Learning of Dynamical Systems under Adversarial Attacks

Han Feng and Javad Lavaei

Industrial Engineering and Operations Research, University of California, Berkeley

***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 sufficient conditions for the recovery of the system matrices by solving a constrained lasso-type optimization problem. In the settings without control input or when the input is sub-Gaussian with a known matrix B, we characterize the type of disturbance that does not affect the estimation of the matrix A. We furthermore analyze the case when A and B are estimated simultaneously, and study how to design the input of the system to properly excite the system and make the identification possible in the presence of adversarial attacks. We introduced the key notion of Δ-spaced disturbance and element-wise identifiability to study the success of the constrained lasso estimator. The superiority of the proposed technique is demonstrated in numerical experiments.**

I. Introduction

The control of large-scale unknown dynamical systems, such as the power distribution networks, calls for an accurate model of the system. Recent interests in data-driven control and non-asymptotic analysis of statistical estimators provide a wealth of frameworks and tools applicable to the control of unknown dynamical systems [1], [2]. Although learning an accurate dynamical model is not necessary to achieve the control objectives, as recent exciting progress in deep reinforcement learning shows [3], a state-space model has the advantage of being applicable to many control tasks and objectives. The issue is particularly salient in the operation of safety-critical systems, where a robust design of control laws is necessary [4], [5].

This paper focuses on the identification of a linear dynamical system where the states can be perfectly measured but are subject to unknown disturbance, accounting for adversarial attacks or faults. We prove that a type of identification scheme based on constrained lasso can perfectly recover the system matrices when the state disturbance is sparse. The issue of robustness in identification has a long history. Dating back to Tukey [6] which made the observation that a small deviation from the model assumption could have dramatic effects on estimation and prediction, there have since been many attempts to robustify the M-estimators and to use regularization to achieve robustness. The work [7] showed the equivalence of robust optimization and l_1-regularization for support vector machines and further attributed generalization ability to robustness against local disturbance. The more recent study [8] significantly extended the connection between robustification and regularization in regression problems. However, there is a lack of study in the problem of system identification under disturbance of arbitrary magnitudes.

Email: han_feng@berkeley.edu, lavaei@berkeley.edu
This work was supported by grants from ARO, ONR, AFOSR, and NSF.

A. Related Works

To situate the paper in the broader context, we discuss related works on robust regression and system identification. The paper [9] studied the related problem of outlier detection in linear regression. It proved the equivalence of adding a penalty to the least-squares loss function and using an alternative loss function to the least-squares loss. In particular, it noted that l_1 regularization is equivalent to using the Huber loss and that Huber loss may not be the best choice for guaranteeing robustness in many cases — a non-convex loss function may be more appropriate. However, unless in very specialized settings, the theoretical justifications of non-convex estimators are rare, and the computation of non-convex estimators is not well-understood [10], [11].

The work [12] solved the problem of regression with sparse disturbance via iterative hard thresholding. There has been a flurry of recent papers on robust training [13]–[15]. Nevertheless, the independence assumption between samples renders them inapplicable to system identification — the state measurements are dependent and cannot be re-ordered. Transforming the data samples to deal with missing data in linear regression does not directly translate to the system identification case due to the need to measure several trajectories or solve nonlinear optimization problems. It is undesirable to reset the system in practical applications. Furthermore, it is unclear how identification can be achieved robustly in an online fashion.

The papers [16] and [17] studied the system identification problem subject to a sparsity assumption on the A and B matrices and derived improved sample complexity bounds. However, their models were based on Gaussian disturbance that is not applicable to adversarial analysis. The recent work [18] studied the identification problem using a conic relaxation, which linearizes the problem at the expense of increasing the problem dimension. More recently, [19] proved finite-time identification bounds for linear dynamical systems without control input. The identification method is based on ordinary least-squares, which succeeds under the important assumption of regular matrices. Concurrently, [20] proved non-asymptotic bounds for system identification with Markov parameters, which are estimated using least-squares and the Kalman-Ho algorithm. It is not obvious how to generalize those algorithms to the case when the samples are missing or when they are corrupted. The set-membership estimator can deal with missing samples and is consistent [21], but the disturbance is assumed to be bounded. The existing literature lacks a fundamental study of the system identification problem in an adversarial setting. This will be the focus of the current paper.

B. Contribution

The paper provides consistency guarantees for a constrained lasso estimator when the system is subject to sparse state disturbances. After formulating the problem in Section II, we introduce the key notion of Δ-spaced disturbance. Section III is devoted to the study of the system identification problem without control input. Section IV studies the case with control inputs. Both sections make extensive use of our notion of element-wise identifiability. The problem of designing the input to assist with the system identification problem is discussed in Section V. Section VI illustrates the results with numerical simulations. Section VII makes concluding remarks.

II. Problem Formulation

Consider the linear time-invariant dynamical system over the time horizon $[0, T]$:

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

where $\bar{A} \in \mathbb{R}^{n\times n}, \bar{B} \in \mathbb{R}^{n\times m}$ are unknown matrices in the state space model to be estimated and $\bar{b}_t$'s are unknown disturbances. 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 $\bar{b}_0, ..., \bar{t}_{T_1}$ model anomalies in the system, such as attacks on the input data 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 $\bar{b}_t$. We will make the assumption that the disturbance signal is *sparse*, meaning that only a small subset of the vectors $\bar{b}_0, ..., \bar{b}_{T-1}$ are possibly non-zero. This is a common model for stealth attacks. The locations of non-zero disturbance vectors are not known and need to be inferred from the states $x_0, \ldots, x_T$ and control inputs $u_0, \ldots, u_{T-1}$. We introduce the notion of disturbance sparsity below.

Definition 1: Given a nonnegative integer Δ, the disturbance sequence $\{\bar{b}_i\}_{i=0}^{T-1}$ is said to be Δ-spaced if for every integer $i \in \{0, \ldots, T-\Delta-1\}$ such that $\bar{b}_i \neq 0$, we have $\bar{b}_j = 0$, for all $j \in \{i+1, \ldots, i+\Delta\}$.

