Abstract — This paper studies the system identification problem for linear discrete-time systems under adversaries and analyzes two lasso-type estimators. We study both asymptotic and non-asymptotic properties of these estimators in two separate scenarios, corresponding to deterministic and stochastic models for the attack times. Since the samples collected from the system are correlated, the existing results on lasso are not applicable. We show that when the system is stable and the attacks are injected periodically, the sample complexity for the exact recovery of the system dynamics is linear in terms of the dimension of the states. When the adversarial attacks occur at each time instance with probability p, the required sample complexity for the exact recovery scales logarithmically in the dimension of the states and polynomial in probability p. This result implies the almost sure convergence to the true system dynamics under the asymptotic regime. As a by-product, our estimators still learn the system correctly even when more than half of the data is compromised. We highlight that the attack vectors are allowed to be correlated to each other in this work whereas we make some assumptions about the times at which the attacks happen. This paper provides the first mathematical guarantee in the literature on learning from correlated data for dynamical systems in the case when there is less clean data than corrupt data.

Index Terms — System Identification, Robust Control, Statistical Learning, Linear Systems, Uncertain Systems

I. INTRODUCTION

Dynamic systems are the building blocks of reinforcement learning and control systems. The system dynamics may not be known exactly when the system is complex. Therefore, learning the underlying system dynamics, named the system identification problem, and using the data collected from the system is essential in robotics, control, time-series, and reinforcement learning applications. Despite being ubiquitously studied, most results in system identification were focused on the asymptotic properties of the proposed estimators until recently. Nonetheless, the non-asymptotic analysis of the system identification problem gained popularity in recent years [1]–[4]. Although the non-asymptotic analysis is harder because of the correlation between the sample points, it is crucial to understand the required sample complexity for online control problems.

The robust learning of dynamical systems is also crucial for safety-critical applications, such as autonomous driving, unmanned aerial vehicles, and biomedical applications. Although there are several recent papers on online non-asymptotic control of linear time-invariant (LTI) systems, their methods are often designed to only handle small noise in the measurements and do not consider large magnitudes of noise corresponding to adversarial attacks or other types of data corruption. [5]–[7]. Least-square estimators are the main tool in those works, which are susceptible to outliers and large noise in the system. Consequently, we propose two new estimators inspired by the lasso problem and robust regression literature [8]. We study the required sample complexity for the exact recovery of LTI systems using these estimators when there are sporadic large disturbance injections to the system.

The robust regression and learning problems under adversaries are ubiquitously studied in the literature [9]–[12]. However, existing methods for analyzing the estimators cannot be directly generalized to control problems due to the correlation between the samples. Therefore, different strategies were developed recently to tackle this challenge. Firstly, the system is initiated multiple times and the data point at the end of each run is used to obtain uncorrelated data points as in [13]. However, obtaining multiple trajectories is not viable and cost-efficient for most safety-critical applications. One of the methods with a single trajectory relies on the persistent excitation of the states so that the dynamics can be explored thoroughly. This is achieved by injecting Gaussian noise input into the system. The small ball techniques are used to analyze the properties of the estimator [5], [14], [15]. This technique uses normalized martingale bounds for the estimation error when the excitation is large enough [5].

Unlike the non-asymptotic analysis of correlated data, the least-squares estimator offers a closed-form solution when the system is subjected to small white noise [16]–[18]. When systems are not injected with large and sparse noise vectors, the least-squares estimator performs relatively well. However, it is not robust to adversarial attacks and the literature on robust learning of dynamical systems is limited. The work by [19] defines null space property (NSP) to analyze a lasso-type estimator for the system. It provides necessary and sufficient conditions for exact recovery when NSP is satisfied, which is NP-hard to check. To circumvent the computational complexity, we build upon [19] and study robust estimators from a non-asymptotic point of view under generic assumptions such as

This work was supported by grants from AFOSR, ARO, ONR, and NSF.
B. Yalcin and J. Lavaei are with the Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA, USA, 94720 (e-mail: baturalp_yalcin@berkeley.edu; lavaei@berkeley.edu).
M. Arcak is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA, 94720 (e-mail: arcak@berkeley.edu).
the system being stable and the attacks being sub-Gaussian.

Contributions: We study discrete-time linear time-invariant systems of the form \( x_{t+1} = Ax_t + Bu_t + d_t, i = 0, 1, \ldots, T - 1 \), where \( A \in \mathbb{R}^{n \times n} \) and \( B \in \mathbb{R}^{n \times m} \) are unknown matrices of the model. We aim to learn these matrices from the samples \( \{ x_i, u_i \}_{i=0}^{T-1} \) when the disturbance vectors \( d_i \) are adversarial. If \( d_i \) is zero, there is no attack at time \( i \). If \( d_i \) is nonzero, then an attack has occurred at time \( i \) and we have no information on the value of \( d_i \), which is designed by an attacker to adversely affect the states of the system.

i) We study two convex estimators based on the minimization of the \( \ell_2 \) and \( \ell_1 \) norm of the estimated disturbance vectors, \( \sum_{i=0}^{T-1} \| d_i \|_2 \) and \( \sum_{i=0}^{T-1} \| d_i \|_1 \), with the variables \( A \) and \( B \) subject to \( x_{t+1} = Ax_t + Bu_t + d_t \), given the samples \( \{ x_i, u_i \}_{i=0}^{T-1} \). The \( \ell_2 \) norm estimator is most effective when the set of attack times is sparse while the vector \( d_i \) at each attack time \( i \) could have several nonzero entries (meaning that most entries of the state \( x_{i+1} \) are affected by the attack). In contrast, the \( \ell_1 \) norm estimator is preferable when the vector \( d_i \) at each attack time is sparse.

ii) We first consider the case when the attacks happen periodically over time with the period \( \Delta \). We show that both of our estimators exactly recover the true system matrices \( A \) and \( B \) when the system is stable and the number of samples, \( n \), is larger than \( n + \Delta \).

iii) We then consider a probabilistic model for the attack occurrences in which there is an attack at each time \( t \) with probability \( p \), independently of previous time periods. Nevertheless, we allow attack vectors to be dependent. We show that our estimators find the true system matrices almost surely when the attack vectors are stealthy.

iv) We study the required sample complexity of our estimators for exact recovery. Let \( \lambda_i \) be the eigenvalues of \( A \). Suppose that the adversarial noise and the input sequence are sub-Gaussian random vectors and possibly dependent. Then, the estimators achieve exact recovery with probability at least \( 1 - \delta \) if

\[
T \geq \max \left\{ \frac{p}{(1-p)^2} \max_i \frac{1}{|\lambda_i|^2(1-|\lambda_i|)^2} \log(n^2/\delta), \frac{1}{(1-p)^2} \log(mn/\delta) \right\}
\]

This paper is organized as follows. In Sections 2 and 3, we introduce the notations used in the paper and formulate the problem, respectively. In Section 4, we study the convergence and sample complexity properties of our estimators in the case when the system is autonomous. In Section 5, we generalize the results to non-autonomous systems. In Section 6, we demonstrate the results on a biomedical system that models the blood sugar level with the injection of bolus insulin. This work provides the first bound in the literature on sample complexity for dynamical systems under adversaries and its techniques can be adopted to study other robust online learning problems.

II. Notation

For a matrix \( Z \), \( \| Z \|_F \) denotes the Frobenius norm of a matrix. For a vector \( z \), \( \| z \|_1 \) and \( \| z \|_2 \) denote its \( \ell_1 \) and \( \ell_2 \) norms, respectively. Moreover, for a given function \( f \), \( \partial f(z) \) shows the subdifferential for the function \( f \) at the point \( z \). Given two functions \( f \) and \( g \), the relations \( f(x) \lesssim g(x) \) and \( f(x) \gtrsim g(x) \) mean that there exist universal constants \( c_1 \) and \( c_2 \) such that \( f(x) \leq c_1 g(x) \) and \( f(x) \geq c_2 g(x) \), respectively. For two vectors \( v \) and \( w \), \( \langle v, w \rangle \) is the inner product between those vectors in their respective vector space. Furthermore, we use the notation \( v \otimes w = vw^T \) to denote the outer product. \( \mathbb{P}(\cdot) \) denote the probability of an event and the expectation of a random variable. A Gaussian random variable \( X \) with mean \( \mu \) and covariance matrix \( \Sigma \) is written as \( X \sim N(\mu, \Sigma) \). A random variable \( X \sim G(\sigma) \) is sub-Gaussian with parameter \( \sigma \) and \( X \sim sE(\nu, \alpha) \) is sub-exponential with parameters \( \nu \) and \( \alpha \). Given a time-dependent variable \( z(t) \), \( \dot{z}(t) \) represents its derivative with respect to time \( t \). \( |S| \) shows the cardinality of a given set \( S \).

III. Problem Formulation

We consider a linear time-invariant dynamical system over the time horizon \([0, T]\), \( x_{t+1} = Ax_t + Bu_t + d_t, i = 0, 1, \ldots, T - 1 \), where \( A \in \mathbb{R}^{n \times n} \) and \( B \in \mathbb{R}^{n \times m} \) are system matrices, and \( d_t \in \mathbb{R}^n \) are unknown system disturbances. Given the set of state measurements \( \{ x_i \}_{i=0}^T \) and the set of inputs \( \{ u_i \}_{i=0}^{T-1} \), the goal is to estimate the unknown system matrices \( A \) and \( B \).

In ordinary systems without attacks, the disturbance vectors \( d_i \) represent small disturbances on the input and small modeling errors. However, the disturbance vectors \( \{ d_i \}_{i=0}^{T-1} \) can be engineered to be large if there is an outside attack on the system from an agent or there is a sensor/actuation fault. Note that the purpose of the paper is to establish the exact recovery of the proposed estimators when there are large disturbances in the system and there is no small measurement noise. In that case, the least-squares method cannot achieve the exact recovery and this fact can be easily verified from its closed-form solution. Define the matrices \( X := [x_0, \ldots, x_{T-1}] \) and \( D := [d_0, \ldots, d_{T-1}] \). The solution for the least-squares problem is \( \hat{A} = (AX + DX)^T X(XX^T)^{-1} \) in the absence of the input sequence \( \{ u_i \}_{i=0}^{T-1} \). Thus, the estimation error is \( \| D^T X(XX^T)^{-1} \|_2 \), which is nonzero and arbitrarily large in the presence of adversarial disturbances. A similar calculation can be made in the presence of an input sequence.

Therefore, we observe a plateau in the estimation error of the least-squares estimator in the numerical experiment section. Define \( D := [d_0, \ldots, d_{T-1}] \), as well as \( \|D\|_{1, col} := \sum_i \|d_i\|_1 \), and \( \|D\|_{2, col} := \sum_i \|d_i\|_2 \). To exactly recover the system matrices \( A \) and \( B \), we analyze the following convex optimization problems:

\[
\min_{A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m}} \|D\|_{2, col}
\]

\[
\text{s.t. } x_{i+1} = Ax_i + Bu_i + d_i, \quad i = 0, \ldots, T - 1,
\]

and

\[
\min_{A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m}} \|D\|_{1, col}
\]

\[
\text{s.t. } x_{i+1} = Ax_i + Bu_i + d_i, \quad i = 0, \ldots, T - 1,
\]

where the states \( \{ x_i \}_{i=0}^T \) are generated according to \( x_{i+1} = Ax_i + Bu_i + d_i, \quad i = 0, \ldots, T - 1 \). The difference between problems
where the samples are assumed to be independently generated. Theorem 1: Consider the convex optimization problems \((\text{CO-L2})\) and \((\text{CO-L1})\) and let \((\text{CO-L2})\) and \((\text{CO-L1})\) be their objective functions. Note that the system is stable and it is initialized at the origin throughout this section.

