IEEE Control Systems Letters paper presented at
2023 American Control Conference (ACC)
San Diego, CA, USA. May 31 - June 2, 2023

Learning of Dynamical Systems under Adversarial Attacks - Null Space Property Perspective

Han Feng, Baturalp Yalcin, and Javad Lavaei

Abstract— **We study the identification of a linear time-invariant dynamical system affected by large-and-sparse disturbances modeling adversarial attacks or faults. Under the assumption that the states are measurable, we develop necessary and sufficient conditions for the recovery of the system matrices by solving a constrained lasso-type optimization problem. In addition, we provide an upper bound on the estimation error whenever the disturbance sequence is a combination of small noise values and large adversarial values. Our results depend on the null space property that has been widely used in the lasso literature, and we investigate under what conditions this property holds for linear time-invariant dynamical systems. Lastly, we further study the conditions for a specific probabilistic model and support the results with numerical experiments.**

I. INTRODUCTION

The identification of linear time-invariant (LTI) systems is a classic problem in control theory that has been studied extensively. Despite the long history of this problem and its application in a wide range of real-world systems, the non-asymptotic analysis of the system identification problem has gained popularity in recent years, which targets the sample complexity of the problem [1], [2]. With the growing popularity of safety-critical applications, such as autonomous driving and unmanned aerial vehicles, the design of a system identification framework that is robust against adversarial attacks is crucial [3].

In this paper, we consider LTI systems for which the states are under a sequence of unknown disturbances, some of which take small values modeling noise and the remaining ones take strategic values due to adversarial attacks or faults in the system. We study the lasso-based optimization problem recently proposed in [4]. It can recover the exact system dynamics uniquely whenever adversarial attacks occur intermittently with enough time separation. In [4], some adversarial attacks that do not influence the estimation are studied, whereas this paper improves those results by providing a necessary and sufficient condition for recovery as well as non-asymptotic bounds on the error. Our approach is based on defining a null space property that is analogous to the null space property condition for the lasso problem [5], which is required to guarantee the exact recovery.

The robustness analysis of estimators has a long history, dating back to the seminal paper [6]. It is known that a small disturbance on the estimation problem, such as the perturbation of a data point, could lead to significant changes in the outcome of the estimator. This has led to a major effort on the robustification estimators via regularizers. The works [7] and [8] have found a strong relationship between the robust estimation and regularization in regression problems by showing the equivalence of these two problems.

The recent papers [9] and [10] on robust estimation of linear measurement models have considered a framework with two types of noise: small measurement noise and large intermittent noise. They have developed necessary and sufficient conditions for the exact recovery when a column-wise summable norm is used to minimize the error. We focus on this type of norm in this paper, which will be defined as the sum of ℓ_2 norms of the columns of a matrix. Nevertheless, our analysis is for the more challenging problem of system identification where the parameters are correlated over time. Our results indirectly provide a guideline on how to design an effective input sequence to learn system dynamics faster.

The recent papers [11] and [12] on system identification have studied the problem of learning a sparse and structured state-space model, and provided bounds on the required sample size, i.e., sample complexity bounds. However, none of the aforementioned works are applicable to adversarial attacks since their noise/disturbance model is Gaussian. The more recent work [13] has utilized a conic relaxation, which significantly increases the problem dimension and is not directly applicable to dynamical systems. It estimates how many erroneous measurements or adversarial attacks can be handled by the estimator without causing a nonzero estimation error. There are some other works in the literature that provide non-asymptotic error bounds for the linear system identification problem when the ordinary least-squares estimation method and Kalman-Ho algorithm are used [14], [15]. However, these methods are not particularly efficient for robust estimation whenever the data is corrupted in an adversarial way. Membership estimators are also utilized to show a consistent estimation of linear systems [16]. Unfortunately, they do not provide non-asymptotic bounds. The work [17] has studied the scenario where the attack is executed on the outputs rather than the states. Unlike the attack on the outputs, the effect of the attack on the states propagates over time. Lastly, some other related works on robust estimation are resilient state estimation [18], [19] and Byzantine fault tolerance [20], [21].

One could place the system identification problem into the broader context of robust regression to gain some valuable insight on robust estimation. It is well-known that least-squares methods are not robust to outliers. The work [22] has studied the identification of outliers in linear regression. It is shown that a non-convex loss function outperforms the

This work was supported by grants from ARO, AFOSR, ONR, and NSF.
The authors are with the University of California, Berkeley. E-mail:
{han_feng, byalcin, lavaei}@berkeley.edu

ℓ_1 regularization of the least-squares function. Nevertheless, it is not always justifiable to solve large-scale non-convex problems instead of convex ones unless the landscape of the non-convex optimization problem can be shown to be benign (e.g., it does not have a spurious solution). Nonetheless, this is problem specific and not understood thoroughly [23], [24]. Another mainly used estimator for sparse estimation is the hard thresholding estimator. The work [25] has proposed an iterative scheme based on this estimator and analyzed it on regression with sparse disturbances. There have been several other works on robust estimation and training [26], [27], [28]. The major difference between those works and the system identification is that the states or the training data are not independent over time. Hence, they cannot be re-ordered, which makes it challenging to exploit the existing results in robust statistics. A possible solution to this is resetting the system and using the last available data point from each trajectory. However, this is not a feasible approach for common real-life applications due to its complexity. Also, it is often desired to identify the system in an online fashion to benefit from the available data, but it is not well understood how this can be achieved for robust estimation.

