + \frac{k-j+1}{2}\left(b^2 - \frac{\alpha^2}{2}\right),$$

with $c_2 = \frac{k-1}{2k}$ and $d_2 = \frac{k-1}{2}\left(b^2 - \frac{\alpha^2}{2}\right)$. Note that $c_2 < 1/2$ and $c_j \leq c_{j-1}$. By unrolling the recursions, we obtain the conclusion.

Appendix D. Refinement of Assumption 6

We saw that when Assumption 6 offers a lower bound

$$\lambda_{\min}\left(\sum_{t=0}^{t_0-1} \phi(\mathbf{x}_t, \mathbf{u}_t)\phi(\mathbf{x}_t, \mathbf{u}_t)^\top\right) \geq 1 + c_2 b_w^2 L^2 \left(\sum_{i=0}^{H-1} \gamma^i\right)\left(d + k\log(b_\phi^2 T) + \log\left(\frac{\pi^2 T^2}{6\delta}\right)\right)$$

Algorithm 1 with parameter $\beta = c_4\left(d + k\log(\beta_\phi^2 T) + \log(\pi^2 T^2/(6\delta))\right)^{-1}$ is guaranteed to collect measurements $(\phi_t, \mathbf{x}_{t+1})$ such that

$$\lambda_{\min}\left(\sum_{t=0}^{T+t_0} \phi_t \phi_t^\top\right) \geq \frac{\alpha^2}{32}\left(\frac{T}{H} - N_e(T)\right) - \frac{k-1}{2} b_\phi^2. \tag{21}$$

We wrote $N_e(T)$ because N_e is a function of β and β is a function of T. Then, let us consider $T_\star$ to be the smallest value such that

$$\frac{\alpha^2}{32}\left(\frac{T_\star}{H} - N_e(T_\star)\right) - \frac{k-1}{2} b_\phi^2 \geq 1 + c_2 b_w^2 L^2 \left(\sum_{i=0}^{H-1} \gamma^i\right)\left(d + k\log(2b_\phi^2 T_\star) + \log\left(\frac{\pi^2 T_\star^2}{3\delta}\right)\right). \tag{22}$$

Such $T_\star$ exists because $N_e(T_\star) = \mathcal{O}(\log(T_\star))$. Moreover, if (22) holds, any $T \geq T_\star$ satisfies (22).

Now, suppose Assumption 6 is satisfied with T replaced by $T_\star$. Then, we can set

$$\beta = c_4\left(d + k\log(\beta_\phi^2 T_\star) + \log(\pi^2 T_\star^2/(6\delta))\right)^{-1}$$

for all iterations of Algorithm 1 with $t \leq T_\star$. When $t > T_\star$ inequalities (21) and (22) guarantee that we can update β to be equal to

$$c_4\left(d + k\log(2\beta_\phi^2 T_\star) + \log(2\pi^2 T_\star^2/(3\delta))\right)^{-1}$$

and that the minimal eigenvalue of $\sum_{t=0}^{T} \phi_t \phi_t^\top$ is sufficiently large as long as $T \leq 2T_\star$. Therefore, we can update β in epochs of doubling lengths and have Theorem 1 hold as long as Assumption 6 holds with T replaced by $T_\star$.

Appendix E. Detailed Pseudo-code of Algorithm 1

Algorithm 2 Active learning for nonlinear system identification

Require: Parameters: the feature map ϕ, initial trajectory $\mathcal{D}$, and parameters T, α, and β.

Initialize Φ to have rows $\phi(\mathbf{x}_j, \mathbf{u}_j)^\top$ and Y to have rows $(\mathbf{x}_{j+1})^\top$, for $(\mathbf{x}_j, \mathbf{u}_j, \mathbf{x}_{j+1}) \in \mathcal{D}$.
Set $\widehat{A} \leftarrow Y^\top \Phi (\Phi^\top \Phi)^{-1}$, i.e. the OLS estimate according to $\mathcal{D}$.
Set $t \leftarrow t_0$.
while $t \leq T + t_0$ **do**
  Set $\mathbf{x}_0^R \leftarrow \mathbf{x}_t$,
  Set v to be a minimal eigenvector of $\Phi^\top \Phi$, with $\|v\| = 1$.
  Trajectory planning: find inputs $\mathbf{u}_0^R$, $\mathbf{u}_1^R$, …, $\mathbf{u}_r^R$, with $\|\mathbf{u}_j^R\| \leq b_u$ and $r \leq H$, such that

