Efficient Secure State Estimation against Sparse Integrity Attack for System with Non-derogatory Dynamics

Zishuo Li and Yilin Mo [*][†]

Abstract: We consider the problem of estimating the state of a time-invariant linear Gaussian system in the presence of integrity attacks. The attacker can compromise p out of m sensors, the set of which is fixed over time and unknown to the system operator, and manipulate the measurements arbitrarily. Under the assumption that all the unstable eigenvalues of system matrix A have geometric multiplicity 1 (unstable part of A is non-derogatory), we propose a secure estimation scheme that is resilient to integrity attack as long as the system is $2p$-sparse detectable, which is proved to be the fundamental limit of secure dynamic estimation. In the absence of attack, the proposed estimation coincides with Kalman estimation with a certain probability that can be adjusted to trade-off between performance with and without attack. Furthermore, the detectability condition checking in the designing phase and the estimation computing in the online operating phase are both computationally efficient. A numerical example is provided to corroborate the results and illustrate the performance of the proposed estimator.

Keywords: secure estimation, sparse integrity attack, Kalman filter, sparse observability

1   Introduction

As the confluence of sensors, platforms, and networks increases, the already widespread applications of Cyber-Physical System (CPS) and Internet of Things (IoT) are expected to continue to emerge and expand [1]. They play an increasingly important role in critical infrastructures and everyday life, while the cyber-security risks and attack surfaces are also increasing [2]. However, CPS is vulnerable to a variety of cyber attacks since it usually relies on remote sensing devices, communication channels, and spatially distributed processors, which are prone to failures when exposed to unintentional faults and malicious attacks. Failure of CPS may cause severe damage to industrial infrastructures, economic order, and environmental systems, e.g., the Stuxnet launched on Iran's nuclear facilities[3], power blackouts in Ukraine [4], North America and Europe [5], etc. The research community has recognized the importance of CPS security, especially the design of secure detection, estimation, and control strategy[2].

Recently, substantial research efforts have been devoted to secure state estimation against various types of attacks, such as deception (integrity) attacks [6, 7] and denial-of-service (DoS) attacks [8, 9]. The integrity attacks focus on destructing system data integrity by stealthily manipulating the transmitted data, whereas DoS attacks jeopardize the availability of data resources by blocking the communication channels. A review of the secure estimation against various attacks is referred to [10]. This paper focuses on secure estimation against sparse integrity attack where an unknown subset of sensors is compromised by the adversary. The measurements from those corrupted sensors can be manipulated arbitrarily by the adversary. In order to identify the sparse malicious sensors and mitigate the impact of manipulated measurements, the main research paths include error correction approach based on compressed sensing

and switch estimation approach based on fault identification. The error correction approach usually takes measurements in a finite time window and adopts a sparsity-inducing optimization to handle the outliers. For example, minimizing the ℓ_0 norm or its convex relaxation ℓ_1 norm for lower computational complexity [11, 12]. Fawzi et al. [11] derive the fundamental limit for state reconstruction in the absence of noise and increase the number of correctable errors by state feedback. This result is further generalized to the scenario where the set of attacked nodes can change over time in [13]. For the scenario with bounded noise, Pajic et al. [12] provides rigorous analytic bounds on the estimation errors for ℓ_0 and ℓ_1-based estimation procedures. Similarly, Shoukry and Tabuada [14] adopt a 2-norm batch optimization approach for state estimation and a customized gradient descent algorithm to solve it efficiently. These works provide fundamental limits for estimation against integrity attack, i.e., proves that $2p$-sparse observability is necessary for secure state recovery against p compromised sensors. However, the sensory data out of the window are discarded in the finite time window approach, which may cause performance degradation and estimation delay.

Another solution is the switch estimator[15, 16, 17, 18, 19, 20] where the system operator switches between multiple estimate candidates[15, 18] or sensor subset combinations [16, 17, 19, 20] based on the evaluation of their reliability by consistency checking or malicious detection algorithms. However, the combinatorial nature of candidate estimates or sensor combinations poses challenges for storage or computation capability, and various solutions are proposed. Shoukry et al. [16, 17] aim at reducing searching complexity by Satisfiability Modulo Theory, and [20] reduces the number of candidates with the help of a set cover approach. In view of the computational problems, Liu et al. [21] propose a secure estimation scheme based on decomposing Kalman filters into local estimators whose weighted sum recovers the Kalman estimate with a certain probability in the absence of attack. The local estimates are fused securely by a quadratic programming problem with an ℓ_1 term to handle the sparse outliers. However, in the designing phase, the sufficient condition for estimation resiliency is computationally hard to validate. Moreover, the sufficient condition has a gap from $2p$-sparse observability, which is the fundamental limit [14] for state reconstruction. Similar to [21], other results in the literature[11, 22] also impose more restrictive conditions than $2p$-sparse observability and are NP-hard to validate. Besides the research paths aforementioned, cost function-based adaptive observers provide resilient estimation for scenarios with time-varying set of malicious sensors[23] and unknown inputs [24]. Arpan and Urbashi[23] propose a filtering and learning algorithm with adaptive gain, minimizing the cost function based on estimation error in normal operation and under attack on various sensor subsets. For the scenario with both unknown inputs and corrupted sensors, Wu et al. [24] propose a projected sliding-mode observer-based estimation algorithm to minimize the reconstruction error. However, the convergence of the descent algorithm used to minimize cost function does not necessarily imply the estimation error is bounded, which means secure guarantee is not easily obtained in these approaches.

In this paper, we focus on LTI system with Gaussian noise and intend to propose an estimation scheme that is secure (has bounded estimation error covariance) to p corrupted senors as long as the system is $2p$-sparse detectable, under the assumption that the stable part of A is non-derogatory. This achieves the fundamental limit for dynamic state estimation since it is proved that if the system is not $2p$-sparse detectable, there is no secure estimator [18]. Moreover, by introducing a non-derogatory assumption, the sparse detectability index can be computed computational easily. For general system matrix A with geometric multiplicity of unstable eigenvalues larger than 1, it has been proved that computing sparse observability is an NP-hard problem[25], and there is no computational efficient solution unless P = NP. Therefore, our proposed scheme reduces the computational complexity significantly when possible and achieves the fundamental limit of secure dynamic estimation. Preliminary versions of some of the results have been presented in [26]. This paper is significantly expanded from the previous work in the following points:

- The estimator proposed in [26] is secure if the system is $2p$-sparse observable. We design a novel estimator to secure the stable states in this paper, and the condition is relaxed to $2p$-sparse detectable and thus achieves the fundamental limit of secure dynamic estimation.

- This paper quantifies the probability of recovering the Kalman estimation in the absence of attack

2

and the estimation error upper bound under attack, which is not provided in [26]. Moreover, all the proofs of results in this paper are provided.

- The system of this paper is more general than [21, 26] since control input $u(k)$ is considered.

In summary, in this paper, we propose a secure dynamic estimation scheme for linear Gaussian systems, and it has the following merits:

- In the presence of p compromised sensors, the proposed estimation is secure if the system is $2p$-sparse detectable, which achieves the fundamental limit of secure dynamic state estimation.

- In the absence of attack, the proposed estimation coincides with Kalman estimation with certain probability, which can be adjusted to trade-off the performance with and without attack.

- During the designing phase, the sparse detectability index can be computed with low complexity. Moreover, during the algorithm operating phase, the proposed estimation is formulated as the solution of a convex optimization problem based on LASSO [27], which can be computed efficiently.

Organization: We introduce the problem formulation and preliminary results in Section 2. The main results are provided in Section 3 and collaborated by numerical simulation in Section 4. Section 5 finally concludes the paper.

Notations: Cardinality of a set $\mathcal{S}$ is denoted as $|\mathcal{S}|$. A' represents conjugate transpose of matrix A. Diagonal matrix with diagonal elements $A_1, \cdots, A_k$ is denoted as $\mathrm{diag}(A_1, \cdots, A_k)$. Denote the span of row vectors of matrix A as $\mathrm{rowspan}(A)$. The trace of matrix A is represented as $\mathrm{tr}(A)$. All-one vector with size $m \times 1$ is denoted as $\mathbf{1}_m$. I_n is the identity matrix with size $n \times n$. $\mathbb{C}^{m \times n}$ ($\mathbb{R}^{m \times n}$) represents the set of complex (real) matrices with m rows and n columns. The i-th entry of a vector x is represented by x_i or $[x]_i$. $\| \cdot \|_q$ represents the vector q-norm or (induced) matrix q-norm which is clear according to the context.