In Section II, we introduce the main notations used in the paper. Section III considers a particular type of l_1 minimization problem and formulates the problem. In Section IV, we derive sufficient conditions for exact recovery in finite time when we have exact measurements of the states that are influenced by the adversarial attacks. The noisy case is studied in Section V, where we provide an error bound on the estimation error based on the noise intensity. The conditions are based on the null space property (NSP), which is hard to verify directly. We derive sufficient conditions for NSP in Section VI and then show in Section VII that NSP holds for a particular attack model where the input is Gaussian and the adversary injects disturbances intermittently with a fixed policy based on the states and input measurements. The proofs are provided in the technical report [29].

II. NOTATIONS

For a given matrix Z, the i-th largest singular value of Z is denoted by $\sigma_i(Z)$, and the minimal and maximum singular values of Z are shown by $\sigma_{\min}(Z)$ and $\sigma_{\max}(Z)$, respectively. For a matrix Z, $\|Z\|_F$ denotes its Frobenius norm and for a vector z, $\|z\|_2$ denotes its ℓ_2 norm. $Z \succ 0$ and $Z \succeq 0$ denote a square symmetric matrix Z that is positive definite and positive semidefinite, respectively. The function $\mathrm{tr}(\cdot)$ stands for the trace of a square matrix. The $n \times n$ identity matrix is denoted as $\mathbf{I}_n$. The Minkowski sum of two sets $\mathscr{E}$ and $\mathscr{F}$ is denoted by $\mathscr{E} \oplus \mathscr{F} = \{e+f : e \in \mathscr{E}, f \in \mathscr{F}\}$. The sum with the inverse of the set is denoted by $\mathscr{E} \ominus \mathscr{F} = \{e-f : e \in \mathscr{E}, f \in \mathscr{F}\}$. For two vectors v and w, $\langle v, w \rangle$ is the inner product between those vectors in their respective vector space. $\mathbb{P}(\cdot)$ and $\mathbb{E}[\cdot]$ denote the probability of an event and the expectation of a random variable. A Gaussian random variable X with mean μ and covariance matrix Σ is written as $X \sim N(\mu, \Sigma)$. $|S|$ shows the cardinality of a given set S.

III. PROBLEM FORMULATION

Consider an LTI dynamical system over the time horizon $[0, T]$:

$$x_{t+1} = \bar{A}x_t + \bar{B}u_t + \bar{d}_t, \quad t = 0, 1, \ldots, T-1,$$

where $\bar{A} \in \mathbb{R}^{n \times n}$ and $\bar{B} \in \mathbb{R}^{n \times m}$ are unknown matrices in the state-space model to be estimated and $\bar{d}_t$'s are unknown disturbances. Throughout the paper, the bar over each parameter of interest (such as $\bar{A}$) indicates the unknown ground truth. The goal is to find the matrices $\bar{A}$ and $\bar{B}$ from the state measurements $x_0, ..., x_T \in \mathbb{R}^n$ and input data $u_0, ..., u_{T-1} \in \mathbb{R}^m$. The disturbances $\bar{d}_0, ..., \bar{d}_{T-1}$ model both noise and anomalies in the system, such as attacks or actuator's faults. Without any assumptions on the disturbance, the identification problem is not well-defined due to the impossibility of separating $\bar{A}x_t + \bar{B}u_t$ from the disturbance $\bar{d}_t$. For instance, if $\bar{d}_t = A'x_t + B'u_t$ for some matrices A' and B', then the system evolves as if the system matrices are $(\bar{A}+A', \bar{B}+B')$ and the disturbance is zero, which makes the identification problem have non-unique solutions. We will make certain sparsity assumptions on the disturbance in the noiseless case, and generalize the result to the noisy case.

To formulate the problem, we introduce the matrix notation $X = [x_0, \ldots, x_{T-1}]$, $U = [u_0, \ldots, u_{T-1}]$, and $D = [d_0, \ldots, d_{T-1}]$. The last state x_T appears in our optimization problem, but it is not a column in the matrix notation. The attack D is assumed to be restricted to a set $\mathscr{D} \subseteq \mathbb{R}^{n \times T}$. The set $\mathscr{D}$ captures the user's belief of possible times of attack and their values.

