Exact Recovery Guarantees for Parameterized Non-linear System Identification Problem under Adversarial Attacks

Haixiang Zhang
Department of Math
UC Berkeley
Berkeley, CA 94720
haixiang_zhang@berkeley.edu

Baturalp Yalcin
Department of IEOR
UC Berkeley
Berkeley, CA 94720
baturalp_yalcin@berkeley.edu

Javad Lavaei
Department of IEOR
UC Berkeley
Berkeley, CA 94720
lavaei@berkeley.edu

Eduardo Sontag
Department of ECE and BioE
Northeastern University
Boston, MA 02115
sontag@gmail.com

Abstract

In this work, we study the system identification problem for parameterized non-linear systems using basis functions under adversarial attacks. Motivated by the LASSO-type estimators, we analyze the exact recovery property of a non-smooth 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 non-smooth estimator for the non-linear system identification problem.

1 Introduction

Dynamical systems are the foundation of 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 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. These adversarial disturbance or attacks often have a dependent temporal structure, which makes the majority of the existing literature on system identification inapplicable.

Preprint. Under review.

The system identification literature initially focused on the asymptotic properties of the least-squares estimator (LSE) [7, 18, 19, 4], 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 [27]. While early non-asymptotic analyses centered on linear-time invariant (LTI) systems with i.i.d. noise using mixing arguments [17, 22], recent research employs martingale and small-ball techniques to provide sample complexity guarantees for LTI systems [25, 11, 26]. For non-linear systems, recent studies investigated parameterized models [20, 21, 14, 23, 32], 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 non-smooth estimators for both linear and non-linear systems [12, 13, 31], particularly in handling dependent and adversarial noise vectors, are limited. Robust regression techniques utilizing regularizers have been developed [30, 5, 16], 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 A.

In this paper, we study the system identification problem for parameterized non-linear systems in the presence of adversarial attacks. We model the unknown non-linear 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:

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

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., $\{x_0, \ldots, x_T\}$. 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 $\{x_0, \ldots, x_T\}$. 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 [31] 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 [31].

Motivated by the exact recovery property of non-smooth 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 non-smooth objective) and, therefore, it can be solved efficiently by various existing optimization solvers. The estimator (2) is closely related to the LASSO estimator in the sense that the loss function in (2) can be viewed as a generalization of the ℓ_1-loss function. More specifically, in the case when $n = 1$, the estimator (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 (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 (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 (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 questions i) and ii). 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 iii). 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 non-linear 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 $\langle Z_1, Z_2\rangle = \mathrm{Tr}(Z_1^T Z_2)$ 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 (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

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

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

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

$$\sum_{t\in\mathcal{K}} f_t^T Z^T f(x_t) \le \sum_{t \in \mathcal{K}^c} \|Z^T f(x_t)\|_2, \quad \forall Z \in \mathbb{R}^{m\times n}, \tag{4}$$

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 (3). The condition (4) is established by applying the generalized Farkas' lemma, which avoids the inner approximation of the ℓ_2-ball by an ℓ_∞-ball in [31]. As a result, the sample complexity bounds to be obtained in this work are stronger than those in [31] 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 necessary conditions and sufficient conditions for the optimality of $\bar{A}$.

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

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

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

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

$$\left\|\sum_{t\in\mathcal{K}} f(x_t) f_t^T\right\|_F \le \sum_{t\in\mathcal{K}^c} \|f(x_t)\|_2. \tag{6}$$

In the case when $m = 1$, condition (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* (6) *reduces to*

$$\left|\sum_{t\in\mathcal{K}} f_t x_t\right| \le \sum_{t\in\mathcal{K}^c} |x_t|,$$

which is the same as Theorem 1 in [12].

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 [31]. By considering the attack period starting at the time step t_1, a sufficient condition to guarantee condition* (4) *is given by*

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

where we denote $f := f_{t_1}$ for simplicity. Let $F \in \mathbb{R}^{n\times(n-1)}$ be the matrix of orthonormal bases of the orthogonal complementary space of f, namely,

$$F^T f = 0, \quad F^T F = I_{n-1}, \quad FF^T = I_n - ff^T.$$

Then, we can calculate that

$$\left\|Z\bar{A}^t\bar{d}_{t_1}\right\|_2^2 \ge \left(Z\bar{A}^t\bar{d}_{t_1}\right)^T ff^T \left(Z\bar{A}^t\bar{d}_{t_1}\right),$$

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

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

Condition (8) *leads to a better sufficient condition than that in [31]. 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* (8) *becomes*

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

which is a stronger condition than that in [31]. When the attack period Δ is large, we approximately have $|\lambda| \le 2 - 2^{1-\Delta}$, which is a better condition than that in Figure 1 of [31].

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

$$\sum_{t\in\mathcal{K}} f_t^T Z^T f(x_t) = \sum_{t\in\mathcal{K}^c} \|Z^T f(x_t)\|_2 \implies \sum_{t\in\mathcal{K}} \left|f_t^T Z^T f(x_t)\right| < \sum_{t\in\mathcal{K}} \|Z^T f(x_t)\|_2, \tag{9}$$

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 (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* (4) *holds. If it holds that*

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

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

Proof. Under condition (10), the condition on the left hand-side of (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 [12].

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*

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

As a comparison, the sufficient condition in Theorem 1 in [12] is

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

Since $|f_t| = 1$ for all $t \in \mathcal{K}$, our results (11)*, as well as Theorem 2, are more general and stronger than that in [12].*

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 (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 non-smooth 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 [6, 8]. To state the stealthy condition, we define the filtration

$$\mathcal{F}_t := \sigma\left\{x_0, x_1, \dots, x_t\right\}.$$

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 $f_t = \bar{d}_t / \|\bar{d}_t\|_2$ 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*

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

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 (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 (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 (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 (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 (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 (3) *are not unique with probability at least* $\max\left\{1 - 2\exp\left(-m/3\right), \delta\right\}$.

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. 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 the 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 $\ell_t f_t$, where*

1. *$f_t \in \mathbb{R}^n$ and $\ell_t \in \mathbb{R}$ are independent conditional on $\mathcal{F}_t$ and $\bar{d}_t \neq 0_n$;*
2. *f_t is a zero-mean unit vector, namely, $\mathbb{E}(f_t \mid \mathcal{F}_t, \bar{d}_t \neq 0_n) = 0_n$ and $\|f_t\|_2 = 1$;*
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, f_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 f_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 $\kappa := \sigma L/\lambda \ge 1$. If the sample complexity T satisfies*

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

then $\bar{A}$ is the unique global solution to problem (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 $T \ge 1$, the ground truth $\bar{A}$ is not a global solution to problem* (3) *with probability at least $p[1 - (1-p)^{T-1}]$.*

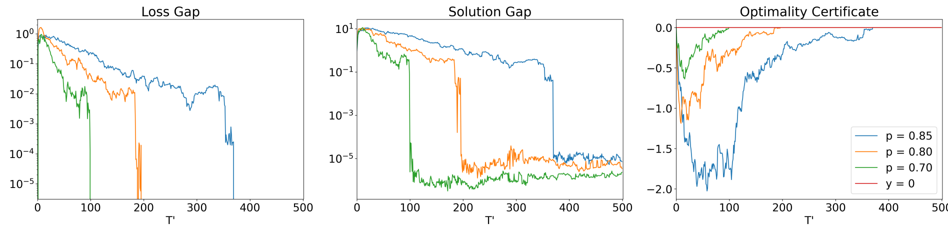

Figure 1: Loss gap, solution gap and optimality certificate of the Lipschitz basis function case with attack probability $p = 0.7, 0.8$ and 0.85.

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 C. More specifically, we illustrate the convergence of estimator (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.

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 [15]. 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

$$\min_{Z \in \mathbb{R}^{m \times n}} \sum_{t \in \mathcal{K}^c} \|Z^T f(x_t)\|_2 - \sum_{t \in \mathcal{K}} f_t^T Z^T f(x_t) \quad \text{s.t. } \|Z\|_F \leq 1,$$

  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 (2) and the proposed optimality conditions. For each choice of parameters, we independently generate 10 trajectories using the dynamics (1) and compute the average of the three metrics.

Results. Since we need to solve estimator (2) many times (for different trajectories and steps T'), we consider relatively small-scale problems. In practice, the estimator (2) is only required for $T' = T$ and we only need to solve a single optimization problem. As a result, estimator (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 (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 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 (2) exactly recovers the ground truth $\bar{A}$ when there exist is 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

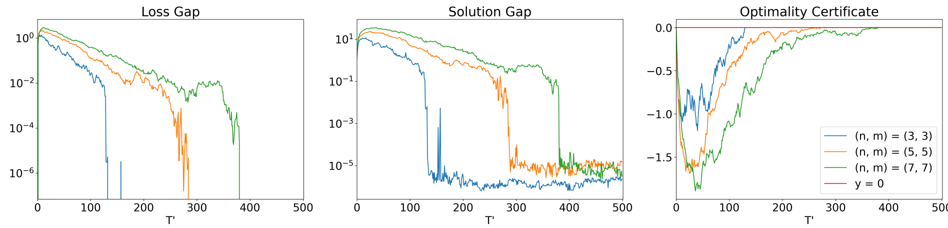

Figure 2: Loss gap, solution gap and optimality certificate of the Lipschitz basis function case with dimension $(n, m) = (3, 3), (5, 5)$ and $(7, 7)$.

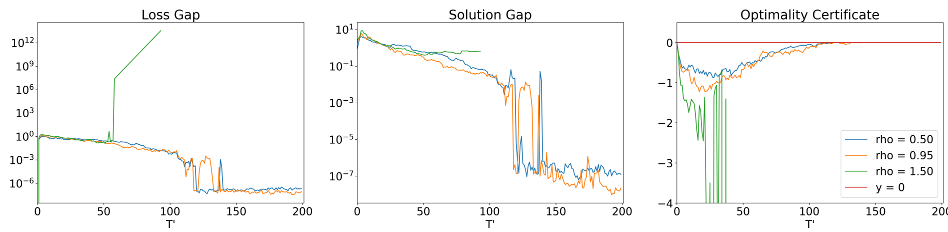

Figure 3: Loss gap, solution gap and optimality certificate of the Lipschitz basis function case with spectral norm $\rho = 0.5, 0.95$ and 1.5.

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 (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 $\|x_t\|_2 \geq 10^{15}$, 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 (2). With that said, it is possible that estimator (2) still achieves the exact recovery with large values of ρ, when a stable numerical method is applied to compute the estimator (2). This does not contradict with our theory since Theorem 5 only serves as a sufficient condition for the exact recovery.

7 Conclusion and Future Works

This paper is concerned with the parameterized non-linear system identification problem with adversarial attacks. The non-smooth estimator (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 (2) and the uniqueness of optimal solutions to the embedded optimization problem (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

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

[2] Laurent Bako. On a Class of Optimization-Based Robust Estimators. *IEEE Transactions on Automatic Control*, 62(11):5990–5997, November 2017.

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

[4] 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.

[5] 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.

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

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

[8] 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.

[9] 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.

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

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

[12] Han Feng and Javad Lavaei. Learning of dynamical systems under adversarial attacks. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 3010–3017, 2021.

[13] Han Feng, Baturalp Yalcin, and Javad Lavaei. Learning of dynamical systems under adversarial attacks - null space property perspective. In *2023 American Control Conference (ACC)*, pages 4179–4184, 2023.

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

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

[16] Dong Huang, Ricardo Cabral, and Fernando De la Torre. Robust regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):363–375, 2016.

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

[18] L Ljung, R Pintelon, and J Schoukens. System identification: Wiley encyclopedia of electrical and electronics engineering. In *System identification: a frequency domain approach*, number 2012, pages 1–19. John Wiley & Sons Cham, 1999.

[19] 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.

[20] 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.

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

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

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

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

[25] 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*, pages 439–473. PMLR, 2018.

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

[27] 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.

[28] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

[29] Martin J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.

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

[31] 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.

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

A Literature Overview

The literature on the system identification problem focused on the asymptotic properties of the least-squares estimator (LSE) until recently [7, 18, 19, 4]. With the growing popularity of statistical learning theory [28, 29], understanding the required number of samples for a certain error threshold for the system identification problem has gained significant importance. For an overview of results and proof techniques, the reader is referred to the survey paper [27]. The literature on the non-asymptotic analysis mainly focused on the linear-time invariant (LTI) system identification problem with i.i.d. noise. The earlier research used the mixing arguments that heavily rely on the stability of the system [17, 22]. The most recent studies used martingale and small-ball techniques to provide sample complexity guarantees for least-squares estimators applied to LTI systems [25, 11, 26]. These works showed that the LSE converges to the true system parameters with the rate $T^{-1/2}$, where T is the number of samples. This result was applied to the linear-quadratic regulator problem using adaptive control to obtain optimal regret bounds [9, 1, 10].

The non-linear system identification problem is vastly studied [20, 21]. Yet, the research on the non-asymptotic analysis of the non-linear system identification is in its infancy and is mostly focused on parameterized non-linear systems. Recursive and gradient algorithms designed for the least-squares loss function converge to the true system parameters with the rate $T^{-1/2}$ for non-linear systems with a known link function ϕ of the form $\phi(\bar{A}x_t)$ using martingale techniques [14] and mixing time arguments [23]. Most recently, [32] provided sample complexity guarantees for non-parametric learning of non-linear system dynamics, which scales with $T^{-1/(2+q)}$. Here, q scales with the size of the function class in which we search the true dynamics. Existing studies on both linear and non-linear system identification analyzed the problem under i.i.d. (sub)-Gaussian noise structures.