2 Problem Formulation and Preliminary Results

2.1 Secure dynamic state estimation

In this paper, we consider the linear time-invariant system with Gaussian noise:

$$x(k+1) = Ax(k) + Bu(k) + w(k), \tag{1}$$
$$y(k) = Cx(k) + v(k) + a(k), \tag{2}$$

where $x(k) \in \mathbb{R}^n$ is the system state, $w(k) \sim N(0, Q)$ and $v(k) \sim N(0, R)$ are i.i.d. Gaussian process noise and measurement noise with zero mean and covariance matrix Q and R. Vector $u(k) \in \mathbb{R}^d$ is the external input. The vector $y(k) \in \mathbb{R}^m$ is the collection of measurement from all m sensors, and i-th entry $y_i(k)$ is the measurement from sensor i. The vector $a(k)$ denotes the bias injected by an adversary and $a_i(k)$ is the attack on sensor i. Define

$$z(k) = Cx(k) + v(k)$$

as the true measurements without the attack. The initial state $x(0) \sim N(0, \Sigma)$ is assumed to be zero mean Gaussian and is independent from the process noise $\{w(k)\}$.

The secure dynamic estimation problem aims at recovering system state $x(k)$ at every time k based on all historical observations and inputs $\{y(t), u(t)|0 \le t \le k\}$, where $y(k)$ has been partly manipulated by the malicious attacker. It is conventional in the literature [11, 16] that the attacker can only compromise a fixed subset of sensors with known maximum cardinality. Denote the index set of all sensors as $\mathcal{J} \triangleq \{1, 2, \ldots, m\}$. For any index set $\mathcal{I} \subseteq \mathcal{J}$, define the complement set to be $\mathcal{I}^c \triangleq \mathcal{J} \backslash \mathcal{I}$. Define the support of vector $a \in \mathbb{R}^n$ as $\mathrm{supp}(a) \triangleq \{i | 1 \le i \le n, a_i \neq 0\}$ where a_i is the i-th entry of vector a. We have the following assumptions on the malicious adversary.

Definition 1 (Sparse Attack). *The attack called a (p, m)-sparse attack if the vector sequence $a(k)$ satisfy that, there exists a time invariant index set $\mathcal{I} \subseteq \mathcal{J}$ with $|\mathcal{I}| = p$ such that $\bigcup_{k=1}^{\infty} \text{supp} \{a(k)\} = \mathcal{I}$.*

Closely related to the sparse attack, we introduce the notion of sparse observability (detectability) that characterizes the system observability (detectability) in the presence of attack.

Definition 2 (Sparse observable / detectable). *The sparse observability (detectability) index of system (1)-(2) is the largest integer s such that system[1] $(A, C_{\mathcal{J} \setminus \mathcal{I}})$ is observable (detectable) for any set of sensors $\mathcal{I} \subset \mathcal{J}$ with cardinality $|\mathcal{I}| = s$. When the sparse observability (detectability) index is s, we say that the system with pair (A, C) is s-sparse observable (detectable).*

Remark 1. *Since one can do Kalman decomposition for uncontrollable or unobservable systems and obtain a minimal realization which is both controllable and observable, we assume the system (1)-(2) is controllable and observable without loss of generality. Thus, the sparse observability (detectability) index is a non-negative interger.*

Define $y(k_1 : k_2)$ as the sequence $\{y(k_1), y(k_1 + 1), \cdots, y(k_2)\}$. Similar notation is also applied on $z(k), u(k)$. An estimator is an infinite sequence of mappings $g = \{g_k\}_{k=1}^{\infty}$ where g_k is a mapping from all the historical outputs and inputs to a state estimation at time k:

$$g_k(y(0 : k), u(0 : k)) = \hat{x}(k).$$

It is written as $g_k(y, u) = \hat{x}(k)$ for notation simplicity. For linear Gaussian noise system, the estimation is secure if the estimation error covariance is bounded by a constant term irrelevant to the attack.

Definition 3 (Secure estimator). *Define the estimation difference introduced by attack as*

$$
\begin{aligned}
q_k &\triangleq g_k(z, u) - g_k(y, u) \\
&= g_k(z, u) - g_k(z + a, u).
\end{aligned}
$$

The estimator is said to be secure against (p, m)-sparse attack if the following holds:

$$\sup_{k \in \mathbb{Z}^+} \mathbb{E}\left[\text{tr}(q_k q_k')\right] < \infty,$$

where $\mathbb{E}$ is the expectation with respect to the probability measure generated by the Gaussian noise $\{w(k)\}$ and $\{v(k)\}$.

If all sensors are benign, i.e., $a(k) = \mathbf{0}$ for all k, the optimal state estimator is the classical Kalman filter:

$$
\begin{aligned}
\hat{x}(k) &= \hat{x}(k|k-1) + K(k)[y(k) - C\hat{x}(k|k-1)], \\
P(k) &= P(k|k-1) - K(k)CP(k|k-1),
\end{aligned}
$$

where

$$\hat{x}(k|k-1) = A\hat{x}(k-1) + Bu(k), P(k|k-1) = AP(k-1)A' + Q,$$

$$K(k) = P(k|k-1)C'\left(CP(k|k-1)C' + R\right)^{-1},$$

with initial condition $\hat{x}(0|-1) = 0$, $P(0|-1) = \Sigma$. It is well-known that for observable system, the estimation error covariance matrices $P(k)$ and the gain $K(k)$ will converge to

$$P \triangleq \lim_{k \to \infty} P(k), \ P_+ = APA' + Q, \ K \triangleq P_+ C'\left(CP_+ C' + R\right)^{-1}.$$

Since typically the control system will be running for an extended period of time, we focus on the case where the Kalman filter is in steady state, and thus the Kalman filter reduces to the following fixed-gain linear estimator:

$$\hat{x}(k+1) = (I - KC)(A\hat{x}(k) + Bu(k)) + Ky(k+1). \tag{3}$$

Before introducing our work, we first recall some results in the previous work that decomposes the fix gain Kalman filter to local estimates and recovers it securely by an optimization problem.

---

[1] The matrix $C_{\mathcal{J} \setminus \mathcal{I}}$ represents the matrix composed of rows of C with row index in $\mathcal{J} \setminus \mathcal{I}$.

2.2 Preliminary Results

We introduce some preliminaries in this subsection which are fundamental to main results in this paper. The following assumption is introduced to prevent system degradation.

Assumption 1. *The matrix A is invertible; $A - KCA$ has n distinct eigenvalues. Moreover, $A - KCA$ and A do not share any eigenvalue.*

Remark 2. *Since the invertibility of A implies that (A, CA) is also observable, we can freely assign the poles of $A - KCA$ by choosing a proper gain K. Hence, $A - KCA$ can satisfy the condition in Assumption 1 with a small estimation performance loss.*

Since $A - KCA$ has distinct eigenvalues, it can be diagonalized as:

$$A - KCA = V\Pi V^{-1}. \tag{4}$$

Define the eigenvalues of $A - KCA$ as $\pi_1, \cdots, \pi_n$. Consider local estimation $\zeta_i(k)$ which is the system response of sensor i. The local estimator satisfies the following dynamic:

$$\zeta_i(k+1) = \Pi\zeta_i(k) + \mathbf{1}_n y_i(k+1) + (G_i - \mathbf{1}_n C_i) Bu(k), \tag{5}$$

where C_i is i-th row of matrix C, and G_i is defined as

$$G_i \triangleq \begin{bmatrix} C_i A (A - \pi_1 I)^{-1} \\ \vdots \\ C_i A (A - \pi_n I)^{-1} \end{bmatrix}. \tag{6}$$

The following lemma shows the relationship between $\zeta_i(k)$ and $G_i x(k)$.

