EXACT RECOVERY GUARANTEES FOR PARAMETERIZED NONLINEAR SYSTEM IDENTIFICATION PROBLEM UNDER ADVERSARIAL ATTACKS

Anonymous authors
Paper under double-blind review

ABSTRACT

In this work, we study the system identification problem for parameterized nonlinear systems using basis functions under adversarial attacks. Motivated by the LASSO-type estimators, we analyze the exact recovery property of a nonsmooth estimator, which is generated by solving an embedded ℓ_1-loss minimization problem. First, we derive necessary and sufficient conditions for the well-specifiedness of the estimator and the uniqueness of global solutions to the underlying optimization problem. Next, we provide exact recovery guarantees for the estimator under two different scenarios of boundedness and Lipschitz continuity of the basis functions. The non-asymptotic exact recovery is guaranteed with high probability, even when there are more severely corrupted data than clean data. Finally, we numerically illustrate the validity of our theory. This is the first study on the sample complexity analysis of a nonsmooth estimator for the nonlinear system identification problem.

1 INTRODUCTION

Dynamical systems are the foundation for the areas of sequential decision-making, reinforcement learning, control theory, and recurrent neural networks. They are imperative for modeling the mechanics governing the system and for predicting the states of a system. However, it is cumbersome to exactly model these systems due to the growing complexity of contemporary systems. Thus, the learning of these system dynamics is essential for an accurate decision-making. The problem of estimating the dynamics of a system using past information collected from the system is called the *system identification* problem. This problem is ubiquitously studied in the control theory literature for systems under relatively small independent and identically distributed (i.i.d.) noise due to modeling, measurement, and sensor errors. Nevertheless, safety-critical applications, such as power systems, autonomous vehicles, and unmanned aerial vehicles, require the robust estimation of the system due to the possible presence of adversarial disturbance, such as natural disasters and data manipulation through cyberattacks and system hacking. Although machine learning techniques have been successful in addressing a wide range of problems, such as computer vision and language processing, their application in safety-critical systems has been extremely limited due to the lack of theoretical guarantees. This paper offers a strong result in this regard, which is concerned with studying dynamical systems via machine learning techniques.

As a motivating example, we consider the dynamical system corresponding to a power system (e.g., the U.S. electrical grid or a regional interconnect of the grid), where the states capture various physical parameters such as voltage magnitudes and frequencies in different parts of the system. With the objective of increasing sustainability, resiliency and efficiency of energy systems, modern power systems include a large volume of wind turbines, solar panels, and electric cars. The operation of power systems is further complicated by the fact that people have started to play an active role by observing electricity prices and taking strategic actions in response to the price signals. On the other hand, sensors have been widely installed across the grid to collect data to enable data-driven grid operation. This has raised a major concern since a small strategic data manipulation would cause power suppliers to over-supply or under-supply power electricity, which could lead to a system-wide blackout. This case can be modeled as a nonlinear dynamical system where the input of the system is subject to stealth attacks at various locations leading to injecting wrong values of electricity into the system. Given the presence of a large set of new devices in the system coupled by strategic human behavior, the power operators do not have a complete model of the dynamical system. Therefore, they may need to learn the system and any possible adversarial attack simultaneously to be able to nullify the attack and restore the system's operation. When an attack to the input is not controlled, it will affect the transient behavior and make the signals unstable , which leads to a cascading failure across the grid.

The prior system identification literature has mainly focused on attacks on measurements, where the goal is to extract knowledge from noisy and corrupted measurements (such as the matrix sensing problem in machine learning). However, this paper considers an emerging and overlooked type of attack for safety-critical systems, where the attack on input data or actuators leads to injecting a wrong input signal into the system which affects the states of the system and makes them unstable. More relevant literature focused on the asymptotic properties of the least-squares estimator (LSE) Chen & Guo (2012); Ljung et al. (1999); Ljung & Wahlberg (1992); Bauer et al. (1999), and with the emergence of statistical learning theory, this area evolved into studying the necessary number of samples for a specific error threshold to be met Tsiamis et al. (2023). While early non-asymptotic analyses centered on linear-time invariant (LTI) systems with i.i.d. noise using mixing arguments Kuznetsov & Mohri (2017); Rostamizadeh & Mohri (2007), recent research employs martingale and small-ball techniques to provide sample complexity guarantees for LTI systems Simchowitz et al. (2018); Faradonbeh et al. (2018); Tsiamis & Pappas (2019). For nonlinear systems, recent studies investigated parameterized models Noël & Kerschen (2017); Nowak (2002); Foster et al. (2020); Sattar & Oymak (2022); Ziemann et al. (2022), showing convergence of recursive and gradient algorithms to true parameters with a rate of $T^{-1/2}$ using martingale techniques and mixing time arguments. Furthermore, efforts towards nonsmooth estimators for both linear and nonlinear systems Feng & Lavaei (2021); Feng et al. (2023); Yalcin et al. (2023), particularly in handling dependent and adversarial noise vectors, are limited. Robust regression techniques utilizing regularizers have been developed Xu et al. (2009); Bertsimas & Copenhaver (2018); Huang et al. (2016), yet non-asymptotic analysis on sample complexity remains sparse, especially for dynamical systems due to sample auto-correlation. A more detailed literature review is provided in Appendix.