Despite the growing interest on non-asymptotic system identification, the literature on the system identification problem with non-smooth estimators that can handle dependent and adversarial noise vectors is limited to linear systems. The studies [12] and [13] considered a non-smooth convex estimator in the form of least absolute deviation estimator and analyzed the required conditions for the exact recovery of the system dynamics using the KKT conditions and Null Space Property from the LASSO literature. Later, [31] showed that the exact recovery of the system parameters is attainable with high probability even when more than half of the data are corrupted. This provides a further avenue of research for the adversarially robust system identification problem. [31] was the first paper that employed a non-smooth estimator for non-linear system identification.

On the other hand, robust regression techniques have been developed using regularizers in the objective function [30, 5, 16]. In addition, the robust estimation literature provided multiple non-smooth estimators, such as M-estimators, least absolute deviation, convex estimators, least median squares, and least trimmed squares [24]. The convex estimator (2) was proposed in [3, 2] in the context of robust regression. They showed that the estimator can achieve the exact recovery when we have infinitely many samples. However, the study lacks a non-asymptotic analysis on the sample complexity. Additionally, the analysis techniques cannot be applied to the analysis of dynamical systems due to the auto-correlation among the samples.

B Proofs

B.1 Proof of Theorem 1

Proof of Theorem 1. Since problem (3) is convex in A, the ground truth matrix $\bar{A}$ is a global optimum if and only if

$$0 \in \sum_{t\in\mathcal{K}^c} f(x_t) \otimes \partial\|\mathbf{0}_n\|_2 + \sum_{t\in\mathcal{K}} f(x_t) \otimes f_t. \tag{15}$$

Using the form of the subgradient of the ℓ_2-norm, condition (15) holds if and only if there exist vectors

$$g_t \in \mathbb{R}^n, \quad \forall t \in \mathcal{K}^c$$

such that

$$\sum_{t\in\mathcal{K}^c} f(x_t)g_t^T + \sum_{t\in\mathcal{K}} f(x_t)f_t^T = \mathbf{0}_{n\times n}, \quad \|g_t\|_2 \le 1, \quad \forall t \in \mathcal{K}^c. \tag{16}$$

Define the matrices

$$B := [f(x_t) \quad \forall t \in \mathcal{K}^c] \in \mathbb{R}^{m\times(T-|\mathcal{K}|)}, \quad V := [f(x_t) \quad \forall t \in \mathcal{K}] \in \mathbb{R}^{m\times|\mathcal{K}|},$$
$$G := [g_t \quad \forall t \in \mathcal{K}^c] \in \mathbb{R}^{n\times(T-|\mathcal{K}|)}, \quad F := [f_t \quad \forall t \in \mathcal{K}] \in \mathbb{R}^{n\times|\mathcal{K}|}.$$

Condition (16) can be written as a combination of second-order cone constraints and linear constraints:

$$\exists G \in \mathbb{R}^{n\times(T-|\mathcal{K}|)}, s, r \in \mathbb{R} \quad \text{s.t. } BG^T + VF^T = \mathbf{0}_{m\times n}, \quad \|G_{:,t}\|_2 \leq s, \; \forall t,$$
$$s + r = 1, \quad s, r \geq 0, \tag{17}$$

where $G_{:,t}$ is the t-th column of G for all $t \in \{1, \dots, T - |\mathcal{K}|\}$. We define the closed convex cone

$$\mathcal{S} := \left\{ z \in \mathbb{R}^{(T-|\mathcal{K}|)n+2} \middle| \sqrt{\sum_{i=1}^{n} z^2_{(T-|\mathcal{K}|)i+t}} \leq z_{(T-|\mathcal{K}|)n+1}, \; \forall t \in \{0, \dots, T - |\mathcal{K}| - 1\}, \right.$$
$$\left. z_{(T-|\mathcal{K}|)n+1}, z_{(T-|\mathcal{K}|)n+2} \geq 0 \right\},$$

and we define the matrix and vector

$$\mathcal{A} := \begin{bmatrix} I_n \otimes B & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \in \mathbb{R}^{(mn+1)\times[(T-|\mathcal{K}|)n+2]}, \quad b := \begin{bmatrix} -(VF^T)_{:,1} \\ -(VF^T)_{:,2} \\ \vdots \\ -(VF^T)_{:,n} \\ 1 \end{bmatrix} \in \mathbb{R}^{mn+1},$$

where $(VF^T)_{:,i}$ is the i-th column of VF^T. Then, condition (17) can be equivalently written as

$$\exists z \in \mathbb{R}^{(T-|\mathcal{K}|)n+2} \quad \text{s.t. } \mathcal{A}z = b, \quad z \in \mathcal{S}. \tag{18}$$

Since the cone $\mathcal{S}$ is closed and convex, we can apply the *generalized Farka's lemma* to conclude that condition (18) is equivalent to

$$\forall y \in \mathbb{R}^{mn+1}, \quad \left(\mathcal{A}^T y \in \mathcal{S}^* \implies b^T y \geq 0\right), \tag{19}$$

where $\mathcal{S}^*$ is the dual cone of $\mathcal{S}$. It can be verified that the dual cone is

$$\mathcal{S}^* = \left\{ z \in \mathbb{R}^{(T-|\mathcal{K}|)n+2} \middle| \sum_{t=0}^{T-|\mathcal{K}|-1} \sqrt{\sum_{i=1}^{n} z^2_{(T-|\mathcal{K}|)i+t}} \leq z_{(T-|\mathcal{K}|)n+1}, \right.$$
$$\left. z_{(T-|\mathcal{K}|)n+1}, z_{(T-|\mathcal{K}|)n+2} \geq 0 \right\}.$$

We can equivalently write condition (19) as

$$\forall Z \in \mathbb{R}^{m\times n}, \; p \in \mathbb{R}, \quad \left(\|Z^T B\|_{2,1} \leq p, \quad p \geq 0 \implies \langle VF^T, Z\rangle \leq p\right),$$

By eliminating variable p, we get

$$\langle VF^T, Z\rangle \leq \|Z^T B\|_{2,1}, \quad \forall Z \in \mathbb{R}^{m\times n},$$

where the $\ell_{2,1}$-norm is defined as

$$\|M\|_{2,1} := \sum_{j=1}^{n} \sqrt{\sum_{i=1}^{m} M^2_{ij}}, \quad \forall M \in \mathbb{R}^{m\times n}.$$

The above condition is equivalent to condition (4), and this completes the proof. □

B.2 Proof of Corollary 1

Proof of Corollary 1. The sufficient condition follows from the fact that $\|f_t\|_2 = 1$ and

$$f_t^T Z^T f(x_t) \leq \|Z^T f(x_t)\|_2, \quad \forall t \in \mathcal{K}.$$

This completes the proof. □

B.3 Proof of Corollary 2

Proof of Corollary 2. We choose

$$Z := \frac{\sum_{t\in\mathcal{K}} f(x_t) f_t^T}{\left\|\sum_{t\in\mathcal{K}} f(x_t) f_t^T\right\|_F}.$$

Then, condition (4) implies

$$\left\|\sum_{t\in\mathcal{K}} f(x_t) f_t^T\right\|_F = \sum_{t\in\mathcal{K}} f_t^T Z^T f(x_t) \leq \sum_{t\in\mathcal{K}^c} \|Z^T f(x_t)\|_2 \leq \sum_{t\in\mathcal{K}^c} \|f(x_t)\|_2,$$

where the last step is because $\|Z^T\|_2 \leq \|Z\|_F = 1$. Now, suppose that the basis dimension $m = 1$. In this case, we have

$$\sum_{t\in\mathcal{K}} f_t^T Z^T f(x_t) \leq \left(\sum_{t\in\mathcal{K}} f(x_t) f_t\right)^T Z \leq \left\|\sum_{t\in\mathcal{K}} f(x_t) f_t\right\|_F \|Z\|_2,$$

$$\sum_{t\in\mathcal{K}^c} \|Z^T f(x_t)\|_2 = \sum_{t\in\mathcal{K}^c} |f(x_t)|\|Z\|_2 = \sum_{t\in\mathcal{K}^c} \|f(x_t)\|_2 \|Z\|_2.$$

Combining the above two inequalities shows that condition (6) is also a sufficient condition. □

B.4 Proof of Theorem 2

We establish the sufficient and the necessary parts of Theorem 2 by the following two lemmas.

Lemma 1 (Sufficient condition for uniqueness)**.** *Suppose that condition* (4) *holds. If for every nonzero $Z \in \mathbb{R}^{m\times n}$ such that*

$$\sum_{t\in\mathcal{K}} f_t^T Z^T f(x_t) = \sum_{t\in\mathcal{K}^c} \|Z^T f(x_t)\|_2,$$

it holds that

$$\sum_{t\in\mathcal{K}} \left|f_t^T Z^T f(x_t)\right| < \sum_{t\in\mathcal{K}} \|Z^T f(x_t)\|_2.$$

Then, the ground truth matrix $\bar{A}$ is the unique global solution to problem (3).

Proof. The ground truth $\bar{A}$ is the unique solution if and only if for every matrix $A \in \mathbb{R}^{n\times m}$ such that $A \neq \bar{A}$, the loss function of A is larger than that of $\bar{A}$, namely,

$$\sum_{t\in\mathcal{K}} \|\bar{d}_t\|_2 < \sum_{t\in\mathcal{K}^c} \|(\bar{A} - A) f(x_t)\|_2 + \sum_{t\in\mathcal{K}} \|(\bar{A} - A) f(x_t) + \bar{d}_t\|_2. \tag{20}$$

Denote

$$Z := (A - \bar{A})^T \in \mathbb{R}^{m\times n}.$$

The inequality (20) becomes

$$\sum_{t\in\mathcal{K}^c} \| - Z^T f(x_t)\|_2 + \sum_{t\in\mathcal{K}} \left(\| - Z^T f(x_t) + \bar{d}_t\|_2 - \|\bar{d}_t\|_2\right) > 0. \tag{21}$$

Since problem (3) is convex in A, it is sufficient to guarantee that $\bar{A}$ is a strict local minimum. Therefore, the uniqueness of global solutions can be formulated as

$$\text{condition (21) holds}, \quad \forall Z \in \mathbb{R}^{m\times n} \quad \text{s.t. } 0 < \|Z\|_F \leq \epsilon, \tag{22}$$

where $\epsilon > 0$ is a sufficiently small constant. In the following, we fix the direction Z and discuss two different cases.

Case I. We first consider the case when condition (4) holds strictly, namely,

$$\sum_{t\in\mathcal{K}^c}\|Z^T f(x_t)\|_2-\sum_{t\in\mathcal{K}} f_t^T Z^T f(x_t)>0.$$

Since the ℓ_2-norm is a convex function, it holds that

$$\|-Z^T f(x_t)+\bar{d}_t\|_2-\|\bar{d}_t\|_2\geq\left\langle\partial\|\bar{d}_t\|_2,-Z^T f(x_t)\right\rangle=-f_t^T Z^T f(x_t).$$

Therefore, we get

$$\begin{aligned}&\sum_{t\in\mathcal{K}^c}\|-Z^T f(x_t)\|_2+\sum_{t\in\mathcal{K}}\left(\|-Z^T f(x_t)+\bar{d}_t\|_2-\|\bar{d}_t\|_2\right)\\ \geq&\sum_{t\in\mathcal{K}^c}\|-Z^T f(x_t)\|_2+\sum_{t\in\mathcal{K}}-f_t^T Z^T f(x_t)>0,\end{aligned}$$

which exactly leads to inequality (21).

Case II. Next, we consider the case when

$$\sum_{t\in\mathcal{K}} f_t^T Z^T f(x_t)=\sum_{t\in\mathcal{K}^c}\|Z^T f(x_t)\|_2,\quad \sum_{t\in\mathcal{K}}\left|f_t^T Z^T f(x_t)\right|<\sum_{t\in\mathcal{K}}\|Z^T f(x_t)\|_2. \tag{23}$$

Since ϵ is a sufficiently small constant, we know

$$\bar{d}_t^\alpha:=-\alpha Z^T f(x_t)+\bar{d}_t\neq 0,\quad \forall\alpha\in[0,1],$$

and the ℓ_2-norm is second-order continuously differentiable in an open set that contains the line. Therefore, the *mean value theorem* implies that there exists $\alpha\in[0,1]$ such that for each $t\in\mathcal{K}$, it holds

$$\begin{aligned}\|-Z^T f(x_t)+\bar{d}_t\|_2-\|\bar{d}_t\|_2=&\left\langle f_t,-Z^T f(x_t)\right\rangle \qquad (24)\\ &+\frac{1}{2}\left[-Z^T f(x_t)\right]^T\left(\frac{I}{\|\bar{d}_t^\alpha\|_2}-\frac{\bar{d}_t^\alpha\left(\bar{d}_t^\alpha\right)^T}{\|\bar{d}_t^\alpha\|_2^3}\right)\left[-Z^T f(x_t)\right].\end{aligned}$$

We can calculate that

$$\begin{aligned}&\left[-Z^T f(x_t)\right]^T\left(\frac{I}{\|\bar{d}_t^\alpha\|_2}-\frac{\bar{d}_t^\alpha\left(\bar{d}_t^\alpha\right)^T}{\|\bar{d}_t^\alpha\|_2^3}\right)\left[-Z^T f(x_t)\right] \qquad (25)\\ =&\frac{\left\|Z^T f(x_t)\right\|_2^2}{\|\bar{d}_t^\alpha\|_2}-\frac{\left\langle\bar{d}_t^\alpha,Z^T f(x_t)\right\rangle^2}{\|\bar{d}_t^\alpha\|_2^3}\geq 0,\end{aligned}$$

where the equality holds if and only if $Z^T f(x_t)$ is parallel with $\bar{d}_t^\alpha$. By the definition of $\bar{d}_t^\alpha$, the equality holds if and only if $Z^T f(x_t)$ is parallel with $\bar{d}_t$, which is further equivalent to

$$\left|\left\langle f_t,Z^T f(x_t)\right\rangle\right|=\left\|Z^T f(x_t)\right\|_2.$$

Substituting (24) and (25) into (21), we have

$$\begin{aligned}&\sum_{t\in\mathcal{K}^c}\|-Z^T f(x_t)\|_2+\sum_{t\in\mathcal{K}}\left(\|-Z^T f(x_t)+\bar{d}_t\|_2-\|\bar{d}_t\|_2\right)\\ \geq&\sum_{t\in\mathcal{K}^c}\|Z^T f(x_t)\|_2-\sum_{t\in\mathcal{K}}\left\langle f_t,Z^T f(x_t)\right\rangle=0,\end{aligned}$$

where the equality holds if and only if

$$\left|\left\langle f_t,Z^T f(x_t)\right\rangle\right|=\left\|Z^T f(x_t)\right\|_2,\quad\forall t\in\mathcal{K}.$$

Considering the second condition in (23), the above equality condition is violated by some $t\in\mathcal{K}$. Therefore, we have proven that condition (21) holds strictly.