Define the sum of norm error $\|D\|_{2,col} := \sum_i \|d_i\|_2$, where the index is over the columns of D. The (column-wise) support of D is defined as $\mathrm{supp}(D) = \{i \in \{0, \ldots, T-1\} : d_i \neq 0\}$. For each subset of indices $I \subseteq \{0, 1, \ldots, T-1\}$, the complement of I is defined as $I^c = \{i \in \{0, \ldots, T-1\} : i \notin I\}$. For a matrix $Z \in \mathbb{R}^{n \times T}$, the *projection* $\Pi_I Z$ is a matrix whose columns are zero except for those in I, i.e.,

$$(\Pi_I Z)_i = \begin{cases} z_i, & \text{if } i \in I \\ 0, & \text{otherwise} \end{cases},$$

where z_i denotes the i-th column of Z. Define Z_I as a submatrix of Z of size $n \times |I|$, that includes only those columns of Z in the index set I. We use the shorthand notations $Z_{\neq i}$ and $Z_{\notin I}$ to denote $Z_{\{0,\ldots,T-1\}\setminus\{i\}}$ and $Z_{\{0,\ldots,T-1\}\setminus I}$, respectively. The range of Z is defined as $\mathscr{R}(Z) = \{\sum_i \lambda_i z_i : \lambda_i \in \mathbb{R}\}$.

To recover the system matrices A and B, we analyze the following convex optimization problem:

$$\min_{\substack{A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m}, \\ D \in \mathscr{D}}} \sum_{i=0}^{T-1} \|d_i\|_2 \tag{1}$$
$$s.t. \quad x_{i+1} = Ax_i + Bu_i + d_i, \quad i = 0, \ldots, T-1,$$

where the states $x_i, i \in \{0, \ldots, T\}$, are generated according to

$$x_{i+1} = \bar{A}x_i + \bar{B}u_i + \bar{d}_i, \quad i = 0, \ldots, T-1 \tag{2}$$

and they are measured perfectly. The control inputs $u_i, i \in \{0, \ldots, T-1\}$, may be designed but are fixed in the optimization problem (1). This problem differs from the classical

l_1 minimization (basis pursuit) problem that aims to find a ground truth vector $\bar{z}$ via

$$\begin{aligned} \min_{z} \quad & \|z\|_1 \\ s.t. \quad & \Phi\bar{z} = \Phi z, \end{aligned}$$

for a given matrix Φ for multiple reasons. First, we apply the l_1 norm at the group level to the disturbances $d_1,\dots,d_{T-1}$, because we only assume sparsity in the occurrence of the disturbance but not its value. The vector d_i is not sparse if there is an attack at time i. Second, we do not attempt to minimize the l_1 norm of all the unknown parameters. In particular, the system matrices A and B are not assumed to be sparse. Moreover, the states $x_0,\dots,x_{T-1}$ appear in the constraints and depend on the input u_i and the disturbance d_i. Since the states are correlated, we cannot independently rescale them, as is commonly done in the analysis of l_1 optimization problems.

IV. THE NOISELESS CASE

This section studies the noiseless case, where each disturbance $\bar{d}_i$ is either zero or designed by an attacker to disturb the operation of the system. We aim to understand how to design the input of the system so that the identification of the excited system in the presence of adversarial disturbances is possible. We use $S = \mathrm{supp}(\bar{D})$ to denote the time stamps of actual attacks. The set of possible disturbances $\mathscr{D}$ is assumed to be closed under the projection onto S.

Assumption 1. *The set of disturbances $\mathscr{D}$ is convex and contains 0 in its interior. $\Pi_S(D) \in \mathscr{D}$ for all $D \in \mathscr{D}$.*

A key step in the study of problem (1) is the Null Space Property, which is formalized below.

Definition 1. *Let $c > 0$, $S \subseteq \{0,\dots,T-1\}$, and $\mathscr{R}$ be a subset of $\mathbb{R}^{n\times T}$. The matrix $\begin{bmatrix} X^T & U^T \end{bmatrix}^T \in \mathbb{R}^{(n+m)\times T}$ is said to satisfy the Null Space Property (NSP) with the constant c, index set S, and range set $\mathscr{R}$ ($(c,S,\mathscr{R})$-NSP) if, for all matrices $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ such that $-AX - BU \in \mathscr{R}$ and (A,B) is not zero, it holds that*

$$\left\| [A,B] \begin{bmatrix} X_S \\ U_S \end{bmatrix} \right\|_{2,col} < c \left\| [A,B] \begin{bmatrix} X_{S^c} \\ U_{S^c} \end{bmatrix} \right\|_{2,col}. \tag{3}$$

When the set S or $\mathscr{R}$ is clear from the context, we omit them and use c-NSP or (c,S)-NSP to highlight the parameters of interest. NSP has been used in [30], [31] to prove an exact recovery result for ℓ_1-norm minimization for compressed sensing and has been further studied in [32]. Nevertheless, the above definition and the theorems in this paper for system identification are novel ideas for the system identification problem. The following theorem states that 1-NSP is sufficient for the exact recovery of all sparse disturbances.