This paper paves the way for the area of online optimal control in presence of adversaries and the first step is to learn the dynamics of the system, known as the system identification problem. More specifically, we study the system identification problem for parameterized nonlinear systems in the presence of adversarial attacks. We model the unknown nonlinear functions describing the system via a linear combination of some given basis functions, by taking advantage of their representation properties. Our goal is to learn the parameters of these basis functions that govern the updates of the dynamical system. Mathematically, we consider the following autonomous dynamical system:

$$x_0 = 0_n, \quad x_{t+1} = \bar{A}f(x_t) + \bar{d}_t, \quad \forall t \in \{0, \dots, T-1\}, \tag{1}$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ is a combination of m known basis functions and $\bar{A} \in \mathbb{R}^{n \times m}$ is the unknown matrix of parameters. In addition, the system trajectory is attacked by the adversarial noise or disturbance $\bar{d}_t \in \mathbb{R}^n$, which is unknown to the system operator. At any time instance that the system is not attacked, we have $\bar{d}_t = 0$. In other words, the noise only stems from adversarial attacks. The goal of the system identification problem is to recover the ground truth matrix $\bar{A}$ using observations from the states of the system, i.e., $\{x_0, \dots, x_T\}$. The adversarial noise $\bar{d}_t$'s are designed by an attacker to maximize the impact as much as possible and yet keep the attacks undetectable to the system operator. The underlying assumptions about the noise model will be given later.

One of the main challenges of this estimation problem is the time dependence of the collected samples. As opposed to the empirical risk minimization problem, there exists auto-correlation among the samples $\{x_0, \dots, x_T\}$. As a result, the common assumption that the samples are i.i.d. instances of the data generation distribution is violated. The existence of the auto-correlation imposes significant challenges on the theoretical analysis, and we address it in this work by proposing a novel and non-trivial extension of the area of exact recovery guarantees to the system identification problem. Since the adversarial attacks $\bar{d}_t$ are unknown to the system operator, it is necessary to utilize estimators to the ground truth $\bar{A}$ that are robust to the noise $\bar{d}_t$ and converge to $\bar{A}$ when the sample size T is large enough. Our work is inspired by Yalcin et al. (2023) that studies the above problem for linear systems. The linear case is noticeably simpler than the nonlinear system identification problem since each observation x_t becomes a linear function of previous disturbances. In the nonlinear case, the relationship between the measurements and the disturbances are highly sophisticated, which requires significant technical developments compared to the linear case in Yalcin et al. (2023).

Motivated by the exact recovery property of nonsmooth loss functions (e.g., the ℓ_1-norm and the nuclear norm), we consider the following estimator:

$$\hat{A} \in \arg\min_{A\in\mathbb{R}^{n\times m}} \sum_{t=0}^{T-1} \|x_{t+1} - Af(x_t)\|_2. \tag{2}$$

We note that the optimization problem on the right hand-side is convex in A (while having a nonsmooth objective) and, therefore, it can be solved efficiently by various existing optimization solvers. The estimator equation 2 is closely related to the LASSO estimator in the sense that the loss function in equation 2 can be viewed as a generalization of the ℓ_1-loss function. More specifically, in the case when $n = 1$, the estimator equation 2 reduces to

$$\hat{A} \in \arg\min_{A\in\mathbb{R}^{1\times m}} \sum_{t=0}^{T-1} |x_{t+1} - Af(x_t)|,$$

which is the auto-correlated linear regression estimator with the ℓ_1-loss function.

In this work, the goal is to prove the efficacy of the above estimator by obtaining mild conditions under which the ground truth $\bar{A}$ can be *exactly recovered* by the estimator equation 2. More specifically, we focus on the following questions:

i) What are the *necessary and sufficient* conditions such that $\bar{A}$ is an optimal solution to the optimization problem in equation 2 or the unique solution?
ii) What is the required number of samples such that the above necessary and sufficient conditions are satisfied with high probability under certain assumptions?

In this work, we provide answers to the above questions. In Section 2, we first analyze the necessary and sufficient conditions for the global optimality of $\bar{A}$ for the problem in equation 2. Then, in Section 3, we establish the necessary and sufficient conditions such that $\bar{A}$ is the unique solution. The results in these two sections provide an answer to question (i). Finally, in Sections 4 and 5, we derive lower bounds on the number of samples T such that $\bar{A}$ is the unique solution with high probability in the case when the basis function f is bounded or Lipschitz continuous, respectively. These results serve as an answer to question (ii). We provide numerical experiments that support the theoretical results throughout the paper in Section 6. This work provides the first non-asymptotic sample complexity analysis to the exact recovery of the nonlinear system identification problem.

Notation. For a positive integer n, we use 0_n and I_n to denote the n-dimensional vector with all entries being 0 and the n-by-n identity matrix. For a matrix Z, $\|Z\|_F$ denotes its Frobenius norm and $\mathbb{S}_F$ is the unit sphere of matrices with Frobenius norm $\|Z\|_F = 1$. For two matrices Z_1 and Z_2, we use $\langle Z_1, Z_2\rangle = \mathrm{Tr}(Z_1^\top Z_2)$ to denote the inner-product. For a vector z, $\|z\|_2$ and $\|z\|_\infty$ denote its ℓ_2- and ℓ_∞-norms, respectively. Moreover, $\mathbb{S}^{n-1}$ is the unit ball $\{z \in \mathbb{R}^n | \|z\|_2 = 1\}$. Given two functions f and g, the notation $f(x) = \Theta[g(x)]$ means that there exist universal positive constants c_1 and c_2 such that $c_1 g(x) \le f(x) \le c_2 g(x)$. The relation $f(x) \lesssim g(x)$ holds if there exists a universal positive constant c_3 such that $f(x) \le c_3 g(x)$ holds with high probability when T is large. The relation $f(x) \gtrsim g(x)$ holds if $g(x) \lesssim f(x)$. $|S|$ shows the cardinality of a given set S. $\mathbb{P}(\cdot)$ and $\mathbb{E}(\cdot)$ denote the probability of an event and the expectation of a random variable. A Gaussian random vector X with mean μ and covariance matrix Σ is written as $X \sim \mathcal{N}(\mu, \Sigma)$.