III. The Case Without Control Input

We first study the case without control input for three reasons. First, we do not need to distinguish the input $\bar{B}u_t$ from the disturbance $\bar{b}_t$, making it possible to analyze only the effect of sparse disturbance. Second, any estimation techniques for the no-input case can be adapted to solve the case with *sparse* input. More precisely, one can define $\tilde{b}_t$ as $\bar{B}u_t + \bar{b}_t$, and then find $(\bar{A}, \bar{B})$ in three steps: (i) identify $\bar{A}$ from the measurement equations $x_{t+1} = \bar{A}x_t + \tilde{b}_t$, (ii) obtain the new disturbances from the equation $\tilde{b}_t = x_{t+1} - \bar{A}x_t$, (iii) solve a regression problem for the model $\tilde{b}_t = \bar{B}u_t + \bar{b}_t$ to find $\bar{B}$. Finally, the study of the no-input case provides insights into how the lasso estimator, which is widely used for rejecting outliers in machine learning with uncorrelated data, would perform on dynamical systems for which there is correlation over time.

Consider the following lasso-type estimator

$$\begin{aligned} \min_{A,b} \quad & \sum_{i=0}^{T-1} \|b_i\|_2 \\ s.t \quad & x_{i+1} = Ax_i + b_i, \quad i = 0, \ldots, T-1, \end{aligned} \tag{1}$$

where the measurements $x_0, \ldots, x_T$ are generated according to the ground truth

$$x_{i+1} = \bar{A}x_i + \bar{b}_i.$$

We use

$$K = \{i \mid \bar{b}_i \neq 0, i \in \{0, 1, ..., T-1\}\}$$

to denote the time instances of non-zero disturbance vectors. For clarity, when summing over the indices, we use the shorthand notation $\sum_{i\notin K}$ instead of $\sum_{0\le i\le T-1, i\notin K}$. In what follows, we develop conditions for the perfect identification of the system matrices. We will first study the one-dimensional case, where we derive sufficient conditions for the uniqueness of the Lasso solution. We will then generalize the results to systems of arbitrary dimensions. Throughout the paper, we use $\text{sgn}(x)$ to denote the sub-differential of the 2-norm $\|x\|_2$ and use $\langle\cdot,\cdot\rangle$ to denote the inner product of two vectors. The notation $(x)_j$ extracts the j-th entry of a vector x. For a real number z, we use $|z|$ to denote its absolute value.

A. One-dimensional Case

We study the Lasso-typo estimator (1) below.

Theorem 1: Consider the convex optimization problem (1) and assume that $n = 1$. It holds that

- If $\sum_{i\notin K}|x_i| \geq \left|\sum_{i\in K}\langle x_i, \text{sgn}(\bar{b}_i)\rangle\right|$, then $\bar{A}$ is a solution to (1).
- If $\sum_{i\notin K}|x_i| > \sum_{i\in K}|x_i|$, then $\bar{A}$ is the unique solution.

Proof: The first-order necessary condition states that

$$\begin{aligned} &\lambda_i \in \text{sgn}(b_i), \quad i = 0, 1, \ldots, T-1, \\ &\sum_{i=0}^{T-1} x_i\lambda_i^T = 0, \\ &x_{i+1} - Ax_i - b_i = 0, \quad i = 0, \ldots, T-1. \end{aligned}$$

Since $n = 1$, the conditions are simplified to

$$0 \in \sum_i x_i \text{sgn}(x_{i+1} - Ax_i).$$

Note that the right-hand side of the above relation is a set. On the other hand,

$$x_i = \bar{A}^i x_0 + \sum_{k\in K} \bar{A}^{(i-1-k)+}\bar{b}_k, \quad i = 0, \ldots, T, \tag{2}$$

where

$$A^{(i)_+} = \begin{cases} 0, & \text{if } i < 0, \\ 1, & \text{if } i = 0, \\ A^i, & \text{if } i > 0. \end{cases}$$

The first-order condition can be simplified to

$$0 \in \sum_{i=0}^{T-1} \Big\langle \bar{A}^i x_0 + \sum_{k \in K} \bar{A}^{(i-1-k)_+} \bar{b}_k, \ \mathrm{sgn}((\bar{A} - A)\bar{A}^i x_0 + \sum_{k \in K} (\bar{A}^{(i-k)_+} - A\bar{A}^{(i-1-k)_+})\bar{b}_k) \Big\rangle,$$

which is equivalent to

$$0 \in \sum_{i=0}^{T-1} \Big\langle \bar{A}^i x_0 + \sum_{k \in K} \bar{A}^{(i-1-k)_+} \bar{b}_k, \mathrm{sgn}((\bar{A} - A)(\bar{A}^i x_0 + \sum_{k \in K} \bar{A}^{(i-1-k)_+} \bar{b}_k) + \sum_{k \in K} (\bar{A}^{(i-k)_+} - \bar{A}\bar{A}^{(i-1-k)_+})\bar{b}_k) \Big\rangle.$$

By substituting back the expression of x_i together with the observations $x_i \,\mathrm{sgn}(ax_i) = |x_i| \,\mathrm{sgn}(a)$ and $\sum_{k \in K} (\bar{A}^{(i-k)_+} - \bar{A}\bar{A}^{(i-1-k)_+})\bar{b}_k = \bar{b}_i$ for all $i \in K$, the first-order necessary condition can be reduced to

$$0 \in \sum_{\substack{0 \le i \le T-1 \\ i \notin K}} |x_i| \,\mathrm{sgn}(\bar{A} - A) + \sum_{i \in K} \langle x_i, \mathrm{sgn}((\bar{A} - A)x_i + \bar{b}_i) \rangle.$$

The proof of the theorem is completed by noting that

- If a matrix $A_* \neq \bar{A}$ is a solution, then

$$\sum_{i \notin K} |x_i| = \Big| \sum_{i \in K} \langle x_i, \mathrm{sgn}((\bar{A} - A_*)x_i + \bar{b}_i) \Big| \le \sum_{i \in K} |x_i|.$$

- $\bar{A}$ is a solution if and only if

$$\sum_{i \notin K} |x_i| \ge \Big| \sum_{i \in K} \langle x_i, \mathrm{sgn}(\bar{b}_i) \rangle \Big|.$$

■

Remark 1: The conditions in Theorem 1 show that the absolute magnitude of individual disturbances does not directly affect perfect recovery, as long as the relative magnitude of states is well-controlled. Furthermore, if there is a non-zero disturbance at the end of the horizon, namely $\bar{b}_{T-1} \neq 0$, it may cause the first condition of Theorem 1 to be violated, and the system identification will fail.