Combining the two cases, we complete the proof. □

Next, we prove that the condition in Lemma 1 is also necessary for the uniqueness.

Lemma 2 (Necessary condition for uniqueness)**.** *Suppose that condition* (4) *holds. If the ground truth matrix* $\bar{A}$ *is the unique global solution to problem* (3)*, then for every nonzero* $Z \in \mathbb{R}^{m\times n}$*, we have*

$$\sum_{t\in\mathcal{K}} f_t^T Z^T f(x_t) < \sum_{t\in\mathcal{K}^c} \|Z^T f(x_t)\|_2 \quad \textit{or} \quad \sum_{t\in\mathcal{K}} \left|f_t^T Z^T f(x_t)\right| < \sum_{t\in\mathcal{K}} \|Z^T f(x_t)\|_2. \tag{26}$$

Proof. Assume conversely that there exists a nonzero $Z \in \mathbb{R}^{m\times n}$ such that

$$\sum_{t\in\mathcal{K}} f_t^T Z^T f(x_t) = \sum_{t\in\mathcal{K}^c} \|Z^T f(x_t)\|_2, \quad \sum_{t\in\mathcal{K}} \left|f_t^T Z^T f(x_t)\right| = \sum_{t\in\mathcal{K}} \|Z^T f(x_t)\|_2. \tag{27}$$

Without loss of generality, we assume that

$$0 < \|Z\|_2 \leq \epsilon$$

for a sufficiently small ϵ. In this case, the second condition in (27) implies that

$$\left|f_t^T Z^T f(x_t)\right| = \|Z^T f(x_t)\|_2, \text{ and } \quad Z^T f(x_t) \text{ is parallel with } \bar{d}_t, \quad \forall t \in \mathcal{K}.$$

Therefore, when ϵ is sufficiently small, equations (25) and (23) lead to

$$\| - Z^T f(x_t) + \bar{d}_t\|_2 - \|\bar{d}_t\|_2 = -\left\langle f_t, Z^T f(x_t)\right\rangle, \quad \forall t \in \mathcal{K}.$$

We now show that condition (21) fails:

$$\begin{aligned}&\sum_{t\in\mathcal{K}^c} \| - Z^T f(x_t)\|_2 + \sum_{t\in\mathcal{K}} \left(\| - Z^T f(x_t) + \bar{d}_t\|_2 - \|\bar{d}_t\|_2\right) \\ =&\sum_{t\in\mathcal{K}} \left\langle f_t, Z^T f(x_t)\right\rangle - \sum_{t\in\mathcal{K}} \left\langle f_t, Z^T f(x_t)\right\rangle = 0.\end{aligned}$$

This contradicts with the assumption that $\bar{A}$ is the unique solution to problem (3). □

Combining Lemmas 1 and 2, we have the following necessary and sufficient condition for the uniqueness of the ground truth solution $\bar{A}$.

B.5 Proof of Theorem 3

Proof of Theorem 3. Since both sides of inequality (10) are affine in Z, it suffices to prove that

$$\mathbb{P}\left[f_1(Z) - f_2(Z) < 0, \, \forall Z \in \mathbb{S}_F\right] \geq 1 - \delta, \tag{28}$$

where $\mathbb{S}_F$ is the Frobenius-norm unit sphere in $\mathbb{R}^{m\times n}$ and

$$f_1(Z) := \sum_{t\in\mathcal{K}} \langle Z, f(x_t) f_t^T\rangle, \quad f_2(Z) := \sum_{t\in\mathcal{K}^c} \|Z^T f(x_t)\|_2.$$

The proof is divided into two steps.

Step 1. First, we fix the vector $Z \in \mathbb{S}_F$ and prove that

$$\mathbb{P}\left[f_1(Z) - f_2(Z) < -\theta\right] \geq 1 - \delta,$$

holds for some constant $\theta > 0$. Using Markov's inequality, it is sufficient to prove that for some $\nu > 0$, it holds that

$$\mathbb{E}\left[\exp\left(\nu\left[f_1(Z) - f_2(Z)\right]\right)\right] \leq \exp(-\nu\theta)\delta. \tag{29}$$

We focus on the case when $\mathcal{K}$ is not empty, which happens with high probability. The proof of this step is also divided into two sub-steps.

Step 1-1. We first analyze the term $f_1(Z)$. Let T' be the last attack time instance, i.e.,

$$T' := \max\{t \mid t \in \mathcal{K}\}.$$

Then, we have

$$\mathbb{E}\left[\exp\left[\nu f_1(Z)\right]\right] = \mathbb{E}\left[\exp\left(\nu \sum_{t\in\mathcal{K}\setminus\{T'\}} \left\langle Z, f(x_t) f_t^T \right\rangle\right) \times \mathbb{E}\left[\exp\left[\nu \left\langle Z, f(x_{T'}) f_{T'}^T \right\rangle\right] \mid \mathcal{F}_{T'}\right]\right]. \tag{30}$$

According to Assumption 2, the direction $f_{T'}$ is a unit vector. Since

$$\begin{aligned}\left|\left[Z^T f(x_{T'})\right]^T f_{T'}\right| &\leq \|Z^T f(x_{T'})\|_2 \leq \|Z\|_2 \|f(x_{T'})\|_2 \\ &\leq \|Z\|_F \sqrt{m} \|f(x_{T'})\|_\infty \leq \sqrt{m} B,\end{aligned}$$

the random variable $\left[Z^T f(x_{T'})\right]^T f_{T'}$ is sub-Gaussian with parameter mB^2. Therefore, the property of sub-Gaussian random variables implies that

$$\mathbb{E}\left[\exp\left[\nu \left\langle Z, f(x_{T'}) f_{T'}^T \right\rangle\right] \mid \mathcal{F}_{T'}\right] \leq \exp\left(\frac{\nu^2 \cdot mB^2}{2}\right).$$

Substituting into (30), we get

$$\mathbb{E}\left[\exp\left[\nu f_1(Z)\right]\right] \leq \mathbb{E}\left[\exp\left(\nu \sum_{t\in\mathcal{K}\setminus\{T'\}} \left\langle Z, f(x_t) f_t^T \right\rangle\right)\right] \cdot \exp\left(\frac{\nu^2 \cdot mB^2}{2}\right).$$

Continuing this process for all $t \in \mathcal{K}$, it follows that

$$\mathbb{E}\left[\exp\left[\nu f_1(Z)\right]\right] \leq \exp\left(\frac{\nu^2 \cdot mB^2 |\mathcal{K}|}{2}\right). \tag{31}$$

Step 1-2. Now, we consider the second term in (29), namely, $-f_2(Z)$. Define

$$\mathcal{K}' := \{t \mid 1 \leq t \leq T,\ t \in \mathcal{K}^c,\ t-1 \in \mathcal{K}\}.$$

With probability at least $1 - \exp[-\Theta[p(1-p)T]]$, we have

$$|\mathcal{K}'| = \Theta[p(1-p)T].$$

Therefore, $\mathcal{K}'$ is non-empty with high-probability. Since $\|Z^T f(x_t)\|_2 \geq 0$ for all $t \in \mathcal{K}^c$, we have

$$\mathbb{E}\left[\exp\left[-\nu f_2(Z)\right]\right] \leq \mathbb{E}\left[\exp\left(-\nu \sum_{t\in\mathcal{K}'} \|Z^T f(x_t)\|_2\right)\right] \tag{32}$$

$$= \mathbb{E}\left[\exp\left(-\nu \sum_{t\in\mathcal{K}'\setminus\{T'\}} \|Z^T f(x_t)\|_2\right) \times \mathbb{E}\left[\exp\left(-\nu \|Z^T f(x_{T'})\|_2\right) \mid \mathcal{F}_{T'}\right]\right],$$

where T' is the last time instance in $\mathcal{K}'$, namely,

$$T' := \max\{t \mid t \in \mathcal{K}'\}.$$

By Bernstein's inequality [29], we can estimate that

$$\begin{aligned}&\mathbb{E}\left[\exp\left(-\nu \|Z^T f(x_{T'})\|_2\right) \mid \mathcal{F}_{T'}\right] \\ \leq &\exp\left[-\nu \mathbb{E}\left(\|Z^T f(x_{T'})\|_2 \mid \mathcal{F}_{T'}\right) + \frac{\nu^2}{2} \mathbb{E}\left(\|Z^T f(x_{T'})\|_2^2 \mid \mathcal{F}_{T'}\right)\right] \\ \leq &\exp\left[-\frac{\nu}{\sqrt{m}B} \mathbb{E}\left(\|Z^T f(x_{T'})\|_2^2 \mid \mathcal{F}_{T'}\right) + \frac{\nu^2}{2} \mathbb{E}\left(\|Z^T f(x_{T'})\|_2^2 \mid \mathcal{F}_{T'}\right)\right],\end{aligned}$$

where the last inequality is from

$$\|Z^T f(x_{T'})\|_2 \leq \sqrt{m} B.$$

Assumption 3 implies that

$$\mathbb{E}\left(\|Z^T f(x_{T'})\|_2^2 \mid \mathcal{F}_{T'}\right) = \left\langle ZZ^T, \mathbb{E}\left[f(x_{T'})f(x_{T'})^T \mid \mathcal{F}_{T'}\right]\right\rangle \geq \lambda^2\|Z\|_F^2 = \lambda^2.$$

If we choose ν such that

$$0 < \nu < \frac{2}{\sqrt{m}B}, \tag{33}$$

we have

$$\mathbb{E}\left[\exp\left(-\nu\|Z^T f(x_{T'})\|_2\right) \mid \mathcal{F}_{T'}\right] \leq \exp\left[\left(\frac{\nu^2}{2} - \frac{\nu}{\sqrt{m}B}\right)\lambda^2\right].$$

Substituting into inequality (32), it follows that

$$\begin{aligned}&\mathbb{E}\left[\exp\left[-\nu f_2(Z)\right]\right]\\ \leq&\mathbb{E}\left[\exp\left(-\nu\sum_{t\in\mathcal{K}'\backslash\{T'\}}\|Z^T f(x_t)\|_2\right)\times\exp\left[\left(\frac{\nu^2}{2} - \frac{\nu}{\sqrt{m}B}\right)\lambda^2\right]\right].\end{aligned}$$

Continuing this process for all $t \in \mathcal{K}'$, we have

$$\mathbb{E}\left[\exp\left[-\nu f_2(Z)\right]\right] \leq \exp\left[\left(\frac{\nu^2}{2} - \frac{\nu}{\sqrt{m}B}\right)\lambda^2|\mathcal{K}'|\right]. \tag{34}$$

Combining the inequalities (31) and (34), we have

$$\mathbb{E}\left[\exp\left(\nu\left[f_1(Z) - f_2(Z)\right]\right)\right] \leq \exp\left[\frac{m\nu^2B^2}{2}|\mathcal{K}| + \left(\frac{\nu^2}{2} - \frac{\nu}{\sqrt{m}B}\right)\lambda^2|\mathcal{K}'|\right].$$

We choose

$$\theta := \frac{\lambda^2 p(1-p)T}{4\sqrt{m}B}.$$

In order to satisfy condition (29), it is equivalent to have

$$\frac{m\nu^2B^2}{2}|\mathcal{K}| + \left(\frac{\nu^2}{2} - \frac{\nu}{\sqrt{m}B}\right)\lambda^2|\mathcal{K}'| + \frac{\lambda^2\nu p(1-p)T}{4\sqrt{m}B} \leq \log(\delta). \tag{35}$$

Now, we consider the fact that $\mathcal{K}$ is generated by the probabilistic attack model. Using the Bernoulli bound, it holds with probability at least $1 - \exp[-\Theta[p(1-p)T]]$ that

$$|\mathcal{K}| \leq 2pT, \quad |\mathcal{K}'| \geq \frac{p(1-p)T}{2}. \tag{36}$$