2 Global Optimality of Ground Truth

In this section, we derive conditions under which the ground truth $\bar{A}$ is a global minimizer to the optimization problem in equation 2. By the system dynamics, the optimization problem is equivalent to

$$\min_{A\in\mathbb{R}^{n\times m}} \sum_{t=0}^{T-1} \|(\bar{A} - A)f(x_t) + \bar{d}_t\|_2, \tag{3}$$

where $x_0, \dots, x_T$ are generated according to the unknown system under adversaries. We define the set of attack times as $\mathcal{K} := \{t \mid \bar{d}_t \neq 0\}$ and the normalized attacks as

$$\hat{d}_t := \bar{d}_t/\|\bar{d}_t\|_2, \quad \forall t \in \mathcal{K}.$$

The following theorem provides a necessary and sufficient condition for the global optimality of ground truth matrix $\bar{A}$ in problem equation 3.

Theorem 1 (Necessary and sufficient condition for optimality)**.** *The ground truth matrix $\bar{A}$ is a global solution to problem equation 3 if and only if*

$$\sum_{t\in\mathcal{K}} \hat{d}_t^\top Z f(x_t) \leq \sum_{t\in\mathcal{K}^c} \|Zf(x_t)\|_2, \quad \forall Z \in \mathbb{R}^{n\times m}, \tag{4}$$

where $\mathcal{K}^c := \{0,\dots,T-1\}\backslash\mathcal{K}$.

Theorem 1 provides a necessary and sufficient condition for the well-specifiedness of optimization problem equation 3. Intuitively, we can view the left-hand side as the impact of noisy attacks and the right-hand side as the normal dynamics. If the impact from noise does not override the correct system dynamics, then the predictor is able to recover the ground truth system dynamics. The condition equation 4 is established by applying the generalized Farkas' lemma, which avoids the inner approximation of the ℓ_2-ball by an ℓ_∞-ball in Yalcin et al. (2023). As a result, the sample complexity bounds to be obtained in this work are stronger than those in Yalcin et al. (2023) when specialized to the setting of linear systems; see Sections 4 and 5 for more details.

Using the condition in Theorem 1, we can derive the necessary conditions and sufficient conditions for the optimality of $\bar{A}$.

Corollary 1 (Sufficient condition for optimality)**.** *If it holds that*

$$\sum_{t\in\mathcal{K}} \|Zf(x_t)\|_2 \leq \sum_{t\in\mathcal{K}^c} \|Zf(x_t)\|_2, \quad \forall Z \in \mathbb{R}^{n\times m}, \tag{5}$$

then the ground truth matrix $\bar{A}$ is a global solution to problem equation 3.

Corollary 2 (Necessary condition for optimality)**.** *If the ground truth matrix $\bar{A}$ is a global solution to problem equation 3, then it holds that*

$$\left\|\sum_{t\in\mathcal{K}} f(x_t)\hat{d}_t^\top\right\|_F \leq \sum_{t\in\mathcal{K}^c} \|f(x_t)\|_2. \tag{6}$$

In the case when $m = 1$, condition equation 6 is necessary and sufficient.

The proof of Corollaries 1 and 2 is provided in the appendix. The above conditions are more general than many existing results in literature; see the following two examples.

Example 1 (First-order systems)**.** *In the special case when $n = m = 1$ and the basis function is $f(x) = x$, condition equation 6 reduces to*

$$\left|\sum_{t\in\mathcal{K}} \hat{d}_t x_t\right| \leq \sum_{t\in\mathcal{K}^c} |x_t|,$$

which is the same as Theorem 1 in Feng & Lavaei (2021).

Example 2 (Linear systems)**.** *We consider the case when $m = n$ and the basis function is $f(x) = x$. We also assume the Δ-spaced attack model; see the definition in Yalcin et al. (2023). By considering the attack period starting at the time step t_1, a sufficient condition to guarantee condition equation 4 is given by*

$$\hat{d}^\top Z\bar{A}^{\Delta-1}\bar{d}_{t_1} \leq \sum_{t=0}^{\Delta-2} \left\|Z\bar{A}^t\bar{d}_{t_1}\right\|_2, \quad \forall Z \in \mathbb{R}^{n\times n}, \tag{7}$$

where we denote $\hat{d} := \hat{d}_{t_1}$ for simplicity. Let $\hat{D} \in \mathbb{R}^{n\times(n-1)}$ be the matrix of orthonormal bases of the orthogonal complementary space of f, namely, $\hat{D}^\top\hat{d} = 0$, $\hat{D}^\top\hat{D} = I_{n-1}$, and $\hat{D}\hat{D}^\top = I_n - \hat{d}\hat{d}^\top$. Then, we can calculate that

$$\left\|Z\bar{A}^t\bar{d}_{t_1}\right\|_2^2 \geq \left(Z\bar{A}^t\bar{d}_{t_1}\right)^\top \hat{d}\hat{d}^\top \left(Z\bar{A}^t\bar{d}_{t_1}\right),$$

where the equality holds when $\hat{D}^\top Z\bar{A}^t\bar{d}_{t_1} = 0$, i.e., $Z\bar{A}^t\bar{d}_{t_1}$ is parallel with $\hat{d}$. Therefore, for condition equation 7 to hold, it is equivalent to consider Z with the form $Z = \hat{d}z^\top$ for some vector $z \in \mathbb{R}^n$. In this case, condition equation 7 reduces to

$$z^\top \bar{A}^{\Delta-1}\bar{d}_{t_1} \leq \sum_{t=0}^{\Delta-2} \left| z^\top \bar{A}^t \bar{d}_{t_1} \right|, \quad \forall z \in \mathbb{R}^n. \tag{8}$$

Condition equation 8 leads to a better sufficient condition than that in Yalcin et al. (2023). To illustrate the improvement, we consider the special case when the ground truth matrix is $\bar{A} = \lambda I_n$ for some $\lambda \in \mathbb{R}$. Then, condition equation 8 becomes

$$|\lambda|^{\Delta-1} \leq \sum_{t=0}^{\Delta-2} |\lambda|^t = \frac{1-|\lambda|^{\Delta-1}}{1-|\lambda|} \quad \text{, which is further equivalent to } |\lambda| + |\lambda|^{1-\Delta} \leq 2,$$

which is a stronger condition than that in Yalcin et al. (2023). When the attack period Δ is large, we approximately have $|\lambda| \leq 2 - 2^{1-\Delta}$, which is a better condition than that in Figure 1 of Yalcin et al. (2023).