It is desirable to understand what types of systems satisfy the conditions of Theorem 1. We will show that these conditions are satisfied in at least two scenarios. Define

$$s(a, k) = \sum_{i=0}^{k-1} a^i = \begin{cases} \frac{1-a^k}{1-a}, & \text{if } a \neq 1 \\ ka, & \text{if } a = 1. \end{cases}$$

Proposition 1: For $n = 1$, if the disturbance sequence satisfies

$$\sum_{i \notin K} r^i |x_0| - \sum_{i \in K} r^i |x_0| > \sum_{k \in K} s(r, T-k-1) |\bar{b}_k|, \quad (3)$$

then $\bar{A}$ is the unique solution to the optimization problem (1).

Proof: It suffices to show that the condition in Theorem 1 is satisfied. From (2), we have

$$\sum_{i \in K} |x_i| \le \sum_{i \in K} |\bar{A}^i x_0| + \sum_{i \in K, k \in K} \Big|\bar{A}^{(i-1-k)_+}\Big| |\bar{b}_k| \le \sum_{i \in K} |\bar{A}^i x_0| + \sum_{i \in K, k \in K} r^{(i-1-k)_+} |\bar{b}_k|.$$

Moreover,

$$\sum_{i \notin K} |x_i| \ge \sum_{i \notin K} |\bar{A}^i x_0| - \sum_{i \notin K, k \in K} \Big|\bar{A}^{(i-1-k)_+}\Big| |\bar{b}_k| \ge \sum_{i \notin K} |\bar{A}^i x_0| - \sum_{i \notin K, k \in K} r^{(i-1-k)_+} |\bar{b}_k|.$$

As a result, the condition in Theorem (1) is satisfied if

$$\sum_{i \notin K} |\bar{A}^i x_0| - \sum_{i \in K} |\bar{A}^i x_0| \ge \sum_{i=0}^{T-1} \sum_{k \in K} r^{(i-1-k)_+} |\bar{b}_k|.$$

This inequality is equivalent to (3). ■

Proposition 1 implies that if the disturbances are small, then the system identification via a Lasso-typo estimator is successful. Consider now the opposite case where the disturbances are Δ-spaced and large enough to drive the system.

Proposition 2: Assume that the disturbance sequence is Δ-spaced. If

$$\sum_{i \in K} |\bar{b}_i| > \frac{s(r, \Delta+1)}{s(r, \Delta)} \sum_{i \in K} |x_i|,$$

where $r = |\bar{A}|$, then $\bar{A}$ is the unique solution to the optimization problem (1).

Proof: Consider a time $i \in K$ such that $\bar{b}_i \neq 0$. Then, $\bar{b}_j = 0$ for all $j \in \{i+1, ..., i+\Delta\}$. As a result,

$$x_j = \bar{A}^{j-i} x_i + \bar{A}^{(j-1-i)_+} \bar{b}_i.$$

One can write

$$\sum_{j \notin K} |x_j| \ge \sum_{i \in K} \sum_{i+1 \le j \le i+\Delta} |x_j| \ge \sum_{i \in K} \sum_{i+1 \le j \le i+\Delta} \Big| \big|\bar{A}^{(j-1-i)_+} \bar{b}_i\big| - \big|\bar{A}^{j-i} x_i\big| \Big|.$$

Hence, the condition in Theorem 1 holds if

$$\sum_{i \in K} \sum_{i+1 \le j \le i+\Delta} \Big|\bar{A}^{(j-1-i)_+} \bar{b}_i\Big| - \big|\bar{A}^{j-i} x_i\big| > \sum_{i \in K} |x_i|.$$

After rewriting it with $r = |\bar{A}|$, we obtain the condition in the proposition. ■

B. High-dimensional Case

In this part, we generalize the results of the previous section to systems of arbitrary dimensions. We use the notation $a \otimes b = ab^T$.

Theorem 2: The following statements hold:

- If there exist vectors $e_i,$ for all $i \notin K$, of length at most 1 such that
$$\sum_{i \notin K} x_i \otimes e_i = \sum_{i \in K} x_i \otimes b_i / \|b_i\|, \tag{4}$$
then $\bar{A}$ is a solution to the optimization problem (1).
- If $\bar{A}$ is not the unique solution to (1), then there exits a non-zero matrix E such that
$$\sum_{i \notin K} x_i \otimes \text{sgn}(Ex_i) + \sum_{i \in K} x_i \otimes \text{sgn}(Ex_i + \bar{b}_i) \ni 0.$$

Proof: Similar to the one-dimensional case, the first-order necessary condition becomes
$$0 \in \sum_{i \notin K} x_i \otimes \text{sgn}((\bar{A} - A)x_i) + \sum_{i \in K} x_i \otimes \text{sgn}((\bar{A} - A)x_i + \bar{b}_i).$$
Let A_* be a solution of (1). If $A_* \neq \bar{A}$, then we set $E = \bar{A} - A_*$. The rest of the proof closely follows the proof of Theorem 1. ∎

To understand what types of systems satisfy the conditions of Theorem 2. We introduce the notion of element-wise identifiability below.

Definition 2: Given $A \in \mathbb{R}^{n \times n}$, $y \in \mathbb{R}^n$, and $z \in \mathbb{R}^n$, the triplet (A, y, z) is said to be Δ-spaced element-wise identifiable if either $z = 0$ or
$$y \in \left\{ \sum_{i=1}^{\Delta} g_i (A^i y + A^{i-1} z) \;\middle|\; -1 \le g_i \le 1 \text{ for } 1 \le i \le \Delta \right\}. \tag{5}$$

Theorem 3: Assume that $\{\bar{b}_k\}_{k=0}^{T-1}$ is Δ-spaced and that the triplet $(\bar{A},\, x_k,\, \bar{b}_k)$ is Δ-spaced element-wise identifiable for $k \in \{0, 1, \ldots, T-1\}$. Then, $\bar{A}$ is a solution to the optimization problem (1).