Thus, with the same probability, we have the estimation

$$\begin{aligned}&\frac{m\nu^2B^2}{2}|\mathcal{K}| + \left(\frac{\nu^2}{2} - \frac{\nu}{\sqrt{m}B}\right)\lambda^2|\mathcal{K}'| + \frac{\lambda^2\nu p(1-p)T}{4\sqrt{m}B}\\ \leq&\frac{m\nu^2B^2}{2}\cdot 2pT + \left(\frac{\nu^2}{2} - \frac{\nu}{2\sqrt{m}B}\right)\lambda^2\cdot\frac{p(1-p)T}{2}.\end{aligned}$$

Choosing

$$\nu := \frac{\lambda^2(1-p)}{2\sqrt{m}B[4mB^2 + \lambda^2(1-p)]},$$

we get

$$\frac{m\nu^2B^2}{2}|\mathcal{K}| + \left(\frac{\nu^2}{2} - \frac{\nu}{\sqrt{m}B}\right)\lambda^2|\mathcal{K}'| + \frac{\lambda^2\nu p(1-p)T}{4\sqrt{m}B} \leq -\frac{p(1-p)^2}{16m\kappa^2(4m\kappa^2 + 1 - p)}\cdot T,$$

where we define $\kappa := B/\lambda \geq 1$. Note that our choice of ν satisfies the condition (33). Therefore, in order for inequality (35) to hold, the sample complexity should satisfy

$$T \geq \frac{16m\kappa^2(4m\kappa^2 + 1 - p)}{p(1-p)^2}\log\left(\frac{1}{\delta}\right).$$

By considering the Bernoulli bound (36), the sample complexity bound becomes

$$\begin{aligned}T &\geq \Theta\left[\max\left\{\frac{m\kappa^2(m\kappa^2 + 1 - p)}{p(1-p)^2}, \frac{1}{p(1-p)}\right\}\log\left(\frac{1}{\delta}\right)\right] \\ &= \Theta\left[\frac{m^2\kappa^4}{p(1-p)^2}\log\left(\frac{1}{\delta}\right)\right].\end{aligned} \tag{37}$$

Step 2. Next, we establish the bound (28) by discretization techniques. More specifically, suppose that $\epsilon > 0$ is a constant and $\{Z^1, \dots, Z^N\} \subset \mathbb{S}_F$ is an ϵ-net of the sphere $\mathbb{S}_F$ under the Frobenius norm, where we can bound

$$\log(N) \le mn \cdot \log\left(1 + \frac{2}{\epsilon}\right).$$

Then, for every $Z \in \mathbb{S}_F$, we can find a point in the ϵ-net, denoted as Z', such that

$$\|Z - Z'\|_F \le \epsilon.$$

Now, we upper bound the difference $f(Z) - f(Z')$, where we define the function

$$f(Z) := f_1(Z) - f_2(Z), \quad \forall Z \in \mathbb{R}^{m\times n}.$$

We can calculate that

$$\begin{aligned}
f(Z) - f(Z') &= \sum_{t\in\mathcal{K}} f_t (Z - Z')^T f(x_t) - \sum_{t\in\mathcal{K}^c} \left(\|Z^T f(x_t)\|_2 - \|(Z')^T f(x_t)\|_2\right) \\
&\le \sum_{t\in\mathcal{K}} f_t (Z - Z')^T f(x_t) + \sum_{t\in\mathcal{K}^c} \|(Z - Z')^T f(x_t)\|_2 \\
&\le \sum_{t\in\mathcal{K}} \|Z - Z'\|_F \|f(x_t) f_t^T\|_F + \sum_{t\in\mathcal{K}^c} \|Z - Z'\|_2 \|f(x_t)\|_2 \\
&\le \sum_{t\in\mathcal{K}} \|Z - Z'\|_F \|f(x_t)\|_2 + \sum_{t\in\mathcal{K}^c} \|Z - Z'\|_F \|f(x_t)\|_2 \\
&\le T \cdot \epsilon\sqrt{m} B = \sqrt{m} TB \cdot \epsilon.
\end{aligned}$$

We choose

$$\epsilon := \frac{\theta}{\sqrt{m}TB} = \Theta\left[\frac{p(1-p)}{m\kappa^2}\right].$$

Therefore, under the event that

$$f(Z^i) < -\theta, \quad \forall i = 1, \dots, N, \tag{38}$$

we have

$$f(Z) < -\theta + \sqrt{m}TB \cdot \epsilon = 0, \quad \forall Z \in \mathbb{S}_F.$$

Hence, it suffices to estimate the probability that event (38) happens. To bound the failing probability, we replace δ with δ/N in (37) and it follows that

$$\mathbb{P}\left[f(Z^i) < -\theta\right] \ge 1 - \frac{\delta}{N}, \quad \forall i = 1, \dots, N.$$

Applying the union bound over all $i \in \{1, \dots, N\}$, the event (38) happens with probability at least $1 - \delta$, namely,

$$\mathbb{P}\left[f(Z^i) < -\theta,\ \forall i = 1, \dots, N\right] \ge 1 - \delta.$$

With this choice of δ, the sample complexity should be at least

$$\begin{aligned}
T &\ge \Theta\left[\frac{m^2\kappa^4}{p(1-p)^2} \log\left(\frac{N}{\delta}\right)\right] \\
&= \Theta\left[\frac{m^2\kappa^4}{p(1-p)^2}\left[mn\log\left(\frac{m\kappa}{p(1-p)}\right) + \log\left(\frac{1}{\delta}\right)\right]\right].
\end{aligned}$$

This completes the proof. □

B.6 Proof of Theorem 4

Proof of Theorem 4. We only need to show that condition (9) fails with probability at least $1 - \exp(-m/3)$. We choose the matrix

$$\bar{A} := \begin{bmatrix} 1 & 0_{1\times(m-1)} \\ 0_{n-1} & 0_{(n-1)\times(m-1)} \end{bmatrix} \in \mathbb{R}^{n\times m}.$$

As a result, the last $n-1$ elements of $\bar{A}f(x)$ are zero for every state $x \in \mathbb{R}^n$. Moreover, we will choose the basis function f such that its values will only depend on the first element of state $x \in \mathbb{R}^n$. With these definitions, the dynamics of x_t reduces to the dynamics of its first element $(x_t)_1$. Hence, we can assume without loss of generality that $n = 1$ in the remainder of the proof.

We define the basis function $f : \mathbb{R} \mapsto \mathbb{R}^m$ as

$$\tilde{f}(x) := \begin{bmatrix} \frac{x}{\max\{|x|,1\}} & \sin(x) & \sin(2x) & \cdots & \sin[(m-1)x] \end{bmatrix}, \quad \forall x \in \mathbb{R}.$$

Under the above definitions, it is straightforward to show that the following properties hold and we omit the proof:

$$f(0) = 0_m, \quad f\left[\bar{A}f(x)\right] = f(x), \quad \forall x \in \mathbb{R}. \tag{39}$$

Finally, the attack vector is defined as

$$\bar{d}_t|\mathcal{F}_t \sim \text{Uniform}\left\{[-(|x_t|+2\pi), -(|x_t|+\pi)] \cup [|x_t|+\pi, |x_t|+2\pi]\right\}, \quad \forall t \in \mathcal{K}.$$

The remainder of the proof is divided into three steps.

Step 1. In the first step, we prove that Assumptions 1-3 hold. By the definition of $f(x)$, we have

$$\|f(x)\|_\infty = \max\left\{\frac{|x|}{\max\{|x|,1\}}, |\sin(x)|, \ldots, |\sin[(m-1)x]|\right\} \leq 1, \quad \forall x \in \mathbb{R},$$

which implies that Assumption 1 holds with $B = 1$. Moreover, the stealthy condition (Assumption 2) is a result of the symmetric distribution of $\bar{d}_t|\mathcal{F}_t$.

Finally, we prove that Assumption 3 holds. For the notational simplicity, in this step, we omit the subscript t, the conditioning on the filtration $\mathcal{F}_t$ and the event $t \in \mathcal{K}$. The model of attack d implies that

$$|x+d| \geq |d| - |x| \geq \pi > 1.$$

Therefore, we have

$$f(x+d) = \begin{bmatrix} \frac{x+d}{|x+d|} & \sin[(x+d)] & \cdots & \sin[(m-1)(x+d)] \end{bmatrix}.$$

For any vector $\nu \in \mathbb{R}^m$, we want to estimate

$$\nu^T \mathbb{E}\left[f(x+d)f(x+d)^T\right]\nu = \mathbb{E}\left[\nu_1 \frac{x+d}{|x+d|} + \sum_{i=1}^{m-1} \nu_{i+1}\sin[i(x+d)]\right]^2.$$

First, we can calculate that

$$\mathbb{E}\left(\nu_1\frac{x+d}{|x+d|}\right)^2 = \nu_1^2, \ \mathbb{E}\left[\nu_{i+1}\sin[i(x+d)]\right]^2 = \nu_{i+1}^2 \cdot \frac{1}{2}, \quad \forall i \in \{1, \ldots, m-1\}. \tag{40}$$

Then, for every $i \in \{1, \ldots, m-1\}$, we have

$$\begin{aligned} &\mathbb{E}\left[\nu_1 \frac{x+d}{|x+d|} \cdot \nu_{i+1}\sin[i(x+d)]\right] \\ =&\nu_1\nu_{i+1}\left[\int_{-|x|-2\pi}^{-|x|-\pi} \frac{x+d}{|x+d|}\sin[i(x+d)]\,\mathrm{d}d + \int_{|x|+\pi}^{|x|+2\pi} \frac{x+d}{|x+d|}\sin[i(x+d)]\,\mathrm{d}d\right] \\ =&\nu_1\nu_{i+1}\left[\int_{-|x|-2\pi}^{-|x|-\pi} -\sin[i(x+d)]\,\mathrm{d}d + \int_{|x|+\pi}^{|x|+2\pi} \sin[i(x+d)]\,\mathrm{d}d\right] = 0. \end{aligned} \tag{41}$$

For every $i, j \in \{1, \ldots, m-1\}$ such that $i \neq j$, it holds that

$$\begin{aligned} &\mathbb{E}\left[\nu_{i+1}\sin[i(x+d)] \cdot \nu_{j+1}\sin[j(x+d)]\right] \\ =&\nu_{i+1}\nu_{j+1}\left[\int_{-|x|-2\pi}^{-|x|-\pi} \sin[i(x+d)]\sin[j(x+d)]\,\mathrm{d}d \right.\\ &\left. + \int_{|x|+\pi}^{|x|+2\pi} \sin[i(x+d)]\sin[j(x+d)]\,\mathrm{d}d\right] = 0. \end{aligned} \tag{42}$$

Combining equations (40)-(42), it follows that

$$\nu^T \mathbb{E}\left[f(x+d)f(x+d)^T\right]\nu = \nu_1^2 + \frac{1}{2}\sum_{i=1}^{m-1}\nu_{i+1}^2 \geq \frac{1}{2}\|\nu\|_2^2,$$

which implies that Assumption 3 holds with $\lambda^2 = 1/2$.

Step 2. In this step, we prove that the linear space spanned by the set of vectors

$$\mathcal{F}^c := \{f(x_t) \mid t \in \mathcal{K}^c\}$$

has dimension at most $m-1$ with probability at least $1-\delta$. By the second property in (39), the subspace spanned by $\mathcal{F}^c$ is equivalent to that spanned by

$$\mathcal{F}' := \{f(x_t) \mid t \in \mathcal{K}'\},$$

where we define

$$\mathcal{K}' := \{t \mid t-1 \in \mathcal{K},\ t \in \mathcal{K}^c\}.$$

Therefore, the dimension of the subspace is at most $|\mathcal{K}'|$.