Theorem 1. *The following statements are equivalent:*

a) $\begin{bmatrix} X \\ U \end{bmatrix}$ *satisfies the $(1,S,\mathscr{D}\ominus\mathscr{D})$-NSP where $S = \mathrm{supp}(\bar{D})$.*

b) $(\bar{A},\bar{B},\bar{D})$ *is the unique solution to problem* (1).

The paper [9] has shown that 1-NSP is necessary for the exact recovery of all instances of a certain class of robust regression problems. However, because $x_0,\dots,x_{T-1}$ appear on both sides of the constraint, the system identification problem under study is structured and is only a subset of all instances of the regression problems.

Remark 1. *The case without control input is a special case of problem* (1)*, for which the $(c,S,\mathscr{D}\ominus\mathscr{D})$-NSP becomes*

$$\|AX_S\|_{2,col} < c\|AX_{S^c}\|_{2,col} \tag{4}$$

for all nonzero matrices $A \in \mathbb{R}^{n\times n}$ such that $-AX \in \mathscr{D}\ominus\mathscr{D}$. The NSP property with $c = 1$ ensures that $\bar{A}$ is the unique solution to the optimization problem

$$\min_{\substack{A\in\mathbb{R}^{n\times n},\\ D\in\mathscr{D}}} \sum_{i=0}^{T-1} \|d_i\|_2 \tag{5a}$$

$$s.t. \quad x_{i+1} = Ax_i + d_i, \quad i = 0,\dots,T-1, \tag{5b}$$

where the states $x_i, i \in \{0,\dots,T\}$, are generated according to

$$x_{i+1} = \bar{A}x_i + \bar{d}_i, \quad i = 0,\dots,T-1.$$

V. THE NOISY CASE

This section studies the noisy case, where some of the disturbances $d_0,\dots,d_{T-1}$ represent regular noise values and others are engineered by an attacker. Let S denote the attack times, meaning that d_i represents an attack if $i \in S$. For $i \in S^c$, the parameter d_i represents noise and its value is often small in practice. The next theorem provides an error bound for estimating the matrices A and B.

Theorem 2. *Assume that $T > (m+n)$ and that the matrix $\begin{bmatrix} X \\ U \end{bmatrix}$ has full row rank. If (X,U) satisfies the $(c,S,\mathscr{D}\ominus\mathscr{D})$-NSP with $c < 1$, then each solution $(\hat{A},\hat{B},\hat{D})$ to the optimization problem* (1) *satisfies*

$$\|[\hat{A}-\bar{A},\hat{B}-\bar{B}]\|_F \le 2\frac{1+c}{1-c} \times \frac{\|\bar{D}_{S^c}\|_{2,col}}{\sigma_{\min}\left(\begin{bmatrix} X \\ U \end{bmatrix}\right)}.$$

The term $2\|\bar{D}_{S^c}\|_{2,col}$ on the right-hand side of Theorem 2 can be improved to $2\|\bar{D}_{S^c}\|_{2,col} - \|\bar{D}\|_{2,col} + \|\hat{D}\|_{2,col}$ using similar techniques as those in basis pursuit; see for example [31, Theorem 4.14] that has proven an equivalence to c-NSP for basis pursuit problems. This term shows that the intensity of noise is zero in the noiseless case subject to attacks. Problem (1) is a special case of basis pursuit where measurements are correlated. The bound, including the constant $\frac{1+c}{1-c}$, could potentially be improved with more knowledge about the constraints (see [33] for a similar scenario).

VI. SATISFACTION OF NSP

After observing the states and input sequence, condition (3) enables certifying whether one can recover the true dynamics using problem (1). Theorem 2 has shown that the NSP condition is useful in obtaining a bound of the identification error. The following lemmas attempt to derive

stronger conditions that are more tractable than $(c,S,\mathscr{D})$-NSP. They can be combined with the results of the previous two sections to understand how to design the input to improve the likelihood of successfully recovering the system matrices through the convex optimization problem (1).

Lemma 1. *If $T \geq (m+n)$ and*

$$\sqrt{|S|}\sigma_{\max}\begin{bmatrix} X_S \\ U_S \end{bmatrix} < c \times \sigma_{\min}\begin{bmatrix} X_{S^c} \\ U_{S^c} \end{bmatrix}, \tag{6}$$

where $S = \operatorname{supp}(\bar{D})$ and $|S^c| \geq m+n$, then $\begin{bmatrix} X^T & U^T \end{bmatrix}^T$ satisfies the $(c,S,\mathscr{R})$-NSP for every range set $\mathscr{R}$.