Assumption 1: Given an autonomous system \(x_{i+1} = Ax_i + d_i\) for \(i = 0, \ldots, T - 1\) with dimension \(n\), assume that \(x_0 = 0\) and all eigenvalues of \(\bar{A}\) are inside the unit circle. The stability assumption is generic in system identification problems to avoid unbounded growth of the states. Without loss of generality, we initialize the trajectories at the origin, and initialization at other points affects the results only with a constant factor. We consider noiseless systems to obtain exact recovery results, meaning that if there is no attack at time period \(i\), then \(d_i = 0\). The noisy case when \(d_i\) is small can be addressed using our framework via perturbation analysis which allows us to bound how far the recovered solution is away from the true solution in terms of the values of small noise vectors. We denote the set of time instances at which the attack occurs with \(\mathcal{K}\). Mathematically, it is represented as \(\mathcal{K} = \{i | d_i \neq 0, i \in \{0, 1, \ldots, T - 1\}\}\). As a result, we are interested in recovering the system matrix \(\bar{A}\) using the following convex optimization problems for autonomous systems:

\[
\min_{\bar{A} \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m}} \sum_{i=0}^{T-1} \|x_{i+1} - \bar{A}x_i - Bu_i\|_o
\]

This is equivalent to an empirical risk minimization problem for which the loss function is the \(\ell_1\) and \(\ell_2\) norm depending on the choice of \(o\). However, we remark that the classical statistics theory on empirical risk minimization is not valid due to the correlated data at each time instance. If we represent the data points \(X_i\) as the tuple of \((x_{i+1}, x_i, \mu_i)\), it is impossible to claim that \(X_i\) and \(X_{i+1}\) are independent, which is the main assumption in the empirical risk minimization literature. Thus, KKT conditions will be used to analyze the properties of these estimators.

Since \((\text{CO-L2})\) and \((\text{CO-L1})\) are convex optimization problems with linear equalities, the Karush-Kuhn-Tucker (KKT) conditions are sufficient to guarantee optimality, as stated below.

**Theorem 1:** Consider the convex optimization problems \((\text{CO-L2})\) and \((\text{CO-L1})\) and let \(o \in \{1, 2\}\). Given a pair of matrices \((\bar{A}, \bar{B})\), if the following conditions hold simultaneously

\[
0 \in \sum_{i \in \mathcal{K}} x_i \otimes \partial \|(\bar{A} - \hat{A})x_i + (\bar{B} - \hat{B})u_i\|_o + \sum_{i \in \mathcal{K}} x_i \otimes \partial \|\bar{A} - \hat{A}\|_o x_i + (\bar{B} - \hat{B})u_i + d_i\|_o (1)
\]

\[
0 \in \sum_{i \notin \mathcal{K}} u_i \otimes \partial \|(\bar{A} - \hat{A})x_i + (\bar{B} - \hat{B})u_i\|_o + \sum_{i \notin \mathcal{K}} u_i \otimes \partial \|\bar{A} - \hat{A}\|_o x_i + (\bar{B} - \hat{B})u_i + d_i\|_o (2)
\]

then \((\bar{A}, \bar{B})\) is a solution to \((\text{CO-L1})\) when \(o = 1\) and a solution to \((\text{CO-L2})\) when \(o = 2\).

The proof for the KKT conditions when \(o = 2\) is given in [20] and the proof for the case \(o = 1\) can be done similarly. We will utilize the conditions above to show when the exact recovery is possible. As a simple corollary to Theorem 1, we can state that \((\bar{A}, \bar{B})\) is a solution to our estimator if the following condition holds:

\[
0 \in \sum_{i \notin \mathcal{K}} x_i \otimes \partial \|0\|_o + \sum_{i \notin \mathcal{K}} x_i \otimes \partial \|d_i\|_o
\]

**A. Single State Case**

We first assume that the problem is one-dimensional. When \(n = 1\), the problems \((\text{CO-L1-Aut})\) and \((\text{CO-L2-Aut})\) are equivalent and, therefore, we only focus on \((\text{CO-L2-Aut})\). After establishing the optimality conditions for these problems, we will examine two types of attack structures. The first one is a deterministic attack model for which attacks occur at every \(\Delta\) time period. Later, we investigate a probabilistic attack structure where each attack may occur with the probability \(p\) at each time instance. We first define the deterministic attack model, which is borrowed from [20].

**Definition 1:** Given a non-negative integer \(\Delta\), the disturbance sequence \(\{\hat{d}_i\}_{i=0}^{T-1}\) is said to be \(\Delta\)-spaced if for every \(i \in \{0, 1, \ldots, T - \Delta - 1\}\) such that \(\hat{d}_i \neq 0\), we have \(\hat{d}_j = 0\) for all \(j \in \{i + 1, \ldots, i + \Delta - 1\}\) and \(\hat{d}_{i+\Delta} \neq 0\).

We will show that the convex formulation \((\text{CO-L2-Aut})\) exactly recovers \(\hat{A}\) in the case of \(\Delta\)-spaced disturbance sequence with \(\Delta \geq 2\).

**Proposition 1:** Consider a one-dimensional autonomous system with \(\Delta\)-spaced disturbance sequence with \(\Delta \geq 2\).
Then, the convex formulation (CO-L2-Aut) (or equivalently (CO-L1-Aut)) has the unique solution $\hat{A}$ as long as the sample complexity $T \geq \Delta + 1$.

**Proof:** The proof of Proposition 1 is established based on Lemma 1 below.

**Lemma 1:** (Theorem 1 in [20]) Consider the convex optimization problem (CO-L2-Aut) and assume that $n = 1$. If $\sum_{i \in \mathcal{I}} |x_i| > \sum_{i \in \mathcal{I}} |x_i|$, then $\hat{A}$ is the unique solution to the problem.

The proof of the lemma is derived from the KKT conditions of the problem provided earlier. Let $i_1, i_2, \ldots$ be the set of attack times over time horizon $T$. Therefore, $\mathcal{X} = \{i_1, i_2, \ldots\}$. Due to $\Delta$-spaced attack model, the first attack time must be smaller than $\Delta$, i.e., $i_1 \leq \Delta$. Since $x_0 = 0$, we have $x_{t} = 0$ for $t = 0, 1, \ldots, i_1$. Define $N$ as the set of natural numbers. We can utilize Lemma 1 to show that $\hat{A}$ is the unique solution. Using these facts, we can decompose the sum of the magnitudes of the states at non-attack times as

$$\sum_{i \in \mathcal{I}} |x_i| = \sum_{i \in \mathcal{I}, t \geq i_1} |x_i| = \sum_{i \in \mathcal{I}} |x_i| + \sum_{i \in \mathcal{I}} |x_i|,$$

where $\mathcal{I} = N \setminus (\mathcal{I} \cup \{0, 1, \ldots, i_1 - 1\})$, $\mathcal{I}' = \mathcal{I} \setminus \mathcal{I}''$, and $\mathcal{I}'' = \{i_2 - i_3, i_3 - 1, \ldots\}$. The second term on the right-hand side is the sum of magnitudes at the time step just before the attack while the first term covers the rest of the magnitudes of the states. In addition, the magnitudes of the states at attack times can be written as

$$\sum_{i \in \mathcal{I}} |x_i| = \sum_{i \in \mathcal{I}} |x_i| = \sum_{i \in \mathcal{I}} |\bar{A}x_i| = \sum_{i \in \mathcal{I}} |\bar{A}| |x_i|.$$

The second equality follows from the fact that $x_{i_1} = \bar{A}x_{i_1 - 1}$ due to lack of attack. We compare the sum of the magnitudes of the states at attack times for the non-attack times to check if the condition in Lemma 1 holds:

$$\sum_{i \in \mathcal{I}} |x_i| - \sum_{i \in \mathcal{I}} |x_i| = \sum_{i \in \mathcal{I}} |x_i| + \sum_{i \in \mathcal{I}} |x_i| - \sum_{i \in \mathcal{I}} |\bar{A}| |x_i| = \sum_{i \in \mathcal{I}} |x_i| + (1 - |\bar{A}|) \sum_{i \in \mathcal{I}} |x_i| > 0. \tag{4}$$

Note that the term $\sum_{i \in \mathcal{I}} |x_i|$ becomes positive at time period $i_1 + 1$ while $\sum_{i \in \mathcal{I}} |x_i|$ is positive first time at time step $i_2$. Consequently, the strict inequality holds for $\bar{A}$ for every time step after $i_1$ because $(1 - |\bar{A}|) > 0$ by assumption. As a result, we have a unique and exact recovery for every time period $T \geq \Delta + 1$.

Note that Proposition 1 does not make any assumption on the vector set $\{\bar{d}_i : i \in \mathcal{I}\}$ and it could be arbitrarily large and correlated. As a result, regardless of the severity of the attack, an exact recovery is guaranteed for (CO-L1-Aut) and (CO-L2-Aut). One important implication of Proposition 1 is that whenever we have $\Delta$-spaced disturbance sequence with $\Delta = 2$, it implies that half of the observations are corrupted. In the robust regression estimation literature, the exact recovery is possible only if the number of attacked observations is less than half of the total observations. The main difference between robust regression and system identification problems is that the observations are correlated with each other in the latter. This enables the exact recovery for the convex formulation even if half of the data is wrong.

Next, it is natural to ask whether it is possible to learn the system when there is more corrupt data than clean data. We cannot use a $\Delta$-spaced disturbance sequence model because the minimum value of $\Delta$ is 2, which does not allow the size of corrupt data to exceed the size of clean data. To address this, we consider a probabilistic attack model for which there is a parameter $p$ specifying the probability of having an attack at each time instance. Specifically, given a time instance $i$, $\bar{d}_i$ is nonzero with probability $p$ and this is independent of all previous and future time instances. As a result, the event of having an attack at each time is identically and independently distributed with Bernoulli distribution with parameter $p$. Nevertheless, the attack vectors are allowed to be correlated with each other. Our goal is to discover the properties of (CO-L1-Aut) and (CO-L2-Aut) when $p > 0.5$.

The next theorem states that as long as the attacks have the same probability of being negative or positive, the estimators recover the true system matrix $\bar{A}$.

**Theorem 2:** Consider a one-dimensional autonomous system, and suppose that there is an attack at each time $i$, i.e., $\bar{d}_i$ in $\mathcal{I}$, with probability $p$ and this is independent of the other time periods. Assume that the attack vectors $\bar{d}_i$ are possibly dependent and come from an arbitrary probability distribution such that $\mathbb{P}(\bar{d}_i < 0) = \mathbb{P}(\bar{d}_i > 0)$ for all $i \in \mathcal{I}$. Then, the solution $\hat{A}$ is almost surely a solution to the convex optimization (CO-L2-Aut), or equivalently (CO-L1-Aut).

**Proof:** One can write the KKT condition for the solution $\hat{A}$ as follows:

$$0 \in \sum_{i \in \mathcal{I}} x_i \partial \|\hat{A} - \bar{A}\|_2 + \sum_{i \in \mathcal{I}} x_i \partial \|(\bar{A} - \bar{A})x_i + \bar{d}_i\|_2.$$ 

Based on the KKT condition and the convexity of the problem, $\bar{A}$ is the solution to the problem if and only if

$$0 \in \sum_{i \in \mathcal{I}} x_i \partial \|0\|_2 + \sum_{i \in \mathcal{I}} x_i \partial \|\bar{d}_i\|_2.$$ 

Our goal is to show that 0 is included in the expectation of the terms on the right-hand side. The randomness of the system stems from the set of attack vectors $\bar{d}_i, i \in \mathcal{I}$. Define $S := |\mathcal{I}|$ and $\mathcal{I} := \{i_1, i_2, \ldots, i_S\}$. Let the set of attack vectors be $\bar{d}_\mathcal{I} = \{\bar{d}_{i_1}, \bar{d}_{i_2}, \ldots, \bar{d}_{i_S}\}$. Taking expectations with respect to the random vectors $\bar{d}_i, i \in \mathcal{I}$, gives the following:

$$\mathbb{E}_{\bar{d}_\mathcal{I}} \left[ \sum_{i \in \mathcal{I}} x_i \partial \|0\|_2 + \sum_{i \in \mathcal{I}} x_i \partial \|\bar{d}_i\|_2 \right]$$

$$= \sum_{i \in \mathcal{I}} \mathbb{E}_{\bar{d}_i} \left[ x_i \partial \|0\|_2 \right] + \mathbb{E}_{\bar{d}_\mathcal{I}} \left[ \sum_{i \in \mathcal{I}} x_i \partial \|\bar{d}_i\|_2 \right]$$

$$= \sum_{i \in \mathcal{I}} \mathbb{E}_{\bar{d}_i} \left[ x_i \partial \|0\|_2 \right] + \mathbb{E}_{\bar{d}_\mathcal{I}} \left[ \sum_{i \in \mathcal{I}} x_i \partial \|\bar{d}_i\|_2 \right] \mathbb{P}(\bar{d}_\mathcal{I} = 0) x_i \partial \|0\|_2$$

The conditional expectation is taken in the first line. Given all the attack vectors until the last attack vector, every state
$x_i$ is deterministic until $x_{i_0}$. Since $\mathbb{P}(\bar{d}_i < 0) = \mathbb{P}(\bar{d}_i > 0)$, we have $\mathbb{E}[x_i \partial \|\bar{d}_i\|_2] = \mathbb{P}(\bar{d}_i = 0)x_i \partial \|0\|_2$. If the conditional expectation is taken iteratively, we obtain that

$$E_{\bar{d}_i} \sum_{i \in \mathcal{X}} |x_i| \partial \|0\|_2 + \sum_{i \in \mathcal{X}} x_i \partial \|\bar{d}_i\|_2 \] = \sum_{i \in \mathcal{X}} E_{\bar{d}_i} [x_i | \partial |0|_2 + \sum_{i \in \mathcal{X}} \mathbb{P}(\bar{d}_i = 0)x_i \partial |0|_2].$$

For the subderivative of $\|\cdot\|_2$, it is known that $0 \in \partial \|0\|_2$. Consequently, we have

$$0 \in \mathbb{E} \left[ \sum_{i \in \mathcal{X}} |x_i| \partial \|0\|_2 + \sum_{i \in \mathcal{X}} x_i \partial \|\bar{d}_i\|_2 \right].$$

Therefore, $\bar{A}$ is a solution almost surely due to the KKT condition and the strong law of large numbers.