To estimate the cardinality of $\mathcal{K}'$, we divide $\mathcal{K}'$ into the following two disjoint sets:

$$\mathcal{K}_1' := \{2t+1 \mid 2t \in \mathcal{K},\ 2t+1 \in \mathcal{K}^c\}, \quad \mathcal{K}_2' := \{2t \mid 2t-1 \in \mathcal{K},\ 2t \in \mathcal{K}^c\}.$$

The size of $\mathcal{K}_1'$ is the summation of $\lceil T/2 \rceil$ independent Bernoulli random variables with parameter $p(1-p)$. Therefore, the Chernoff bound implies

$$\mathbb{P}\left[|\mathcal{K}_1'| \leq 2p(1-p)\cdot\left\lceil\frac{T}{2}\right\rceil\right] \geq 1 - \exp\left[-\frac{p(1-p)}{3}\cdot\left\lceil\frac{T}{2}\right\rceil\right]. \tag{43}$$

Similarly, the size of $\mathcal{K}_2'$ is the summation of $\lfloor T/2 \rfloor$ independent Bernoulli random variables with parameter $p(1-p)$. Therefore, the Chernoff bound implies

$$\mathbb{P}\left[|\mathcal{K}_2'| \leq 2p(1-p)\cdot\left\lfloor\frac{T}{2}\right\rfloor\right] \geq 1 - \exp\left[-\frac{p(1-p)}{3}\cdot\left\lfloor\frac{T}{2}\right\rfloor\right]. \tag{44}$$

Combining the bounds (43) and (44) and applying the union bound, it holds that

$$\begin{aligned}\mathbb{P}\left[|\mathcal{K}'| \leq 2p(1-p)T\right] &\geq 1 - \exp\left[-\frac{p(1-p)}{3}\cdot\left\lceil\frac{T}{2}\right\rceil\right] - \exp\left[-\frac{p(1-p)}{3}\cdot\left\lfloor\frac{T}{2}\right\rfloor\right]\\ &\geq 1 - 2\exp\left[-\frac{p(1-p)T}{3}\right],\end{aligned}$$

where the last inequality is because $\lfloor T/2 \rfloor \leq \lceil T/2 \rceil \leq T$. Since

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

we know

$$\mathbb{P}\left[|\mathcal{K}'| < m\right] \geq 1 - 2\exp\left(-m/3\right). \tag{45}$$

In addition, when $\mathcal{K}$ is the empty set $\emptyset$ or the full set $\{0,\dots,T-1\}$, the set $\mathcal{K}'$ is an empty set, which implies that $|\mathcal{K}'|$ is smaller than m. This event happens with probability

$$p^T + (1-p)^T \geq 2[p(1-p)]^{T/2}.$$

Combining with inequality (45), we get

$$\mathbb{P}\left[|\mathcal{K}'| < m\right] \geq \max\left\{1 - 2\exp\left(-m/3\right), 2[p(1-p)]^{T/2}\right\}.$$

Step 3. Finally, we prove that if the dimension of the subspace spanned by $\mathcal{F}^c$ is smaller than m, the condition (9) cannot hold. Since the dimension of the subspace is at most $m-1$, there exists $Z \in \mathbb{R}^m$ such that

$$Z^T f(x_t) = 0, \quad \forall t \in \mathcal{K}^c.$$

With this choice of Z, the condition on the left hand-side of (9) holds while the strict inequality on the right hand-side fails. Therefore, we know that $\bar{A}$ is not the unique global solution to (3). $\square$

B.7 Proof of Theorem 5

Proof of Theorem 5. The proof is similar to that of Theorem 3. Since both sides of inequality (10) are affine in Z, it suffices to prove that

$$\mathbb{P}\left[f_1(Z) - f_2(Z) < 0,\ \forall Z \in \mathbb{S}_F\right] \geq 1 - \delta,$$

where $\mathbb{S}_F$ is the Frobenius-norm unit sphere in $\mathbb{R}^{m\times n}$ and

$$f_1(Z) := \sum_{t\in\mathcal{K}} \left\langle Z, f(x_t) f_t^T \right\rangle, \quad f_2(Z) := \sum_{t\in\mathcal{K}^c} \left\| Z^T f(x_t) \right\|_2.$$

The proof is divided into two steps.

Step 1. First, we fix the vector $Z \in \mathbb{S}_F$ and prove that

$$\mathbb{P}\left[f_1(Z) - f_2(Z) < -\theta\right] \geq 1 - \delta,$$

holds for some constant $\theta > 0$. The proof of this step is divided into two steps.

Step 1-1. We first analyze the term $f_1(Z)$. For each $k \in \mathcal{K}$, we define the following attack vectors:

$$\bar{d}_t^k := \begin{cases} \bar{d}_t & \text{if } t \leq k, \\ 0_n & \text{otherwise,} \end{cases} \quad \forall t \in \{0, \dots, T-1\}.$$

Then, we define the trajectory generated by the above attack vectors:

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

Let

$$\mathcal{K} = \{k_1, \dots, k_{|\mathcal{K}|}\},$$

where the elements are sorted as $k_1 < k_2 < \cdots < k_{|\mathcal{K}|}$. Under the above definition, we know $x_t^{k_{|\mathcal{K}|}} = x_t$ for all t. We define

$$g_t^{k_j} := \begin{cases} f(x_t^{k_j}) - f(x_t^{k_{j-1}}) & \text{if } j > 1, \\ f(x_t^{k_1}) & \text{if } j = 1, \end{cases} \quad \forall j \in \{1, \dots, |\mathcal{K}|\}.$$

We note that $g_t^{k_j}$ is measurable on $\mathcal{F}_{k_j}$. Using these introduced notations, we can write $f_1(Z)$ as

$$f_1(Z) = \sum_{j=1}^{|\mathcal{K}|} \left\langle Z, f(x_{k_j}) f_{k_j}^T \right\rangle = \sum_{j=1}^{|\mathcal{K}|} \left\langle Z, \sum_{\ell=1}^{j-1} g_{k_j}^{k_\ell} f_{k_j}^T \right\rangle = \sum_{\ell=1}^{|\mathcal{K}|} \sum_{j=\ell+1}^{|\mathcal{K}|} f_{k_j}^T Z^T g_{k_j}^{k_\ell}.$$

Then, Assumption 6 implies that $\bar{d}_t$ is sub-Gaussian with parameter σ conditional on $\mathcal{F}_t$. Now, we estimate the expectation

$$\mathbb{E}\left[\exp\left[\nu f_1(Z)\right]\right],$$

where $\nu \in \mathbb{R}$ is an arbitrary constant. First, for each $\ell \in \{1, \dots, |\mathcal{K}| - 1\}$, we estimate the following probability:

$$\mathbb{P}\left(\left| \sum_{j=\ell+1}^{|\mathcal{K}|} f_{k_j}^T Z^T g_{k_j}^{k_\ell} \right| \geq \epsilon \,\middle|\, \mathcal{F}_{k_\ell} \right).$$

Since f_{k_j} is a unit vector and $\|Z\|_F = 1$, we know

$$\left\| f_{k_j}^T Z^T \right\|_2 \leq \|f_{k_j}^T\|_2 \|Z^T\|_2 \leq \|f_{k_j}^T\|_2 \|Z^T\|_F = 1. \tag{46}$$

Moreover, we can estimate that

$$\begin{aligned} \left\| g_{k_j}^{k_\ell} \right\|_2 &= \left\| f(x_{k_j}^{k_\ell}) - f(x_{k_j}^{k_{\ell-1}}) \right\|_2 \leq L \left\| x_{k_j}^{k_\ell} - x_{k_j}^{k_{\ell-1}} \right\|_2 \\ &= L \left\| \bar{A}\left[f\left(x_{k_j-1}^{k_\ell}\right) - f\left(x_{k_j-1}^{k_{\ell-1}}\right) \right] \right\|_2 \leq \rho L \left\| f\left(x_{k_j-1}^{k_\ell}\right) - f\left(x_{k_j-1}^{k_{\ell-1}}\right) \right\|_2 \\ &\leq L(\rho L) \left\| x_{k_j-1}^{k_\ell} - x_{k_j-1}^{k_{\ell-1}} \right\|_2 \leq \cdots \leq L(\rho L)^{k_j - k_\ell - 1} \left\| x_{k_\ell+1}^{k_\ell} - x_{k_\ell+1}^{k_{\ell-1}} \right\|_2 \\ &= L(\rho L)^{k_j - k_\ell - 1} \|\bar{d}_{k_\ell}\|_2, \end{aligned} \tag{47}$$

where the first inequality holds because f has Lipschitz constant L, the second inequality is from $\|\bar{A}\|_2 \leq \rho$ and the last equality holds because

$$x_{k_\ell+1}^{k_\ell} = \bar{A} f\left(x_{k_\ell}^{k_\ell}\right) + \bar{d}_{k_\ell}, \quad x_{k_\ell+1}^{k_{\ell-1}} = \bar{A} f\left(x_{k_\ell}^{k_{\ell-1}}\right) = \bar{A} f\left(x_{k_\ell}^{k_\ell}\right).$$

By the sub-Gaussian assumption (Assumption 6), it holds that

$$\mathbb{P}\left(\|\bar{d}_{k_\ell}\|_2 \geq \eta \,\middle|\, \mathcal{F}_{k_\ell}\right) \leq 2\exp\left(-\frac{\eta^2}{2\sigma^2}\right), \quad \forall \eta \geq 0. \tag{48}$$

Combining inequalities (46)-(48), we get

$$\begin{aligned}
\mathbb{P}&\left(\left|\sum_{j=\ell+1}^{|\mathcal{K}|} f_{k_j}^T Z^T g_{k_j}^{k_\ell}\right| \geq \epsilon \,\middle|\, \mathcal{F}_{k_\ell}\right) \leq \mathbb{P}\left(\sum_{j=\ell+1}^{|\mathcal{K}|} \left\|g_{k_j}^{k_\ell}\right\|_2 \geq \epsilon \,\middle|\, \mathcal{F}_{k_\ell}\right) \\
&\leq \mathbb{P}\left(\sum_{j=\ell+1}^{|\mathcal{K}|} L(\rho L)^{k_j - k_\ell - 1} \|\bar{d}_{k_\ell}\|_2 \geq \epsilon \,\middle|\, \mathcal{F}_{k_\ell}\right) \\
&\leq \mathbb{P}\left(\frac{L(\rho L)^{\Delta_j}}{1-\rho L} \|\bar{d}_{k_\ell}\|_2 \geq \epsilon \,\middle|\, \mathcal{F}_{k_\ell}\right) \leq 2\exp\left[-\frac{(1-\rho L)^2 \epsilon^2}{2\sigma^2 L^2 (\rho L)^{2\Delta_j}}\right],
\end{aligned} \tag{49}$$

where $\Delta_j := k_j - k_{j-1} - 1$ and the second last inequality is from

$$\sum_{j=\ell+1}^{|\mathcal{K}|} L(\rho L)^{k_j - k_\ell - 1} < \sum_{i=\Delta_j}^{\infty} L(\rho L)^i = \frac{L(\rho L)^{\Delta_j}}{1-\rho L}.$$

Since

$$\mathbb{E}\left(\sum_{j=\ell+1}^{|\mathcal{K}|} f_{k_j}^T Z^T g_{k_j}^{k_\ell} \,\middle|\, \mathcal{F}_{k_\ell}\right) = 0,$$

inequality (49) implies that the random variable $\sum_{j=\ell+1}^{|\mathcal{K}|} f_{k_j}^T Z^T g_{k_j}^{k_\ell}$ is zero-mean and sub-Gaussian with parameter $\sigma L/(1-\rho L)$ conditional on $\mathcal{F}_{k_\ell}$. By the property of sub-Gaussian random variables, we have

$$\mathbb{E}\left[\exp\left(\nu \sum_{j=\ell+1}^{|\mathcal{K}|} f_{k_j}^T Z^T g_{k_j}^{k_\ell}\right) \,\middle|\, \mathcal{F}_{k_\ell}\right] \leq \exp\left[\frac{\nu^2 \sigma^2 L^2 (\rho L)^{2\Delta_j}}{2(1-\rho L)^2}\right], \quad \forall \nu \geq 0.$$

Finally, utilizing the tower property of conditional expectation, we have

$$\begin{aligned}
\mathbb{E}\left[\exp\left[\nu f_1(Z)\right]\right] &= \mathbb{E}\Bigg[\exp\left(\nu \sum_{\ell=1}^{|\mathcal{K}|-2} \sum_{j=\ell+1}^{|\mathcal{K}|} f_{k_j}^T Z^T g_{k_j}^{k_\ell}\right) \\
&\qquad \times \mathbb{E}\left[\exp\left(\nu \sum_{j=|\mathcal{K}|}^{|\mathcal{K}|} f_{k_j}^T Z^T g_{k_j}^{k_\ell}\right) \,\middle|\, \mathcal{F}_{k_{|\mathcal{K}|-1}}\right]\Bigg] \\
&\leq \mathbb{E}\left[\exp\left(\nu \sum_{\ell=1}^{|\mathcal{K}|-2} \sum_{j=\ell+1}^{|\mathcal{K}|} f_{k_j}^T Z^T g_{k_j}^{k_\ell}\right) \times \exp\left[\frac{\nu^2 \sigma^2 L^2 (\rho L)^{2\Delta_j}}{2(1-\rho L)^2}\right]\right] \\
&\leq \cdots \leq \exp\left[\frac{\nu^2 \sigma^2 L^2}{2(1-\rho L)^2} \sum_{j \in \mathcal{K}} (\rho L)^{2\Delta_j}\right], \quad \forall \nu \geq 0.
\end{aligned} \tag{50}$$

Since the random variable $(\rho L)^{\Delta_j}$ is bounded in $[0,1]$ and thus, it is sub-Gaussian with parameter $1/2$. Therefore, with constant number of samples, the mean of $(\rho L)^{2\Delta_j}$ will concentrate around its expectation, which is approximately

$$\sum_{\Delta=0}^{\infty} p(1-p)^{2\Delta} (\rho L)^{2\Delta} = \frac{p}{1-(1-p)^2(\rho L)^2} \leq \frac{p}{1-\rho L}.$$

Then, the bound in (50) becomes

$$\mathbb{E}\left[\exp\left[\nu f_1(Z)\right]\right] \lesssim \exp\left[\frac{\nu^2\sigma^2 L^2 p|\mathcal{K}|}{2(1-\rho L)^3}\right], \quad \forall \nu \geq 0. \tag{51}$$

Applying Chernoff's bound to (51), we get

$$\mathbb{P}\left[f_1(Z) \leq \epsilon\right] \geq 1 - \exp\left[-\frac{(1-\rho L)^3}{2\sigma^2 L^2 p|\mathcal{K}|} \cdot \epsilon^2\right], \quad \forall \epsilon \geq 0. \tag{52}$$

Step 1-2. Next, we analyze the term $f_2(Z)$. Define the set

$$\mathcal{K}' := \{t \mid 1 \leq t \leq T,\ t \in \mathcal{K}^c,\ t-1 \in \mathcal{K}\}.$$

With probability at least $1 - \exp[-\Theta[p(1-p)T]]$, we have

$$|\mathcal{K}'| = \Theta[p(1-p)T].$$

Therefore, $\mathcal{K}'$ is non-empty with high-probability. Since $\|Z^T f(x_t)\|_2 \geq 0$ for all $t \in \mathcal{K}^c$, we know

$$f_2(Z) \geq \sum_{k \in \mathcal{K}'} \|Z^T f(x_t)\|_2.$$

To establish a high-probability lower bound of $\|Z^T f(x_t)\|_2$, we prove the following lemma.