3 Uniqueness of Global Solutions

In this section, we derive conditions under which the ground truth solution $\bar{A}$ is the unique solution to problem equation 3. We obtain the following necessary and sufficient condition on the uniqueness of global solutions, which is an extension of Theorem 1.

Theorem 2 (Necessary and sufficient condition for uniqueness)**.** *Suppose that condition equation 4 holds. The ground truth $\bar{A}$ is the unique global solution to problem equation 3 if and only if for every nonzero $Z \in \mathbb{R}^{n\times m}$, it holds that*

$$\sum_{t\in\mathcal{K}} \hat{d}_t^\top Z f(x_t) = \sum_{t\in\mathcal{K}^c} \|Zf(x_t)\|_2 \implies \sum_{t\in\mathcal{K}} \left| \hat{d}_t^\top Z f(x_t) \right| < \sum_{t\in\mathcal{K}} \|Zf(x_t)\|_2, \tag{9}$$

meaning that whenever the left-hand side equality holds, the right-hand side inequality should be implied.

Based on the above theorem, the following corollary provides a sufficient condition for the uniqueness of $\bar{A}$, which is easier to verify in practice compared to equation 9. Note that the corollary also generalizes the sufficiency part of Corollary 2 to the multi-dimensional case.

Corollary 3 (Sufficient condition for uniqueness)**.** *Suppose that condition equation 4 holds. If it holds that*

$$\sum_{t\in\mathcal{K}} \hat{d}_t^\top Z f(x_t) < \sum_{t\in\mathcal{K}^c} \|Zf(x_t)\|_2, \quad \forall Z \in \mathbb{R}^{n\times m} \quad \text{s.t. } Z \neq 0, \tag{10}$$

then the ground truth matrix $\bar{A}$ is the unique global solution to problem equation 3.

Proof. Under condition equation 10, the condition on the left hand-side of equation 9 cannot hold and thus, Theorem 2 implies the uniqueness of $\bar{A}$ as a global solution. □

Similar to the optimality conditions in Section 2, Theorem 2 improves and generalizes the results for first-order systems, namely, Theorem 1 in Feng & Lavaei (2021).

Example 3 (First-order linear systems)**.** *In the case when $m = n = 1$ and $f(x) = x$, our results state that the uniqueness of global solutions is equivalent to*

$$\left| \sum_{t\in\mathcal{K}} \hat{d}_t x_t \right| < \sum_{t\in\mathcal{K}^c} |x_t|. \tag{11}$$

As a comparison, the sufficient condition in Theorem 1 in Feng & Lavaei (2021) is

$$\sum_{t\in\mathcal{K}} |x_t| < \sum_{t\in\mathcal{K}^c} |x_t|.$$

Since $|\hat{d}_t| = 1$ for all $t \in \mathcal{K}$, our results equation 11, as well as Theorem 2, are more general and stronger than that in Feng & Lavaei (2021).

4 BOUNDED BASIS FUNCTION

In the next two sections, we provide lower bounds on the sample complexity T such that the ground truth $\bar{A}$ is the unique solution to problem equation 3. We focus on the following probabilistic attack model:

Definition 1 (Probabilistic attack model). *For each time instance t, the attack vector $\bar{d}_t$ is nonzero with probability $p \in (0, 1)$, which is also independent with other time instances.*

Note that the attack vectors $\bar{d}_t$'s are allowed to be correlated over time and Definition 1 is only about the times at which an attack happens. Recall that we define $\mathcal{K} := \{t \mid \bar{d}_t \neq 0\}$. Then, with probability at least $1 - \exp[-\Theta(pT)]$, it holds that $|\mathcal{K}| = \Theta(pT)$. The probabilistic attack model can be viewed as a measure of the sparseness of attacks in the time horizon, since the parameter p reflects the probability that there exists an attack at a given time. Therefore, under the probabilistic attack model, it is natural to utilize the nonsmooth ℓ_1-loss function to achieve the exact recovery of $\bar{A}$. Our model allows p to be close to 1, meaning that the system is under attack frequently and, thus, most of the collected data is corrupted.

In this section, we consider the case when the basis function f is bounded.

Assumption 1 (Bounded basis function). *The basis function $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ satisfies*

$$\|f(x)\|_\infty \leq B, \quad \forall x \in \mathbb{R}^n,$$

where $B > 0$ is a constant.

Moreover, to avoid the bias in estimation, we assume the following stealthy condition on the attack. Note that a similar condition is assumed in literature Candès et al. (2011); Chen et al. (2021). To state the stealthy condition, we define the filtration $\mathcal{F}_t := \sigma\{x_0, x_1, \ldots, x_t\}$.

Assumption 2 (Stealthy condition). *Conditional on the past information $\mathcal{F}_t$ and the event that $\bar{d}_t \neq 0_n$, the attack direction $\hat{d}_t = \bar{d}_t / \|\bar{d}_t\|_2$ is zero-mean.*

If an attack is not stealth, the operator can quickly detect and nullify it. Therefore, the stealth condition is necessary for making the system identification problem meaningful. Note that we do not assume that the probability distribution or model generating the attack is known. Finally, to avoid the degenerate case, we assume that the norm of basis function is lower bounded under conditional expectation after an attack.

Assumption 3 (Non-degenerate condition). *Conditional on the past information $\mathcal{F}_t$ and the event that $\bar{d}_t \neq 0_n$, the attack vector and the basis function satisfy*

$$\lambda_{min}\left[\mathbb{E}\left[f(x + \bar{d}_t) f(x + \bar{d}_t)^\top \mid \mathcal{F}_t, \bar{d}_t \neq 0_n\right]\right] \geq \lambda^2, \quad \forall x \in \mathbb{R}^n,$$

where $\lambda_{min}(F)$ is the minimal eigenvalue of matrix F and $\lambda > 0$ is a constant.