$$\left|\langle \phi(\mathbf{x}_r^R, \mathbf{u}_r^R), v \rangle\right| \geq \frac{\alpha}{2} \text{ or } \phi(\mathbf{x}_r^R, \mathbf{u}_r^R)^\top (\Phi^\top \Phi)^{-1} \phi(\mathbf{x}_r^R, \mathbf{u}_r^R) \geq \beta,$$

  where $\mathbf{x}_{j+1}^R = \widehat{A}\phi(\mathbf{x}_j^R, \mathbf{u}_j^R)$ for all $j \in \{0, 1, \ldots, r-1\}$.
  Trajectory tracking:
  for $j = 0, \ldots, r$ **do**
    if $\max_{\mathbf{u} \in \mathbb{B}_{r_u}} \phi(\mathbf{x}_t, \mathbf{u})^\top (\Phi^\top \Phi)^{-1} \phi(\mathbf{x}_t, \mathbf{u}) \geq \beta$ **then**
      Set $\mathbf{u}_t \in \arg\max_{\mathbf{u} \in \mathbb{B}_{r_u}} \phi(\mathbf{x}_t, \mathbf{u})^\top (\Phi^\top \Phi)^{-1} \phi(\mathbf{x}_t, \mathbf{u})$,
      Input $\mathbf{u}_t$ into the real system and observe the next state: $\mathbf{x}_{t+1} = A_\star \phi(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{w}_t$,
      $t \leftarrow t + 1$,
      break.
    else if $j \leq r - 1$ **then**
      Set $\mathbf{u}_t \in \arg\min_{\mathbf{u} \in \mathbb{B}_{r_u}} \|\widehat{A}(\phi(\mathbf{x}_t, \mathbf{u}) - \phi(\mathbf{x}_j^R, \mathbf{u}_j^R))\|$.
    else
      $\mathbf{u}_t = \mathbf{u}_r^R$.
    end if
    Input $\mathbf{u}_t$ into the real system and observe the next state: $\mathbf{x}_{t+1} = A_\star \phi(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{w}_t$,
    $t \leftarrow t + 1$.
  end for
  Set $\Phi^\top \leftarrow [\phi_0, \phi_1, \ldots, \phi_{t-1}]$ and $Y^\top \leftarrow [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t]$, where $(\phi_j, \mathbf{x}_{j+1})$ are all feature-state transitions observed so far.
  Re-estimate: $\widehat{A} \leftarrow Y^\top \Phi (\Phi^\top \Phi)^{-1}$.
end while
Output the last estimate $\widehat{A}$.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in Neural Information Processing Systems*, 2011.

Pieter Abbeel and Andrew Y Ng. Exploration and apprenticeship learning in reinforcement learning. *International Conference on Machine Learning*, pages 1–8, 2005.

Karl Johan Åström and Peter Eykhoff. System identification—a survey. *Automatica*, 7(2):123–162, 1971.

Sohail Bahmani and Justin Romberg. Convex programming for estimation in nonlinear recurrent models. *Journal of Machine Learning Research*, 21(235):1–20, 2020.

Xavier Bombois, Michel Gevers, Roland Hildebrand, and Gabriel Solari. Optimal experiment design for open and closed-loop system identification. *Communications in Information and Systems*, 11(3):197–224, 2011.

Francesco Borrelli, Alberto Bemporad, Michael Fodor, and Davor Hrovat. An MPC/hybrid system approach to traction control. *IEEE Transactions on Control Systems Technology*, 14(3):541–552, 2006.

Francesco Borrelli, Alberto Bemporad, and Manfred Morari. *Predictive control for linear and hybrid systems*. Cambridge University Press, 2017.

Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

James R. Bunch, Christopher P. Nielsen, and Danny C. Sorensen. Rank-one modification of the symmetric eigenproblem. *Numerische Mathematik*, 31(1):31–48, 1978.

Eduardo F. Camacho, Daniel R. Ramírez, Daniel Limón, David Muñoz de la Peña, and Teodoro Alamo. Model predictive control techniques for hybrid systems. *Annual Reviews in Control*, 34 (1):21–31, 2010.

Marco C. Campi and Erik Weyer. Finite sample properties of system identification methods. *IEEE Transactions on Automatic Control*, 47(8):1329–1334, 2002.

Alessandro Chiuso and Gianluigi Pillonetto. System identification: a machine learning perspective. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:281–304, 2019.

Munther A. Dahleh, Theodore V. Theodosopoulos, and John N. Tsitsiklis. The sample complexity of worst-case identification of FIR linear systems. *Systems & Control Letters*, 20(3):157–166, 1993.