Proof: Consider an index $k \in K$ and, without loss of generality, suppose that $\|\bar{b}_k\| = 1$ in equation (4) of Theorem 2. The assumption of Δ-spaced element-wise identifiability implies the following relation:
$$x_k \in \left\{ \sum_{i=1}^{\Delta} g_i (A^i x_{k+i} + A^{i-1} \bar{b}_k) \;\middle|\; -1 \le g_i \le 1 \text{ for } 1 \le i \le \Delta \right\}.$$
For any $j \in \{1, \ldots, n\}$, the relation implies that the vector $\bar{b}_{kj} x_k$ can be expressed as a linear combination $e_{(k+1)j} x_{k+1} + \cdots + e_{(k+\Delta)j} x_{k+\Delta}$, where the real number $\bar{b}_{kj}$ denotes the j-th entry of $\bar{b}_k$ and the real numbers e_{ij} satisfy $|e_{ij}| \le |\bar{b}_{kj}|$ for all $i \in [k+1, k+\Delta]$. As a result, $\sum_{i=1}^{\Delta} x_{k+i} \otimes e_{k+i} = x_k \otimes \bar{b}_k$, where $\|e_i\|^2 \le \sum_{j=1}^{\Delta} e_{ij}^2 \le \sum_{j=1}^{\Delta} \bar{b}_{kj}^2 = \|\bar{b}_k\|^2 \le 1$. Applying the argument to all $k \in K$ proves that the condition (4) of Theorem 2 is satisfied. ∎

The proof of Theorem 3 shows that element-wise identifiability is stronger than the condition (4) of Theorem 2. The merit of this concept lies in the fact that the satisfaction of Δ-space element-wise identifiability can be captured by the spectrum of $\bar{A}$, as described below.

Theorem 4: Let $\bar{A} = P^{-1} \Lambda P$ be an eigen-decomposition of $\bar{A}$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal real matrix. Given $k \in K$, the triplet $(\bar{A},\, x_k,\, b_k)$ is Δ-spaced element-wise identifiable if
$$|\lambda_j| s(|\lambda_j|, \Delta) \ge \left| \frac{(Px_k)_j}{(P(\lambda_j x_k + b_k))_j} \right| \quad \forall j \in \{1, 2, \ldots, n\}.$$

Proof: Using the eigen-decomposition, we can rewrite condition (5) as
$$Px_k \in \left\{ \sum_{i=1}^{\Delta} g_i \Lambda^{i-1} (\Lambda P x_k + P b_k) \;\middle|\; -1 \le g_i \le 1, i \in [1, \Delta] \right\}.$$
The diagonalizability assumption allows us to rewrite the condition (5) as
$$\frac{(Px_k)_j}{(P(\lambda_j x_k + b_k))_j} \in \left\{ \sum_{i=1}^{\Delta} g_i \lambda_j^i : -1 \le g_i \le 1 \right\},\ \forall j \in [1, n]. \tag{6}$$
The set on the right-hand side of (6) is a convex set, and its boundary points are obtained by setting $g_i = \text{sgn}(\lambda_j^i)$ or $-\text{sgn}(\lambda_j^i),$ for $i \in \{1, \ldots, \Delta\}$. The proof is completed by noting that (6) is equivalent to
$$-|\lambda_j| s(|\lambda_j|, \Delta) \le \frac{(Px_k)_j}{(P(\lambda_j x_k + b_k))_j} \le |\lambda_j| s(|\lambda_j|, \Delta).$$
∎

Remark 2: Theorem 4 states that if the disturbance does not nullify the state at the time of disturbance, then identifiability is met.

IV. The Case with Control Input

In this section, we broaden the analysis to include the control input in the identification problem. In particular, we aim to understand how to design the input of the system (in case that is an option) so that the identification of the excited system in the presence of adversarial disturbances is possible. Consider the constrained optimization problem:
$$\min_{A,B,b} \sum_{i=0}^{T-1} \|b_i\|_2 \tag{7}$$
$$s.t \quad x_{i+1} = Ax_i + Bu_i + b_i, \quad i = 0, \ldots, T-1,$$
where the data are generated according to
$$x_{i+1} = \bar{A} x_i + \bar{B} u_i + \bar{b}_i, \quad i = 0, \ldots, T-1. \tag{8}$$
We first address the case where $\bar{B}$ is known, for which the generalization of element-wise identifiability is straightforward.

A. Known Matrix $\bar{B}$

We first derive the first-order optimality conditions.

Theorem 5: Consider the convex optimization problem (7) after fixing the parameter B at the known matrix $\bar{B}$. The following statements hold:

- If there exist vectors $e_i,$ for all $i \notin K$, of length at most 1 such that
$$\sum_{i \notin K} x_i \otimes e_i = \sum_{i \in K} x_i \otimes \bar{b}_i / \|\bar{b}_i\|, \tag{9}$$
then $\bar{A}$ is a solution to the optimization problem (7).
- If $\bar{A}$ is not the unique solution, then there exits a non-zero matrix E such that
$$\sum_{i \notin K} x_i \otimes \text{sgn}(Ex_i) + \sum_{i \in K} x_i \otimes \text{sgn}(Ex_i + \bar{b}_i) \ni 0.$$

Proof: The first-order necessary condition states that

$$\operatorname{sgn}(b_i) \ni \lambda_i, \quad i = 0, 1, \ldots, T-1,$$
$$\sum_{i=0}^{T-1} x_i \lambda_i^T = 0,$$
$$x_{i+1} - Ax_i - \bar{B}u_i - b_i = 0, \quad i = 0, \ldots, T-1,$$

which is simplified to

$$\sum_i x_i \otimes \operatorname{sgn}(x_{i+1} - Ax_i - \bar{B}u_i) \ni 0.$$

Using $x_{i+1} = \bar{A}x_i + \bar{B}u_i + \bar{b}_i$, the first-order condition can be written as

$$\sum_{i \notin K} x_i \otimes \operatorname{sgn}((\bar{A} - A)x_i) + \sum_{i \in K} x_i \otimes \operatorname{sgn}((\bar{A} - A)x_i + \bar{b}_i) \ni 0$$

The two conditions of the theorem follow from the examination of the above equation. ∎

Now, we study the satisfaction of the first condition of Theorem 5 via the notion of Δ-spaced disturbance.

Definition 3: Given $\bar{A} \in \mathbb{R}^{n \times n}$, $\bar{B} \in \mathbb{R}^{n \times m}$, $y \in \mathbb{R}^n$, and $z \in \mathbb{R}^n$, the quadruplet $(\bar{A}, \bar{B}, y, z)$ is said to be Δ-spaced element-wise identifiable if either $z = 0$ or there exist vectors $w_0, \ldots, w_{\Delta-1} \in \mathbb{R}^m$ such that

$$y \in \left\{ \sum_{i=1}^{\Delta} g_i \left[\bar{A}^i y + \bar{A}^{i-1} z + \sum_{0 \le j < i} \bar{A}^{i-j} \bar{B} w_j \right] : g_i \in [-1, 1] \right\}. \tag{10}$$

The sequence inputs $w_1, \ldots, w_\Delta$ that makes (10) hold is said to be *adaptive* to $(\bar{A}, \bar{B}, y, z)$

Theorem 6: Consider the convex optimization problem (7) after fixing the parameter B at the known matrix $\bar{B}$. Assume that $\{\bar{b}_k\}_{k=0}^{T-1}$ is Δ-spaced. If for all $k \in \{0, \ldots, T - \Delta - 1\}$, the quadruplet ($\bar{A}$, $\bar{B}$, x_k, $\bar{b}_k$) is Δ-spaced element-wise identifiable and the sequence of inputs $(u_k, \ldots, u_{k+\Delta-1})$ is adaptive to ($\bar{A}$, $\bar{B}$, x_k, $\bar{b}_k$) in the sense of (10), then $\bar{A}$ is a solution to (7).