Definition 2. *Given a matrix $V = [v_0,\dots,v_{T-1}]$ and a natural number s, V is said to be s-self-decomposable if for all indices $I \subseteq \{0,1,\dots,T-1\}$ of size $|I| = s$, it holds that $V_i \in \operatorname{range}(V_{\notin I})$ for all $i \in I$. The s-self-decomposable amplitude is defined as*

$$\xi_s(V) := \max_{\substack{I\subseteq\{0,\dots,T-1\}\\ |I|=s}} \min_{\substack{\Gamma_I\in\mathbb{R}^{(T-s)\times s}\\ \Gamma_I=[\gamma_i]_{i\in I}}} \left\{ \sum_{k\in I} \|\gamma_k\|_\infty : V_I = V_{\notin I}\Gamma_I \right\}. \tag{7}$$

If U is s-self-decomposable, by definition it is also t-self-decomposable for $t < s$. We are particularly interested in the cases when $s = 1$ and $s = |S|$.

Lemma 2. *If $\begin{bmatrix} X^T & U^T \end{bmatrix}^T$ has full row rank and is s-self-decomposable where $s = |S|$, then it satisfies the $(c,S,\mathscr{R})$-NSP for every $c > \xi_s(\begin{bmatrix} X^T & U^T \end{bmatrix}^T)$.*

Lemma 3. *Given $S = \operatorname{supp}(\bar{D})$ with $|S| > 1$, assume that $\begin{bmatrix} X^T & U^T \end{bmatrix}^T$ has full row rank and is 1-self-decomposable where*

$$\xi_1 := \xi_1(\begin{bmatrix} X^T & U^T \end{bmatrix}^T) \leq \frac{1}{|S|-1}.$$

Then, it satisfies the $(c,S,\mathscr{R})$-NSP and

$$c > \frac{|S|\xi_1}{1-(|S|-1)\xi_1}.$$

Remark 2. *Lemma 3 implies that when $\xi_1 < \frac{1}{2|S|-1}$, the ground truth $(\bar{A},\bar{B},\bar{D})$ is recoverable through problem* (1).

We have yet not answered how the control input sequence $u_0,\dots,u_{T-1}$ affects the satisfaction of NSP. This will be achieved after we consider a probabilistic model for the disturbances.

VII. A PROBABILISTIC MODEL

The results of the previous section are applicable only when the state and input sequences have been observed — they cannot be directly used to find a concrete input design scheme that achieves exact recovery in the noiseless case or asymptotic recovery in the noisy case. This section considers a particular type of random input. Our approach relies on the observation that, despite the attacker's attempt, one can apply the block martingale small ball condition [34] to obtain a probabilistic estimate on $\sigma_{\min}\left(\begin{bmatrix} X^T & U^T \end{bmatrix}^T\right)$ for sub-Gaussian random matrices.

We first restate the block martingale small ball (BMSB) condition [34]. Define the filtration $\mathscr{F}_t$ as the smallest σ-algebra obtained by the data available up to time t, i.e. $x_0,\dots,x_t,u_0,\dots,u_t,d_0,\dots,d_{t-1}$, so that the vector-valued process $\begin{bmatrix} x_t^T & u_t^T \end{bmatrix}^T$ with $t \geq 0$ is $\{\mathscr{F}_t\}_{t\geq 0}$ adapted.

Definition 3. *Given a filtration $\{\mathscr{F}_t\}_{t\geq 0}$ and a vector-valued process $V_t \in \mathbb{R}^d$ with $t \geq 0$, the process is said to satisfy (k,Γ_{sb},p)-BMSB for some matrix $\Gamma_{sb} \succ 0$ if*

$$\frac{1}{k}\sum_{i=1}^{k} \mathbb{P}\left(|\langle w, V_{j+i}\rangle|^2 \geq w^\top \Gamma_{sb} w | \mathscr{F}_j\right) \geq p, \text{ almost surely}$$

for all fixed vectors $w \in \mathbb{R}^d$ with $\|w\|_2 = 1$ and all $j \geq 0$.

We make the following two assumptions about the input and attack model. They ensure that the process is Gaussian and we can apply the BMSB conditions to estimate σ_{min}.

Assumption 2. *The input sequence $u_0,\dots,u_{T-1}$ are independent and identically distributed Gaussian random vectors with $N(0,\sigma^2\mathbf{I}_m)$.*

Assumption 3. *The attack model satisfies the following:*

- *The set of attack times $S \subseteq \{0,\dots,T-1\}$ is fixed.*
- *For every $t \notin S$, d_t is some noise that follows the distribution $N(0,\varepsilon^2\mathbf{I}_n)$ for some positive number ε.*
- *For every $t \in S$, d_t can be expressed as $Px_t + Qu_t + e_t$, where P and Q are constant matrices of compatible size. The matrices P and Q are not dependent on $\mathscr{F}_t$. Also, e_t is a random variable that follows the Gaussian distribution $N(0,\varepsilon^2\mathbf{I}_n)$ and is independent of $\mathscr{F}_t$.*