Lemma 1. *$\zeta_i(k)$ is stable estimation of $G_i x(k)$. Define their difference as $\epsilon_i(k) \triangleq \zeta_i(k) - G_i x(k)$, then $\epsilon_i(k)$ satisfies the following dynamics:*

$$\epsilon_i(k+1) = \Pi\epsilon_i(k) - (G_i - \mathbf{1}_n C_i) w(k) + \mathbf{1}_n v_i(k+1) + \mathbf{1}_n a_i(k+1). \tag{7}$$

Proof. According to the definition of $\zeta_i(k+1)$, one obatins

$$\begin{aligned} \epsilon_i(k+1) =& \Pi\zeta_i(k) + \mathbf{1}_n \left[C_i \left(Ax(k) + Bu(k) + w(k) \right) + v_i(k+1) + a_i(k+1) \right] \\ & - (G_i - \mathbf{1}_n C_i) Bu(k) - G_i \left(Ax(k) + Bu(k) + w(k) \right) \\ =& \Pi\zeta_i(k) - (G_i A - \mathbf{1}_n C_i A) x(k) - (G_i - \mathbf{1}_n C_i) w(k) + \mathbf{1}_n \left(v_i(k+1) + a_i(k+1) \right) \end{aligned}$$

Since it has been proved in [21] Corollary 1 that $G_i A - \mathbf{1}_n C_i A = \Pi G_i$, one can verify that equation (7) holds. $\qquad\square$

Define $\tilde{Q} \in \mathbb{C}^{mn \times mn}$ as the covariance of noise term $(G_i - \mathbf{1}_n C_i) w(k) - \mathbf{1}_n v_i(k+1)$ for all i, i.e.,

$$\tilde{Q} \triangleq \begin{bmatrix} G_1 - \mathbf{1}_n C_1 \\ \vdots \\ G_m - \mathbf{1}_n C_m \end{bmatrix} Q \begin{bmatrix} G_1 - \mathbf{1}_n C_1 \\ \vdots \\ G_m - \mathbf{1}_n C_m \end{bmatrix}' + R \otimes \mathbf{1}_{n \times n}, \tag{8}$$

where $\otimes$ is the Kronecker product. Define $\tilde{\Pi} \in \mathbb{C}^{mn \times mn}$ as

$$\tilde{\Pi} \triangleq \begin{bmatrix} \Pi & & \\ & \ddots & \\ & & \Pi \end{bmatrix}.$$

5

The stable covariance of $\epsilon(k) \triangleq [\epsilon_1(k)', \cdots, \epsilon_m(k)']'$ is the solution $\tilde{W}$ of the following Lyapunov equation:

$$\tilde{W} = \tilde{\Pi}\tilde{W}\tilde{\Pi}' + \tilde{Q}.$$

The matrix $\tilde{W}$ is well-defined since Π is strictly stable. As a result, the secure estimation can be recovered by the solution of the following optimization problem where $\zeta(k) \triangleq [\zeta_1(k)', \cdots, \zeta_m(k)']'$ and $G \triangleq [G'_1, \cdots, G'_m]'$.

$$\underset{\check{x}(k),\mu(k),\nu(k)}{\text{minimize}} \quad \frac{1}{2}\mu(k)'\tilde{W}^{-1}\mu(k) + \gamma\|\nu(k)\|_1 \tag{9a}$$

$$\text{subject to} \quad \zeta(k) = G\check{x}(k) + \mu(k) + \nu(k). \tag{9b}$$

The parameter γ is a non-negative constant chosen by the system operator. The following theorem from Liu et al.[21] proves that the solution $\check{x}(k)$ to problem (9) is a secure estimation under specific condition.

Theorem 1 (Liu et al.[21]). *In the presence of (p,m)-sparse attack, the state estimation $\check{x}(k)$ is secure if the following inequality holds for all $x \neq \mathbf{0}$, $x \in \mathbb{R}^n$:*

$$\sum_{i \in \mathcal{I}} \|G_i x\|_1 < \sum_{i \in \mathcal{I}^c} \|G_i x\|_1, \quad \forall\, \mathcal{I} \subset \mathcal{J}, |\mathcal{I}| \leq p. \tag{10}$$

Even though Theorem 1 establishes the sufficient condition of the estimation to be secure, this condition can be improved in the following two aspects.

(1) Validating condition (10) is NP-hard. The computational complexity can be significantly reduced by introducing the non-derogatory assumption and further analysis on matrix G_i.

(2) Condition (10) does not achieve the fundamental limit. It is more restrictive that $2p$-sparse observability and has a gap from $2p$-sparse detectablity.

In the following section, we proposed a secure estimation scheme that improves the aforementioned two points. Under assumption on unstable eigenvalues of A, the sufficient condition of the estimation to be secure is proved to be $2p$-sparse detectable, which is easily validated and achieves the fundamental limit of secure dynamic estimation.

3   Secure Estimation with Sparse Detectability

In this section, under the assumption that all the unstable eigenvalues of A have geometric multiplicity 1, we design a state estimator that is secure in the presence of (p,m)-sparse attack as long as the system (A, C) is $2p$-sparse detectable. Moreover, we prove that if the system is not $2p$-sparse detectable, there exists an attack strategy under which no estimator can be secure. We first introduce the following assumption on unstable[2] eigenvalues of A.

Assumption 2. *All the unstable eigenvalues of A have geometric multiplicity 1.*

Since we can perform invertible linear transformation T on state x and study the following system instead:

$$\bar{x}(k) = \bar{A}\bar{x}(k) + TB\bar{x}(k) + Tw(k),$$
$$y(k) = CT^{-1}\bar{x}(k) + v(k) + a(k),$$

where $\bar{A} \triangleq TAT^{-1}$ is similar to A and $\bar{x} = Tx$, we can assume that A is in the following Jordan form without loss of generality:

$$A = \begin{pmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & A_2 \end{pmatrix},$$

---

[2]Unstable eigenvalues are those eigvalues that satisfy $|\lambda_i| \geq 1$.

where block A_1 is composed of the Jordan blocks with unstable eigenvalues and A_2 is composed of the Jordan blocks with stable eigenvalues.

Denote the number of unstable eigenvalues of A (counted with repetition) as n_u and number of stable eigenvalues as n_s. In order to analyze stable and unstable states separately with simple notation, we denote the index set of unstable entries of state as $\mathcal{U} \triangleq \{1, 2, \cdots, n_u\}$ and index set of stable entries as $\mathcal{S} \triangleq \{n_u + 1, \cdots, n\}$. Furthermore, a matrix X can be divided vertically to two sub-matrices:

$$X = \left[\ X^{\mathcal{U}} \mid X^{\mathcal{S}}\ \right],$$

where $X^{\mathcal{U}}$ is the matrix composed of first n_u columns of matrix X and $X^{\mathcal{S}}$ is composed of last n_s columns of X. Define the observable matrix of system (A, C_i) as

$$O_i \triangleq \left[\ C_i' \mid (C_i A)' \mid \cdots \mid \left(C_i A^{n-1}\right)'\ \right]'. \tag{11}$$

Therefore, rowspan$(O_i^{\mathcal{U}})$ is the observable subspace of sensor i corresponding to unstable states.

3.1   Canonical form of G_i

We prove in this subsection that the row span of $G_i^{\mathcal{U}}$ coincides with rowspan$(O_i^{\mathcal{U}})$, which implies that the matrix $G_i^{\mathcal{U}}$ has a canonical form under row operations. Before continuing on, we need the following notation of state-sensor observability. Define $\mathcal{E}_j$ as the index set of sensors that can observe state j, i.e.

$$\mathcal{E}_j \triangleq \{i \in \mathcal{J} \mid O_i'e_j \neq \mathbf{0}\}, \tag{12}$$

where $\mathcal{J} \triangleq \{1, 2, \cdots, m\}$ is the index set of all sensors and e_j is the n-dimensional canonical basis vector with 1 on the j-th entry and 0 on the other entries. We have the following theorem characterizing the structure of G_i.

Theorem 2. *Assume system matrix A satisfies Assumption 2, then the following equation holds:*

$$\mathrm{rowspan}(G_i^{\mathcal{U}}) = \mathrm{rowspan}(O_i^{\mathcal{U}}) = \mathrm{rowspan}(H_i^{\mathcal{U}}), \tag{13}$$

where $H_i^{\mathcal{U}}$ is the following $n \times n_u$ matrix

$$H_i^{\mathcal{U}} \triangleq \left[\begin{array}{c} \begin{matrix} \mathbb{I}_{i \in \mathcal{E}_1} & & \\ & \ddots & \\ & & \mathbb{I}_{i \in \mathcal{E}_{n_u}} \end{matrix} \\ \hline \mathbf{0}_{(n-n_u) \times n_u} \end{array}\right],$$

and $\mathbb{I}_{\mathscr{E}}$ is the indicator function that takes the value 1 when $\mathscr{E}$ is true and value 0 when $\mathscr{E}$ is not. Therefore, there exists an invertible $n \times n$ matrix P_i such that $P_i G_i = H_i$ and H_i is in the following form:

$$H_i = P_i G_i = \left[\ H_i^{\mathcal{U}} \mid P_i G_i^{\mathcal{S}}\ \right].$$

Proof of Theorem 2 is provided in Appendix A. After transformation P_i, matrix $G_i^{\mathcal{U}}$ is transformed into canonical form $H_i^{\mathcal{U}}$ whose rows are either canonical basis vectors or zero vectors. The non-zero entries of H_i records the state observability of sensor i. Therefore, the sparse observability/detectability index can be directly obtained from $H_i^{\mathcal{U}}$.