Lemma 3. *For each $t \in \mathcal{K}'$, it holds that*

$$\mathbb{P}\left[\|Z^T f(x_t)\|_2 \geq \frac{\lambda}{2} \,\middle|\, \mathcal{F}_t\right] \geq \frac{c\lambda^4}{\sigma^4 L^4},$$

where $c := 1/1058$ is an absolute constant.

For each $t \in \mathcal{K}'$, let $\mathbf{1}_t$ be the indicator of the event that $\|Z^T f(x_t)\|_2$ is larger than the $\frac{c\lambda^4}{\sigma^4 L^4}$-quantile conditional on $\mathcal{F}_t$. Then, it holds that

$$\mathbb{P}(\mathbf{1}_t = 1 \mid \mathcal{F}_t) = 1 - \mathbb{P}(\mathbf{1}_t = 0 \mid \mathcal{F}_t) = \frac{c\lambda^4}{\sigma^4 L^4}.$$

Therefore, we know

$$\left\{\mathbf{1}_t - \frac{c\lambda^4}{\sigma^4 L^4},\ t \in \mathcal{K}'\right\}$$

is a martingale with respect to filtration set $\{\mathcal{F}_t,\ t \in \mathcal{K}'\}$. Applying Azuma's inequality, it holds with probability at least $1 - \exp[-\Theta(\frac{\lambda^4|\mathcal{K}'|}{\sigma^4 L^4})]$ that

$$\sum_{t \in \mathcal{K}'} \mathbf{1}_t \geq \frac{c\lambda^4|\mathcal{K}'|}{2\sigma^4 L^4},$$

which means that for at least $\frac{c\lambda^4|\mathcal{K}'|}{2\sigma^4 L^4}$ elements in $\mathcal{K}'$, the event that $\|Z^T f(x_t)\|_2$ is larger than the $\frac{c\lambda^4}{\sigma^4 L^4}$-quantile conditional on $\mathcal{F}_t$ happens. Using the lower bound on the quantile in Lemma 3, we know

$$\sum_{t \in \mathcal{K}'} \|Z^T f(x_t)\|_2 \geq \frac{c\lambda^4|\mathcal{K}'|}{2\sigma^4 L^4} \cdot \frac{\lambda}{2} + \left(|\mathcal{K}'| - \frac{c\lambda^4|\mathcal{K}'|}{2\sigma^4 L^4}\right) \cdot 0 = \frac{c\lambda^5|\mathcal{K}'|}{4\sigma^4 L^4} \tag{53}$$

holds with the same probability.

Combining inequalities (52) and (53), we get

$$\mathbb{P}\left[f(Z) \leq \epsilon - \frac{c\lambda^5|\mathcal{K}'|}{4\sigma^4 L^4}\right] \geq 1 - \exp\left[-\frac{(1-\rho L)^3}{2\sigma^2 L^2 p|\mathcal{K}|} \cdot \epsilon^2\right] - \exp\left[-\Theta\left(\frac{\lambda^4|\mathcal{K}'|}{\sigma^4 L^4}\right)\right],$$

where we define $f(Z) := f_1(Z) - f_2(Z)$. Choosing

$$\epsilon := \frac{c\lambda^5|\mathcal{K}'|}{8\sigma^4 L^4},$$

it follows that

$$\mathbb{P}\left[f(Z) \le -\frac{c\lambda^5|\mathcal{K}'|}{8\sigma^4 L^4}\right] \tag{54}$$
$$\ge 1 - \exp\left[-\Theta\left(\frac{(1-\rho L)^3\lambda^{10}|\mathcal{K}'|^2}{\sigma^{10}L^{10}p|\mathcal{K}|}\right)\right] - \exp\left[-\Theta\left(\frac{\lambda^4|\mathcal{K}'|}{\sigma^4 L^4}\right)\right].$$

By the definition of the probabilistic attack model, it holds with probability at least $1-\exp[-\Theta[p(1-p)T]]$ that

$$|\mathcal{K}| \le 2pT, \quad |\mathcal{K}'| \ge \frac{p(1-p)T}{2}. \tag{55}$$

Therefore, the probability bound in (54) becomes

$$\mathbb{P}\left[f(Z) \le -\frac{c\lambda^5 p(1-p)T}{16\sigma^4 L^4}\right] \ge 1 - \exp\left[-\Theta\left(\frac{(1-\rho L)^3\lambda^{10}(1-p)^2 T}{\sigma^{10}L^{10}}\right)\right]$$
$$- \exp\left[-\Theta\left(\frac{\lambda^4 p(1-p)T}{\sigma^4 L^4}\right)\right] - \exp[-\Theta[p(1-p)T]].$$

Now, if the sample complexity satisfies

$$T \ge \Theta\left[\max\left\{\frac{\kappa^{10}}{(1-\rho L)^3(1-p)^2}, \frac{\kappa^4}{p(1-p)}\right\}\log\left(\frac{1}{\delta}\right)\right], \tag{56}$$

we know

$$\mathbb{P}\left[f(Z) \le -\theta\right] \ge 1-\delta, \tag{57}$$

where we define

$$\kappa := \frac{\sigma L}{\lambda}, \quad \theta := \frac{c\lambda^5 p(1-p)T}{16\sigma^4 L^4}.$$

Step 2. In the second step, we apply discretization techniques to prove that condition (57) holds for all $Z \in \mathbb{S}_F$. For a sufficiently small constant $\epsilon > 0$, let

$$\{Z^1, \dots, Z^N\}$$

be an ϵ-cover of the unit ball $\mathbb{S}_F$. Namely, for all $Z \in \mathbb{S}_F$, we can find $r \in \{1, 2, \dots, N\}$ such that $\|Z - Z^r\|_F \le \epsilon$. It is proved in [29] that the number of points N can be bounded by

$$\log(N) \le mn\log\left(1 + \frac{2}{\epsilon}\right).$$

Now, we estimate the Lipschitz constant of $f(Z)$ and construct a high-probability upper bound for the Lipschitz constant. For all $Z, Z' \in \mathbb{R}^{m\times n}$, we can calculate that

$$f(Z) - f(Z') = \sum_{t\in\mathcal{K}}\left\langle Z - Z', f(x_t)f_t^T\right\rangle - \sum_{t\in\mathcal{K}^c}\left(\left\|Z^T f(x_t)\right\|_2 - \left\|(Z')^T f(x_t)\right\|_2\right)$$
$$\le \|Z - Z'\|_F \sum_{t\in\mathcal{K}}\left\|f(x_t)f_t^T\right\|_F + \|Z - Z'\|_2 \sum_{t\in\mathcal{K}^c}\|f(x_t)\|_2$$
$$\le \|Z - Z'\|_F \sum_{t=0}^{T-1}\|f(x_t)\|_2. \tag{58}$$

Using the decomposition in **Step 1-1**, we have

$$f(x_t) = \sum_{\ell=1}^{j} g_t^{k_\ell},$$

where k_j is the maximal element in $\mathcal{K}$ such that $k_j < t$. Therefore, we can calculate that

$$\sum_{t=0}^{T-1}\|f(x_t)\|_2 \le \sum_{j=1}^{|\mathcal{K}|}\sum_{t=k_j+1}^{T-1}\left\|g_t^{k_j}\right\|_2. \tag{59}$$

For each $j \in \{1, \dots, |\mathcal{K}|\}$, we can prove in the same way as (47) that

$$\left\| g_t^{k_j} \right\|_2 \leq L(\rho L)^{k_j - t - 1} \|\bar{d}_{k_j}\|_2, \quad \forall t > k_j.$$

Substituting into inequality (59), it follows that

$$\sum_{t=0}^{T-1} \|f(x_t)\|_2 \leq \sum_{j=1}^{|\mathcal{K}|} \sum_{t=k_j+1}^{T-1} L(\rho L)^{k_j - t - 1} \|\bar{d}_{k_j}\|_2 \leq \frac{L}{1 - \rho L} \sum_{j=1}^{|\mathcal{K}|} \|\bar{d}_{k_j}\|_2.$$

Using Assumption 6 and the same technique as in (50), we know

$$\mathbb{P}\left(\sum_{j=1}^{|\mathcal{K}|} \|\bar{d}_{k_j}\|_2 \leq \eta \right) \geq 1 - 2\exp\left(-\frac{\eta^2}{2\sigma^2 |\mathcal{K}|} \right) \geq 1 - 2\exp\left(-\frac{\eta^2}{4\sigma^2 pT} \right),$$

where the second inequality is from the high probability bound in (55). Hence, it holds that

$$\mathbb{P}\left(\sum_{t=0}^{T-1} \|f(x_t)\|_2 \leq \eta \right) \geq 1 - 2\exp\left(-\frac{\eta^2 (1 - \rho L)^2}{4\sigma^2 L^2 pT} \right), \tag{60}$$

Choosing

$$\eta := \frac{\theta}{2\epsilon},$$

the bound in (60) becomes

$$\begin{aligned} \mathbb{P}\left(\sum_{t=0}^{T-1} \|f(x_t)\|_2 \leq \frac{\theta}{2\epsilon} \right) &\geq 1 - 2\exp\left(-\frac{(1 - \rho L)^2}{4\sigma^2 L^2 pT\epsilon^2} \cdot \theta^2 \right) \\ &= 1 - 2\exp\left[-\Theta\left[\frac{(1 - \rho L)^2}{4\sigma^2 L^2 pT\epsilon^2} \cdot \left(\frac{\lambda^5 p(1-p)T}{\sigma^4 L^4} \right)^2 \right] \right] \\ &= 1 - 2\exp\left[-\Theta\left[\frac{(1 - \rho L)^2 \kappa^{10} p(1-p)^2 T}{\epsilon^2} \right] \right]. \end{aligned} \tag{61}$$

We set

$$\epsilon := \Theta\left[\sqrt{(1 - \rho L)^2 \kappa^{10} p(1-p)^2} \right].$$

Then, it follows that

$$\exp\left[-\Theta\left[\frac{(1 - \rho L)^2 \kappa^{10} p(1-p)^2 T}{\epsilon^2} \right] \right] = \exp\left[-\Theta(T) \right] \leq \frac{\delta}{4},$$

where the last inequality is from the choice of T in (56). Substituting back into (61), we get

$$\mathbb{P}\left(\sum_{t=0}^{T-1} \|f(x_t)\|_2 \leq \frac{\theta}{2\epsilon} \right) \geq 1 - \frac{\delta}{2}. \tag{62}$$

Under the event in (62), for all $Z \in \mathbb{S}_F$, there exists an element Z^r in the ϵ-net such that

$$f(Z) \leq f(Z^r) + \epsilon \cdot \sum_{t=0}^{T-1} \|f(x_t)\|_2 \leq f(Z^r) + \frac{\theta}{2}.$$

If we replace δ with $\delta/(2N)$ in (57) and choose $Z = Z^r$ for all $r \in \{1, \dots, N\}$, the union bound implies that

$$\mathbb{P}\left[f(Z^r) \leq -\theta,\ r = 1, \dots, N \right] \geq 1 - \frac{\delta}{2}. \tag{63}$$

Under the above condition, we have

$$f(Z) \leq f(Z^r) + \frac{\theta}{2} \leq -\frac{\theta}{2} < 0.$$

To satisfy condition (63), the sample complexity bound (56) becomes

$$\begin{aligned}
T &\geq \Theta\left[\max\left\{\frac{\kappa^{10}}{(1-\rho L)^3(1-p)^2}, \frac{\kappa^4}{p(1-p)}\right\}\log\left(\frac{2N}{\delta}\right)\right] \\
&= \Theta\Bigg[\max\left\{\frac{\kappa^{10}}{(1-\rho L)^3(1-p)^2}, \frac{\kappa^4}{p(1-p)}\right\} \\
&\qquad\qquad \times\left[mn\log\left(\frac{1}{(1-\rho L)\kappa p(1-p)}\right)+\log\left(\frac{1}{\delta}\right)\right]\Bigg],
\end{aligned}$$

which is the desired sample complexity bound in the theorem.

Lower bound of κ. Before we close the proof, we provide a lower bound of $\kappa = \sigma L/\lambda$. Equivalently, we provide an upper bound on λ^2, which is at most the minimal eigenvalue of

$$\mathbb{E}\left[f(x+\bar{d}_t)f(x+\bar{d}_t)^T \mid \mathcal{F}_t, \bar{d}_t \neq 0_n\right].$$

Let $\nu \in \mathbb{R}^m$ be a vector satisfying

$$\|\nu\|_2 = 1, \quad \nu^T f(x) = 0.$$

Then, we know

$$\begin{aligned}
\nu^T f(x+\bar{d}_t)f(x+\bar{d}_t)^T\nu &= \nu^T\left[f(x+\bar{d}_t)-f(x)\right]\left[f(x+\bar{d}_t)-f(x)\right]^T\nu \\
&= \left[\left[f(x+\bar{d}_t)-f(x)\right]^T\nu\right]^2 \leq \left\|f(x+\bar{d}_t)-f(x)\right\|_2^2 \\
&\leq L^2\|\bar{d}_t\|_2^2,
\end{aligned} \tag{64}$$

where the last inequality is from the Lipschitz continuity of f. Using the sub-Gaussian assumption, it follows that

$$\mathbb{E}\left[\|\bar{d}_t\|_2^2 \mid \mathcal{F}_t,\ \bar{d}_t \neq 0_n\right] \leq \sigma^2, \tag{65}$$

where we utilize the fact that the standard deviation of sub-Gaussian random variables with parameter σ is at most σ. Combining inequalities (64) and (65), it follows that

$$\nu^T\mathbb{E}\left[f(x+\bar{d}_t)f(x+\bar{d}_t)^T \mid \mathcal{F}_t, \bar{d}_t \neq 0_n\right]\nu \leq \sigma^2 L^2.$$

Therefore, it holds that

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

which further leads to

$$\kappa = \frac{\sigma L}{\lambda} \geq 1.$$

This completes the proof. □

B.8 Proof of Lemma 3

Proof of Lemma 3. Let

$$\delta := \frac{c\lambda^4}{\sigma^4 L^4}, \quad \theta_t := \left\|Z^T f\left[\bar{A}f(x_{t-1})\right]\right\|_2.$$

We finish the proof by discussing two cases.