We assume a Gaussian attack model because stealthy and covert attacks happen with symmetric probability distributions with zero mean to stay undetected. Therefore, the Gaussian distribution assumption best fits this type of attack. The attack dynamics cannot be augmented with the system's dynamics since $d_t = 0$ if $t \notin S$. It is desirable to bound the Gramian matrix, which is identified as a key measure of sample complexity in [34]. Define the following constants:

$$\begin{aligned} \alpha_{\min} &= \min(\sigma_{\min}(A+P), \sigma_{\min}(A)) \\ \alpha_{\max} &= \max(\sigma_{\max}(A+P), \sigma_{\max}(A)) \\ \beta_{\max} &= \max(\sigma_{\max}(B+Q), \sigma_{\max}(B)). \end{aligned}$$

Lemma 4. *Let $\Gamma_t = \mathbb{E}\left[x_t x_t^\top\right]$ for $t = 0,\dots,T-1$. We have*

$$\begin{aligned} \Gamma_t &\succeq \alpha_{\min}^2\Gamma_{t-1} + \varepsilon^2\mathbf{I}_n, \\ \Gamma_t &\preceq \alpha_{\max}^2\Gamma_{t-1} + \left(\varepsilon^2 + \beta_{\max}^2\right)\mathbf{I}_n. \end{aligned}$$

In particular,

$$\begin{aligned} \Gamma_t &\succeq \sum_{i=0}^{T-1} \alpha_{\min}^{2i}\varepsilon^2\mathbf{I}_n \\ \Gamma_t &\preceq \Gamma_t^{\max} := \alpha_{\max}^{2t}\Gamma_0 + \sum_{i=0}^{T-1} \alpha_{\max}^{2i}\left(\varepsilon^2+\beta_{\max}^2\right)\mathbf{I}_n. \end{aligned}$$

Let $\Gamma := \operatorname{diag}(\varepsilon^2\mathbf{I}_n, \sigma^2\mathbf{I}_\mathbf{m})$. The next lemma confirms that the BMSB condition can be leveraged for our problems.

Lemma 5. *Under Assumptions 2 and 3, for every sequence of indices $0 \le s_0 < s_1 < s_2, \ldots$, the sub-process $\begin{bmatrix} x_{s_t}^T & u_{s_t}^T \end{bmatrix}^T$ with $t \ge 0$ satisfies the $(k, \frac{1}{2}\Gamma, \frac{1}{12})$-BMSB condition.*

The BMSB condition provides a non-asymptotic bound on the singular value of $\begin{bmatrix} X^T & U^T \end{bmatrix}^T$.

Proposition 1. *Under Assumptions 2 and 3, define $C(I) := \left(m\sigma^2|I| + \sum_{i\in I} tr(\Gamma_i^{\max})\right)$, where $\Gamma_i^{\max}$ is given in Lemma 4. For every subset $I \subseteq \{0,1,\ldots,T-1\}$, we have*

$$\mathbb{P}\left(\sigma_{\max}\left(\begin{bmatrix} X_I \\ U_I \end{bmatrix}\right) > \sqrt{\frac{C(I)}{\eta}}\right) \le \eta \tag{8}$$

and

$$\begin{aligned}\mathbb{P}&\left(\sigma_{\min}\left(\begin{bmatrix} X_I \\ U_I \end{bmatrix}\right) < \min(\varepsilon,\sigma)\sqrt{\frac{k\lfloor |I|/k \rfloor p^2}{16}}\right) \le \eta \\ &+ \exp\left(-\frac{|I|p^2}{10k} + 2(m+n)\log(10/p)\right. \\ &\left.+\frac{1}{2}(m+n)\log\left(\frac{C(I)}{\min(\varepsilon,\sigma)^2 \frac{k\lfloor |I|/k \rfloor p^2}{16}\eta^2}\right)\right)\end{aligned} \tag{9}$$

The proof is a direct consequence of the covering argument in [34, Section D]. We are now able to provide a sufficient condition for the satisfaction of NSP in our attack model.

Theorem 3. *Assume that $\alpha_{\max} < 1$. Given $c, \eta > 0$, there exist constants $N \in \mathbb{N}$ and $h > 0$ such that $\begin{bmatrix} X^T & U^T \end{bmatrix}^T$ is c-NSP with probability at least $1 - 3\eta$ as long as $|S|^2 < h|S^c|$ and $|S^c| > N$.*