Proof: The assumption implies that the sequence of inputs causes the system states to satisfy

$$x_k \in \left\{ \sum_{i=1}^{\Delta} g_i x_{k+i} : -1 \le g_i \le 1, \forall i \in \{1, \ldots, \Delta\} \right\}, \forall k \in K.$$

In particular, for any $k \in K$, we can select the vectors $e_{k+1}, \ldots, e_{k+\Delta}$ from the same procedure of Theorem 3 to achieve the equality

$$\sum_{i=1}^{\Delta} x_{k+i} \otimes e_{k+i} = x_k \otimes b_k / \|b_k\|_2, \quad \forall k \in K.$$

Because the disturbance sequence is Δ-spaced, we can piece together the vectors e_i, for all $i \notin K$, from the above construction so that (9) is satisfied. ∎

As before, the merit of element-wise identifiability lies in the fact that it is easily verifiable and guides the design of the input to enable the identification of the system.

Theorem 7: Suppose that $\bar{A} = P^{-1} \Lambda P$ is the eigen-decomposition of $\bar{A}$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is a real diagonal matrix. For any $k \in K$, let all entries of the input vectors $u_k, \ldots, u_{k+\Delta-1}$ be independent and identically distributed (i.i.d.) sub-Gaussian random variables with parameter σ^2. Then, the inputs $(u_k, \ldots, u_{k+\Delta-1})$ are adaptive to (A, B, x_k, b_k) with probability at least

$$1 - \sum_{l=1}^{n} \exp\left(- \frac{\left(|(Px_k)_l| + \sum_{i=1}^{\Delta} \left| \lambda_l^{i-1} (P(\lambda_l x_k + b_k))_l \right| \right)^2}{2\sigma^2 \sum_{q=1}^{m} (P\bar{B})_{lq}^2 \sum_{j=0}^{\Delta-1} |\lambda_l|^2 s(|\lambda_l|, \Delta - j)^2} \right).$$

Proof: We first rewrite the definition of element-wise identification in the spirit of Theorem 4. Since $b_k \neq 0$, the condition (10) can be written as

$$\sum_{i=k+1}^{k+\Delta} \left| \lambda_l^{i-k-1} (P(\lambda_l x_k + b_k))_l + \sum_{k \le j < k+i} \lambda_l^{i-j} (P\bar{B}u_j)_l \right| \ge |(Px_k)_l|,$$

for all indices $l \in \{1, \ldots, n\}$. The inequality can be strengthened to

$$\begin{aligned} &\sum_{j=0}^{\Delta-1} \left| (P\bar{B}u_{k+j})_l \right| \sum_{j < i \le \Delta} \left| \lambda_l^{i-j} \right| \\ &\quad = \sum_{i=1}^{\Delta} \sum_{0 \le j < i} \left| \lambda_l^{i-j} \right| \left| (P\bar{B}u_{k+j})_l \right| \\ &\quad \ge |(Px_k)_l| + \sum_{i=1}^{\Delta} \left| \lambda_l^{i-1} (P(\lambda_l x_k + b_k))_l \right|, \end{aligned}$$

for all indices $l \in \{1, \ldots, n\}$. The left-hand side is sub-Gaussian with parameter $\sigma^2 \sum_{j=0}^{\Delta-1} \sum_{q=1}^{m} (P\bar{B})_{lq}^2 |\lambda_l|^2 s(|\lambda_l|, \Delta - j)^2$. To complete the proof, we use the tail bounds for sub-Gaussian random variable for every $l \in \{1, \ldots, n\}$ and use the union bound. ∎

Remark 3: In the case when $\bar{B}$ is known, the bound in Theorem 7 shows that, as long as the disturbance-state pair is such that the numerator is non-zero and that the system matrix $\bar{A}$ has no zero mode, then a sub-Gaussian random input with small variance σ can achieve perfect identification of $\bar{A}$ with high probability.

B. Unknown Matrix $\bar{B}$

We now study the challenging case where $\bar{A}$ and $\bar{B}$ are both unknown.

Theorem 8: Consider the optimization problem (7). The following statements hold:

- If there exist vectors e_i, for all $i \notin K$, of length at most 1 such that

$$\sum_{i \notin K} x_i \otimes e_i = \sum_{i \in K} x_i \otimes b_i / \|b_i\|,$$
$$\sum_{i \notin K} u_i \otimes e_i = \sum_{i \in K} u_i \otimes b_i / \|b_i\|,$$

  then ($\bar{A}$, $\bar{B}$) is a solution to (7).
- If the optimization problem (7) has a solution pair (A_*, B_*) that is not equal to $(\bar{A}, \bar{B})$, then there exist

matrices E and F that are not both zero such that

$$\sum_{i\notin K} x_i \otimes \operatorname{sgn}(Ex_i + Fu_i) + \sum_{i\in K} x_i \otimes \operatorname{sgn}(Ex_i + Fu_i + \bar{b}_i) \ni 0$$
$$\sum_{i\notin K} u_i \otimes \operatorname{sgn}(Ex_i + Fu_i) + \sum_{i\in K} u_i \otimes \operatorname{sgn}(Ex_i + Fu_i + \bar{b}_i) \ni 0.$$

Proof: The first-order necessary condition states that

$$\operatorname{sgn}(b_i) \ni \lambda_i, \quad i = 0, 1, \ldots, T-1,$$
$$\sum_{i=0}^{T-1} x_i \lambda_i^T = 0, \sum_{i=0}^{T-1} u_i \lambda_i^T = 0,$$
$$x_{i+1} - Ax_i - Bu_i - b_i = 0, \quad i = 0, \ldots, T-1.$$