Varsha Dani, Thomas P. Hayes, and Sham M. Kakade. Stochastic linear optimization under bandit feedback. *Conference on Learning Theory*, pages 355–366, 2008.

Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. Finite time identification in unstable linear systems. *Automatica*, 96:342–353, 2018.

Salar Fattahi, Nikolai Matni, and Somayeh Sojoudi. Learning sparse dynamical systems from a single sample trajectory. *IEEE Conference on Decision and Control*, pages 2682–2689, 2019.

Dylan J Foster, Alexander Rakhlin, and Tuhin Sarkar. Learning nonlinear dynamical systems from a single trajectory. *Learning for Dynamics and Control*, 2020.

Tobias Geyer, Georgios Papafotiou, and Manfred Morari. Hybrid model predictive control of the step-down DC–DC converter. *IEEE Transactions on Control Systems Technology*, 16(6):1112–1124, 2008.

Weiqiao Han and Russ Tedrake. Feedback design for multi-contact push recovery via LMI approximation of the piecewise-affine quadratic regulator. *IEEE-RAS International Conference on Humanoid Robotics*, pages 842–849, 2017.

Moritz Hardt, Tengyu Ma, and Benjamin Recht. Gradient descent learns linear dynamical systems. *Journal of Machine Learning Research*, 19(1):1025–1068, 2018.

Elad Hazan, Karan Singh, and Cyril Zhang. Learning linear dynamical systems via spectral filtering. *Advances in Neural Information Processing Systems*, 2017.

Elad Hazan, Holden Lee, Karan Singh, Cyril Zhang, and Yi Zhang. Spectral filtering for general linear dynamical systems. *Advances in Neural Information Processing Systems*, 2018.

Maurice Heemels, Bart De Schutter, and Alberto Bemporad. Equivalence of hybrid dynamical models. *Automatica*, 37(7):1085–1091, 2001.

Xia Hong, Richard J. Mitchell, Sheng Chen, Chris J. Harris, Kang Li, and George W. Irwin. Model selection approaches for non-linear system identification: a review. *International Journal of Systems Science*, 39(10):925–946, 2008.

Chi Jin, Akshay Krishnamurthy, Max Simchowitz, and Tiancheng Yu. Reward-free exploration for reinforcement learning. *International Conference on Machine Learning*, 2020.

Anatoli Juditsky, Håkan Hjalmarsson, Albert Benveniste, Bernard Delyon, Lennart Ljung, Jonas Sjöberg, and Qinghua Zhang. Nonlinear black-box models in system identification: mathematical foundations. *Automatica*, 31(12):1725–1750, 1995.

Sham Kakade, Akshay Krishnamurthy, Kendall Lowrey, Motoya Ohnishi, and Wen Sun. Information theoretic regret bounds for online nonlinear control. In *Advances in Neural Information Processing Systems*, 2020.

Lydia E. Kavraki, Petr Švestka, Jean-Claude Latombe, and Mark H. Overmars. Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE Transactions on Robotics and Automation*, 12(4):566–580, 1996.

Mohammad Khosravi and Roy S Smith. Convex nonparametric formulation for identification of gradient flows. *IEEE Control Systems Letters*, 5(3):1097–1102, 2020a.

Mohammad Khosravi and Roy S Smith. Nonlinear system identification with prior knowledge on the region of attraction. *IEEE Control Systems Letters*, 5(3):1091–1096, 2020b.

Steven M LaValle and James J. Kuffner Jr. Randomized kinodynamic planning. *The International Journal of Robotics Research*, 20(5):378–400, 2001.

Lennart Ljung. System identification: theory for the user. *Prentice Hall*, 1987.

Lennart Ljung, Tianshi Chen, and Biqiang Mu. A shift in paradigm for system identification. *International Journal of Control*, 93(2):173–180, 2020.

Tobia Marcucci, Robin Deits, Marco Gabiccini, Antonio Bicchi, and Russ Tedrake. Approximate hybrid model predictive control for multi-contact push recovery in complex environments. *IEEE-RAS International Conference on Humanoid Robotics*, pages 31–38, 2017.

Andrew Y. Ng, Adam Coates, Mark Diel, Varun Ganapathi, Jamie Schulte, Ben Tse, Eric Berger, and Eric Liang. Autonomous inverted helicopter flight via reinforcement learning. *Experimental Robotics IX, Springer*, pages 363–372, 2006.