VIII. NUMERICAL EXPERIMENTS

We provide numerical experiments to support the theoretical results obtained in this paper. For illustration purposes, we focus on an autonomous system of the form $x_{t+1} = \bar{A}x_t + \bar{d}_t$. It is a special case with $\bar{B} = 0$. We generate a diagonal matrix Λ whose diagonal entries are uniformly distributed between 0 and 1. In addition, we generate a Gaussian matrix P whose entries are independent and identically distributed (i.i.d.) with $N(0,1)$. Then, we set the matrix $\bar{A}$ to be $P\Lambda P^{-1}$. The vector x_0 is generated randomly based on a Gaussian distribution with i.i.d $N(0,1)$ entries. At each time instance, an adversarial attack occurs with the probability $p = 0.3$. Thus, $\mathbb{E}[|S|] = Tp$. An adversarial attack d_i at time i is distributed as $N(0, 100\mathbf{I}_n)$. In the noisy case, a Gaussian noise vector with i.i.d. $N(0,1)$ entries are added at each time instance while they are omitted in the noiseless case. We compare the estimation error of problem (1) with the estimation error of the least-squares problem:

$$\min_{A\in\mathbb{R}^{n\times n}} \sum_{t=0}^{T-1} \|x_{i+1} - Ax_i\|_2^2.$$

Given an estimate $\hat{A}$, the error is calculated as $\|\hat{A} - \bar{A}\|_F$. We plot the estimation errors of problem (1) and the least-squares method with respect to time. In Figure 1, we report the results for the noiseless case. It is observed that after a sufficient number of data points are obtained, problem (1) exactly identifies the system as supported by Theorems 1 and 3. On the other hand, the least-squares estimation errors reach a plateau due to adversarial attacks. In Figure 2, we implement a similar analysis for the noisy case. It is seen that the estimation errors decrease significantly faster than those for the least-squares estimation when our convex optimization formulation is used as supported by Theorem 2. Thus, we can conclude that problem (1) provides a more accurate estimation of the system dynamics. Additional experimental results in [29] results indicate that as the probability of an attack p increases, the exact recovery requires more iterations. We leave the theoretical analysis of the effect of p on the effect recovery as further study.

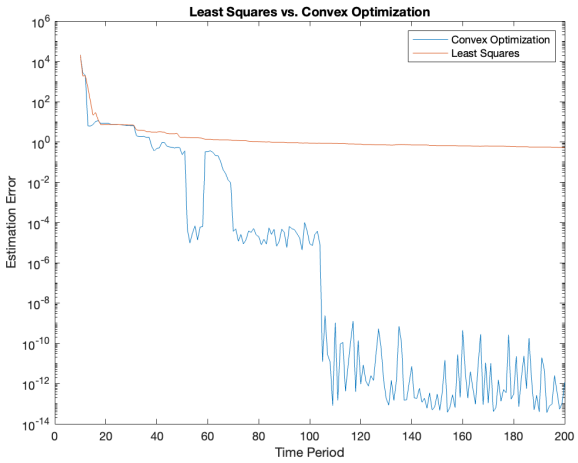

Fig. 1. Estimation Error Comparison for Noiseless Case with $\bar{A} \in \mathbb{R}^{10\times 10}$ with Time Horizon $T = 200$

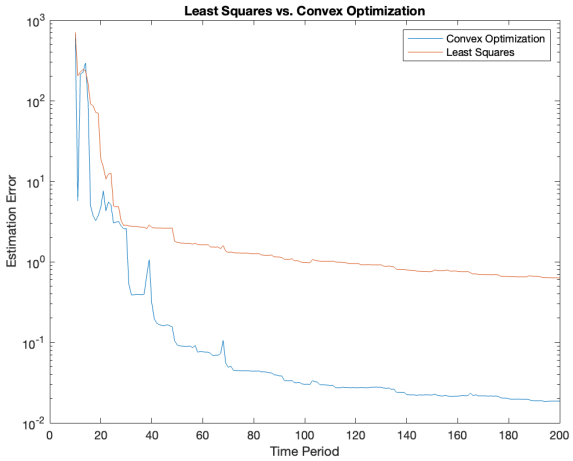

Fig. 2. Estimation Error Comparison for Noisy Case with $\bar{A} \in \mathbb{R}^{10\times 10}$ with Time Horizon $T = 200$

IX. CONCLUSION

We studied an l_1-based identification scheme for a fully observable LTI system affected by sparse state disturbances. We find that as long as the attack is not too frequent, even assuming that the attack can take the form of a linear state

and input feedback, an accurate state-space representation can be obtained. We derive some inequalities in the form of the null space property serving as conditions for the exact recovery of the model and develop a bound on the estimation error. It is intriguing to study when consistency and error bounds hold for other models of attack.

REFERENCES

[1] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "Regret bounds for robust adaptive control of the linear quadratic regulator," *arXiv preprint arXiv:1805.09388*, 2018.

[2] S. Fattahi, N. Matni, and S. Sojoudi, "Efficient learning of distributed linear-quadratic control policies," *SIAM Journal on Control and Optimization*, vol. 58, no. 5, pp. 2927–2951, 2020.

[3] J. F. Fisac, A. K. Akametalu, M. N. Zeilinger, S. Kaynama, J. Gillula, and C. J. Tomlin, "A General Safety Framework for Learning-Based Control in Uncertain Robotic Systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 7, pp. 2737–2752, Jul. 2019.

[4] H. Feng and J. Lavaei, "Learning of dynamical systems under adversarial attacks," in *2021 60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 3010–3017.