Corollary 1. *The sparse observability index of system (A, C) is $\min_{j \in \{1, 2, \cdots, n\}} |\mathcal{E}_j| - 1$. The sparse detectability index of system (A, C) is $\min_{j \in \mathcal{U}} |\mathcal{E}_j| - 1$ if $\mathcal{U} \neq \varnothing$ and is $m - 1$ if $\mathcal{U} = \varnothing$.*

Proof. Define $s \triangleq \min_{j \in \{1, 2, \cdots, n\}} |\mathcal{E}_j| - 1$. For arbitrary $\bar{s}$ that satisfy $\bar{s} \geq s + 1$, there exists a state index j^* and a sensor index set $\mathcal{I}^*$ with $|\mathcal{I}^*| = \bar{s}$ such that $\mathcal{E}_{j^*} \cap (\mathcal{J} \setminus \mathcal{I}^*) = \varnothing$. As a result, state j^* cannot be observed by any sensor in $\mathcal{J} \setminus \mathcal{I}^*$, i.e.,

$$e_{j^*} \notin \mathrm{rowspan}(O_i), \ \forall i \in \mathcal{J} \setminus \mathcal{I}^*,$$

and thus system $(A, C_{\mathcal{J} \setminus \mathcal{I}^*})$ is not observable. For arbitrary $\underline{s}$ that satisfies $\underline{s} \leq s$, arbitrary j and arbitrary $\mathcal{I}$ with $|\mathcal{I}| = \underline{s}$, one obtains $\mathcal{E}_{j^*} \cap (\mathcal{J} \setminus \mathcal{I}^*) \neq \varnothing$, which means for all j, there exists $i^* \in \mathcal{J} \setminus \mathcal{I}$ such that: $e_j \in \mathrm{rowspan}(O_{i^*})$. Therefore, system $(A, C_{\mathcal{J} \setminus \mathcal{I}})$ is observable. According to Definition 2, the system is s-sparse observable. The detectability index is obtained in the same way by considering unstable subsystem when $\mathcal{U} \neq \varnothing$. When $\mathcal{U} = \varnothing$(i.e., A is stable), system is always detectable as long as $\mathcal{J} \neq \varnothing$ and thus the sparse detectability is $m - 1$ according to definition 2. $\qquad\square$

In conclusion, under Assumption 2, the matrix $G_i^{\mathcal{U}}$ has a canonical form which $H_i^{\mathcal{U}}$ is determined by state-sensor observability. Leveraging upon the canonical form $H_i^{\mathcal{U}}$, we will propose an estimation scheme that is secure in the presence of (p, m)-sparse attack as long as the system is $2p$-sparse detectable.

3.2 Secure Estimation Design

Recalling the transformation P_i introduced in Theorem 2, define $\tilde{P} \triangleq \mathrm{diag}\,(P_1, \cdots, P_m)$, $\tilde{M} \triangleq \tilde{P} \tilde{W} \tilde{P}'$ and

$$Y(k) \triangleq \begin{bmatrix} P_1 \zeta_1(k) \\ \vdots \\ P_m \zeta_m(k) \end{bmatrix} \in \mathbb{C}^{mn \times 1}, \ H \triangleq \begin{bmatrix} H_1 \\ \vdots \\ H_m \end{bmatrix} \in \mathbb{C}^{mn \times n}. \tag{14}$$

Define the following matrix

$$N \triangleq I_m \otimes \begin{bmatrix} \mathbf{0}_{n_s \times n_u} & I_{n_s} \end{bmatrix} \in \mathbb{R}^{mn_s \times mn}.$$

Consider the following least square problem.

$$\underset{\tilde{x}_{\mathrm{ls}}(k), \varphi(k)}{\mathrm{minimize}} \quad \frac{1}{2} \begin{bmatrix} \varphi(k) \\ NH\tilde{x}_{\mathrm{ls}}(k) \end{bmatrix}' \mathcal{W} \begin{bmatrix} \varphi(k) \\ NH\tilde{x}_{\mathrm{ls}}(k) \end{bmatrix} \tag{15a}$$

$$\text{subject to} \quad Y(k) = H\tilde{x}_{\mathrm{ls}}(k) + \varphi(k). \tag{15b}$$

where

$$\mathcal{W} \triangleq \begin{bmatrix} \tilde{M}^{-1} + N'N & N' \\ N & I \end{bmatrix}. \tag{16}$$

Notice that $\mathcal{W}$ is positive definite since $\tilde{M}^{-1} \succ 0$. Define[3] $F_i \triangleq V \mathrm{diag}(V^{-1} K_i)$, $F = \begin{bmatrix} F_1 & \cdots & F_m \end{bmatrix}$, where V is defined in (4). Recall that $\epsilon(k) \triangleq [\epsilon_1(k)', \cdots, \epsilon_m(k)']'$ and $\epsilon_i(k) = \zeta_i(k) - G_i x(k)$ from Lemma 1 and fix gain Kalman estimation $\hat{x}(k)$ from (3).

Lemma 2. *In the absence of attack, the solution to least square problem* (15) *coincides with the Kalman estimation and satisfies the following:*

$$\tilde{x}_{ls}(k) = \hat{x}(k), \ \varphi(k) = (I - GF)\epsilon(k).$$

---

[3] $\mathrm{diag}(V^{-1} K_i)$ is a $n \times n$ diagonal matrix whose diagonal with the j-th diagonal entry equals to j-th element of vector $V^{-1} K_i$.

Proof. Consider the following least square problem

$$\underset{\tilde{x}_{\mathrm{ls}}(k)}{\text{minimize}} \quad \frac{1}{2} \left(Y(k) - H\tilde{x}_{\mathrm{ls}}(k) \right)' \tilde{M}^{-1} \left(Y(k) - H\tilde{x}_{\mathrm{ls}}(k) \right). \tag{17}$$

Based on Theorem 2 in [21], the solution to problem (17) is equivalent to Kalman estimation. It is sufficient to prove that problems (15) and (17) are equivalent. Define

$$\mathcal{M} \triangleq \begin{bmatrix} I_{mn} & \mathbf{0}_{mn \times mn_s} \\ N & I_{mn_s} \end{bmatrix} \in \mathbb{R}^{m(n+n_s) \times m(n+n_s)}.$$

Consider the objective function of problem (17) added by a constant term[4]:

$$\frac{1}{2}(Y - H\tilde{x}_{\mathrm{ls}})'\tilde{M}^{-1}(Y - H\tilde{x}_{\mathrm{ls}}) + \frac{1}{2}Y'N'NY = \frac{1}{2}\begin{bmatrix} Y - H\tilde{x}_{\mathrm{ls}} \\ NY \end{bmatrix}' \begin{bmatrix} \tilde{M}^{-1} & \mathbf{0} \\ \mathbf{0} & I \end{bmatrix} \begin{bmatrix} Y - H\tilde{x}_{\mathrm{ls}} \\ NY \end{bmatrix}. \tag{18}$$

Notice that

$$\begin{bmatrix} Y - H\tilde{x}_{\mathrm{ls}} \\ NY \end{bmatrix} = \mathcal{M} \begin{bmatrix} Y - H\tilde{x}_{\mathrm{ls}} \\ NH\tilde{x}_{\mathrm{ls}} \end{bmatrix},$$

and (18) can be written as

$$\frac{1}{2}\left(\mathcal{M}\begin{bmatrix} Y - H\tilde{x}_{\mathrm{ls}} \\ NH\tilde{x}_{\mathrm{ls}} \end{bmatrix}\right)'\begin{bmatrix} \tilde{M}^{-1} & \mathbf{0} \\ \mathbf{0} & I \end{bmatrix}\left(\mathcal{M}\begin{bmatrix} Y - H\tilde{x}_{\mathrm{ls}} \\ NH\tilde{x}_{\mathrm{ls}} \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} Y - H\tilde{x}_{\mathrm{ls}} \\ NH\tilde{x}_{\mathrm{ls}} \end{bmatrix}' \mathcal{W} \begin{bmatrix} Y - H\tilde{x}_{\mathrm{ls}} \\ NH\tilde{x}_{\mathrm{ls}} \end{bmatrix}. \tag{19}$$

Substituting φ in (15) with $Y - H\tilde{x}_{\mathrm{ls}}$ leads to (19). Thus, optimizing objective function (17) is equivalent to optimizing (19), and the latter is equivalent to problem (15). $\qquad\square$

Based on least square problem (15), we present the following optimization problem whose solution $\tilde{x}(k)$ is our proposed secure estimation. The constant γ is a non-negative adjustable parameter.

$$\underset{\tilde{x}(k),\mu(k),\nu(k)}{\text{minimize}} \quad \frac{1}{2}\begin{bmatrix} \mu(k) \\ NH\tilde{x}(k) \end{bmatrix}' \mathcal{W} \begin{bmatrix} \mu(k) \\ NH\tilde{x}(k) \end{bmatrix} + \gamma \left\| \nu(k) \right\|_1 \tag{20a}$$

$$\text{subject to} \quad Y(k) = H\tilde{x}(k) + \mu(k) + \nu(k). \tag{20b}$$

The following theorem characterizes the performance of our proposed estimator when the attacker is absent. The proof is provided in Appendix C.