Intuitively, the non-degenerate assumption allows the exploration of the trajectory in the state space. More specifically, it is necessary that the matrix

$$[f(x_t),\ t \in \mathcal{K}^c] \in \mathbb{R}^{m \times (T - |\mathcal{K}|)} \tag{12}$$

is rank-m for the condition equation 10 to hold; see the proof of Theorem 4 for more details. The non-degenerate assumption guarantees that the basis function $f(x + \bar{d}_t)$ spans the whole state space in expectation and thus, the matrix equation 12 is full-rank with high probability when T is large.

The following theorem proves that when the sample complexity is large enough, the estimator equation 2 exactly recovers the ground truth $\bar{A}$ with high probability.

Theorem 3 (Exact recovery for bounded basis function). *Suppose that Assumptions 1-3 hold and define $\kappa := B/\lambda \geq 1$. For all $\delta \in (0, 1]$, if the sample complexity T satisfies*

$$T \geq \Theta\left[\frac{m^2\kappa^4}{p(1-p)^2}\left[mn \log\left(\frac{m\kappa}{p(1-p)}\right) + \log\left(\frac{1}{\delta}\right)\right]\right], \tag{13}$$

then $\bar{A}$ is the unique global solution to problem equation 3 with probability at least $1 - \delta$.

The above theorem provides a non-asymptotic bound on the sample complexity for the exact recovery with a specified probability $1-\delta$. The lower bound grows with $m^3 n$, which implies that the required number of samples increases when the number of states n and the number of basis functions m is larger. In addition, the sample complexity is larger when B is larger or λ is smaller. This is also consistent with the intuition that B reflects the size of the space spanned by the basis function and λ measures the "speed" of exploring the spanned space.

For the dependence on attack probability p, we show in the next theorem that the dependence on $1/[p(1-p)]$ is inevitable under the probabilistic attack model. In addition, the theorem also establishes a lower bound on the sample complexity that depends on m and $\log(1/\delta)$.

Theorem 4. *Suppose that the sample complexity satisfies*

$$T < \frac{m}{2p(1-p)}.$$

Then, there exists a basis function $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ and an attack model such that Assumptions 1-3 hold and the global solutions to problem equation 3 are not unique with probability at least $\max\left\{1 - 2\exp\left(-m/3\right), 2[p(1-p)]^{T/2}\right\}$. Furthermore, given a constant $\delta \in (0,1]$, if

$$T < \max\left\{\frac{m}{2p(1-p)}, \frac{2}{-\log[p(1-p)]}\log\left(\frac{2}{\delta}\right)\right\},$$

then the global solutions to problem equation 3 are not unique with probability at least $\max\left\{1 - 2\exp\left(-m/3\right), \delta\right\}$.

Remark 1. *The main goal of the paper is to show that exact recovery is possible when more than half of the data are arbitrarily corrupted. We provide an upper bound on the required time horizon in Theorem 3. This result has a major implication for real-world systems. On the other hand, the lower bound in Theorem 4 is mainly a theoretical result. Unlike machine learning problems where the problem size is possibly on the order of tens of millions, the number of states for many real-world systems is much lower and less than several thousands. For that reason, our upper bound is already a practical number and improving the lower bound may have a marginal practical value, although tightening the lower bound is a relevant and interesting theoretical problem.*

5 LIPSCHITZ BASIS FUNCTION

In this section, we consider the case when the basis function $f(x)$ is Lipschitz continuous in x. More specifically, we make the following assumption.

Assumption 4 (Lipschitz basis function)**.** *The basis function $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ satisfies*

$$f(0_n) = 0_m \quad \text{and} \quad \|f(x) - f(y)\|_2 \le L\|x - y\|_2, \quad \forall x, y \in \mathbb{R}^n,$$

where $L > 0$ is the Lipschitz constant.

As a special case of Assumption 4, the basis function of a linear system is $f(x) = x$, which is Lipschitz continuous with Lipschitz constant 1.

Remark 2. *Note that the assumptions of boundedness or Lipschitz continuity are always satisfied for dynamical systems since the user has the choice to select appropriate basis functions to satisfy them. More concretely, the user can select any arbitrary set of basis functions to approximate the unknown function as a linear combination of the bases. This is different from classical machine learning problems where a model is trained to learn the function and there is not control on the Lipschitzness. On the other hand, if the user is not allowed to use unbounded basis functions or functions with a high Lipschitz constant, then the number of basis functions used to approximate the unknown function may be higher. However, many real-world dynamical systems, from robotics to energy systems, are obtained from physical laws where the unknown dynamics is well behaved due to the smoothness of laws of physics, such as Newtonian laws and Kirchhoffs laws of electrical circuits. This is different from various machine learning problems for which the targeted optimal policy could be inevitably nonsmooth and highly complicated.*

In addition, we assume that the spectral norm of $\bar{A}$ is bounded.

Assumption 5 (System stability)**.** *The ground truth $\bar{A}$ satisfies*

$$\rho := \|\bar{A}\|_2 < \frac{1}{L}.$$

We note that Assumption 5 is related to the asymptotic stability of the dynamic system and is sufficient to avoid the finite-time explosion of the dynamics. We show in Theorem 6 that Assumption 5 may be necessary for exact recovery. Finally, we make the assumption that the attack is sub-Gaussian.