Even in the case when the probability $p$ is close to 1, leading to much more corrupt data than clean data, Theorem 2 guarantees the asymptotic exact recovery when there are a sufficient number of samples. Note that the symmetricity assumption on the disturbance vectors is non-restrictive and corresponds to stealth attacks. If this does not hold, the attacks could be detected easily and their effects could be nullified. For an attack to be stealthy, its value should be zero on expectation and our assumption has a similar flavor.

Although almost sure convergence is a desirable property, the number of required samples for exact recovery may be large. Therefore, we want to quantify the sample complexity, the number of required samples for exact recovery. Note that the symmetricity assumption on the disturbance vectors is non-restrictive and corresponds to stealth attacks. If this does not hold, the attacks could be detected easily and their effects could be nullified. For an attack to be stealthy, its value should be zero on expectation and our assumption has a similar flavor.

$A^{(t)}$ is defined as

$$A^{(t)} := \begin{cases} 0, & \text{if } i \leq 0, \\ I, & \text{if } i = 0, \\ A^t, & \text{if } i > 0 \end{cases}$$

where $I$ is the identity matrix. Therefore, substituting this into the KKT condition yields

$$0 \in \sum_{i \in \mathcal{X}} \sum_{k \in \mathcal{K}} A^{(i-k)} \partial \|\bar{d}_k\|_2 + \sum_{i \in \mathcal{X}} \sum_{k \in \mathcal{K}} A^{(i-k)} \partial \|\bar{d}_k\|_2.$$
where \( \sigma^2(\bar{A}) \)

\[
\sqrt{\left( \sum_{k \in \mathcal{K}} \left( \sum_{i > k} |\bar{A}^{(i-1-k)}_i| \right)^2 + \left( \sum_{i > k} |\bar{A}^{(i-1-k)}_i| \right)^2 \right)} \sigma^2.
\]

Here, \( c_1 \) is a positive real constant. In addition, \( \mathbb{E}[^{|d_i|}] = c_2 \sigma \) with some positive constant \( c_2 \) since \( d_i \) is sub-Gaussian with parameter \( \sigma \). Thus, the expectations of \( L \) and \( U \) can be written as

\[
\mathbb{E}[U] = -\mathbb{E}[L] = c \sigma \left( \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{K}} |\bar{A}^{(i-1-k)}_i| \right).
\]

Since \( L \) and \( U \) are sub-Gaussian, we have the following concentration bounds:

\[
P(L > \mathbb{E}[L] + t) \leq e^{-\frac{t^2}{2\sigma^2(\bar{A})}},
\]

and

\[
P(U < \mathbb{E}[U] - t) \leq e^{-\frac{t^2}{2\sigma^2(\bar{A})}}.
\]

Therefore, \( 1 - \mathbb{P}(L < 0 < U) \) can be upper-bounded using the union bound by substituting \( t = \mathbb{E}[L] \) and \( t = \mathbb{E}[U] \):

\[
1 - \mathbb{P}(L < 0 < U) \leq e^{-\frac{\mathbb{E}[L]^2}{2\sigma^2(\bar{A})}} + e^{-\frac{\mathbb{E}[U]^2}{2\sigma^2(\bar{A})}} \leq 2e^{-\frac{\mathbb{E}[L]^2}{2\sigma^2(\bar{A})}}.
\]

where \( \sigma^2(\bar{A}) = \max\{\sigma^2(\bar{A}), \sigma^2(\bar{A})\} \). Moreover, we have

\[
\mathbb{E}[U] = c \sigma \left( \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{K}} |\bar{A}^{(i-1-k)}_i| \right)
\]

\[
\geq c \sigma (1 - p) T |\bar{A}|,
\]

and

\[
\sigma^2(\bar{A}) \leq \left( \sum_{k \in \mathcal{K}} \left( \sum_{i \in \mathcal{K}} |\bar{A}^{(i-1-k)}_i| \right)^2 \right)^{\frac{1}{2}} \leq \left( \sum_{i > k} |\bar{A}^{(i-1-k)}_i| \right)^{\frac{1}{2}} \leq \frac{p T \sigma^2}{(1 - |\bar{A}|)^2}.
\]

As a result,

\[
1 - \mathbb{P}(L < 0 < U) \leq 2 \exp \left\{ -\frac{(1 - p)^2}{p} (1 - |\bar{A}|)^2 T \right\} = \delta.
\]

If \( 1 - \mathbb{P}(L < 0 < U) \leq \delta \), solving for \( T \) implies that if \( T \geq \frac{\log(2/\delta)}{(1 - p)^2 (1 - |\bar{A}|)^2} \), \( \bar{A} \) is a solution to the problem with probability at least \( 1 - \delta \).

If we have sub-Gaussian attack values, the sample complexity can be obtained if \( |\bar{A}| < 1 \) meaning that the system is stable. Due to the logarthmic probability bound, Theorem [3] implies asymptotic convergence as well. The required amount of data increases with the value \( p/(1 - p)^2 \). Hence, as \( p \) increases, the number of samples for an exact recovery with high probability blows up.

### B. General Case with State Size \( n \)

We first generalize the result obtained in Proposition 1 to generic autonomous dynamical systems with an arbitrary \( n \) under a \( \Lambda \)-spaced disturbance sequence with \( \Lambda \geq n + 1 \).

**Proposition 2:** Consider an autonomous system with dimension \( n \) under a \( \Lambda \)-spaced disturbance sequence with \( \Lambda \geq n + 1 \). Suppose that \( \bar{A} \) has \( n \) eigenvalues, \( \lambda_j, j = 1, 2, \ldots, n \), with linearly independent eigenvectors such that

\[
\bar{A}d_i \neq \lambda_j d_i, \forall i \in \mathcal{K}, \quad j = 1, 2, \ldots, n.
\]

\( \bar{A} \) is a solution to the convex formulation (CO-L2-Aut) if \( T \geq n + \Delta \), provided that

\[
|\bar{A}^{(k_1, \ldots, k_n)}| \leq \sum_{i=0}^{\Lambda - n - 1} \sum_{k_1 + \cdots + k_n = \Delta} \bar{A}(k_1, \ldots, k_n),
\]

where the notation \( \bar{A}(k_1, \ldots, k_n) \) denotes \( \bar{A}_1^{k_1} \times \bar{A}_2^{k_2} \times \cdots \times \bar{A}_n^{k_n} \).

**Proof:** By using [3], the necessary and sufficient condition for this problem is

\[
0 \in \sum_{i \in \mathcal{K}} x_i \otimes \partial \| (\bar{A} - A) x_i \|_2 + \sum_{i \in \mathcal{K}} x_i \otimes \partial \| (\bar{A} - A) x_i + d_i \|_2.
\]

Then, \( \bar{A} \) is a solution to the problem if and only if

\[
0 \in \sum_{i \in \mathcal{E}_x} x_i \otimes \partial \| 0 \|_2 + \sum_{i \in \mathcal{K}} x_i \otimes \partial \| d_i \|_2.
\]

Let \( i_1 \) be the time stamp of the first attack time. Then, we have \( i_1 \in \{1, \ldots, \Delta\} \). The set of attack times is \( \mathcal{K} = \{i_1, i_1 + \Delta, i_1 + 2\Delta, i_1 + 3\Delta, \ldots\} \). Since \( x_0 = 0 \), we have \( x_i = 0 \) whenever \( t = 0, 1, \ldots, i_1 \) and \( x_{i_1 + 1} = d_i \). Let \( T = \Delta + i_1 \), i.e., the time step at which a cycle of disturbance is completed. In this case, the necessary and sufficient condition [3] can be written as

\[
0 \in \sum_{i=0}^{\Delta - 1} x_{i + \Delta} \otimes \partial \| 0 \|_2 + x_{i + \Delta} \otimes \partial \| d_{i + \Delta} \|_2,
\]

\[
= \sum_{i=0}^{\Delta - 2} \bar{A} \Delta \otimes \partial \| 0 \|_2 + \bar{A} \Delta \otimes \partial \| d_{i + \Delta} \|_2.
\]

The matrix \( 0 \) belongs to the right-hand side term for arbitrary \( d_{i + \Delta} \) if \( \text{span}\{d_{i_1}, \bar{A} \Delta d_{i_1}, \ldots, \bar{A}^{\Delta - 1} d_{i_1}\} \in \mathbb{R}^n \). Hence, we require those vectors to span \( \mathbb{R}^n \). Because \( \bar{A} \) has distinct eigenvalues, it has \( n \) distinct eigenvectors as well. As a result, as long as \( d_{i_1} \) is not a multiple of the eigenvectors of \( \bar{A} \) as stated in the theorem, we have

\[
\text{span}\{d_{i_1}, \bar{A} \Delta d_{i_1}, \ldots, \bar{A}^{\Delta - 1} d_{i_1}\} \in \mathbb{R}^n.
\]

However, this is not sufficient to ensure that KKT condition [7] holds. The reason is that \( \partial \| 0 \|_2 = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\} \). The vectors chosen for \( \partial \| 0 \|_2 \) have a bounded norm. Therefore, we need a condition that bounds the norm of the columns of \( \bar{A} \Delta \otimes \partial \| d_{i + \Delta} \|_2 \), so it can be expressed as a linear combination of the vectors \( \{d_{i_1}, \bar{A} \Delta d_{i_1}, \ldots, \bar{A}^{\Delta - 1} d_{i_1}\} \). Let \( (\lambda_j, v_j) \) be eigenvalue-eigenvector pairs for the matrix \( \bar{A}^\Delta \). Let \( e_1, \ldots, e_{\Delta - 1} \subset \partial \| 0 \|_2 \). Then, the KKT condition can be written as follows after dropping the subindex \( i_1 \):

\[
0 \in e_1 d^T + e_2 d^T \bar{A}^\Delta + \cdots + e_{\Delta - 1} d^T (\bar{A}^\Delta)^{\Delta - 2} + f d^T (\bar{A}^\Delta)^{\Delta - 1},
\]
where \( f = \frac{\hat{d}_i}{\|d_i\|_2} \) and \( \|f\|_2 = 1 \). If we multiply the equation above by the eigenvector \( v_j \) of \( A^T \), we obtain
\[
0 \in e_1 d^T v_j + \cdots + e_{\Delta - 1} d^T (A^T)^{\Delta - 2} v_j + f d^T (A^T)^{\Delta - 1} v_j.
\]
Because the disturbance vectors are not aligned with the eigenvectors of \( A^T \), by assumption, we have \( d^T v_j \neq 0, j = 1, \ldots, n \). Therefore, the KKT condition holds if
\[
0 \in e_1 + \lambda_j e_2 + \cdots + \lambda_j^{\Delta - 2} e_{\Delta - 1} + \lambda_j^{\Delta - 1} f, \quad j = 1, \ldots, n.
\]
There are \((\Delta - 1)n\) free variables and \(n^2\) equations. One can use the substitution to eliminate \( n \) variables, which leads to
\[
\sum_{k_1 + \cdots + k_n = \Delta - n} \lambda(k_1, \ldots, k_n) f = \sum_{0}^{\Delta - n} \sum_{k_1 + \cdots + k_n = t} \lambda(k_1, \ldots, k_n) e_t + 1.
\]
Taking the norm of both sides and using the triangle inequality yields that
\[
\sum_{k_1 + \cdots + k_n = \Delta - n} \lambda(k_1, \ldots, k_n) \|f\|_2 \leq \sum_{0}^{\Delta - n} \sum_{k_1 + \cdots + k_n = t} \lambda(k_1, \ldots, k_n) \|e_t + 1\|_2.
\]
Using the fact that \( \|e_j\|_2 = 1 \) for \( j = 1, \ldots, \Delta - n \) and \( \|f\|_2 = 1 \), we obtain
\[
\left| \sum_{k_1 + \cdots + k_n = \Delta - n} \lambda(k_1, \ldots, k_n) \right| \leq \sum_{0}^{\Delta - n} \sum_{k_1 + \cdots + k_n = t} \lambda(k_1, \ldots, k_n).
\]
This completes the proof for the proposition.