Theorem 3. *In the absence of attack, if the parameter γ in problem (20) satisfies*

$$\left\| \mathcal{W} \begin{bmatrix} (I - GF)\,\epsilon(k) \\ NH\hat{x}(k) \end{bmatrix} \right\|_\infty \le \gamma, \tag{21}$$

then our proposed estimation $\tilde{x}(k)$ is equivalent to the estimation of fixed gain Kalman filter defined in (3), i.e.,

$$\tilde{x}(k) = \hat{x}(k). \tag{22}$$

Noticing that $\epsilon(k)$ converges to a stationary Gaussian process, and $\hat{x}(k)$ is a Gaussian random variable, the probability that (21) holds is determined only by system parameter A, B, C, Q, R, γ given input $u(k)$, and can be explicitly calculated given these parameters. By tuning design parameter γ, the probability of recovering the Kalman estimation can be adjusted.

---

[4]$Y(k)$ is fixed for each k in the optimization problem and thus is treated as a constant. For legibility, the time index (k) is omitted.

In order to quantify the estimation difference between the attack is absent and present, we consider the following local estimation and Kalman estimation without attack:

$$\zeta_i^o(k+1) = \Pi\zeta_i^o(k) + \mathbf{1}_n z_i(k+1) + (G_i - \mathbf{1}_n C_i)Bu(k), \tag{23}$$

$$\hat{x}^o(k+1) = (I - KC)(A\hat{x}^o(k) + Bu(k)) + Kz(k+1), \tag{24}$$

where $z(k) = Cx(k) + v(k)$ is the original (unmanipulated) measurement. Define $\epsilon_i^o(k)$ correspondingly as $\epsilon_i^o(k) \triangleq \zeta_i^o(k) - G_i x(k)$. The following theorem quantifies the estimation error introduced by the attack.

Theorem 4. *In presence of arbitrary (p,m)-sparse attack, if the system (A,C) is $2p$-sparse detectable, then the estimation difference between $\tilde{x}(k)$ solved from (20) and oracle Kalman estimation $\hat{x}^o(k)$ satisfies*

$$|[\tilde{x}(k)]_j - [\hat{x}^o(k)]_j| \leq \begin{cases} \max\limits_{i_1,i_2 \in \mathcal{E}_j} \left| \left[P_{i_1}\zeta_{i_1}^o(k)\right]_j - \left[P_{i_2}\zeta_{i_2}^o(k)\right]_j \right| + (\gamma + \gamma^o(k))\|\mathcal{F}\|_\infty, \ j \in \mathcal{U}, \\ \gamma \cdot \|\mathcal{F}\|_\infty + |[\hat{x}^o(k)]_j|, \ j \in \mathcal{S}, \end{cases} \tag{25}$$

where

$$\gamma^o(k) \triangleq \left\| \mathcal{W} \begin{bmatrix} (I - GF)\epsilon^o(k) \\ NH\hat{x}^o(k) \end{bmatrix} \right\|_\infty,$$

$$\mathcal{F} \triangleq \left(\begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & \mathcal{L}H'N' \end{bmatrix} \mathcal{W} \begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & NH\mathcal{L}' \end{bmatrix} \right)^{-1} \begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & \mathcal{L}H' \end{bmatrix},$$

$$\mathcal{L} \triangleq \begin{bmatrix} \mathbf{0}_{n_s \times n_u} & I_{n_s} \end{bmatrix},$$

with $\mathcal{E}_j$ defined in (12) and $[\cdot]_j$ being the j-th element of a vector. Since the oracle Kalman estimation $\hat{x}^o(k)$ is a stable estimation of system state $x(k)$, and the upper bounds have bounded variance for all $k \in \mathbb{Z}^+$, our proposed estimation $\tilde{x}(k)$ is secure.

Under Assumption 2, if the system is $2p$-sparse detectable, our proposed estimator is secure. The maximum estimation difference from oracle Kalman filter is shown in (25). Theorem 4 indicates smaller γ leads to lower estimation difference upper bound in the presence of attack. However, based on Theorem 3, smaller γ decreases probability of recovering the optimal Kalman estimation in the absence of attack. The choice of γ represents the trade-off between the performance in normal operation and the performance under attack.

Moreover, since sparse detectability index only requires simple computation according to Corollary 1, our work reduces the complexity of evaluating system vulnerability significantly under the assumption of geometric multiplicity. For general A that has eigenvalues with geometric multiplicity larger than 1 (A is derogatory), computing sparse observability is an NP-hard problem [25], and there is no computational efficient solution unless P = NP. Besides the computation complexity of off-line designing, for algorithm online operation, the computing of estimation involves solving a convex optimization problem which can be done efficiently.

The following theorem proves that $2p$-sparse detectability is necessary for the existence of a secure estimation, which coincides with the sufficient condition of our estimator to be secure.

Theorem 5. *If the system is not $2p$-sparse detectable, there exists a (p,m)-sparse attack strategy under which no estimator is secure.*

Proof. This proof is based on [18][28]. Since we focus on a different formulation, we reorganize it here for paper self-consistency. If the system is not $2p$-sparse detectable, then there exists a set $\mathcal{A}$ with $\mathcal{A} \subset \mathcal{J}, |\mathcal{A}| = 2p$ and an eigenvector ξ of A that corresponds to a unstable eigenvalue λ ($|\lambda| \geq 1$) such that either one of the following two statements is true:

$\quad$ (1) $A\xi = \lambda\xi, C_{\mathcal{J}\backslash\mathcal{A}}\xi = \mathbf{0},$ if ξ is a real-value vector,
$\quad$ (2) $A\xi = \lambda\xi, A\bar{\xi} = \bar{\lambda}\bar{\xi}, C_{\mathcal{J}\backslash\mathcal{A}}\xi = \mathbf{0}, C_{\mathcal{J}\backslash\mathcal{A}}\bar{\xi} = \mathbf{0},$ if ξ is a complex-value vector,

10

where $\bar{\xi}, \bar{\lambda}$ represent the conjugate of ξ, λ. Based on this result, we intend to prove that the following proposition is true.

Proposition 1. *There exists zero mean Gaussian disturbances* $\{w^{(1)}(k)\}_{k=0}^{\infty}$, $\{v^{(1)}(k)\}_{k=0}^{\infty}$, $\{w^{(2)}(k)\}_{k=0}^{\infty}$, $\{v^{(2)}(k)\}_{k=0}^{\infty}$, *two* (p,m)-*sparse attack sequences* $\{a^{(1)}(k)\}_{k=0}^{\infty}, \{a^{(2)}(k)\}_{k=0}^{\infty}$ *and a pair of initial states* $x^{(1)}(0), x^{(2)}(0)$ *such that the two system trajectories* $\{x^{(1)}, w, y^{(1)}, v, a^{(1)}\}, \{x^{(2)}, w, y^{(2)}, v, a^{(2)}\}$ *satisfy:*