Assumption 6 (Sub-Gaussian attacks)**.** *Conditional on the filtration $\mathcal{F}_t$ and the event that $\bar{d}_t \neq 0_n$, the attack vector $\bar{d}_t$ is defined by the product $\ell_t \hat{d}_t$, where*

1. *$\hat{d}_t \in \mathbb{R}^n$ and $\ell_t \in \mathbb{R}$ are independent conditional on $\mathcal{F}_t$ and $\bar{d}_t \neq 0_n$;*
2. *$\hat{d}_t$ is a zero-mean unit vector, namely, $\mathbb{E}(\hat{d}_t \mid \mathcal{F}_t, \bar{d}_t \neq 0_n) = 0_n$ and $\|\hat{d}_t\|_2 = 1$;*
3. *ℓ_k is zero-mean and sub-Gaussian with parameter σ.*

As a special case, the sub-Gaussian assumption is guaranteed to hold if there is an upper bound on the magnitude of the attack. The bounded-attack case is common in practical applications since real-world systems do not accept inputs that are arbitrarily large. For example, physical devices have a clear limitation on the input size and the attacks cannot exceed that limit. In Assumption 6, $\hat{d}_t$ and ℓ_t play the roles of the direction and intensity (such as magnitude) of the attack, respectively. The parameters ℓ_t's could be correlated over time, while $\hat{d}_t$ and ℓ_t are assumed to be zero-mean to make the attack stealth.

Under the above assumptions, we can also guarantee the high-probability exact recovery when the sample size T is sufficiently large.

Theorem 5 (Exact recovery for Lipschitz basis function)**.** *Suppose that Assumptions 3-6 hold and define $\kappa := \sigma L/\lambda \geq 1$. If the sample complexity T satisfies*

$$T \geq \Theta\left[\max\left\{\frac{\kappa^{10}}{(1-\rho L)^3(1-p)^2}, \frac{\kappa^4}{p(1-p)}\right\} \times \left[mn\log\left(\frac{1}{(1-\rho L)\kappa p(1-p)}\right) + \log\left(\frac{1}{\delta}\right)\right]\right], \tag{14}$$

then $\bar{A}$ is the unique global solution to problem equation 3 with probability at least $1-\delta$.

Theorem 5 provides a non-asymptotic sample complexity bound for the case when the basis function is Lipschitz continuous. As a special case, when the basis function is $f(x) = x$ and the attack vector $\bar{d}_t$ obeys the Gaussian distribution $\mathcal{N}(0_n, \sigma^2 I_n)$ conditional on $\mathcal{F}_t$, we have $\kappa = 1$. Compared with Theorem 3, the dependence on attack probability p is improved from $1/[p(1-p)^2]$ to $1/[p(1-p)]$, which is a result of the stability condition (Assumption 5). In addition, the dependence on the dimension m is improved from m^3 to m. Intuitively, the improvement is achieved by improving the upper bound on the norm $\|f(x_t)\|_2$. In the bounded basis function case, the norm is bounded by $\sqrt{m}B$; while in the Lipschitz basis function case, the norm is bounded by σL with high probability, which is independent from the dimension m. Finally, the sample complexity bound grows with the parameter $\kappa = \sigma L/\lambda$ and the gap $1-\rho L$, which is also consistent with the intuition.

On the other hand, we can construct counterexamples showing that when the stability condition (Assumption 5) is violated, the exact recovery fails with probability at least p.

Theorem 6 (Failure of exact recovery for unstable systems)**.** *There exists a system such that Assumptions 3, 4 and 6 are satisfied but for all $T \geq 1$, the ground truth $\bar{A}$ is not a global solution to problem equation 3 with probability at least $p[1-(1-p)^{T-1}]$.*

6 Numerical Experiments

In this section, we implement numerical experiments for the Lipschitz basis function cases to verify the exact recovery guarantees in Section 5. Due to the page limitation, the descriptions of the basis functions and the results for the bounded basis function case are provided in Appendix. More specifically, we illustrate the convergence of estimator equation 2 with different values of the attack probability p, problem dimension (n, m) and spectral norm ρ. In addition, we numerically verify the necessary and sufficient condition in Section 3.

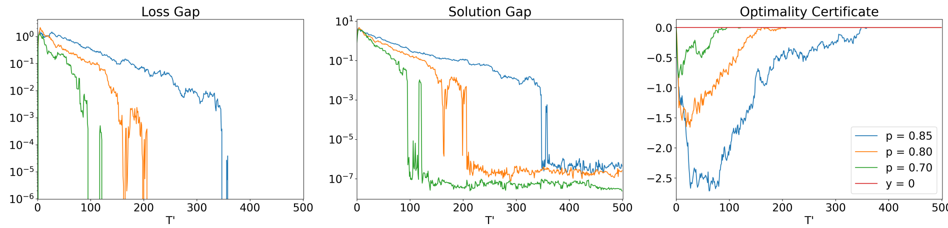

Figure 1: Loss gap, solution gap and optimality certificate of the Lipschitz basis function case with attack probability $p = 0.7, 0.8$ and 0.85.

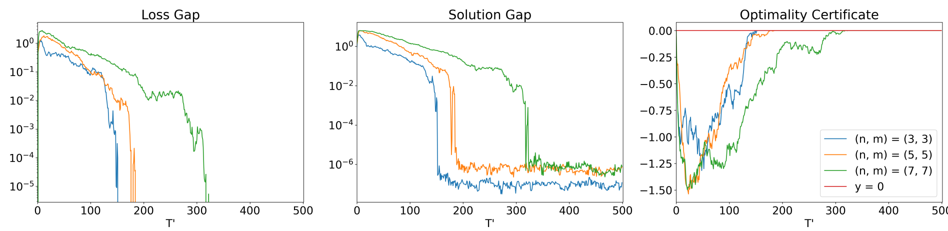

Figure 2: Loss gap, solution gap and optimality certificate of the Lipschitz basis function case with dimension $(n, m) = (3, 3), (5, 5)$ and $(7, 7)$.