This result is the generalization of Proposition 1 in the case of \( n = 1 \). The equation (5) implies that the disturbance vectors are not aligned with the eigenvectors of the matrix \( A \). To gain insight into the condition (6) that involves the product of eigenvalues, consider a special case where \( \lambda \) has the eigenvalue \( \lambda \) with multiplicity \( n \) with \( n \) distinct eigenvectors. In this case, we can simplify (6) as follows. Define \( k := \Delta - n \). Then, (6) is equivalent to
\[
\binom{n + k - 1}{k} |\lambda|^k - \binom{k - 1}{i} \binom{n + i - 1}{i} |\lambda|^i < 0.
\]
This condition is satisfied if \( |\lambda| \leq C_{n,k} \), where \( C_{n,k} \) denotes the upper bound on the eigenvalue magnitudes given the parameters \( n \) and \( k \). Figure 1 summarizes the values of \( C_{n,k} \) for different choices of \( n \) and \( k \). Note that \( C_{n,k} \leq C_{n,k} \) for \( n > m \) and \( C_{n,k} \leq C_{n,l} \) if \( k < l \), due to the definition of \( C_{n,k} \). It can be shown that \( C_{1,k} \to 2 \) as \( k \to \infty \). As a result, \( |\lambda| \leq C_{1,k} \leq C_{1,2} \to 2 \). In addition, whenever \( k = n \) or \( \Delta = 2n \), \( |\lambda| < 1 \) is sufficient for this condition to hold, which in turn is sufficient for exact recovery given that the other conditions in Proposition 2 are satisfied. This conclusion is analogous to the stability of the system. In higher-dimensional cases, Propositions 1 and 2 still apply for problem (CO-L1-Aut). However, the KKT conditions will differ by the subdifferential of the \( \ell_2 \) and \( \ell_1 \) norms. In fact, they both have a similar shape. Therefore, one can show that these propositions hold with the same conditions when we use the convex formulation (CO-L1-Aut) using the \( \ell_1 \) norms of the disturbance vectors. Now that we have established the result for the deterministic attack structure for general autonomous systems, we demonstrate the result for the probabilistic attack structure. The next theorem shows the almost sure convergence for attack vectors with symmetric entries.

**Theorem 4:** Consider an autonomous system with dimension \( n \). Suppose that there is an attack at each time \( t \), i.e., \( i \in \mathcal{X} \), with probability \( p \) and this is independent of the other time periods. Let \( \tilde{d}_j^i \) denote the \( j \)-th entry of the vector \( \tilde{d}_i \) and the attack vectors can be dependent over time. Assume that \( \mathbb{P}(\tilde{d}_j^i < 0) = \mathbb{P}(\tilde{d}_j^i > 0) \) for all \( i \in \mathcal{X} \) and \( j = 1, \ldots, n \). Then, \( \tilde{A} \) is almost surely a solution of convex formulation (CO-L2-Aut).

**Proof:** From Proposition 2, \( \tilde{A} \) is a solution if and only if
\[
0 \in \sum_{i \in \mathcal{X}} x_i \otimes \partial \|0\|_2 + \sum_{i \in \mathcal{X}} x_i \otimes \partial \|\tilde{d}_i\|_2.
\]
Our goal is to show that \( 0 \) is included in the expectation of the right-hand side term. The randomness of the system stems from the set of attack vectors \( \tilde{d}_i, \forall i \in \mathcal{X} \). Define \( S := |\mathcal{X}| \) and \( \mathcal{X} := \{i_1, i_2, \ldots, i_S\} \). Let the set of attack vectors be \( \tilde{d}_\mathcal{X} = \{\tilde{d}_{i_1}, \tilde{d}_{i_2}, \ldots, \tilde{d}_{i_S}\} \). Taking the expectation with respect to the random vectors \( \tilde{d}_i, \forall i \in \mathcal{X} \), give rise to
\[
\mathbb{E}_{\tilde{d}_\mathcal{X}} \left[ \sum_{i \in \mathcal{X}} x_i \otimes \partial \|0\|_2 + \sum_{i \in \mathcal{X}} x_i \otimes \partial \|\tilde{d}_i\|_2 \right] = \sum_{i \in \mathcal{X}} \mathbb{E}_{\tilde{d}_i} \left[ x_i \otimes \partial \|\tilde{d}_i\|_2 \right] + \mathbb{E}_{\tilde{d}_\mathcal{X} \setminus \{i_1\}} \left[ \sum_{i \in \mathcal{X} \setminus \{i_1\}} x_i \otimes \partial \|\tilde{d}_i\|_2 \right] + \mathbb{E}_{\tilde{d}_\mathcal{X} \setminus \{i_1\}} \left[ \sum_{i \in \mathcal{X} \setminus \{i_1\}} x_i \otimes \partial \|\tilde{d}_i\|_2 + x_{i_1} \otimes (\partial \|0\|_2 \otimes \mathbb{P}(\tilde{d}_{i_1} = 0)) \right].
\]
Here, \( \mathbb{P}(\tilde{d}_{i_1} = 0) \) is a vector of dimension \( n \) and the \( j \)-th entry of this vector is equal to \( \mathbb{P}(\tilde{d}_{i_1}^j = 0) \). Since \( \mathbb{P}(\tilde{d}_{i_1}^j < 0) = \mathbb{P}(\tilde{d}_{i_1}^j > 0) \), we have \( \mathbb{E}[x_{i_1} \otimes \partial \|\tilde{d}_{i_1}\|_2 \|\tilde{d}_{i_1} \setminus \{i_1\} \|] = x_{i_1} \otimes (\partial \|0\|_2 \otimes \mathbb{P}(\tilde{d}_{i_1} = 0)) \). If the conditional expectation is taken iteratively, we obtain that
\[
\mathbb{E}_{\tilde{d}_\mathcal{X}} \left[ \sum_{i \in \mathcal{X}} x_i \otimes \partial \|0\|_2 + \sum_{i \in \mathcal{X}} x_i \otimes \partial \|\tilde{d}_i\|_2 \right] = \sum_{i \in \mathcal{X}} \mathbb{E}_{\tilde{d}_i} \left[ x_i \otimes \partial \|0\|_2 + \sum_{i \in \mathcal{X}} x_i \otimes (\partial \|0\|_2 \otimes \mathbb{P}(\tilde{d}_i = 0)) \right].
\]
For the subdifferential of \( \| \cdot \|_2 \), it is known that \( 0 \in \partial \| 0 \|_2 \). Consequently,

\[
0 \in \mathbb{E}_{d_k} \left[ \sum_{i \in \mathcal{K}} x_i \otimes \partial \| 0 \|_2 + \sum_{i \in \mathcal{K}} x_i \otimes \partial \| \tilde{d}_k \|_2 \right].
\]

Therefore, \( \tilde{A} \) is a solution almost surely due to the KKT condition and the strong law of large numbers.

We obtained almost sure convergence to the true solution when each entry of the disturbance vector is symmetric around the origin. One could argue that using the objective \( \| D \|_{2, \text{col}} = \sum_{t=1}^{T-1} \| d_t \|_2 \) could be a better alternative to \( \| D \|_{2, \text{col}} = \sum_{t=0}^{T-1} \| d_t \|_2 \). The results above hold in this case as well since the expectation arguments are still the same despite the change in subdifferentials. Therefore, asymptotic results continue to hold despite this change in the objective function.

The next theorem shows that the sample complexity for the exact recovery grows with \( \log(n) \) and \( p/(1-p)^2 \) similar to the one-dimensional case. We again assume that the attack vectors \( \tilde{d}_k \) are possibly dependent and sub-Gaussian random vectors with parameter \( \sigma \). As a result, \( \partial \| d \|_2 = \frac{d}{|d|_2} \) will have bounded entries between \([-1, 1]\) and the bounded random variables are known to be sub-Gaussian as well. In addition, we will utilize some concentration inequalities for sub-exponential random variables.

**Theorem 5:** Consider an autonomous system with dimension \( n \) and suppose that \( \tilde{A} \) has linearly independent eigenvectors with eigenvalues \( \tilde{\lambda}_i < 1 \) for \( i = 1, \ldots, n \). Assume also that there is an attack at time \( i \), i.e., \( i \in \mathcal{K} \), with probability \( p \) and this is independent of the other time periods. Consider the attack vectors \( \tilde{d}_k \) to be possibly dependent zero mean sub-Gaussian vectors with parameter \( \sigma \). Given a positive \( \delta \), if the time horizon \( T \) satisfies

\[
T \geq \max \left\{ \frac{p}{(1-p)^2} \max_i \left\{ \frac{1}{|\tilde{\lambda}_i|^2 (1-|\tilde{\lambda}_i|)} \right\} \log(n^2/\delta), \frac{1}{(1-p)} \max_i \left\{ \frac{1}{|\tilde{\lambda}_i|} \right\} \log(n^2/\delta) \right\},
\]

then \( \tilde{A} \) is a solution to the convex optimization (CO-L2-Aut) with probability at least \( 1 - \delta \).