- *Two system trajectories both follow dynamics in (1)-(2).*

- $y^{(1)}(k) = y^{(2)}(k), \ \forall k \geq 0.$

- $\|x^{(1)}(k) - x^{(2)}(k)\|_2 \to \infty, \ as \ k \to \infty.$

We construct two trajectories that proves Proposition 1. Divide $\mathcal{A}$ into $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2$ such that $\mathcal{A}_1 \cap \mathcal{A}_2 = \varnothing$ and $|\mathcal{A}_1| = |\mathcal{A}_2| = p$. Define the following two trajectories if ξ is real-value vector:

$$
\textbf{System 1:} \quad x^{(1)}(0) = \mathbf{0}, \qquad\qquad \textbf{System 2:} \quad x^{(2)}(0) = \xi,
$$
$$
w^{(1)}(k) = \mathbf{0}, \ v(k) = \mathbf{0}, \qquad\qquad w^{(2)}(k) = \phi(k) \cdot \xi, \ v(k) = \mathbf{0},
$$
$$
a^{(1)}(k) = \begin{cases} C_i x^{(2)}(k), i \in \mathcal{A}_1 \\ \mathbf{0}, i \in \mathcal{J} \setminus \mathcal{A}_1 \end{cases}, \qquad\qquad a^{(2)}(k) = \begin{cases} -C_i x^{(2)}(k), i \in \mathcal{A}_2 \\ \mathbf{0}, i \in \mathcal{J} \setminus \mathcal{A}_2 \end{cases}.
$$

where $\phi(k)$ is a time-independent, standard Gaussian distributed random scalar. Noticing that $A\xi = \lambda\xi$, $C_{\mathcal{J}\setminus\mathcal{A}}\xi = \mathbf{0}$, one can verify that the output of these two systems are:

$$
\textbf{System 1:} \quad y^{(1)}(k) = \begin{cases} C_i x^{(2)}(k), i \in \mathcal{A}_1 \\ \mathbf{0}, i \in \mathcal{A}_2 \\ \mathbf{0}, i \in \mathcal{J} \setminus \mathcal{A} \end{cases}, \quad \textbf{System 2:} \quad y^{(2)}(k) = \begin{cases} C_i x^{(2)}(k), i \in \mathcal{A}_1 \\ \mathbf{0}, i \in \mathcal{A}_2 \\ \mathbf{0}, i \in \mathcal{J} \setminus \mathcal{A} \end{cases}.
$$

The state sequence of system 2 satisfies

$$
x^{(2)}(k) = \lambda^k \xi + \sum_{t=1}^{k} \lambda^{k-t} \xi \phi(t)
$$

while state sequence of system 1 is $x^{(1)}(k) = \mathbf{0}$. One concludes that $\mathbb{E}\|x^{(1)}(k) - x^{(2)}(k)\|_2^2 \to \infty$ as $k \to \infty$ for both $|\lambda| > 1$ and $|\lambda| = 1$.

If ξ is complex, let $x^{(2)}(0) = \xi + \bar{\xi}$, $w^{(2)}(k) = \phi(k)\left(\xi + \bar{\xi}\right)$ and others hold the same with real-value case. One can verify that

$$
x^{(2)}(k) = \left(\lambda^k \xi + \bar{\lambda}^k \bar{\xi}\right) + \sum_{t=1}^{k} \left(\lambda^{k-t}\xi + \bar{\lambda}^{k-t}\bar{\xi}\right)\phi(t)
$$

and $x^{(1)}(k) = \mathbf{0}$. As a result, $\mathbb{E}\|x^{(1)}(k) - x^{(2)}(k)\|_2^2 \to \infty$ as $k \to \infty$ for both $|\lambda| > 1$ and $|\lambda| = 1$.

However, the system observable output $y^{(1)}(k)$ and $y^{(2)}(k)$ are always the same. As a result, Proposition 1 is proved and there exists no secure estimation under such attack. $\qquad\square$

In view of Theorem 5, our proposed estimator achieves the fundamental limit of secure estimation problem, i.e., provides a secure estimation whenever the system is possible to be securely estimated. This result is stronger than other dynamic estimators in the literature which require $2p$-sparse observability [11, 12, 13, 16, 17, 21]. The performance of our proposed estimator is corroborated by the numerical simulation in the next section.

4   Illustrative Example

We use an inverted pendulum for the numerical simulation[5]. The physical parameters are illustrated in Fig. 1. The control input $u(k)$ is the force applied on the cart, and b_θ, b_x are the friction coefficients at pendulum joint and cart wheels. The state x_1, x_2, x_3, x_4 represent cart position coordinate, cart velocity, pendulum angle from vertical and pendulum angle velocity respectively.
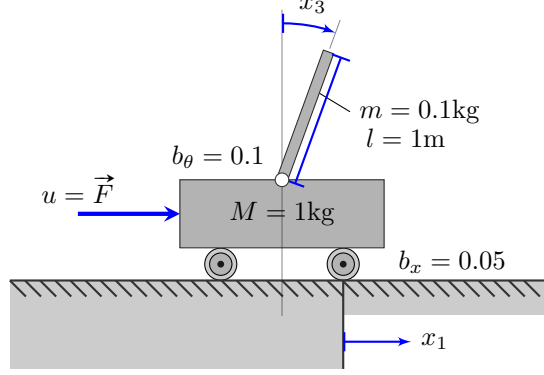

Figure 1: Illustration of the inverted pendulum.

Consider the system linearized at $x_3 = x_4 = 0$, and we sample the continuous-time linear system periodically with sampling interval $T_s = 0.02$ seconds. The system equation is:

$$x(k+1) = \begin{bmatrix} 1 & 2.0 \cdot 10^{-2} & -2.0 \cdot 10^{-4} & 1.9 \cdot 10^{-5} \\ 0 & 1.0 \cdot 10^0 & -2.0 \cdot 10^{-2} & 1.8 \cdot 10^{-3} \\ 0 & 1.0 \cdot 10^{-5} & 1.0 \cdot 10^0 & 2.0 \cdot 10^{-2} \\ 0 & 1.0 \cdot 10^{-3} & 2.1 \cdot 10^{-1} & 9.8 \cdot 10^{-1} \end{bmatrix} x(k) + \begin{bmatrix} 2.0 \cdot 10^{-4} \\ 2.0 \cdot 10^{-2} \\ -2.0 \cdot 10^{-4} \\ -2.0 \cdot 10^{-2} \end{bmatrix} u(k) + w(k).$$

$$y(k) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} x(k) + v(k) + a(k).$$

The system dynamic matrix can be written as the following Jordan canonical form by an invertible linear transformation:

$$\begin{bmatrix} 1.057 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.999 & 0 \\ 0 & 0 & 0 & 0.925 \end{bmatrix},$$

and we consider the system after transformation. System matrix A have four Jordan blocks with size 1×1 and the upper left two blocks have unstable eigenvalues. Therefore, the set of unstable states and stable states are $\mathcal{U} = \{1, 2\}, \mathcal{S} = \{3, 4\}$. The canonical form of $G_i^{\mathcal{U}}$ are

$$H_1^{\mathcal{U}} = H_2^{\mathcal{U}} = H_3^{\mathcal{U}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad H_4^{\mathcal{U}} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{26}$$

Only the first 3 sensors can observe unstable state 1, i.e., $\mathcal{E}_1 = \{1, 2, 3\}$. All the four sensors can observe unstable state 2, i.e., $\mathcal{E}_2 = \{1, 2, 3, 4\}$. Therefore, the system is 2-sparse detectable and our proposed

---

[5]The corresponding code is posted on `https://github.com/zs-li/resilient_dynamic_estimation`.

estimator is secure in the presence of 1 corrupted sensor. In the simulation, the noise covariances of the system are $Q = R = T_s^2 \times \text{diag}(0.1, 0.1, 0.01, 0.01)$. The initial state is assumed to be known by the estimator. The controller of the system is designed as a Linear-Quadratic Regulator (LQR), and the feedback matrix is chosen as $K_{\text{lqr}} = \begin{bmatrix} -8 & -15 & -115 & -32 \end{bmatrix}$.