Evaluation metrics. Given a trajectory $\{x_0, \dots, x_T\}$, we compute the estimators

$$\hat{A}^{T'} \in \arg \min_{A \in \mathbb{R}^{n \times m}} g_{T'}(A), \quad \forall T' \in \{1, \dots, T\},$$

where we define the loss function $g_{T'}(A) := \sum_{t=0}^{T'-1} \|x_{t+1} - Af(x_t)\|_2$. In our experiments, we solve the convex optimization by the CVX solver Grant & Boyd (2014). Then, for each T', we evaluate the recovery quality by the following three metrics:

- The **Loss Gap** is defined as $g_{T'}(\bar{A}) - g_{T'}(\hat{A}_{T'})$. The ground truth $\bar{A}$ is a global solution if and only if the loss gap is 0.
- The **Solution Gap** is defined as $\|\bar{A} - \hat{A}_{T'}\|_F$. The ground truth $\bar{A}$ is the unique solution only if the solution gap is 0.
- The **Optimality Certificate** is defined as

  $$\min_{Z \in \mathbb{R}^{n \times m}} \sum_{t \in \mathcal{K}^c} \|Zf(x_t)\|_2 - \sum_{t \in \mathcal{K}} \hat{d}_t^\top Zf(x_t) \quad \text{s.t. } \|Z\|_F \le 1,$$

  which is a convex optimization problem and can be solved by the CVX solver. The ground truth is a global solution if and only if the optimality certificate is equal to 0.

We note that it is not possible to evaluate these metrics in practice, since we do not have access to the ground truth $\bar{A}$ and the attack vector $\bar{d}_t$. We evaluate the metrics in our experiments to illustrate the performance of the estimator equation 2 and the proposed optimality conditions. For each choice of parameters, we independently generate 10 trajectories using the dynamics equation 1 and compute the average of the three metrics.

Results. Since we need to solve estimator equation 2 many times (for different trajectories and steps T'), we consider relatively small-scale problems. In practice, the estimator equation 2 is only required for $T' = T$ and we only need to solve a single optimization problem. As a result, estimator equation 2 can be solved for large-scale real-world systems since it is convex and should be solved only once.

We first compare the performance of estimator equation 2 under different values of the attack probability p. We choose $T = 500$, $n = 3$ and $p \in \{0.7, 0.8, 0.85\}$. Additionally, we set the upper bound ρ to be 1, which guarantees the stability condition (Assumption 5). The results are plotted

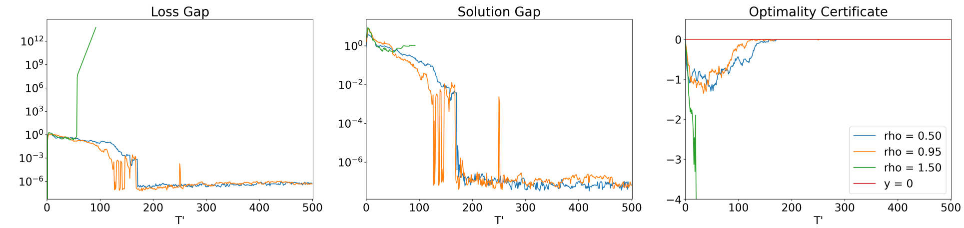

Figure 3: Loss gap, solution gap and optimality certificate of the Lipschitz basis function case with spectral norm $\rho = 0.5, 0.95$ and 1.5.

in Figure 1. It can be observed that both the loss gap and the solution gap converge to 0 when the number of samples T' is large, which implies that the estimator equation 2 exactly recovers the ground truth $\bar{A}$ when there exists a sufficient number of samples. Moreover, the optimality certificate converges to 0 at the same time as the solution gap, which verifies the validity of our necessary and sufficient condition in Sections 2 and 3. Furthermore, the required number of samples increases with probability p, which is consistent with the upper bound in Theorem 5.

Next, we show the performance of estimator equation 2 with different dimensions (n, m). We choose $T = 500$, $p = 0.75$, $\rho = 1$ and $n \in \{3, 5, 7\}$. The results are plotted in Figure 2. We can see that when the problem dimension (n, m) is larger, more samples are required to guarantee the exact recovery. This observation is also consistent with our bound in Theorem 3.

Finally, we illustrate the relation between the sample complexity and the spectral norm ρ. In this experiment, we choose $T = 100$, $p = 0.75$ and $n = 3$. To avoid the randomness in the spectral norm $\|\bar{A}\|_2$, we set singular values of $\bar{A}$ to be $\sigma_1 = \cdots = \sigma_n = \rho \in \{0.5, 0.95, 1.5\}$. For the case when $\rho = 1.5$, we terminate the simulation when $\|x_t\|_2 \geq 10^{14}$, which indicates that the trajectory diverges to infinity and this causes numerical issues for the CVX solver. The results are plotted in Figure 3. We can see that the required sample complexity slightly grows when ρ increases from 0.5 to 0.95, which is consistent with Theorem 5. In addition, the system is not asymptotically stable when $\rho = 1.5$ and Assumption 5 is violated. The explosion of the system (namely, $\|x_t\|_2 \to \infty$) leads to numerical instabilities in computing the estimator equation 2. With that said, it is possible that estimator equation 2 still achieves the exact recovery with large values of ρ, when a stable numerical method is applied to compute the estimator equation 2. This does not contradict with our theory since Theorem 5 only serves as a sufficient condition for the exact recovery.

7 Conclusion

This paper is concerned with the parameterized nonlinear system identification problem with adversarial attacks. The nonsmooth estimator equation 2 is utilized to achieve the exact recovery of the underlying parameter $\bar{A}$. We first provide necessary and sufficient conditions for the well-specifiedness of estimator equation 2 and the uniqueness of optimal solutions to the embedded optimization problem equation 3. Moreover, we provide sample complexity bounds for the exact recovery of $\bar{A}$ in the cases of bounded basis functions and Lipschitz basis functions using the proposed sufficient conditions. For bounded basis functions, the sample complexity scales with $m^3 n$ in terms of the dimension of the problem and with $p^{-1}(1-p)^{-2}$ in terms of the attack probability up to a logarithm factor. As for Lipschitz basis functions, the sample complexity scales with mn in terms of the dimension of the problem and with $\max\{(1-p)^{-2}, p^{-1}(1-p)^{-1}\}$ in terms of the attack probability up to a logarithm factor. Furthermore, if the sample complexity has a smaller order than $p^{-1}(1-p)^{-1}$, the high-probability exact recovery is not attainable. Hence, the term $p^{-1}(1-p)^{-1}$ in our bounds is inevitable. Lastly, numerical experiments are implemented to corroborate our theory.