**Proof:** Since the matrix \( \tilde{A} \) is diagonalizable due to the existence of \( n \) linearly independent eigenvectors, it can be written as \( \tilde{A} = O\tilde{\Lambda}O^T \), where \( O \in \mathbb{R}^n \) is an orthonormal matrix and \( \tilde{\Lambda} \) is the diagonal matrix with diagonal entries \( \tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_n \). From Proposition 2, we know that \( \tilde{A} \) is a solution if and only if

\[
0 \in \sum_{i \in \mathcal{K}} x_i \otimes \partial \| 0 \|_2 + \sum_{i \in \mathcal{K}} x_i \otimes \partial \| \tilde{d}_k \|_2.
\]

Moreover,

\[
x_i = \sum_{k \in \mathcal{K}} \tilde{A}^{(i-1-k)_+} \tilde{d}_k = O \left( \sum_{k \in \mathcal{K}} \tilde{A}^{(i-1-k)_+} \right) O^T \tilde{d}_k
\]

due to the system dynamics, diagonalization of \( \tilde{A} \), and \( x_0 = 0 \). We can substitute this into the KKT condition to arrive at

\[
0 \in O \sum_{i \notin \mathcal{K}} \left( \sum_{k \in \mathcal{K}} \tilde{A}^{(i-1-k)_+} \right) O^T \tilde{d}_k \otimes \partial \| 0 \|_2 + O \sum_{i \in \mathcal{K}} \left( \sum_{k \in \mathcal{K}} \tilde{A}^{(i-1-k)_+} \right) O^T \tilde{d}_k \otimes \frac{\tilde{d}_k}{\| \tilde{d}_k \|_2}.
\]

Since \( O \) is an orthonormal matrix, \( \tilde{d}_k := O^T \tilde{d}_k \) is still a sub-Gaussian vector with the parameter \( \sigma \). In addition, the random variable \( \frac{\tilde{d}_k}{\| \tilde{d}_k \|_2} \) is distributed over the unit sphere \( S_{n-1} \). As a result, the KKT condition can be rewritten as

\[
0 \in \sum_{i \notin \mathcal{K}} \sum_{k \in \mathcal{K}} \tilde{A}^{(i-1-k)_+} \tilde{d}_k \otimes \partial \| 0 \|_2 + \sum_{i \in \mathcal{K}} \sum_{k \in \mathcal{K}} \tilde{A}^{(i-1-k)_+} \tilde{d}_k \otimes \frac{\tilde{d}_k}{\| \tilde{d}_k \|_2}.
\]

Note that the KKT condition is in the matrix form due to the outer product. It is desirable to find the sample complexity for each entry of the matrix, for which we utilize the union bound to guarantee the satisfaction of the KKT condition. The \((l, j)\)-th entry of the matrix in the KKT condition can be written as

\[
0_{l,j} \in \sum_{i \notin \mathcal{K}} \sum_{k \in \mathcal{K}} \tilde{A}^{(i-1-k)_+} \tilde{d}_k \otimes \partial \| 0 \|_2 + \sum_{i \in \mathcal{K}} \sum_{k \in \mathcal{K}} \tilde{A}^{(i-1-k)_+} \tilde{d}_k \otimes \frac{\tilde{d}_k}{\| \tilde{d}_k \|_2},
\]

where \( \tilde{\lambda}_l \) represents the \( l \)-th eigenvalue of the matrix \( \tilde{A} \) and \( \tilde{d}^l \) denotes the \( l \)-th entry of the vector \( \tilde{d}_k \). The right-hand side of the above condition has the minimum value of \( L \) as

\[
\sum_{k \in \mathcal{K}} \left( -\sum_{i \notin \mathcal{K}} \tilde{\lambda}_l^{(i-1-k)_+} \| \partial \| \tilde{d}_k \|_2 + \sum_{i \in \mathcal{K}} \tilde{\lambda}_l^{(i-1-k)_+} \frac{\tilde{d}_k}{\| \tilde{d}_k \|_2} \right) \tilde{d}_k^l,
\]

and the maximum value of \( U \) as

\[
\sum_{k \in \mathcal{K}} \left( \sum_{i \in \mathcal{K}} \tilde{\lambda}_l^{(i-1-k)_+} \| \partial \| \tilde{d}_k \|_2 + \sum_{i \notin \mathcal{K}} \tilde{\lambda}_l^{(i-1-k)_+} \frac{\tilde{d}_k}{\| \tilde{d}_k \|_2} \right) \tilde{d}_k^l.
\]

Since each vector \( \tilde{d}_k \) is sub-Gaussian vector with parameter \( \sigma \), its entries \( \tilde{d}_k^l \) for \( l = 1, \ldots, n \) are sub-Gaussian with parameter \( \sigma \). In addition, we know that \( \frac{\tilde{d}_k^l}{\| \tilde{d}_k \|_2} \) is a bounded random variable between \([-1, 1]\) since it is on the unit sphere. Therefore, \( \frac{\tilde{d}_k^l}{\| \tilde{d}_k \|_2} \) is sub-Gaussian with parameter 1. Consequentially, \( \frac{\tilde{d}_k^l}{\| \tilde{d}_k \|_2} \) is sub-exponential with \( (\nu_1 \sigma, \nu_2 \sigma) \) for some universal constants \( \nu_1 \) and \( \nu_2 \) [21]. Using similar arguments as in the proof of Theorem 3, one can conclude that the expectation of \( L \) and \( U \) are

\[
\mathbb{E}[U] = -\mathbb{E}[L] = c_2 \sigma \left( \sum_{k \in \mathcal{K}} \sum_{i \notin \mathcal{K}} \tilde{\lambda}_l^{(i-1-k)_+} \right).
\]

Since the sum of sub-exponential random variables is sub-exponential as well, it holds that

\[
L \sim U \sim s \mathbb{E}(c_1 \nu_1 \sigma_1 (\tilde{\lambda}_l), \nu_2 \sigma)
\]

\footnote{A random variable \( X \) with mean \( \mu \) is sub-exponential with parameters \((\nu, \alpha)\) if \( \mathbb{E}[e^{\lambda (X - \mu)}] \leq e^{\nu^2 \lambda^2/2}, \forall |\lambda| \leq 1/\alpha \).}
where \( \sigma_1(\tilde{\lambda}_j) \) is defined as
\[
\sigma_1(\tilde{\lambda}_j) = \sqrt{\left( \sum_{k \in \mathcal{X}} \frac{1}{2 (1 - |\tilde{\lambda}_i|)^2 |\tilde{\lambda}_i|^2} \right)^2 + \left( \sum_{k \in \mathcal{X}} \frac{1}{2 (1 - |\tilde{\lambda}_i|)^2 |\tilde{\lambda}_i|^2} \right)^2} \sigma^2
\]
and \( c_1 \) is a positive real constant. Since \( L \) and \( U \) are sub-exponential, we have the following concentration bounds:
\[
P(L > E[L] + t) \leq e^{-\min \left\{ \frac{t^2}{2 |\lambda|^2 (1 - |\lambda_i|)^2} \right\}},
\]
and
\[
P(U < E[U] - t) \leq e^{-\min \left\{ \frac{t^2}{2 |\lambda|^2 (1 - |\lambda_i|)^2} \right\}}.
\]
Therefore, 1 - \( P(L < 0 < U) \) can be upper-bounded using the union bound by substituting \( t = -E[L] \) and \( t = E[U] \):
\[
1 - \text{part of the first minimum in the exponential term results in the sample complexity of}
\]
\[
T \gtrsim \frac{p}{(1 - p)^2 \left( 1 - |\lambda_i| \right)^2 |\lambda_i|^2} \log(2/\delta)
\]
in light of Theorem 3, and the second part of the minimum in the exponential term results in the sample complexity of
\[
T \gtrsim \frac{1}{(1 - p) |\lambda_i|} \log(2/\delta).
\]
Therefore, the \((l, j)\)-th entry of the matrix in the KKT condition is satisfied with probability at least \( 1 - \delta \) if
\[
T \gtrsim \max \left\{ \frac{p}{(1 - p)^2 \left( 1 - |\lambda_i| \right)^2 |\lambda_i|^2} \log(2/\delta), \frac{1}{(1 - p) |\lambda_i|} \log(2/\delta) \right\}.
\]
Using the union-bound over the \( n^2 \) entries of the matrix in the KKT condition, it can be concluded that \( \tilde{A} \) is a solution with the probability at least \( 1 - \delta \) if
\[
T \gtrsim \max \left\{ \frac{p}{(1 - p)^2} \max_i \left\{ \frac{1}{|\lambda_i|^2 (1 - |\lambda_i|)^2} \right\} \log(n^2/\delta), \frac{1}{(1 - p)} \max_i \left\{ \frac{1}{|\lambda_i|} \right\} \log(n^2/\delta) \right\}.
\]
The two different terms for the sample complexity stem from the concentration inequality for the sub-exponential random variables. We require sub-exponential results because the KKT condition involves the multiplication of two sub-Gaussian random variables, which are known to be sub-exponential. The diagonalizability assumption on the matrix \( \tilde{\lambda} \) is not restrictive. Note that when \( \tilde{\lambda} \) is not diagonalizable, for every \( \epsilon > 0 \), it is possible to find a matrix \( \tilde{\lambda} \) such that \( \| \tilde{\lambda} - \tilde{\lambda} \| < \epsilon \). This is equivalent to slightly perturbing the eigenvalues of the matrix \( \tilde{\lambda} \) an infinitesimal amount to obtain a diagonalizable matrix. Therefore, we can push \( \epsilon \) to zero and remove the assumption of diagonalizability from the theorems. The results continue to hold because the optimization solution is continuous in its parameters and \( \tilde{\lambda} \) is exactly recoverable. We can obtain a similar result if one prefers to use the problem \((\text{CO-L1-Aut})\) to recover the system matrix \( \tilde{\lambda} \).

Theorem 6: Under the assumption of Theorem 5, if the time horizon \( T \) satisfies
\[
T \gtrsim \frac{p}{(1 - p)^2} \max_i \left\{ \frac{1}{|\lambda_i|^2 (1 - |\lambda_i|)^2} \right\} \log(n^2/\delta),
\]
then \( \tilde{\lambda} \) is a solution to the convex optimization problem \((\text{CO-L1-Aut})\) with probability at least \( 1 - \delta \).

Proof: We have the same KKT condition as in the previous theorem. However, the subdifferentials of the \( \ell_1 \) and \( \ell_2 \) norms differ. For a vector \( \tilde{d} \), the \( i \)-th entry of subdifferential of its \( \ell_1 \)-norm is given as
\[
[\partial ||\tilde{d}||_1]_i = \begin{cases} \text{sign}(\tilde{d}_i) \tilde{d}_i, & \text{if } \tilde{d}_i \neq 0 \\ [-1, 1], & \text{if } \tilde{d}_i = 0 \end{cases}
\]
Consequently, the \((l, j)\)-th entry of the matrix in the KKT condition can be written as follows using similar arguments as in the proof of Theorem 5:
\[
0_{l, j} \in \sum_{i \not\in \mathcal{X}} \sum_{k \in \mathcal{X}} \tilde{\lambda}_i^{(i-1-k)_+} \tilde{d}_i^{(l-j)_+} \text{sign}(\tilde{d}_k),
\]
and the maximum value of \( L \) as
\[
\sum_{k \in \mathcal{X}} \left( -\sum_{i \not\in \mathcal{X}} \tilde{\lambda}_i^{(i-1-k)_+} |\text{sign}(\tilde{d}_k)| + \sum_{i \not\in \mathcal{X}} \tilde{\lambda}_i^{(i-1-k)_+} |\text{sign}(\tilde{d}_i)| \right) d_k,
\]
This is similar to \( L \) and \( U \) in Theorem 3, where we have \( \tilde{\lambda}_i \) instead of \( \lambda_i \). Therefore, the \((l, j)\)-th entry of the matrix in the KKT condition is satisfied with probability at least \( 1 - \delta \) if
\[
T \gtrsim \frac{p}{(1 - p)^2} \max_i \left\{ \frac{1}{|\lambda_i|^2 (1 - |\lambda_i|)^2} \right\} \log(2/\delta).
\]
Using the union-bound over the \( n^2 \) entries of the matrix in the KKT condition, it can be concluded that \( \tilde{\lambda} \) is a solution with probability at least \( 1 - \delta \) if
\[
T \gtrsim \frac{p}{(1 - p)^2} \max_i \left\{ \frac{1}{|\lambda_i|^2 (1 - |\lambda_i|)^2} \right\} \log(n^2/\delta).
\]
The results on sample complexity are intuitive. As the probability of having an attack increases, we require a larger time horizon for exact recovery. In addition, if the system is barely stable with eigenvalues close to the unit circle, the sample complexity blows up. Furthermore, since the sample complexity scales with the logarithm of the dimension, the proposed estimators are scalable with respect to the dimension.
V. SYSTEMS WITH INPUT SEQUENCE

It is desirable to understand the role of an input sequence in exact recovery because the majority of dynamic systems are controlled by an external input. Since the input sequence is generated by a controller, one can design it in such a way that it accelerates the exact recovery. In the non-autonomous case, the system dynamics is given as $x_{i+1} = Ax_i + Bu_i + d_i$, $i = 0, \ldots, T - 1$, where $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times m}$, similar to the autonomous case, the true system matrices $\hat{A}$ and $\hat{B}$ are not known and the goal is to obtain these matrices using the state trajectory. We will investigate estimators (CO-L2) and (CO-L1) defined earlier. The inputs $u_i, i \in \{0, \ldots, T - 1\}$ are known unlike the disturbances.