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

[6] J. W. Tukey, "The Future of Data Analysis," *The Annals of Mathematical Statistics*, vol. 33, no. 1, pp. 1–67, 1962.

[7] H. Xu, C. Caramanis, and S. Mannor, "Robustness and Regularization of Support Vector Machines." *Journal of machine learning research*, vol. 10, no. 7, 2009.

[8] D. Bertsimas and M. S. Copenhaver, "Characterization of the equivalence of robustification and regularization in linear and matrix regression," *European Journal of Operational Research*, vol. 270, no. 3, pp. 931–942, Nov. 2018.

[9] L. Bako, "On a Class of Optimization-Based Robust Estimators," *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 5990–5997, Nov. 2017.

[10] L. Bako and H. Ohlsson, "Analysis of a nonsmooth optimization approach to robust estimation," *Automatica*, vol. 66, pp. 132–145, Apr. 2016.

[11] S. Fattahi and S. Sojoudi, "Data-Driven Sparse System Identification," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct. 2018, pp. 462–469.

[12] ——, "Sample complexity of sparse system identification problem," *accepted for publication in IEEE Transactions on Control of Network Systems*, 2021.

[13] I. Molybog, R. Madani, and J. Lavaei, "Conic Optimization for Robust Quadratic Regression: Deterministic Bounds and Statistical Analysis," in *2018 IEEE Conference on Decision and Control (CDC)*, Dec. 2018, pp. 841–848.

[14] S. Oymak and N. Ozay, "Non-asymptotic Identification of LTI Systems from a Single Trajectory," in *2019 American Control Conference (ACC)*, Jul. 2019, pp. 5655–5661.

[15] T. Sarkar and A. Rakhlin, "Near optimal finite time identification of arbitrary linear dynamical systems," in *International Conference on Machine Learning*, May 2019, pp. 5610–5618.

[16] P. Hespanhol and A. Aswani, "Statistical Consistency of Set-Membership Estimator for Linear Systems," *IEEE Control Systems Letters*, vol. 4, no. 3, pp. 668–673, Jul. 2020.

[17] M. Showkatbakhsh, P. Tabuada, and S. Diggavi, "System identification in the presence of adversarial outputs," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, Dec. 2016, pp. 7177–7182.

[18] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure Estimation and Control for Cyber-Physical Systems Under Adversarial Attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, Jun. 2014.

[19] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *2015 American Control Conference (ACC)*, Jul. 2015, pp. 2439–2444.

[20] L. Su and S. Shahrampour, "Finite-Time Guarantees for Byzantine-Resilient Distributed State Estimation With Noisy Measurements," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3758–3771, Sep. 2020.

[21] N. Gupta and N. H. Vaidya, "Fault-Tolerance in Distributed Optimization: The Case of Redundancy," in *Proceedings of the 39th Symposium on Principles of Distributed Computing*, ser. PODC '20. New York, NY, USA: Association for Computing Machinery, Jul. 2020, pp. 365–374.

[22] Y. She and A. B. Owen, "Outlier Detection Using Nonconvex Penalized Regression," *Journal of the American Statistical Association*, vol. 106, no. 494, pp. 626–639, Jun. 2011.

[23] C. Josz, Y. Ouyang, R. Y. Zhang, J. Lavaei, and S. Sojoudi, "A theory on the absence of spurious solutions for nonconvex and nonsmooth optimization," 2018. [Online]. Available: https://arxiv.org/abs/1805.08204

[24] I. Molybog, S. Sojoudi, and J. Lavaei, "Role of sparsity and structure in the optimization landscape of non-convex matrix sensing," *Mathematical Programming*, pp. 1–37, 2020.

[25] K. Bhatia, P. Jain, P. Kamalaruban, and P. Kar, "Consistent Robust Regression," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 2110–2119.

[26] I. Diakonikolas, G. Kamath, D. M. Kane, J. Li, J. Steinhardt, and A. Stewart, "Sever: A Robust Meta-Algorithm for Stochastic Optimization," *arXiv:1803.02815 [cs, stat]*, May 2019.

[27] A. Prasad, A. S. Suggala, S. Balakrishnan, and P. Ravikumar, "Robust estimation via robust gradient estimation," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 82, no. 3, pp. 601–627, 2020.

[28] J. Steinhardt, P. W. W. Koh, and P. S. Liang, "Certified Defenses for Data Poisoning Attacks," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 3517–3529.

[29] H. Feng, B. Yalcin, and J. Lavaei, "Learning of dynamical systems under adversarial attacks – null space property perspective," 2022. [Online]. Available: https://arxiv.org/abs/2210.01421

[30] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[31] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, ser. Applied and Numerical Harmonic Analysis. New York, NY: Springer, 2013.

[32] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 211–231, 2009.

[33] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge university press, 2012.

[34] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht, "Learning Without Mixing: Towards A Sharp Analysis of Linear System Identification," in *Conference On Learning Theory*. PMLR, Jul. 2018, pp. 439–473.