References

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

Laurent Bako. On a Class of Optimization-Based Robust Estimators. IEEE Transactions on Automatic Control, 62(11):5990–5997, November 2017. ISSN 0018-9286, 1558-2523. doi: 10.1109/TAC.2017.2703308.

Laurent Bako and Henrik Ohlsson. Analysis of a nonsmooth optimization approach to robust estimation. Automatica, 66:132–145, 2016.

Dietmar Bauer, Manfred Deistler, and Wolfgang Scherrer. Consistency and asymptotic normality of some subspace algorithms for systems without observed inputs. Automatica, 35(7):1243–1254, 1999.

Dimitris Bertsimas and Martin S. Copenhaver. Characterization of the equivalence of robustification and regularization in linear and matrix regression. European Journal of Operational Research, 270 (3):931–942, November 2018. ISSN 0377-2217. doi: 10.1016/j.ejor.2017.03.051.

Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? Journal of the ACM (JACM), 58(3):1–37, 2011.

Han-Fu Chen and Lei Guo. Identification and stochastic adaptive control. Springer Science & Business Media, 2012.

Yuxin Chen, Jianqing Fan, Cong Ma, and Yuling Yan. Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data. Annals of statistics, 49(5):2948, 2021.

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

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. Foundations of Computational Mathematics, 20(4):633–679, 2020.

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

Han Feng and Javad Lavaei. Learning of dynamical systems under adversarial attacks. In 2021 60th IEEE Conference on Decision and Control (CDC), pp. 3010–3017, 2021. doi: 10.1109/CDC45484.2021.9683149.

Han Feng, Baturalp Yalcin, and Javad Lavaei. Learning of dynamical systems under adversarial attacks - null space property perspective. In 2023 American Control Conference (ACC), pp. 4179–4184, 2023. doi: 10.23919/ACC55779.2023.10156016.

Dylan Foster, Tuhin Sarkar, and Alexander Rakhlin. Learning nonlinear dynamical systems from a single trajectory. In Learning for Dynamics and Control, pp. 851–861. PMLR, 2020.

Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `https://cvxr.com/cvx`, March 2014.

Dong Huang, Ricardo Cabral, and Fernando De la Torre. Robust regression. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2):363–375, 2016. doi: 10.1109/TPAMI.2015.2448091.

A Kumar, A Levine, and S Feizi. Policy smoothing for provably robust reinforcement learning. In International Conference on Learning Representations (ICLR), 2022.

Vitaly Kuznetsov and Mehryar Mohri. Generalization bounds for non-stationary mixing processes. Machine Learning, 106(1):93–117, 2017.

L Ljung, R Pintelon, and J Schoukens. System identification: Wiley encyclopedia of electrical and electronics engineering. In System identification: a frequency domain approach, pp. 1–19. John Wiley & Sons Cham, 1999.

Lennart Ljung and Bo Wahlberg. Asymptotic properties of the least-squares method for estimating transfer functions and disturbance spectra. Advances in Applied Probability, 24(2):412–440, 1992.

Thomas M Moerland, Joost Broekens, Aske Plaat, Catholijn M Jonker, et al. Model-based reinforcement learning: A survey. Foundations and Trends® in Machine Learning, 16(1):1–118, 2023.

Jean-Philippe Noël and Gaëtan Kerschen. Nonlinear system identification in structural dynamics: 10 more years of progress. Mechanical Systems and Signal Processing, 83:2–35, 2017.

Robert D Nowak. Nonlinear system identification. Circuits, Systems and Signal Processing, 21: 109–122, 2002.

Afshin Rostamizadeh and Mehryar Mohri. Stability bounds for non-iid processes. In Neural Information Processing Systems (NIPS), 2007.

Yahya Sattar and Samet Oymak. Non-asymptotic and accurate learning of nonlinear dynamical systems. Journal of Machine Learning Research, 23(140):1–49, 2022.

George AF Seber and Alan J Lee. Linear regression analysis. John Wiley & Sons, 2012.

Max Simchowitz, Horia Mania, Stephen Tu, Michael I Jordan, and Benjamin Recht. Learning without mixing: Towards a sharp analysis of linear system identification. In Conference On Learning Theory, pp. 439–473. PMLR, 2018.

Anastasios Tsiamis and George J Pappas. Finite sample analysis of stochastic system identification. In 2019 IEEE 58th Conference on Decision and Control (CDC), pp. 3648–3654. IEEE, 2019.

Anastasios Tsiamis, Ingvar Ziemann, Nikolai Matni, and George J. Pappas. Statistical learning theory for control: A finite-sample perspective. IEEE Control Systems Magazine, 43(6):67–97, 2023. doi: 10.1109/MCS.2023.3310345.

Roman Vershynin. High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. doi: 10.1017/9781108231596.

Martin J. Wainwright. High-Dimensional Statistics: A Non-Asymptotic Viewpoint. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019. doi: 10.1017/9781108627771.

Fan Wu, Linyi Li, Yevgeniy Vorobeychik, Ding Zhao, and Bo Li. Crop: Certifying robust policies for reinforcement learning through functional smoothing. In International Conference on Learning Representations, 2022.

Huan Xu, Constantine Caramanis, and Shie Mannor. Robustness and Regularization of Support Vector Machines. Journal of machine learning research, 10(7), 2009.

Baturalp Yalcin, Javad Lavaei, and Murat Arcak. Exact recovery for system identification with more corrupt data than clean data. arXiv preprint arXiv:2305.10506, 2023.

Ingvar M Ziemann, Henrik Sandberg, and Nikolai Matni. Single trajectory nonparametric learning of nonlinear dynamics. In conference on Learning Theory, pp. 3333–3364. PMLR, 2022.