We again simplify the conditions to arrive at

$$\sum_{i} x_i \otimes \operatorname{sgn}(x_{i+1} - Ax_i - Bu_i) \ni 0,$$
$$\sum_{i} u_i \otimes \operatorname{sgn}(x_{i+1} - Ax_i - Bu_i) \ni 0,$$

which can be combined with the dynamics of the system to obtain

$$\sum_{i\notin K} x_i \otimes \operatorname{sgn}((\bar{A} - A)x_i + (\bar{B} - B)u_i) + \sum_{i\in K} x_i \otimes \operatorname{sgn}((\bar{A} - A)x_i + (\bar{B} - B)u_i + \bar{b}_i) \ni 0,$$
$$\sum_{i\notin K} u_i \otimes \operatorname{sgn}((\bar{A} - A)x_i + (\bar{B} - B)u_i) + \sum_{i\in K} u_i \otimes \operatorname{sgn}((\bar{A} - A)x_i + (\bar{B} - B)u_i + \bar{b}_i) \ni 0.$$

The proof follows from the above equations. ■

Definition 4: The quadruplet $(\bar{A}, \bar{B}, y, z)$ is said to be Δ-spaced element-wise identifiable if either $z = 0$ or there exists input $w_j, j \in \{0, 1, \ldots, \Delta - 1\}$, such that

$$\begin{bmatrix} y \\ z \end{bmatrix} \in \left\{ \sum_{i=1}^{\Delta} g_i \begin{bmatrix} \bar{A}^i y + \bar{A}^{i-1} z + \sum_{0\le j<i} \bar{A}^{i-j} \bar{B} w_j \\ w_i \end{bmatrix}, \text{ where } g_i \in [-1, 1], \forall i \in \{1, \ldots, \Delta\} \right\}. \tag{11}$$

The sequence inputs $w_1, \ldots, w_\Delta$ that make (11) hold is said to be *adaptive* to $(\bar{A}, \bar{B}, y, z)$

Theorem 9: Assume that $\{\bar{b}_k\}_{k=0}^{T-1}$ is Δ-spaced. If for all $k \in \{0, ..., T - \Delta - 1\}$, the quadruplet $(\bar{A}, \bar{B}, x_k, \bar{b}_k)$ is Δ-spaced element-wise identifiable and the sequence of inputs $(u_k, \ldots, u_{k+\Delta-1})$ is adaptive to $(\bar{A}, \bar{B}, x_k, \bar{b}_k)$ in the sense of (11), then the pair $(\bar{A}, \bar{B})$ is a solution to (7).

Proof: The assumption implies that the sequence of inputs causes the system states to satisfy

$$\begin{bmatrix} x_k \\ u_k \end{bmatrix} \in \left\{ \sum_{i=1}^{\Delta} g_i \begin{bmatrix} x_{k+i} \\ u_{k+i} \end{bmatrix} : -1 \le g_i \le 1, \forall i \in \{1, \ldots, \Delta\} \right\},$$

for all $k \in K$. In particular, we can select the vectors $e_{k+1}, \ldots, e_{k+\Delta}$ from the same procedure as in Theorem 3 to achieve the equalities

$$\sum_{i=1}^{\Delta} x_{k+i} \otimes e_{k+i} = x_k \otimes b_k / \|b_k\|_2,$$
$$\sum_{i=1}^{\Delta} u_{k+i} \otimes e_{k+i} = u_k \otimes b_k / \|b_k\|_2,$$

for all $k \in K$. Because the disturbance sequence is Δ-spaced, we can piece together the vectors e_i, for all $i \notin K$, from the above construction so that the condition in Theorem 8 is satisfied. ■

V. The Problem of Input Design

In the case of simultaneous identification of $\bar{A}$ and $\bar{B}$, we require that the input, state and disturbance satisfy the sophisticated Δ-spaced element-wise identifiability condition. In what follows, we provide some insight into how to design the input to assist with the satisfaction of this condition. Let the input of the system be generated according to the dynamics

$$u_{i+1} = Fx_{i+1} + Kx_i + Du_i, \text{ for } i \in \{0, \ldots, T-1\}, \tag{12}$$

where u_0 is arbitrary and the matrices F, K and D are to be designed. We can write the augmented dynamics as

$$\begin{bmatrix} x_{i+1} \\ u_{i+1} \end{bmatrix} = \begin{bmatrix} \bar{A} + \bar{B}F & \bar{B} \\ K & D \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix} + \begin{bmatrix} \bar{b}_i \\ 0 \end{bmatrix}.$$

We write the above expression as $\tilde{x}_{i+1} = \tilde{A}\tilde{x}_i + \tilde{b}_i$, where

$$\tilde{A} = \begin{bmatrix} \bar{A} + \bar{B}F & \bar{B} \\ K & D \end{bmatrix}, \quad \tilde{b}_i = \begin{bmatrix} \bar{b}_i \\ 0 \end{bmatrix}, \tilde{x}_i = \begin{bmatrix} x_i \\ u_i \end{bmatrix}. \tag{13}$$

Note that whenever $\{\bar{b}_i\}_{i=0}^{T-1}$ is Δ-spaced, so is $\{\tilde{b}_i\}_{i=0}^{T-1}$. Therefore, we can use the identification formulation without input (1), replacing (A, b) with $(\tilde{A}, \tilde{b})$ in the problem, and recover the pair $(\tilde{A}, \tilde{b})$ exactly, even though we treat the known lower blocks of $\tilde{A}$ and $\tilde{b}$ as unknowns. Once $\tilde{A}$ and $\tilde{b}$ are recovered, the matrices $\bar{A}, \bar{B}$ can be found with the knowledge of F, K, and D. In summary, the general system identification problem with disturbance can be solved by using a perfect recovery theorem for the case without input and a suitable design of F, K, and D that satisfies the condition of perfect recovery. We illustrate one such design in the following theorem.