Case 1. We first consider the case when

$$\theta_t \geq \frac{\lambda}{2} + \sqrt{2\sigma^2 L^2\log\left(\frac{2}{1-\delta}\right)}.$$

Using the Lipschitz continuity of f, we have

$$\begin{aligned}\left\|Z^T f(x_t)\right\|_2 &= \left\|\left[Z^T f(x_t) - Z^T f\left[\bar{A} f(x_{t-1})\right]\right] + Z^T f\left[\bar{A} f(x_{t-1})\right]\right\|_2 \\ &\geq \left\|Z^T f\left[\bar{A} f(x_{t-1})\right]\right\|_2 - \left\|Z^T f(x_t) - Z^T f\left[\bar{A} f(x_{t-1})\right]\right\|_2 \\ &\geq \theta_t - \|Z\|_2 \left\|f(x_t) - f\left[\bar{A} f(x_{t-1})\right]\right\|_2 \\ &\geq \theta_t - \|Z\|_F \cdot L \left\|\bar{d}_t\right\|_2 \geq \theta_t - L\left\|\bar{d}_t\right\|_2.\end{aligned} \tag{66}$$

By Assumption 6, we know $\left\|\bar{d}_t\right\|_2 = |\ell_t|$ and it follows that

$$\mathbb{P}\left(\left\|\bar{d}_t\right\|_2 \geq \epsilon \mid \mathcal{F}_t\right) \leq 2\exp\left(-\frac{\epsilon^2}{2\sigma^2}\right), \quad \forall \epsilon \geq 0.$$

Therefore, we get the estimation

$$\begin{aligned}\mathbb{P}\left(\|Z^T f(x_t)\|_2 \leq \frac{\lambda}{2} \;\middle|\; \mathcal{F}_t\right) &\leq \mathbb{P}\left(\theta_t - L\left\|\bar{d}_t\right\|_2 \leq \frac{\lambda}{2} \;\middle|\; \mathcal{F}_t\right) \\ &= \mathbb{P}\left(\left\|\bar{d}_t\right\|_2 \geq \frac{\theta_t - \lambda/2}{L} \;\middle|\; \mathcal{F}_t\right) \\ &\leq \mathbb{P}\left(\left\|\bar{d}_t\right\|_2 \geq \sqrt{2\sigma^2 \log\left(\frac{2}{1-\delta}\right)} \;\middle|\; \mathcal{F}_t\right) \leq 1-\delta.\end{aligned}$$

Therefore, we have proved that

$$\mathbb{P}\left(\|Z^T f(x_t)\|_2 \geq \frac{\lambda}{2} \;\middle|\; \mathcal{F}_t\right) \geq \delta.$$

Case 2. Then, we focus on the case when

$$\theta_t \leq \frac{\lambda}{2} + \sqrt{2\sigma^2 L^2 \log\left(\frac{2}{1-\delta}\right)}. \tag{67}$$

Assume conversely that

$$\mathbb{P}\left(\|Z^T f(x_t)\|_2 \geq \frac{\lambda}{2} \;\middle|\; \mathcal{F}_t\right) < \delta. \tag{68}$$

Similar to inequality (66), the Lipschitz continuity of f implies

$$\left\|Z^T f(x_t)\right\|_2 \leq \theta_t + L\left\|\bar{d}_t\right\|_2.$$

Therefore, by applying Assumption 6, we get the tail bound

$$\begin{aligned}&\mathbb{P}\left(\|Z^T f(x_t)\|_2 \geq \theta \mid \mathcal{F}_t\right) \leq \mathbb{P}\left(\theta_t + L\left\|\bar{d}_t\right\|_2 \geq \theta \mid \mathcal{F}_t\right) \\ =&\mathbb{P}\left(\left\|\bar{d}_t\right\|_2 \geq \frac{\theta - \theta_t}{L} \;\middle|\; \mathcal{F}_t\right) \leq 2\exp\left[-\frac{(\theta-\theta_t)^2}{2\sigma^2 L^2}\right], \quad \forall \theta \geq \theta_t.\end{aligned}$$

Define $(x)_+ := \max\{x, 0\}$. The above bound leads to

$$\mathbb{P}\left(\|Z^T f(x_t)\|_2 \geq \theta \mid \mathcal{F}_t\right) \leq 2\exp\left[-\frac{(\theta-\theta_t)_+^2}{2\sigma^2 L^2}\right], \quad \forall \theta \in \mathbb{R}. \tag{69}$$

Using the definition of expectation, we can calculate that

$$\begin{aligned}\mathbb{E}\left[\|Z^T f(x_t)\|_2^2 \mid \mathcal{F}_t\right] &= \int_0^\infty 2\theta \cdot \mathbb{P}\left[\|Z^T f(x_t)\|_2 \geq \theta \mid \mathcal{F}_t\right] \, d\theta \\ &\leq \frac{\lambda^2}{4} + \int_{\lambda/2}^\infty 2\theta \cdot \mathbb{P}\left[\|Z^T f(x_t)\|_2 \geq \theta \mid \mathcal{F}_t\right] \, d\theta.\end{aligned}$$

By condition (68), we get

$$\mathbb{P}\left[\|Z^T f(x_t)\|_2 \geq \theta \mid \mathcal{F}_t\right] \leq \mathbb{P}\left[\|Z^T f(x_t)\|_2 \geq \frac{\lambda}{2} \;\middle|\; \mathcal{F}_t\right] \leq \delta, \quad \forall \theta \geq \frac{\lambda}{2}.$$

Combining with inequality (69), it follows that

$$\mathbb{E}\left[\|Z^T f(x_t)\|_2^2 \mid \mathcal{F}_t\right] \leq \frac{\lambda^2}{4} + \int_{\lambda/2}^{\infty} 2\theta \cdot \min\left\{\delta, 2\exp\left[-\frac{(\theta-\theta_t)_+^2}{2\sigma^2 L^2}\right]\right\} d\theta \tag{70}$$
$$= \frac{\lambda^2}{4} + \delta\left(\theta_1^2 - \frac{\lambda^2}{4}\right) + \int_{\theta_1}^{\infty} 4\theta \exp\left[-\frac{(\theta-\theta_t)^2}{2\sigma^2 L^2}\right] d\theta,$$

where we define

$$\theta_1 := \max\left\{\frac{\lambda}{2}, \theta_t + \sqrt{2\sigma^2 L^2 \log\left(\frac{2}{\delta}\right)}\right\} \geq \theta_t.$$

Using condition (67), we know

$$\theta_1^2 \leq \left(\frac{\lambda}{2} + \sqrt{2\sigma^2 L^2 \log\left(\frac{2}{1-\delta}\right)} + \sqrt{2\sigma^2 L^2 \log\left(\frac{2}{\delta}\right)}\right)^2 \tag{71}$$
$$\leq \left(\frac{\lambda}{2} + 2\sqrt{2\sigma^2 L^2 \log\left(\frac{2}{\delta}\right)}\right)^2 \leq \frac{\lambda^2}{2} + 16\sigma^2 L^2 \log\left(\frac{2}{\delta}\right),$$

where the last inequality is from Cauchy's inequality. Moreover, we can estimate that

$$\int_{\theta_1}^{\infty} 4\theta \exp\left[-\frac{(\theta-\theta_t)^2}{2\sigma^2 L^2}\right] d\theta \leq \int_{\theta_2}^{\infty} 4\theta \exp\left[-\frac{(\theta-\theta_t)^2}{2\sigma^2 L^2}\right] d\theta \tag{72}$$
$$= \int_{\theta_2}^{\infty} 4\theta_t \exp\left[-\frac{(\theta-\theta_t)^2}{2\sigma^2 L^2}\right] d\theta + \int_{\theta_2}^{\infty} 4(\theta-\theta_t) \exp\left[-\frac{(\theta-\theta_t)^2}{2\sigma^2 L^2}\right] d\theta$$
$$= \int_{\theta_2}^{\infty} 4\theta_t \exp\left[-\frac{(\theta-\theta_t)^2}{2\sigma^2 L^2}\right] d\theta + 2\delta\sigma^2 L^2,$$

where we denote $\theta_2 := \theta_t + \sqrt{2\sigma^2 L^2 \log\left(\frac{2}{\delta}\right)} \leq \theta_1$. Utilizing the following bound on the cumulative density function of the standard Gaussian distribution:

$$\int_{\eta}^{\infty} e^{-\frac{x^2}{2}}\, dx \leq \eta^{-1} e^{-\frac{\eta^2}{2}}, \quad \forall \eta > 0,$$

we have

$$\int_{\theta_2}^{\infty} 4\theta_t \exp\left[-\frac{(\theta-\theta_t)^2}{2\sigma^2 L^2}\right] d\theta \leq 4\theta_t \sigma L \cdot \frac{1}{\sqrt{2\log\left(\frac{2}{\delta}\right)}} \cdot \frac{\delta}{2} \leq \sqrt{2}\theta_t \cdot \delta\sigma L.$$

Combining with (72), it follows that

$$\int_{\theta_1}^{\infty} 4\theta \exp\left[-\frac{(\theta-\theta_t)^2}{2\sigma^2 L^2}\right] d\theta \leq \sqrt{2}\theta_t \cdot \delta\sigma L + 2\delta\sigma^2 L^2 \leq 4\delta\theta_t^2 + 4\delta\sigma^2 L^2, \tag{73}$$

where the last inequality is from Cauchy's inequality. Substituting inequalities (71) and (73) back into (70), we get

$$\mathbb{E}\left[\|Z^T f(x_t)\|_2^2 \mid \mathcal{F}_t\right] \leq \frac{\lambda^2}{4} + \delta\left[\frac{\lambda^2}{4} + 16\sigma^2 L^2 \log\left(\frac{2}{\delta}\right)\right] + 4\delta\theta_t^2 + 4\delta\sigma^2 L^2$$
$$\leq \frac{(1+\delta)\lambda^2}{4} + 16\sigma^2 L^2 \cdot \delta \log\left(\frac{2}{\delta}\right) + \delta\left[\frac{\lambda}{2} + \sqrt{2\sigma^2 L^2 \log\left(\frac{2}{1-\delta}\right)}\right]^2 + 4\delta\sigma^2 L^2$$
$$\leq \frac{(1+\delta)\lambda^2}{4} + 16\sigma^2 L^2 \cdot \delta \log\left(\frac{2}{\delta}\right) + \frac{\delta\lambda^2}{2} + 4\sigma^2 L^2 \cdot \delta \log\left(\frac{2}{\delta}\right) + 4\delta\sigma^2 L^2$$
$$\leq \frac{(1+3\delta)\lambda^2}{4} + 24\sigma^2 L^2 \cdot \delta \log\left(\frac{2}{\delta}\right).$$

where the second inequality is from (67) and the last inequality is from Cauchy's inequality and $\delta < 1/2$. On the other hand, Assumption 3 implies that

$$\mathbb{E}\left(\|Z^T f(x_t)\|_2^2 \mid \mathcal{F}_t\right) = \left\langle ZZ^T, \mathbb{E}\left[f(x_t)f(x_t)^T \mid \mathcal{F}_t\right]\right\rangle \geq \lambda^2 \|Z\|_F^2 = \lambda^2.$$

Combining the last two inequalities, we get

$$\lambda^2 \leq \frac{(1+3\delta)\lambda^2}{4} + 24\sigma^2 L^2 \cdot \delta \log\left(\frac{2}{\delta}\right),$$

which is equivalent to

$$\delta \log\left(\frac{2}{\delta}\right) \geq \frac{(3-3\delta)\lambda^2}{96\sigma^2 L^2} \geq \frac{\lambda^2}{23\sigma^2 L^2}.$$

For all $x \in (0,1)$, it holds that $x\log(2/x) < \sqrt{2x}$. Hence, we have

$$\sqrt{2\delta} > \frac{\lambda^2}{23\sigma^2 L^2},$$

which contradicts with our assumption (68).