A. Single State with Single Input Case

We study the non-asymptotic properties of the problems (CO-L2) and (CO-L1) when there is a single state and a single input. These problems are equivalent when the state space is one-dimensional. Thus, we only consider (CO-L2). We assume that an independent and identically distributed sub-Gaussian sequence of inputs is injected into the system at each period. This allows us to obtain a high probability bound for the exact recovery of the matrices $\hat{A}$ and $\hat{B}$. A random input sequence is commonly used in system identification and online learning because it enables the exploration of the system to learn the system dynamics faster. The sub-Gaussian input assumption may sound restrictive. Nevertheless, it is satisfied when $u_i$ is designed in the linear feedback form as $u_i = Kx_i + \omega$. If the states $x_i$ are sub-Gaussian and the input is excited with sub-Gaussian noise $\omega$, the input vector $u_i$ is also sub-Gaussian. Therefore, the most common input sequence used in optimal control satisfies this assumption. Note that the closed loop system can be written as $x_{i+1} = (\hat{A} + BK)x_i + \hat{B} \omega + \hat{d}_i$. Thus, the problem is equivalent to estimating the matrices $(\hat{A} + BK)$ and $\hat{B}$ when the linear feedback control is used. The following theorem follows from these observations.

Theorem 7: Consider a stable single-input single-state system with the dynamics $x_{i+1} = \hat{A}x_i + \hat{B}u_i + \hat{d}_i$ for $i = 0, \ldots, T - 1$. Assume also that there is an attack at time $i$, i.e., $i \in \mathcal{X}$, with probability $p$ and this is independent of the other time periods. Assume that the attack vectors $\hat{d}_i, \forall i \in \mathcal{X}$ are zero-mean sub-Gaussian with parameter $\sigma$ and the inputs are sub-Gaussian with parameter $\epsilon$ and they may be possibly dependent. Given a positive number $\delta$, if the time horizon $T$ satisfies

$$ T \geq \max \left\{ \frac{1}{(1 - p)^2} \log(1/\delta), \right. $$

$$ \left. \frac{p}{(1 - p)^2 (1 - |A|^2 |A|)^2} \log(1/\delta) \right\}, $$

then $(\hat{A}, \hat{B})$ is a solution to (CO-L2) with probability at least $1 - \delta$.

Proof: Due to the system dynamics and given $x_0 = 0$, $x_i$ can be expressed as

$$ x_i = \sum_{k=0}^{T-1} \hat{A}^{(i-k)}, \hat{B}u_k + \sum_{k=0}^{T-1} \hat{A}^{(i-k)}, \hat{d}_k. $$

Since $n = m = 1$, we need to show that

$$ 0 \in \sum_{i \in \mathcal{X}} x_i \partial[0]_2 + \sum_{i \in \mathcal{X}} x_i \partial[\hat{d}_i]_2, \quad (8a) $$

and

$$ 0 \in \sum_{i \in \mathcal{X}} u_i \partial[0]_2 + \sum_{i \in \mathcal{X}} u_i \partial[\hat{d}_i]_2, \quad (8b) $$

hold simultaneously. We first investigate (8a). Substituting the expression for $x_i$ into (8a) results in the relation.

$$ 0 \in \sum_{i \in \mathcal{X}} u_i \partial[0]_2 + \sum_{i \in \mathcal{X}} u_i \partial[\hat{d}_i]_2, \quad (8b) $$

It is known from Theorem 3 that if $T \geq \frac{p}{(1 - p)^2 (1 - |A|^2 |A|)^2} \log(1/\delta)$, then

$$ \mathbb{P}(E_1) \leq \frac{\delta}{4}, $$

where $E_1$ is defined as the following event

$$ E_1 := \{0 \notin \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \hat{A}^{(i-k)}, \hat{d}_k \partial[0]_2 \} $$

$$ + \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \hat{A}^{(i-k)}, \hat{d}_k \partial[\hat{d}_i]_2. $$

Similarly, if $T \geq \frac{p}{(1 - p)^2 (1 - |A|^2 |A|)^2} \log(1/\delta)$, then

$$ \mathbb{P}(E_2) \leq \frac{\delta}{4}, $$

where $E_2$ is defined as the following event

$$ E_2 := \{0 \notin \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \hat{A}^{(i-k)}, \hat{B}u_k \partial[0]_2 \} $$

$$ + \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \hat{A}^{(i-k)}, \hat{B}u_k \partial[\hat{d}_i]_2. $$

Therefore, if $T \geq \frac{p}{(1 - p)^2 (1 - |A|^2 |A|)^2} \log(1/\delta)$, then the first KKT condition (8a) holds with probability at least $1 - \delta/2$ by using the union bound. This can be shown by upper and lower bounding the right-hand side of the term in event $E_2$. This term can be lower-bounded as $L$ and upper-bounded as $U$ as follows:

$$ \sum_{k=0}^{T-1} \left( - \sum_{i \in \mathcal{X}} \hat{A}^{(i-k)}, \hat{B} \partial[u_k]_2 + \sum_{i \in \mathcal{X}} \hat{A}^{(i-k)}, \hat{B} \partial[\hat{d}_i]_2 \right) u_k, $$

and

$$ \sum_{k=0}^{T-1} \left( \sum_{i \in \mathcal{X}} \hat{A}^{(i-k)}, \hat{B} \partial[u_k]_2 + \sum_{i \in \mathcal{X}} \hat{A}^{(i-k)}, \hat{B} \partial[\hat{d}_i]_2 \right) u_k,$$
respectively. We aim to show that $L < 0 < U$ with high probability. Similar to the calculations in the proof of Theorem 3, we have that $u_k \partial |d_i|_2$ and $u_k \partial ||u_k||_2 = |u_k|$ are sub-Gaussian with parameter $\varepsilon$. Therefore, 

$$L \sim U \sim sG(c_1 \varepsilon(\bar{A}, \bar{B})),$$

where $c_1$ is defined as

$$|\bar{B}| \varepsilon \left( \sum_{k=0}^{T-1} \sum_{i \in \mathcal{X}} |\bar{A}^{i-(i-k)}| \right) = \left( \sum_{k=0}^{T-1} \sum_{i \neq k} |\bar{A}^{i-(i-k)}| \right)^2.$$

In addition, $\mathbb{E}[|u_k|] = c_2 \varepsilon$ since $u_k$ is sub-Gaussian with parameter $\varepsilon$. Thus, the expectations of $L$ and $U$ can be written as

$$\mathbb{E}[U] = -\mathbb{E}[L] = c_2 \varepsilon \left( \sum_{k=0}^{T-1} \sum_{i \in \mathcal{X}} |\bar{A}^{i-(i-k)}| \right) |\bar{B}|.$$

Since $L$ and $U$ are sub-Gaussian, the following concentration bounds hold:

$$\mathbb{P}(L > \mathbb{E}[L] + t) \leq e^{-\frac{t^2}{2\varepsilon^2(\bar{A}, \bar{B})}},$$

and

$$\mathbb{P}(U < \mathbb{E}[U] - t) \leq e^{-\frac{t^2}{2\varepsilon^2(\bar{A}, \bar{B})}}.$$

Therefore, $1 - \mathbb{P}(L < 0 < U)$ can be upper-bounded using the union bound by substituting $t = -\mathbb{E}[L]$ and $t = \mathbb{E}[U]$:

$$1 - \mathbb{P}(L < 0 < U) \leq 2e^{-\frac{\mathbb{E}[L]^2}{2\varepsilon^2(\bar{A}, \bar{B})}}.$$

Moreover, 

$$\mathbb{E}[U] = c_2 \varepsilon \left( \sum_{k=0}^{T-1} \sum_{i \in \mathcal{X}} |\bar{A}^{i-(i-k)}| \right) |\bar{B}| \geq \varepsilon(1-p)T|\bar{A}| |\bar{B}|,$$

and

$$\varepsilon^2(\bar{A}, \bar{B}) \leq pT \frac{|\bar{B}|^2}{(1-|\bar{A}|)^2} \varepsilon^2.$$

As a result,

$$1 - \mathbb{P}(L < 0 < U) \leq 2 \exp \left\{ -\frac{(1-p)^2}{p} |\bar{A}|^2 (1-|\bar{A}|)^2 T \right\} \leq \delta / 4.$$

Therefore, if $T \geq \frac{1}{(1-p)^2} \frac{1}{(1-|\bar{A}|)^2} \log(8/\delta)$, then the first KKT condition (8a) is satisfied with probability at least $1 - \delta / 2$.

Checking the second KKT condition (8b) is more straightforward than the first one. Note that lower and upper bounds for (8b) can be expressed as

$$L := - \sum_{i \in \mathcal{X}} |u_i| + \sum_{i \neq k} u_k \partial |d_i|_2,$$

and

$$U := \sum_{i \in \mathcal{X}} |u_i| + \sum_{i \neq k} u_k \partial |d_i|_2.$$

Hence,

$$L \sim U \sim sG(\sqrt{T}),$$

and

$$\mathbb{E}[U] = -\mathbb{E}[L] = c\varepsilon(1-p)T.$$

Then, similar calculations yield the probability bound below:

$$1 - \mathbb{P}(L < 0 < U) \leq 2 \exp \left\{ -\frac{(1-p)^2}{p} \right\} = \delta / 2.$$

Therefore, if $T \geq \frac{1}{(1-p)^2} \log(4/\delta)$, then the second KKT condition (8b) is satisfied with probability at least $1 - \delta / 2$.

As a result, using the union bound, if

$$T \geq \max \left\{ \frac{1}{(1-p)^2} \log(1/\delta), \frac{p}{(1-p)^2 (1-|\bar{A}|)^2 |\bar{A}|} \right\},$$

$(\bar{A}, \bar{B})$ is a solution to the estimator with probability at least $1 - \delta$.

We obtained a high probability bound for the exact recovery of the system matrices $\bar{A}$ and $\bar{B}$. The first term in the sample complexity corresponds to the satisfaction of the KKT condition for the input sequence $\{u_t\}_{t=0}^{T-1}$, whereas the second term corresponds to the satisfaction of the KKT condition for the state measurements $\{x_t\}_{t=0}^{T-1}$. Similar to autonomous systems, the sample complexity increases as the probability of a large disturbance increases. Because we have a logarithmic dependence on the satisfaction of the probability bound, Theorem 7 implies almost sure asymptotic convergence to the correct matrices $\bar{A}$ and $\bar{B}$.

### B. General Case with State Size $n$ and Input Size $m$

In this section, we present our most general results when the state size is $n$ and input size $m$. Our assumptions for the exact recovery are mild: system stability and sub-Gaussian inputs.

**Theorem 8:** Consider a stable system with $n$ states and $m$ inputs with system dynamics $x_{i+1} = \bar{A}x_i + \bar{B}u_i + \bar{d}_i$, for $i = 0, \ldots, T-1$ and suppose that $\bar{A}$ has linearly independent eigenvectors with eigenvalues $|\bar{A}_l| < 1$ for $l = 1, \ldots, n$. Assume also that there is an attack at time $i$, i.e., $i \in \mathcal{X}$, with probability $p$ and this is independent of the other time periods. Assume that the attack vectors $\bar{d}_i, \forall i \in \mathcal{X}$, are zero-mean sub-Gaussian with parameter $\sigma$ and the inputs are sub-Gaussian with parameter $\varepsilon$ and they may be possibly dependent. Given a positive number $\delta$, if the time horizon $T$ is chosen as max $\{T_1, T_2\}$ where

$$T_1 \geq \max \left\{ \frac{p}{(1-p)^2} \frac{1}{|\bar{A}|} \log(n^2/\delta), \frac{1}{(1-p)^2} \log(n^2/\delta) \right\},$$

and

$$T_2 \geq \frac{1}{(1-p)^2} \log(mn/\delta),$$

then $(\bar{A}, \bar{B})$ is a solution to (CO-L2) with probability at least $1 - \delta$. 