Theorem 10: Consider the problem of system identification for the dynamics (8) with the input design (12). Assume that the disturbance sequence is Δ-spaced. Then, we can perfectly recover the pair $(\bar{A}, \bar{B})$ from (13), where $\tilde{A}$ and $\tilde{b}_i$ are the solution to the optimization problem

$$\min_{\tilde{A}, \tilde{b}} \sum_{i=0}^{T-1} \left\| \tilde{b}_i \right\|_2$$
$$s.t \quad \begin{bmatrix} x_{i+1} \\ u_{i+1} \end{bmatrix} = \tilde{A} \begin{bmatrix} x_i \\ u_i \end{bmatrix} + \tilde{b}_i, \quad i = 0, \ldots, T-1$$

if the following conditions hold:

- The matrix $\tilde{A} = \begin{bmatrix} \bar{A} + \bar{B}F & \bar{B} \\ K & D \end{bmatrix}$ is diagonalizable with real eigenvalues;

- $\tilde{A} = \tilde{P}^{-1}\tilde{\Lambda}\tilde{P}$ is an eigen-decomposition of $\tilde{A}$, where $\tilde{\Lambda} = \mathrm{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{m+n})$ is a diagonal real matrix;
- The inequality

$$s(\left|\tilde{\lambda}_j\right|, \Delta) \geq \left| \frac{\left(\tilde{P}\begin{bmatrix} x_k \\ u_k \end{bmatrix}\right)_j}{\left(\tilde{P}\left(\tilde{\lambda}_j \begin{bmatrix} x_k \\ u_k \end{bmatrix} + \begin{bmatrix} \bar{b}_k \\ 0 \end{bmatrix}\right)\right)_j} \right|$$

holds for all k such that $\bar{b}_k \neq 0$ and for all $j \in \{1, 2, \ldots, m+n\}$.

Proof: After applying Theorem 4 to the augmented system $\tilde{x}_{i+1} = \tilde{A}\tilde{x}_i + \tilde{b}_i$, the condition of the theorem states that the extended system is Δ-spaced element-wise identifiable for all time k. Theorem 3 states that $\tilde{A}$ can be perfectly recovered. We can further recover $\bar{A}$ and $\bar{B}$ from (13). ∎

Remark 4: The theorem provides a precise characterization of the type of disturbance that the recovery procedure is robust. Specifically, assume that three properties are satisfied: (1) we can design the input so that the extended system has proper spectral properties, (2) no non-zero the disturbance b_k perfectly aligns with the corresponding state x_k, (3) all stable modes $\tilde{\lambda}_j$ of $\tilde{A}$ satisfy

$$\frac{1}{1 - \left|\tilde{\lambda}_j\right|} > \left| \frac{\left(\tilde{P}\begin{bmatrix} x_k \\ u_k \end{bmatrix}\right)_j}{\left(\tilde{P}\left(\tilde{\lambda}_j \begin{bmatrix} x_k \\ u_k \end{bmatrix} + \begin{bmatrix} \bar{b}_k \\ 0 \end{bmatrix}\right)\right)_j} \right|, \tag{14}$$

for all $k \in K$. Then, as long as the disturbance sequence is Δ-spaced with a long enough spacing Δ, we can perfectly identify the system. Any prior knowledge on $\bar{A}$ and $\bar{B}$ could help with the design of F, K, and D.

Remark 5: Theorem 10 can be extended to the case with complex eigenvalues at the expense of a more complicated characterization of element-wise identifiability.

VI. Numerical Experiments

This section provides numerical simulations to illustrate the efficiency of the identification approach. First, consider the autonomous case where $\bar{B} = 0$. Our baseline for comparison is the least-squares estimator

$$\min_{A} \sum_{t=0}^{T-1} \|x_{i+1} - Ax_i\|_2^2. \tag{15}$$

To obtain the system matrices, we consider the case $n = 5$. We use $N(0, \Sigma)$ to denote the multivariate Gaussian random variable with mean 0 and covariance Σ. We set the spectrum of A to be $\Gamma = \mathrm{diag}(0.9, 0.8, 0.7, 1.1, 0.1)$, and let $A = P\Gamma P^{-1}$, where P is a random matrix whose entries are normally distributed with mean 0 and variance 1. Let x_0 be normally distributed with mean 0 and variance 1. Let the disturbance b_t be non-zero 30% of the time. Moreover, for $t \in K$, let b_t follow the distribution $N(0, 10I_5)$, where I_5 is the 5-by-5 identity matrix. As the horizon T increases from 1 to 50, we compare the constrained Lasso estimator (1) and the least-squares estimator (15) in Figure 1. Due to the frequency and large magnitude of the disturbance, the least-squares estimator never converges to the true system matrix $\bar{A}$. In contrast, the lasso estimator quickly converges to the true system matrix, and after it converges, future disturbance has little effect on the estimation accuracy.

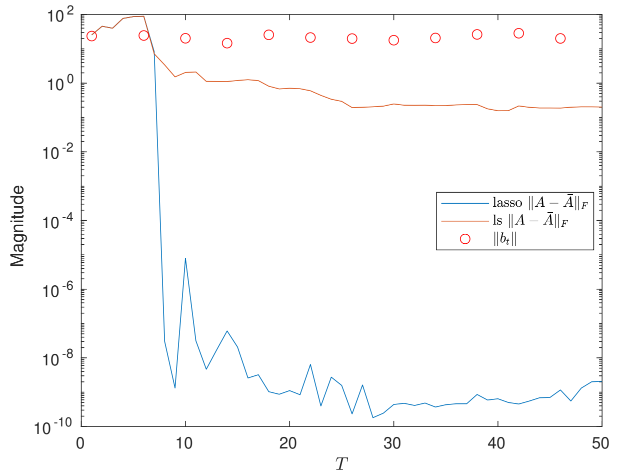

Fig. 1. Comparing the constrained lasso estimator (1) and the least-squares (ls) estimator (15). The circles plot the magnitude of the disturbance b_t when it is non-zero. The difference is measured in the Frobenius norm $\|\cdot\|_F$.

For the second example, we consider the Tennessee Eastman challenge problem. We obtain the A and B matrices from a discretization of the continuous-time LTI model in [22]. The discretization uses zero-order hold with the sampling period being 0.25h. Since the continuous-time model has a large separation between fast and slow modes, the discretized A matrix has four modes close to 0. The values of A and B are provided in (17) and (18). Our baseline for comparison is the least-squares estimator

$$\min_{A,B} \sum_{t=0}^{T-1} \|x_{i+1} - Ax_i - Bu_i\|_2^2. \tag{16}$$

Inspired by Theorem 7, the control inputs come from the distribution $N(0, I_4)$, and the initial state comes from $N(0, I_8)$. The disturbance is generated in the same fashion. Figure 2 shows that the constrained lasso estimator (7) vastly outperforms the least-squares estimator (16). Despite the fact that 30% of the states are disturbed, the identification of both A and B matrices is almost perfect.