We first illustrate the performance of estimation on close-loop system where $u(k) = -K_{\text{lqr}}x(k)$. Fig. 2 presents the performance of the estimation of system states in the absence of attack. Our proposed estimation substantially coincides with the Kalman estimation. The numerical difference attributes to large Gaussian noise that occurs occasionally which violates inequality (21) and error in numerical calculation.
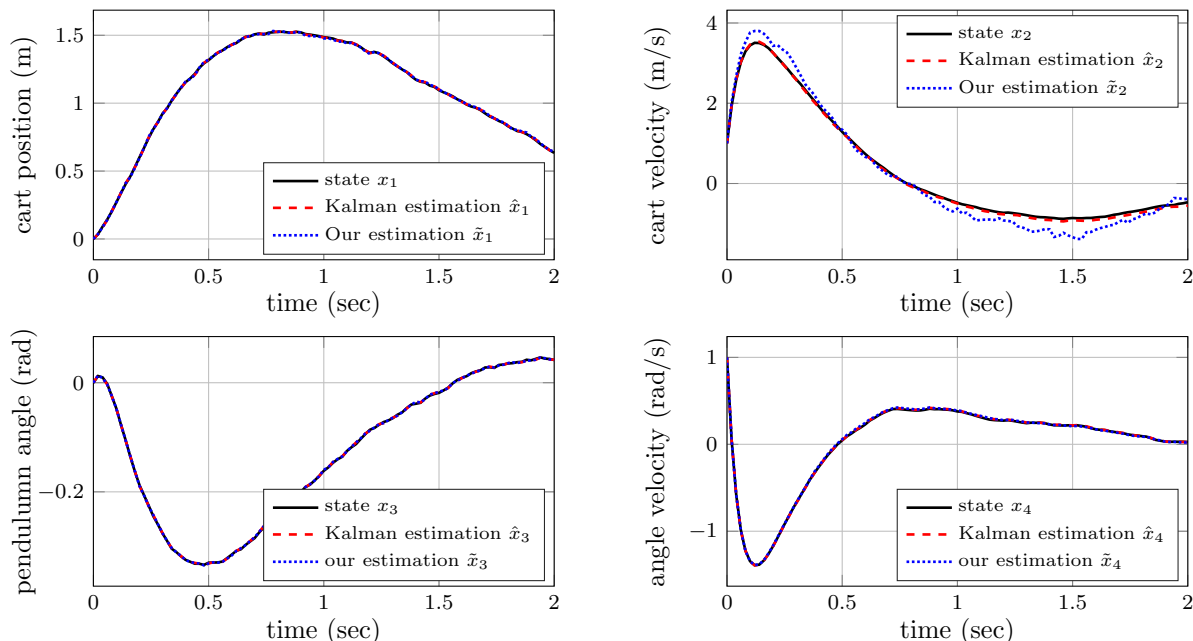

Figure 2: Estimation of states in the absence of attack. The initial state is $x(0) = [0, \ 1, \ 0, \ 1]'$.

Fig. 3 shows the injected attack signal on sensor 3 and the corresponding observations from sensor 3. The attack $a_3(k)$ is a time-independent random value uniformly distributed on interval $(-1, 1)$. Fig. 4 demonstrates the estimation with the attack shown in Fig. 3. As shown in the Fig. 4, Kalman estimation (denoted as red dashed line) has larger estimation error than our proposed estimation under the attack.

Fig. 5 illustrates the estimation mean square error (MSE$= 1/N \sum_{k=1}^{N} \|\hat{x}(k) - x(k)\|_2^2$) of our proposed estimator with varying tuning parameter γ and varying attack magnitude with sensor 3 corrupted. The number of time steps is set as $N = 200$. The attack signal $a_3(k)$ is uniformly distributed in interval $(-\|a\|_\infty, \|a\|_\infty)$. In Fig. 5a, the MSE of the oracle Kalman estimation is illustrated by the red dashed line and that of the Kalman estiamtion under attack is illustrated by the red solid line. As shown in Fig. 5a, by properly choosing γ, the MSE of our proposed estimator is smaller than that of Kalman estimation (blue line is below the red horizontal line), with the cost that MSE without attack being slightly larger. In Fig. 5b, as the magnitude of attack signal increases, MSE of Kalman estimator increases significantly while our proposed secure estimator holds low MSE despite increasing magnitude of injected attack signal.
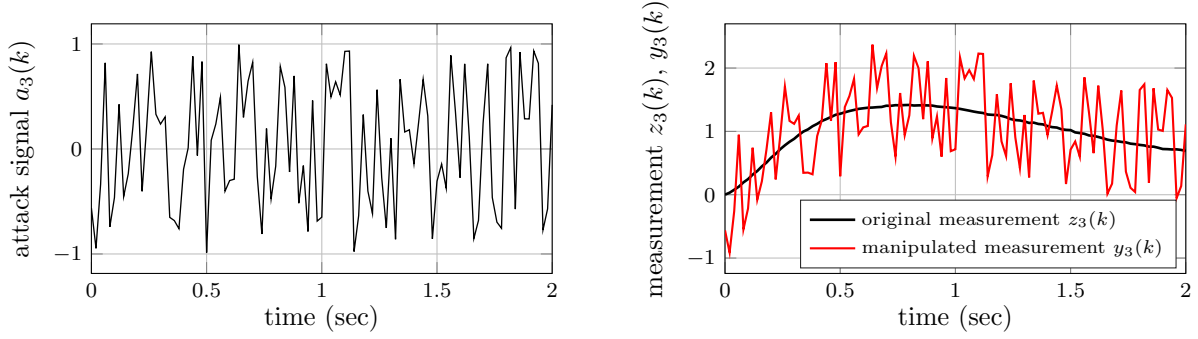
Figure 3: Attack signal and measurements of sensor 3. The attack signal is uniformly distributed in interval $(-1, 1)$.
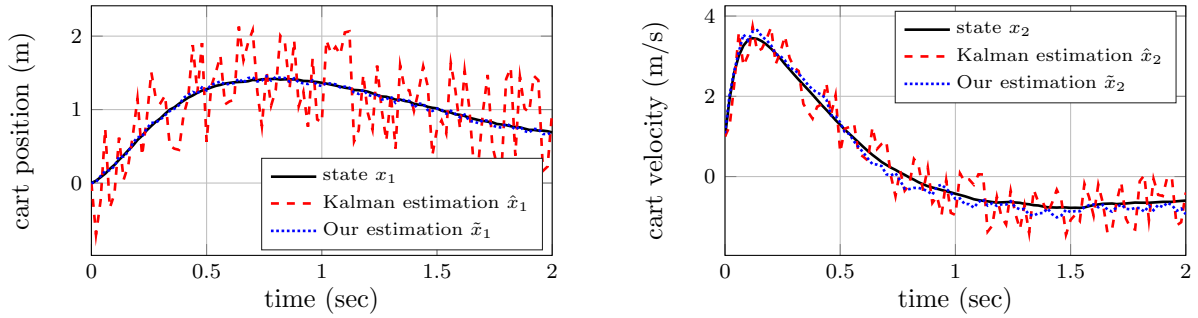

Figure 4: Estimation of states under attack on sensor 3. The attack signal is uniformly distributed in interval $(-1, 1)$.

5 Conclusion

This paper considers LTI system with Gaussian noise against sparse integrity attack on a subset of sensors. Under the geometric multiplicity assumption on stable eigenvalues of A, we propose an estimation scheme that is secure to (p, m)-sparse attack as long as the system is $2p$-sparse detectable. To achieve this, we first prove that the span of the rows of $G_i^{\mathcal{U}}$ is equivalent to the observable space corresponding to unstable states, based on which the canonical form H_i is designed. The proposed estimator is formulated as a convex optimization problem based on H_i, where by careful design, the estimation of stable states is always secured. Moreover, in the absence of attack, the proposed estimation coincides with Kalman estimation for certain probability, which can be adjusted by tuning parameter γ to balance between the performance with and without attack. We further prove that the $2p$-sparse detectability is the fundamental limit of secure dynamic estimation. Our proposed estimator achieves this fundamental limit with good performance in the absence of attack and low computation complexity.

Appendix

A Proof of Theorem 2

Proof. Define the characteristic polynomial of A as $p(x) = a_n x^n + \cdots + a_1 x + a_0$. Define polynomial fraction $q_\pi(x)$ with respect to constant π as $q_\pi(x) = \frac{p(x) - p(\pi)}{x - \pi}$ where $x \neq \pi$. Therefore,

$$q_\pi(A)(A - \pi I) = p(A) - p(\pi)I = -p(\pi)I,$$

14
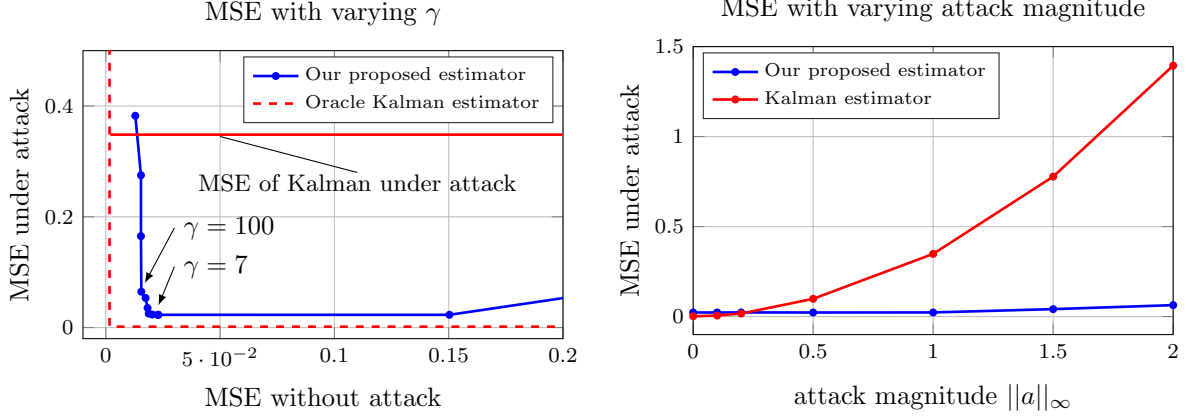
(a) Estimation mean square error (MSE) with varying tuning parameter γ. The attack magnitude is $\|a\|_\infty = 1$.