---

**Note:** The text contains mathematical expressions and theorems that are relevant to system identification and robustness in the presence of disturbances and attacks. Theorems and propositions are presented with clear notations and conditions to ensure clarity and precision in the discussion. The focus is on establishing bounds and conditions for the recovery of system matrices despite adversarial perturbations.
Proof: If the matrices $\tilde{A}$ and $\tilde{B}$ satisfy the KKT conditions provided in Theorem 1, we conclude that $(\tilde{A}, \tilde{B})$ is the solution pair to the CO-L2. The KKT conditions can be written as

\[
0 \in \sum_{i \in \mathcal{X}} x_i \circ \partial \|0\|_2 + \sum_{i \in \mathcal{X}} x_i \circ \partial \|\tilde{d}_i\|_2, 
\tag{10a}
\]

and

\[
0 \in \sum_{i \in \mathcal{X}} u_i \circ \partial \|0\|_2 + \sum_{i \in \mathcal{X}} u_i \circ \partial \|\tilde{d}_i\|_2. 
\tag{10b}
\]

It is desirable to show that (10a) and (10b) are satisfied with probability at least $1 - \delta/2$ when the theorem statements are satisfied separately, which will imply that exact recovery is guaranteed using the union bound with probability at least $1 - \delta$. Due to the system dynamics and given $x_0 = 0$, $x_i$ can be expressed as

\[
x_i = \sum_{k=0}^{T-1} \tilde{A}^{(i-k)} \circ \tilde{B} u_k + \sum_{k \in \mathcal{K}} \tilde{A}^{(i-k)} \circ \tilde{d}_k.
\]

Substituting the expression for $x_i$ into (10a) results in

\[
0 \in \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \tilde{A}^{(i-k)} \circ \tilde{B} u_k \circ \partial \|0\|_2 + \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \tilde{A}^{(i-k)} \circ \tilde{B} u_k \circ \partial \|\tilde{d}_i\|_2 + \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \tilde{A}^{(i-k)} \circ \tilde{d}_k \circ \partial \|0\|_2 + \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \tilde{A}^{(i-k)} \circ \tilde{d}_k \circ \partial \|\tilde{d}_i\|_2.
\]

Since the matrix $\tilde{A}$ is diagonalizable due to the existence of $n$ linearly independent eigenvectors, it can be written as $\tilde{A} = O\Lambda O^T$, where $O \in \mathbb{R}^n$ is an orthonormal matrix and $\Lambda$ is the diagonal matrix with diagonal entries $\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_n$. Similar to the proof of Theorem 5, the $(i,j)$-th entry of the matrix in the KKT condition above can be transformed to the following condition

\[
0_{t,j} \in \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \tilde{\lambda}^{(i-k)} \circ (\tilde{B} u_k)^j \circ \partial \|0\|_2^j + \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \tilde{\lambda}^{(i-k)} \circ (\tilde{B} u_k)^j \circ \partial \|\tilde{d}_i\|_2^j + \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \tilde{\lambda}^{(i-k)} \circ \tilde{d}_k \circ \partial \|0\|_2^j + \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \tilde{\lambda}^{(i-k)} \circ \tilde{d}_k \circ \partial \|\tilde{d}_i\|_2^j,
\]

where $\tilde{B} = O^T \tilde{B}$, and $\tilde{d}_k = O^T \tilde{d}_k$. It is known from Theorem 5 that if

\[
T \geq \max \left\{ \frac{p}{(1-p)^2} \frac{1}{(1 - |\tilde{\lambda}_i|^2)} \log(8/\delta), \right. 
\frac{1}{1 - p} \frac{1}{|\tilde{\lambda}_i|^2} \log(8/\delta) \},
\]

then

\[
\mathbb{P}(E_1) \leq \frac{\delta}{4},
\]

where $E_1$ is defined as the following event

\[
E_1 := 0_{t,j} \not\in \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \tilde{\lambda}^{(i-k)} \circ (\tilde{B} u_k)^j \circ \partial \|0\|_2^j + \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \tilde{\lambda}^{(i-k)} \circ (\tilde{B} u_k)^j \circ \partial \|\tilde{d}_i\|_2^j.
\]

Similarly, when we have the same order of the samples, the following probability bound holds:

\[
\mathbb{P}(E_2) \leq \frac{\delta}{4}, \tag{11}
\]

where $E_2$ is defined as the following event

\[
E_2 := 0_{t,j} \not\in \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \tilde{\lambda}^{(i-k)} \circ (\tilde{B} u_k)^j \circ \partial \|\tilde{d}_i\|_2^j.
\]

Note that $(\tilde{B} u_k)^j$ is a sub-Gaussian random variable with parameter $\varepsilon \|\tilde{B} t\|_2$, where $\tilde{B} t$ denotes the $t$-th row of the $\tilde{B}$. As a result, we can lower and upper bound the term in the probability bound above. The lower bound $L$ is

\[
L := \sum_{k=0}^{T-1} \left[ \sum_{i \in \mathcal{X}} \sum_{k \neq j} \tilde{\lambda}^{(i-k)} \circ \tilde{B} \circ \tilde{u}_k \circ \partial \|\tilde{d}_i\|_2 \right] - \sum_{i \in \mathcal{X}} \sum_{k \neq j} \tilde{\lambda}^{(i-k)} \circ \tilde{B} \circ \tilde{u}_k \circ \partial \|\tilde{d}_i\|_2,
\]

and the upper bound $U$ is

\[
U := \sum_{k=0}^{T-1} \left[ \sum_{i \in \mathcal{X}} \sum_{k \neq j} \tilde{\lambda}^{(i-k)} \circ \tilde{B} \circ \tilde{u}_k \circ \partial \|\tilde{d}_i\|_2 \right] + \sum_{i \in \mathcal{X}} \sum_{k \neq j} \tilde{\lambda}^{(i-k)} \circ \tilde{B} \circ \tilde{d}_k \circ \partial \|\tilde{d}_i\|_2.
\]

By performing similar calculations as in the earlier proofs, one can obtain the sub-exponential parameters for $L$ and $U$, besides the expectation of $L$ and $U$. More precisely,

\[
\mathbb{E}[U] = \mathbb{E}[L] = c_2 \varepsilon \|\tilde{B} t\|_2 \left( \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{X}} \tilde{\lambda}^{(i-k)} \right),
\]

and

\[
L \sim U \sim s \mathbb{E}(c_1 v_1 \|\tilde{B} t\|_2 \varepsilon_1(\tilde{\lambda}_i), v_2 \varepsilon \|\tilde{B} t\|_2)
\]

where $\varepsilon_1(\tilde{\lambda}_i)$ is

\[
\left[ \left( \sum_{k \in \mathcal{K}} \left( \sum_{i \in \mathcal{X}} \tilde{\lambda}^{(i-k)} \right)^2 \right)^\frac{1}{2} + \left( \sum_{i \in \mathcal{X}} \tilde{\lambda}^{(i-k)} \right)^\frac{1}{2} \right] \varepsilon^2
\]
Since $L$ and $U$ are sub-exponential, we have the following concentration bounds:

$$
\mathbb{P}(L > \mathbb{E}[L] + t) \leq e^{-\min\left\{ \frac{t^2}{2\mathbb{E}[\eta_1^2]}, \frac{t^2}{2\mathbb{E}[\eta_1^2]}, \frac{t^2}{2\mathbb{E}[\eta_1^2]} \right\}},
$$

and

$$
\mathbb{P}(U < \mathbb{E}[U] - t) \leq e^{-\min\left\{ \frac{t^2}{2\mathbb{E}[\eta_1^2]}, \frac{t^2}{2\mathbb{E}[\eta_1^2]}, \frac{t^2}{2\mathbb{E}[\eta_1^2]} \right\}}.
$$

Therefore, $1 - \mathbb{P}(L < 0 < U)$ can be upper-bounded using the union bound by substituting $t = -\mathbb{E}[L]$ and $t = \mathbb{E}[U]$:

$$
1 - \mathbb{P}(L < 0 < U) \leq 2e^{-\min\left\{ \frac{\mathbb{E}[\eta_1^2]}{2\mathbb{E}[\eta_1^2]}, \frac{\mathbb{E}[\eta_1^2]}{2\mathbb{E}[\eta_1^2]}, \frac{\mathbb{E}[\eta_1^2]}{2\mathbb{E}[\eta_1^2]} \right\}}.
$$

Solving the probability bound above in terms of the sample complexity $T$ gives that if

$$
T \gtrsim \max\left\{ \frac{p}{(1-p)^2} \left( \frac{1}{\mathcal{A}_1} \right)^2 \log\left( \frac{8}{\delta} \right), \frac{1}{(1-p)\mathcal{A}_1^2} \log\left( \frac{8}{\delta} \right) \right\},
$$

then (11) holds. Hence, by taking the union bound over the $n^2$ entries of the matrix in the first KKT condition, if the sample complexity is

$$
T \gtrsim \max\left\{ \frac{p}{(1-p)^2} \max_i \left( \frac{1}{|\mathcal{A}_i|^2 (1 - |\mathcal{A}_i|)^2} \right) \log\left( \frac{n^2}{\delta} \right), \frac{1}{(1-p)\min|\mathcal{A}_i|} \log\left( \frac{n^2}{\delta} \right) \right\},
$$

then the first KKT condition is satisfied with probability at least $1 - \delta/2$. As a next step, we can show that the second KKT condition (10b) holds with probability at least $1 - \delta/2$ as well. The $(l,j)$-th entry of the second KKT condition above can be written as

$$
0_{l,j} = \sum_{i \in \mathcal{X}} u_i^l \partial \|0\|_2 + \sum_{i \in \mathcal{X}} u_i^l \bar{d}_i^l \|\bar{d}_i\|_2.
$$

We again define lower and upper bounds for the above expression using the flexibility of the subdifferential at the origin, given as

$$
L := -\sum_{i \in \mathcal{X}} |u_i^l| + \sum_{i \in \mathcal{X}} u_i^l \bar{d}_i^l \|\bar{d}_i\|_2,
$$

and

$$
U := \sum_{i \in \mathcal{X}} |u_i^l| + \sum_{i \in \mathcal{X}} u_i^l \bar{d}_i^l \|\bar{d}_i\|_2,
$$

respectively. Note that $L$ and $U$ are sub-exponential random variables:

$$
L \sim U \sim sE(c_1 \nu_1 \varepsilon \sqrt{T}, v_2 \varepsilon).
$$

Moreover, $\mathbb{E}[U] = -\mathbb{E}[L] = c_4 \varepsilon (1 - p) T$. As before, we obtain that

$$
1 - \mathbb{P}(L < 0 < U) \leq e^{-\min\left\{ \frac{\mathbb{E}[\eta_1^2]}{2\mathbb{E}[\eta_1^2]}, \frac{\mathbb{E}[\eta_1^2]}{2\mathbb{E}[\eta_1^2]}, \frac{\mathbb{E}[\eta_1^2]}{2\mathbb{E}[\eta_1^2]} \right\}}.
$$

Therefore, if $T \gtrsim \max\left\{ \frac{1}{(1-p)^2} \log\left( \frac{4}{\delta} \right), \frac{1}{(1-p)} \log\left( \frac{4}{\delta} \right) \right\} = \frac{1}{(1-p)^2} \log\left( \frac{4}{\delta} \right)$, then each element of the second KKT condition is satisfied with probability at least $1 - \delta/2$. If we take the union bound over the $mn$ entries of the second KKT condition, we conclude that the condition (10b) is satisfied with probability at least $1 - \delta/2$ if

$$
T \gtrsim \frac{1}{(1-p)^2} \log(mn/\delta).
$$

Hence, the union bound over the two KKT conditions completes the proof.

Similar to previous theorems for the autonomous case, we require a sample complexity that scales with $p/(1-p)^2$ and terms depending on the eigenvalues of $\tilde{A}$. The sample complexity $T_2$ is needed to satisfy the KKT condition that depends on the input sequence. The number of required samples increases with the logarithm of the dimension of the unknown matrices $n^2$ and $mn$.

**Theorem 9:** Under assumptions of Theorem 8 if the time horizon $T$ satisfies

$$
T \gtrsim \frac{1}{(1-p)^2} \log(mn/\delta),
$$

then $\tilde{A}, \tilde{B}$ is a solution to (CO-L1) with probability at least $1 - \delta$.