VII. Conclusion

This paper studies the identification of linear systems under possible attacks appearing as disturbances to the dynamics. We develop the notion of Δ-spaced disturbance and element-wise identifiability. This leads to sufficient conditions for recovering the exact system dynamics in various scenarios. In particular, we show that if the disturbance occurs infrequently with an arbitrary magnitude (known as stealth attack), then a perfect identification of the parameters of the system is possible in the autonomous case. For the non-autonomous case, we study how to design the input to properly excite

$$A = \begin{bmatrix} 5.4893{\times}10^{-1} & 4.8137{\times}10^{-3} & -1.7226{\times}10^{-1} & -2.4752{\times}10^{-2} & 1.6520{\times}10^{-3} & 3.4343{\times}10^{-4} & -9.6398{\times}10^{-5} & 1.4510{\times}10^{-4} \\ 5.9242{\times}10^{-4} & 9.8284{\times}10^{-1} & 9.9585{\times}10^{-4} & -1.6428{\times}10^{-4} & 5.2225{\times}10^{-5} & 3.6788{\times}10^{-7} & -7.0184{\times}10^{-5} & 9.5650{\times}10^{-7} \\ -4.3298{\times}10^{-1} & 4.0718{\times}10^{-3} & 8.0876{\times}10^{-1} & -2.4586{\times}10^{-2} & 1.8725{\times}10^{-3} & -2.6758{\times}10^{-4} & -5.5680{\times}10^{-5} & 1.4413{\times}10^{-4} \\ 3.1393{\times}10^{-1} & -1.1807{\times}10^{-1} & 5.6784{\times}10^{-2} & 7.5675{\times}10^{-1} & 1.6457{\times}10^{-3} & 1.9424{\times}10^{-4} & -7.5567{\times}10^{-5} & -4.4716{\times}10^{-3} \\ 0 & 0 & 0 & 0 & 6.3656{\times}10^{-40} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6.3656{\times}10^{-40} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6.3656{\times}10^{-40} & 0 \\ 1.7555{\times}10^{-1} & -6.5758{\times}10^{-2} & 3.1911{\times}10^{-2} & 4.2687{\times}10^{-1} & 9.2087{\times}10^{-4} & 1.0861{\times}10^{-4} & -4.2300{\times}10^{-5} & -2.5223{\times}10^{-3} \end{bmatrix} \quad (17)$$

$$B = \begin{bmatrix} 0.2530 & 0.0412 & -0.0138 & -0.0111 \\ 0.0044 & 0.0000 & -0.0063 & -0.0001 \\ 0.2730 & -0.0138 & -0.0101 & -0.0111 \\ 0.0903 & 0.0104 & -0.0042 & 0.6455 \\ 1.0000 & 0 & 0 & 0 \\ 0 & 1.0000 & 0 & 0 \\ 0 & 0 & 1.0000 & 0 \\ 0.0499 & 0.0057 & -0.0023 & -1.0406 \end{bmatrix} \quad (18)$$

the system in order to perfectly recover the model of the system under adversarial attack. The efficacy of the proposed framework is demonstrated in numerical experiments.

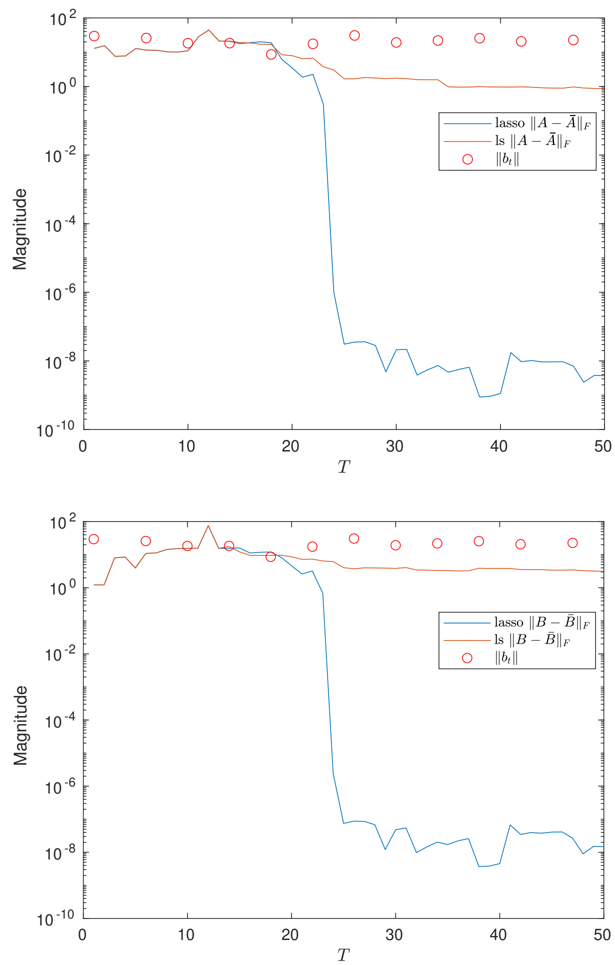

Fig. 2. Comparing the constrained lasso estimator (7) and the least-squares (ls) estimator (16) for the Tennessee Eastman challenge problem. The circles plot the magnitude of the disturbance b_t when it is non-zero. The difference is measured in the Frobenius norm $\|\cdot\|_F$.

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] S. Levine and P. Abbeel, “Learning Neural Network Policies with Guided Policy Search under Unknown Dynamics,” in *Advances in Neural Information Processing Systems 27*, 2014, pp. 1071–1079.

[4] 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, 2019.

[5] M. Jin and J. Lavaei, “Stability-certified reinforcement learning: A control-theoretic perspective,” *IEEE Access*, 2020.

[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] 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, 2011.

[10] 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,” *arXiv preprint arXiv:1805.08204*, 2018.

[11] 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.

[12] K. Bhatia, P. Jain, P. Kamalaruban, and P. Kar, “Consistent Robust Regression,” in *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., 2017, pp. 2110–2119.

[13] I. Diakonikolas, G. Kamath, D. M. Kane, J. Li, J. Steinhardt, and A. Stewart. Sever: A Robust Meta-Algorithm for Stochastic Optimization. [Online]. Available: http://arxiv.org/abs/1803.02815

[14] 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, 2019.

[15] J. Steinhardt, P. W. W. Koh, and P. S. Liang, “Certified Defenses for Data Poisoning Attacks,” in *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., 2017, pp. 3517–3529.

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

[17] S. Fattahi and S. Sojoudi, “Sample complexity of sparse system identification problem,” *accepted for publication in IEEE Transactions on Control of Network Systems*, 2021.

[18] 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)*, 2018, pp. 841–848.

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

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

[21] 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.

[22] N. Lawrence Ricker, “Model predictive control of a continuous, nonlinear, two-phase reactor,” *Journal of Process Control*, vol. 3, no. 2, pp. 109–123, 1993.