□

B.9 Proof of Theorem 6

Proof of Theorem 6. In this proof, we focus on the case when $m = n$ and the counterexample can be easily extended into more general cases. We construct the following system dynamics:

$$\bar{A} := \rho I_n, \quad f(x) := x, \quad \forall x \in \mathbb{R}^n,$$

where $\rho \geq 2 + \sqrt{6}$ is a constant. One can verify Assumption 4 holds with Lipschitz constant $L = 1$. Therefore, the stability condition (Assumption 5) is violated since $\rho > 1/L$. The system dynamics can be written as

$$x_t = \sum_{k \in \mathcal{K}, k < t} \rho^{t-k-1} d_k, \quad \forall t \in \{0, \ldots, T\}. \tag{74}$$

Conditional on $\mathcal{F}_t$ and $t \in \mathcal{K}$, the attack vector is generated as

$$d_t \sim \mathrm{Uniform}(\mathbb{S}^{n-1}),$$

where $\mathbb{S}^{n-1}$ is the unit ball $\{d \in \mathbb{R}^n \mid \|d\|_2 = 1\}$. The attack model satisfies Assumption 3 with $\lambda = 1/\sqrt{n}$ and Assumption 6 with $\sigma = 1/\sqrt{n}$. Define the event

$$\mathcal{E} := \{T-1 \in \mathcal{K}, |\mathcal{K}| > 1\}.$$

By the definition of the probabilistic attack model, we can calculate that

$$\mathbb{P}(\mathcal{E}) = p\left[1 - (1-p)^{T-1}\right].$$

Our goal is to prove that

$$\mathbb{P}\left[f_1(Z) - f_2(Z) > 0 \mid \mathcal{E}\right] = 1,$$

where we define

$$f_1(Z) := \sum_{t \in \mathcal{K}} \left\langle Z, f(x_t) f_t^T \right\rangle, \quad f_2(Z) := \sum_{t \in \mathcal{K}^c} \left\|Z^T f(x_t)\right\|_2.$$

Then, by Theorem 1, we know that $\bar{A}$ is not a global solution to problem (3) with probability at least

$$p\left[1 - (1-p)^{T-1}\right].$$

Let t_1 be the smallest element in $\mathcal{K}$, namely, the first time instance when there is an attack. Under event $\mathcal{E}$, it holds that $t_1 < T-1$. We first prove that

$$x_t \neq 0_n, \quad \forall t \in \{t_1+1, \ldots, T-1\}.$$

By the system dynamics (74) and the triangle inequality, we have

$$\|x_t\|_2 \geq \rho^{t-t_1-1}\|d_{t_1}\|_2 - \sum_{k\in\mathcal{K},t_1<k<t} \rho^{t-k-1}\|d_k\|_2 = \rho^{t-t_1-1} - \sum_{k\in\mathcal{K},t_1<k<t} \rho^{t-k-1}$$

$$\geq \rho^{t-t_1-1} - \sum_{i=0}^{t-t_1-2} \rho^i = \frac{\rho^{t-t_1} - 2\rho^{t-t_1-1} + 1}{\rho - 1} > 0,$$

where the last inequality holds because $\rho \geq 2$. Then, we choose

$$Z := x_{T-1} f_{T-1}^T \neq 0.$$

It follows that

$$f_1(Z) = \sum_{t\in\mathcal{K}} \left\langle Z, f(x_t) f_t^T \right\rangle = \left\| x_{T-1} f_{T-1}^T \right\|_F^2 + \sum_{t\in\mathcal{K},t<T-1} \left\langle x_{T-1} f_{T-1}^T, f(x_t) f_t^T \right\rangle$$

$$\geq \|x_{T-1}\|_2^2 - \sum_{t\in\mathcal{K},t<T-1} \|x_{T-1}\|_2 \|x_t\|_2 ,$$

$$f_2(Z) = \sum_{t\in\mathcal{K}^c} \left\| Z^T f(x_t) \right\|_2 = \sum_{t\in\mathcal{K}^c} \left\| x_{T-1} f_{T-1}^T x_t \right\|_2 \leq \sum_{t\in\mathcal{K}^c} \|x_{T-1}\|_2 \|x_t\|_2 .$$

Combining the above two inequalities, we get

$$f_1(Z) - f_2(Z) \leq \|x_{T-1}\|_2 \left(\|x_{T-1}\|_2 - \sum_{t=0}^{T-2} \|x_t\|_2 \right) = \|x_{T-1}\|_2 \left(\|x_{T-1}\|_2 - \sum_{t=t_1+1}^{T-2} \|x_t\|_2 \right),$$

where the last equality holds because $x_t = 0_n$ for all $t \leq t_1$. Since $\|x_{T-1}\|_2 > 0$, it is sufficient to prove that

$$\|x_{T-1}\|_2 > \sum_{t=t_1+1}^{T-2} \|x_t\|_2 . \tag{75}$$

Considering the system dynamics (74) and the fact that $\|d_k\|_2 = 1$ for all $k \in \mathcal{K}$, we have the estimation

$$\rho^{t-t_1-1} - \sum_{k\in\mathcal{K},t_1<k<t} \rho^{t-k-1} \leq \|x_t\|_2 \leq \sum_{k\in\mathcal{K},k<t} \rho^{t-k-1}.$$

The desired inequality (75) holds if we can show

$$\rho^{T-1-t_1-1} - \sum_{k\in\mathcal{K},t_1<k<T-1} \rho^{T-1-k-1} > \sum_{t=t_1+1}^{T-2} \sum_{k\in\mathcal{K},k<t} \rho^{t-k-1},$$

which is further equivalent to

$$2\rho^{T-t_1-2} > \sum_{t=t_1+1}^{T-1} \sum_{k\in\mathcal{K},k<t} \rho^{t-k-1}$$

$$\Longleftarrow 2\rho^{T-t_1-2} > \sum_{t=t_1+1}^{T-1} \sum_{k=t_1}^{t-1} \rho^{t-k-1} = \sum_{t=t_1+1}^{T-1} \frac{\rho^{t-t_1} - 1}{\rho - 1} = \frac{\rho^{T-t_1} - \rho - (T - t_1 - 1)(\rho - 1)}{(\rho - 1)^2}$$

$$\Longleftarrow 2\rho^{T-t_1-2} \geq \frac{\rho^{T-t_1}}{(\rho - 1)^2} \iff \rho^2 - 4\rho - 2 \geq 0 \Longleftarrow \rho \geq 2 + \sqrt{6}.$$

By our choice of ρ, we know condition (75) holds and this completes our proof.

□

C Numerical Experiments for Bounded Basis Function

In this section, we provide the descriptions of basis functions and analyze the performance of estimator (2) in the case of bounded basis function. We show that the estimator (2) is able to exactly recover the ground truth $\bar{A}$ with different attack probability p and problem dimension (n, m). We utilize the same evaluation metrics as in Section 6 and define the system dynamics as follows.

Lipschitz basis function. Given the state space dimension n, we choose $m = n$ and define the basis function as

$$f(x) := \frac{1}{\sqrt{n}} \begin{bmatrix} \sqrt{\|x - x_1\|_2^2 + 1} - \sqrt{\|x_1\|_2^2 + 1} \\ \vdots \\ \sqrt{\|x - x_n\|_2^2 + 1} - \sqrt{\|x_n\|_2^2 + 1} \end{bmatrix}, \quad \forall x \in \mathbb{R}^n,$$

where $x_1, \dots, x_n \in \mathbb{R}^n$ are instances of i.i.d. standard Gaussian random vectors. We can verify that the basis function is Lipschitz continuous with Lipschitz constant $L = 1$ and thus, it satisfies Assumption 4. For each time instance $t \in \mathcal{K}$, the noise $\bar{d}_t$ is generated by

$$\bar{d}_t := \ell_t f_t, \quad \text{where } \ell_t \sim \mathcal{N}(0, \sigma_t^2),\ f_t \sim \text{uniform}(\mathbb{S}^{n-1}),\ \ell_t \text{ and } f_t \text{ are independent.}$$

Here, we define $\sigma_t^2 := \min\{\|x_t\|_2^2, 1/n\}$. We can verify that the random variable ℓ_t is zero-mean and sub-Gaussian with parameter $\sigma = 1$. In addition, the random vector f_t follows the uniform distribution and therefore, Assumption 6 is satisfied. Note that $\bar{d}_0, \dots, \bar{d}_{T-1}$ are correlated and they violate the i.i.d. assumption in the literature. Our attack model implies that the intensity of an attack (namely, ℓ_t) depends on the current state, which is a function of previous attacks. Since the points $x_1, \dots, x_n$ are randomly generated, the multiquadric radial basis functions are linearly independent[1] with probability 1 and therefore, the non-degenerate assumption (Assumption 3) is satisfied. Finally, the ground truth matrix $\bar{A}$ is constructed as $U\Sigma V^T$, where $U, V \in \mathbb{R}^{n\times n}$ are random orthogonal matrices and $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ is a diagonal matrix. The singular values are generated as follows:

$$\sigma_i \overset{\text{i.i.d.}}{\sim} \text{uniform}(0, \rho), \quad \forall i \in \{1, \dots, n\},$$

where $\rho > 0$ is the upper bound on the spectral norm of $\bar{A}$.

Bounded basis function. Given the state space dimension n, we choose $m = 5n$ and define the basis function as

$$f(x) := \begin{bmatrix} \tilde{f}(x_1) \\ \vdots \\ \tilde{f}(x_n) \end{bmatrix}, \quad \text{where } \tilde{f}(y) := \begin{bmatrix} \sin(y) \\ \vdots \\ \sin(5y) \end{bmatrix}, \quad \forall x \in \mathbb{R}^n,\ y \in \mathbb{R}.$$

The basis function satisfies Assumption 1 with $B = 1$. For each time instance $t \in \mathcal{K}$ and for each $i \in \{1, \dots, n\}$, the noise $\bar{d}_{t,i}$ is independently generated by

$$\bar{d}_{t,i} \sim \text{Uniform}\left(-c_{t,i}\pi, c_{t,i}\pi\right), \quad \text{where } c_{i,t} := \min\{\max\{|x_{t,i}|, 0.1\}, 0.5\}.$$

Note that $\bar{d}_{t,i}$ and $x_{t,i}$ is the i-th component of $\bar{d}_t$ and x_t, respectively. Since the attack is symmetric with respect to the origin, it satisfies Assumption 2. Since the sine functions $\sin(y), \dots, \sin(5y)$ are linearly independent, the non-degenerate assumption (Assumption 3) is satisfied. Finally, the ground truth matrix $\bar{A}$ is constructed such that

$$\bar{A}f(x) = \begin{bmatrix} \sum_{k=1}^5 \bar{a}_{1,k} \sin(kx_1) \\ \vdots \\ \sum_{k=1}^5 \bar{a}_{n,k} \sin(kx_n) \end{bmatrix},$$

where

$$\bar{a}_{i,k} \overset{\text{i.i.d.}}{\sim} \text{Uniform}(-100, 100), \quad \forall i \in \{1, \dots, n\},\ k \in \{1, \dots, 5\}.$$

We note that we choose the upper bound of coefficients $\bar{a}_{i,k}$ to be larger than 1 to show that the stability condition (Assumption 5) is not required in the bounded basis function case.

Results. We first compare the performance of estimator (2) under different attack probability p. We choose $T = 900$, $n = 1$ and $p \in \{0.7, 0.8, 0.85\}$. The results are plotted in Figure 4. We can observe similar behaviors to the Lipschitz basis function case. More specifically, the optimality certificate accurately measures the exact recovery of estimator (2), and the required sample complexity grows with the attack probability p.

Next, we show the performance of estimator (2) with different dimensions (n, m). We choose $T = 500$, $p = 0.7$ and $n \in \{1, 2, 4\}$. The results are plotted in Figure 5. We can see that the exact recovery happens with more samples when (n, m) is larger, which still verifies the results in Theorem 3.

[1] Functions $g_1(y), \dots, g_k(y)$ are said to be linearly independent if there do not exist constants $c_1, \dots, c_k$ such that $\sum_{i=1}^k c_i g_i(y) = 0$ for all y.

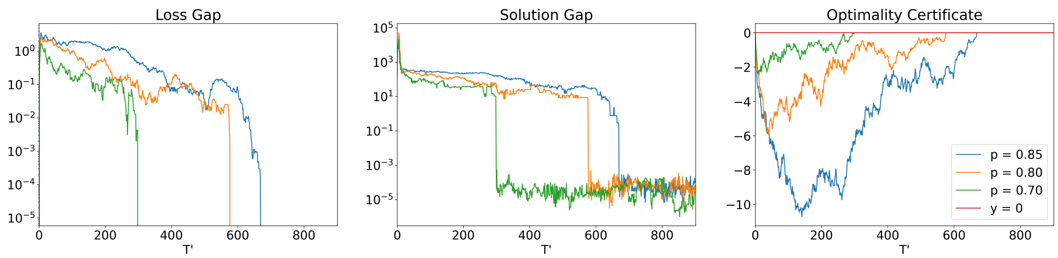

Figure 4: Loss gap, solution gap and optimality certificate of the bounded basis function case with attack probability $p = 0.7, 0.8$ and 0.85.

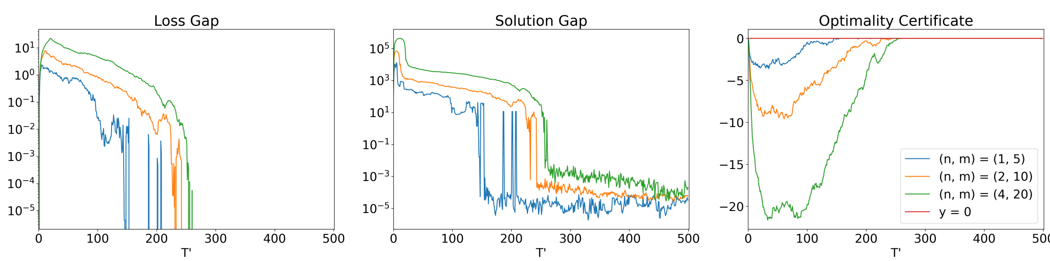

Figure 5: Loss gap, solution gap and optimality certificate of the bounded basis function case with dimension $(n, m) = (1, 5), (2, 10)$ and $(4, 20)$.