Samet Oymak. Stochastic gradient descent learns state equations with nonlinear activations. *Conference on Learning Theory*, 2019.

Samet Oymak and Necmiye Ozay. Non-asymptotic identification of LTI systems from a single trajectory. *American Control Conference*, pages 5655–5661, 2019.

Friedrich Pukelsheim. *Optimal Design of Experiments*. Wiley & Sons, 1993.

Stephane Ross and J. Andrew Bagnell. Agnostic system identification for model-based reinforcement learning. *International Conference on Machine Learning*, 2012.

Paat Rusmevichientong and John N Tsitsiklis. Linearly parameterized bandits. *Mathematics of Operations Research*, 35(2):395–411, 2010.

Sadra Sadraddini and Russ Tedrake. Sampling-based polytopic trees for approximate optimal control of piecewise affine systems. *International Conference on Robotics and Automation*, pages 7690–7696, 2019.

Tuhin Sarkar and Alexander Rakhlin. Near optimal finite time identification of arbitrary linear dynamical systems. *International Conference on Machine Learning*, 2019.

Tuhin Sarkar, Alexander Rakhlin, and Munther A. Dahleh. Nonparametric system identification of stochastic switched linear systems. *IEEE Conference on Decision and Control*, 2019.

Tuhin Sarkar, Alexander Rakhlin, and Munther A Dahleh. Finite time lti system identification. *Journal of Machine Learning Research.*, 22:26–1, 2021.

Shankar Sastry. *Nonlinear Systems: Analysis, Stability, and Control*. Springer, 1999.

Yahya Sattar and Samet Oymak. Non-asymptotic and accurate learning of nonlinear dynamical systems. *arXiv preprint:2002.08538*, 2020.

Johan Schoukens and Lennart Ljung. Nonlinear system identification: a user-oriented road map. *IEEE Control Systems Magazine*, 39(6):28–99, 2019.

Max Simchowitz, Horia Mania, Stephen Tu, Michael I. Jordan, and Benjamin Recht. Learning without mixing: towards a sharp analysis of linear system identification. *Conference on Learning Theory*, pages 439–473, 2018.

Max Simchowitz, Ross Boczar, and Benjamin Recht. Learning linear dynamical systems with semi-parametric least squares. *Conference on Learning Theory*, pages 2714–2802, 2019.

Sumeet Singh, Spencer M Richards, Vikas Sindhwani, Jean-Jacques E Slotine, and Marco Pavone. Learning stabilizable nonlinear dynamics with contraction-based regularization. *The International Journal of Robotics Research*, 40(10-11):1123–1150, 2021.

Jonas Sjöberg, Qinghua Zhang, Lennart Ljung, Albert Benveniste, Bernard Deylon, Pierre-Yves Glorennec, Håkan Hjalmarsson, and Anatoli Juditsky. Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 31:1691–1724, 1995.

Jean-Jacques E. Slotine and Weiping Li. *Applied Nonlinear Control*. Prentice Hall, 1991.

Eduardo Sontag. Nonlinear regulation: The piecewise linear approach. *IEEE Transactions on Automatic Control*, 26(2):346–358, 1981.

Yue Sun, Samet Oymak, and Maryam Fazel. Finite sample system identification: improved rates and the role of regularization. *Learning for Dynamics and Control*, 2020.

Anastasios Tsiamis and George J. Pappas. Finite sample analysis of stochastic system identification. *IEEE Conference on Decision and Control*, 2019.

Anastasios Tsiamis, Nikolai Matni, and George J. Pappas. Sample complexity of kalman filtering for unknown systems. *Learning for Dynamics and Control*, 2019.

Andrew Wagenmaker and Kevin Jamieson. Active learning for identification of linear dynamical systems. *Conference on Learning Theory*, 2020.

Geoffrey Wolfer and Aryeh Kontorovich. Minimax learning of ergodic Markov chains. *International Conference on Algorithmic Learning Theory*, 2019.

Boyan Yordanov, Jana Tumova, Ivana Cerna, Jiří Barnat, and Calin Belta. Temporal logic control of discrete-time piecewise affine systems. *IEEE Transactions on Automatic Control*, 57(6):1491–1504, 2011.

Matt Zucker, Nathan Ratliff, Anca D. Dragan, Mihail Pivtoraiko, Matthew Klingensmith, Christopher M. Dellin, J. Andrew Bagnell, and Siddhartha S. Srinivasa. Chomp: Covariant Hamiltonian optimization for motion planning. *The International Journal of Robotics Research*, 32(9-10): 1164–1193, 2013.