**Proof:** Similar to the proof of Theorem 8, the KKT conditions for problem (CO-L1) can be written as

$$
0 \in \sum_{i \in \mathcal{X}} x_i \otimes \partial \|0\|_2 + \sum_{i \in \mathcal{X}} x_i \otimes \partial \|\tilde{d}_i\|_2, \quad (12a)
$$

$$
0 \in \sum_{i \in \mathcal{X}} u_i \otimes \partial \|0\|_2 + \sum_{i \in \mathcal{X}} u_i \otimes \partial \|\tilde{d}_i\|_2, \quad (12b)
$$

Using similar transformations and arguments, one can show that if the sample complexity is

$$
T \gtrsim \frac{p}{(1-p)^2} \left( \frac{1}{|\mathcal{A}_1|^2 - |\mathcal{A}_1|^2} \right) \log\left( \frac{8}{\delta} \right),
$$

then the following probability bounds hold for each entry of the first KKT condition (12a):

$$
\mathbb{P}(E_1) \leq \frac{\delta}{4},
$$

where $E_1$ is defined as the following event

$$
E_1 := 0_{l,j} \notin \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \mathcal{A}_1^{i-k} \tilde{d}_k^l \partial \|0\|_2^i + \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \mathcal{A}_1^{i-k} \tilde{d}_k^l \partial \|\tilde{d}_k\|_2^i,
$$

and

$$
\mathbb{P}(E_2) \leq \frac{\delta}{4}.
$$
where $E_2$ is defined as the following event
\[
E_2 := 0_{i,j} \not\in \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \lambda_k (i-1-k) \gamma (\bar{B} u_k)^T \partial ||0||_1^T + \sum_{i \in \mathcal{X}} \sum_{k=0}^{T-1} \lambda_k (i-1-k) \gamma (\bar{B} u_k)^T \partial ||d_i ||_1.
\]

Thus, whenever
\[
T \geq \frac{p}{(1-p)^2} \max_i \left\{ \frac{1}{|\lambda_k|^2 (1- |\lambda_k|)^2} \right\} \log(n^2 / \delta),
\]
the condition (12a) is satisfied with probability at least $1 - \delta / 2$. By adopting the same arguments as in the proof of Theorem 8, we can show that each entry of the second KKT condition (12b) is satisfied with probability at least $1 - \delta / 2$ if
\[
T \geq \frac{1}{(1-p)^2} \log(1 / \delta).
\]
Therefore, the condition (12b) is satisfied with probability at least $1 - \delta / 2$ when
\[
T \geq \frac{1}{(1-p)^2} \log(mn / \delta).
\]

Hence, the union bound over the two KKT conditions completes the proof.

We note that when the input sequence $u_t = K x_t$ is used to control the system, this input sequence satisfies the assumptions in the above theorems if $x_t$ are sub-Gaussian. The closed-loop system with the matrix $(A + BK)$ results in the second solution $A' = A + BK$ and $B' = 0$. Nevertheless, the ground-truth system matrix pair $(\bar{A}, \bar{B})$ is also a solution to our estimators. This phenomenon occurs due to the existence of multiple optimal solutions and it could be avoided if the input is excited with a small sub-Gaussian noise in the form of $u_t = K x_t + \omega$.

VI. NUMERICAL EXPERIMENT

We provide a numerical experiment inspired by biomedical applications to demonstrate the results of this paper. We consider a compartmental model of blood sugar and insulin dynamics in the human body; see [22]. It is crucial to accurately estimate the parameters of the dynamics when the blood sugar level is regulated through the injection of a bolus of insulin into the system. Due to the complex structure of the human body, the dynamics are not the same for different individuals. We consider a linear system based on Hovarka’s model given below [23]:

\[
\begin{align*}
\dot{x}_1 &= -k_{a1} x_1 - k_{b1} I + d_1 \\
\dot{x}_2 &= -k_{a2} x_1 - k_{b2} I + d_2 \\
\dot{x}_3 &= -k_{a3} x_1 - k_{b3} I + d_3 \\
\dot{S}_1 &= -S_1 / t_{\text{max},I} + d_4 \\
\dot{S}_2 &= S_1 / t_{\text{max},I} - S_2 / t_{\text{max},I} + d_5 \\
I &= S_2 / (t_{\text{max},I}) - k_I + d_6
\end{align*}
\]

where the states $x_1, x_2, x_3$ represent the influence of insulin on the system of the body. $S_1$ and $S_2$ represent the absorption rate of insulin in the, directly and indirectly, accessible compartment models, respectively. Lastly, the state $I$ stands for the blood-sugar level in the body. The disturbance $d_4$ shows the bolus injection to the body and the remaining disturbance vectors model sudden changes in the body due to diseases such as diabetes. Although the injected insulin amount could be known, the amount of insulin and when it reaches the effective body parts are not known exactly. Hence, $d_i$ values are treated as unknown. Even though the disturbance in this application is not a malicious attack, it has similar features for identification purposes: the arrival time of the bolus is unknown and, once it arrives, it has a large magnitude.

We discretize the continuous time system to obtain an LTI system using $\Delta t = 0.5$. The obtained matrix $\bar{A}$ is stable and our goal is to estimate the parameters $(k_{ai}, k_{bi}, t_{\text{max},I}, V_I, k_e)$ where the true values are obtained from Table 1 in [24]. We model the attacks as zero-mean Gaussian random vectors with identity covariance matrix with variance 10 and we run our model with the probability of attack being $p = 0.2$, $p = 0.4$, and $p = 0.6$. We report the estimation error $||A^* - \bar{A}||_F$ for the least-squares estimator, problem CO-L2 and problem CO-L1.

Figure 4 suggests that our proposed estimators attain the exact recovery while the least-squares estimator fails to do such. As the probability of having an attack $p$ increases, the number of required time periods for exact recovery grows proportional to $p / (1 - p)^2$. Note that there are more corrupted data than clean data in the case of $p = 0.6$. In addition, because there is no sparsity assumption on the attack vectors, CO-L2 performs slightly better than CO-L1. We compare the performance of CO-L2 and CO-L1 by running a similar experiment with and without sparse disturbances. When the disturbances are sparse, $d_1, d_2, d_3, d_5$ are set to zero while $d_4$ and $d_6$ have the same Gaussian distribution as before. Figure 4 shows that the two methods perform similarly when the attack vectors are also sparse.

VII. CONCLUSION AND FUTURE WORK

We studied the problem of learning LTI systems under adversarial attacks by studying two lasso-type estimators. We considered both deterministic and probabilistic attack models in terms of the time occurrence of the attack and developed strong conditions for the exact recovery of the system dynamics. When the attack occurs deterministically at every $\Delta$ period, the exact recovery is possible after $n + \Delta$ time steps. Moreover, if the system is attacked at each time instance with probability $p$, then the system matrices are recovered with high probability when $T$ is on the order of $O(p / (1 - p)^2)$ and the logarithm of the dimension of the problem. We obtained similar results when the system is controlled by an input sequence. The results are corroborated by a numerical experiment in biology that supports the nonasymptotic analytic results. This work provides the first set of mathematical guarantees for the robust non-asymptotic analysis of dynamic systems. One of the possible extensions to this work will be the design of an algorithm to predict and update $(\bar{A}_{t+1}, \bar{B}_{t+1})$ using the latest estimation $(\bar{A}_t, \bar{B}_t)$ and the new data $(x_{t+1}, u_t)$, instead of solving the problem from scratch at each time period. Furthermore, We leave the study of noisy systems and online control of dynamic systems under adversaries as future work.
variables is also sub-Gaussian. Specifically, if the independent Gaussian random variables are dependent, the sum of sub-Gaussian random variables is again sub-Gaussian with mean 0 and parameter \( \sum_{i=1}^{n} \sigma_i^2 \). Therefore, we have the following concentration bound for the sum of sub-Gaussian random variables.

**Lemma 2:** (Hoeffding Bound [25]) Suppose that the variables \( X_i, i = 1, \ldots, n \), are independent, and that \( X_i \) has mean \( \mu_i \) and sub-Gaussian parameter \( \sigma_i \). Then, for all \( t \geq 0 \), we have

\[
\mathbb{P} \left( \sum_{i=1}^{n} (X_i - \mu_i) \geq t \right) \leq \exp \left( -\frac{t^2}{2 \sum_{i=1}^{n} \sigma_i^2} \right).
\]

Since the subgradients involve the absolute values of the sub-Gaussian random variables, it is useful to understand the properties of the random variable \( |X| \) when \( X \) is sub-Gaussian with parameter \( \sigma \). The following lemma characterizes the properties of \( |X| \).

**Lemma 3:** (Definition [21]): Given a zero mean sub-Gaussian random variable \( X \) with mean 0 and parameter \( \sigma \), the random variable \( |X| \) is sub-Gaussian with parameter \( \sigma \). Moreover, there exists a constant \( c \) such that \( \mathbb{E}[|X|] = c \sigma \).

Because subgradients are bounded random variables, we utilize the following Lemma in our proofs.

**Lemma 4:** Suppose that the random variable \( X \) is bounded and belongs to the interval \([a, b]\). Then, \( X \) is sub-Gaussian with parameter \((b - a)/2\).

For high-dimensional problems, we define sub-Gaussian vectors. Each element of a sub-Gaussian vector with parameter \( \sigma \) is also sub-Gaussian with parameter \( \sigma \).

**Definition 3 (Sub-Gaussian Vector [25]):** A random vector \( X \in \mathbb{R}^d \) with zero mean is sub-Gaussian with parameter \( \sigma \) if

\[
\mathbb{E}[e^{\lambda (X - \mu)}] \leq e^{\sigma^2 \lambda^2 / 2}, \quad \forall \lambda \in \mathbb{R}.
\]

It is denoted as \( X \sim sG(\sigma) \).

It is trivial to show that the sum of sub-Gaussian random variables is also sub-Gaussian. Specifically, if the independent random variables \( X_i \) are sub-Gaussian with mean \( \mu_i \) and parameter \( \sigma_i \), then \( \sum_{i=1}^{n} (X_i - \mu_i) \) is sub-Gaussian with mean 0 and parameter \( \sqrt{\sum_{i=1}^{n} \sigma_i^2} \). Furthermore, even if the sub-Gaussian random variables are dependent, the sum of sub-Gaussian random variables is also sub-Gaussian with mean 0 and parameter \( \sum_{i=1}^{n} \sigma_i^2 \). Therefore, we have the following concentration bound for the sum of sub-Gaussian random variables.

**Definition 4 (Sub-exponential Random Variable [25]):** A random variable \( X \) with mean \( \mu = \mathbb{E}[X] \) is sub-exponential with parameters \((\nu, \alpha)\) if

\[
\mathbb{E}[e^{\lambda(X - \mu)}] \leq e^{\nu \lambda^2 / 2}, \quad \forall |\lambda| \leq 1/\alpha.
\]

It is denoted as \( X \sim sE(\nu, \alpha) \).

We have the following tail bound for sub-exponential random variables, which is analogous to the Hoeffding bound for sub-Gaussian random variables.
Lemma 5 (Sub-exponential Tail Bound [25]): Suppose that $X$ is sub-exponential with parameters $(v, \alpha)$. Then,

$$
\mathbb{P}(X - \mu \geq t) \leq \begin{cases} 
    e^{-t^2/2\alpha^2}, & \text{if } 0 \leq t \leq \sqrt{2\alpha}, \\
    e^{-t/2\alpha}, & \text{if } t \geq \sqrt{2\alpha}.
\end{cases}
$$

The sum of independent sub-exponential random variables with mean $\mu_i$ and parameters $(v_i, \alpha_i)$ is also sub-exponential with parameters $\sum_i v_i^2, \max_i \alpha_i$. Furthermore, the sum of possibly dependent sub-exponential random variables with mean $\mu_i$ and parameters $(v_i, \alpha_i)$ is also sub-exponential with parameters $\left( \sum_i v_i^2, \max_i \alpha_i \right)$, where $c$ is a constant real number. We also provide a lemma for the multiplication of two sub-Gaussian random variables.

Lemma 6: [21] Given two independent zero-mean sub-Gaussian random variables $X_1$ and $X_2$ with parameters $\sigma_1$ and $\sigma_2$, there exist scalar constants $c_1$ and $c_2$ such that $X_1 X_2$ is sub-exponential with parameters $(c_1 \sigma_1 \sigma_2, c_2 \sigma_1 \sigma_2)$.

REFERENCES