(b) Estimation mean square error (MSE) with varying attack magnitude. The parameter γ is set as $\gamma = 10$.

Figure 5

where the last equality comes from Cayley-Hamilton Theorem. As a result, when π is not the eigenvalue of A, we have

$$(A - \pi I)^{-1} = -\frac{1}{p(\pi)} q_\pi(A). \tag{27}$$

In order to simplify notations, we define

$$b_{j,k} \triangleq -\frac{1}{p(\pi_j)} \sum_{i=0}^{n-k-1} a_{i+k+1} \pi_j^i, \tag{28}$$

where π_j is the j-th diagonal element of Π, i.e., j-th eigenvalue of $A - KCA$ as defined in (4). According to (27), the j-th row of matrix G_i can be reformulated as

$$C_i A \left(A - \pi_j I\right)^{-1} = \begin{bmatrix} b_{j,0} & b_{j,1} & \cdots & b_{j,n-1} \end{bmatrix} O_i A.$$

Therefore, G_i can be interpreted as follows

$$G_i = \begin{bmatrix} b_{1,0} & b_{1,1} & \cdots & b_{1,n-1} \\ b_{2,0} & b_{2,1} & \cdots & b_{2,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n,0} & b_{n,1} & \cdots & b_{n,n-1} \end{bmatrix} O_i A = \mathcal{D}_1 \mathcal{D}_2 \mathcal{D}_3 O_i A,$$

where $\mathcal{D}_1 \triangleq \mathrm{diag}\left(-\frac{1}{p(\pi_1)}, -\frac{1}{p(\pi_2)}, \cdots, -\frac{1}{p(\pi_n)}\right),$

$$\mathcal{D}_2 \triangleq \begin{bmatrix} \pi_1^{n-1} & \pi_1^{n-2} & \cdots & 1 \\ \pi_2^{n-1} & \pi_2^{n-2} & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ \pi_n^{n-1} & \pi_n^{n-2} & \cdots & 1 \end{bmatrix}, \mathcal{D}_3 \triangleq \begin{bmatrix} a_n & 0 & \cdots & 0 \\ a_{n-1} & a_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \cdots & a_n \end{bmatrix}.$$

According to Assumption 1, all π_j are distinct eigenvalues and they are not the eigenvalues of A, i.e. the diagonal matrix $\mathcal{D}_1$ and the Vandermonde matrix $\mathcal{D}_2$ are invertible. Moreover, $a_n = 1$. Therefore, the

15

lower triangular Toeplitz matrix $\mathcal{D}_3$ is invertible and thus rowspan(G_i) = rowspan(O_iA). We continue to prove rowspan(O_i) = rowspan(O_iA). Considering that $A^n = -a_{n-1}A^{n-1} - \cdots - a_0 I$, one obtains the following equation (29).

$$O_i A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix} O_i. \tag{29}$$

According to Assumption 1, A is invertible and $a_0 = (-1)^n \det(A) \neq 0$, which leads to the equation that rowspan(O_i) = rowspan(O_iA). As a result, rowspan$(G_i^{\mathcal{U}})$=rowspan$(O_i^{\mathcal{U}})$. We continue to prove that rowspan$(O_i^{\mathcal{U}})$ = rowspan$(H_i^{\mathcal{U}})$. Since A_1 is assumed to be in the Jordan canonical form and all eigenvalues have geometric multiplicity 1, one can verify that nonzero columns of $O_i^{\mathcal{U}}$ are linear independent. Therefore, $i \in \mathcal{E}_j$ is equivalent to that j-th column of O_i is non-zero, i.e., $O_i^{\mathcal{U}}$ has the same row-span with the canonical form $H_i^{\mathcal{U}}$.

$\square$

B   Proof of Theorem 3

Proof of Theorem 3. Considering the KKT condition of problem (20), one obtains that if

$$\left\| \mathcal{W} \begin{bmatrix} \mu(k) \\ NH\tilde{x}(k) \end{bmatrix} \right\|_\infty \leq \gamma,$$

then the solution $\nu(k)$ satisfy that $\nu(k) = \mathbf{0}$. In this scenario, solutions to problem (20) and problem (15) are equivalent and the solution $\tilde{x}(k), \mu(k), \nu(k)$ satisfy

$$\tilde{x}(k) = \tilde{x}_{\mathrm{ls}}(k) = \hat{x}(k), \ \mu(k) = \varphi(k), \ \nu(k) = \mathbf{0}. \tag{30}$$

According to Lemma 2, the solution $\varphi(k)$ of problem (15) satisfy the following equation:

$$\tilde{x}_{\mathrm{ls}}(k) = \hat{x}(k), \ \varphi(k) = (I - GF)\epsilon(k), \tag{31}$$

where $\hat{x}(k)$ is the fixed gain Kalman estimation defined in (3). Combining (31) and (30), result in Theorem 3 is obtained. $\square$

C   Proof of Theorem 4

Before proving Theorem 4, we need the following Lemma. Define the number of honest sensors and compromised sensors (w.r.t. compromised set $\mathcal{I}$) that can observe state j as:

$$h_j(\mathcal{I}) \triangleq |\mathcal{E}_j \cap \mathcal{I}^c|, \ c_j(\mathcal{I}) \triangleq |\mathcal{E}_j \cap \mathcal{I}|.$$

We have the following lemma quantifying the property of $h_j(\mathcal{I})$ and $c_j(\mathcal{I})$.

Lemma 3. *The following two propositions are equivalent.*

1. *The system is $2p$-sparse observable.*

2. *For any $\mathcal{I}$ with $|\mathcal{I}| = p$, the inequality $c_j(\mathcal{I}) < h_j(\mathcal{I})$ holds for all $j \in \{1, 2, \cdots, n\}$.*

Proof of Lemma 3. We prove the contrapositive of (1)$\Rightarrow$(2). Supposing that there exists j^* and $\mathcal{I}^*$ with $|\mathcal{I}^*| = p$ such that $c_{j^*}(\mathcal{I}^*) \geq h_{j^*}(\mathcal{I}^*)$, then $h_{j^*}(\mathcal{I}^*) \leq c_{j^*}(\mathcal{I}^*) \leq |\mathcal{I}^*| = p$. Noticing that $c_j(\mathcal{I}) + h_j(\mathcal{I}) = |\mathcal{E}_j|$ holds for all $\mathcal{I}$, we have $|\mathcal{E}_{j^*}| \leq 2p$. There exists set $\mathcal{A}$ that satisfy $\mathcal{A} \supseteq \mathcal{E}_{j^*}$ and $|\mathcal{A}| = 2p$. According to the definition of $\mathcal{E}_{j^*}$, there exists no sensor in set $\mathcal{J} \setminus \mathcal{A}$ who can observe state j^*, i.e.,

$$e_{j^*} \notin \text{rowspan}(O_i), \ \forall i \in \mathcal{J} \setminus \mathcal{A}.$$

As a result, system $(A, C_{\mathcal{J}\setminus\mathcal{A}})$ is not $2p$-sparse observable according to Definition 2.

We proceed to prove (2)$\Rightarrow$(1). Since for any $\mathcal{I}$ with $|\mathcal{I}| = p$, $h_j(\mathcal{I}) > c_j(\mathcal{I}) \geq 0$, the system sparse observability index is at least p. Therefore, for each $j \in \{1, 2, \cdots, n\}$, there exists an $\mathcal{I}^*$ such that $c_j(\mathcal{I}^*) = p$, and thus $|\mathcal{E}_j| = h_j(\mathcal{I}^*) + c_j(\mathcal{I}^*) \geq 2p + 1$. According to the definition $\mathcal{E}_j$, there are at least $2p + 1$ sensors that can observe sensor j, and the system is $2p$-sparse observable.

$\square$

We need the following notations for the proof. Define the unstable part and stable parts of x as the following where $x_u \in \mathbb{C}^{n_u \times 1}$ and $x_s \in \mathbb{C}^{n_s \times 1}$. Similarly, divide matrix H_i into four parts based on Theorem 2 where $H_{uu,i} \in \mathbb{C}^{n_u \times n_u}$ and $H_{ss,i} \in \mathbb{C}^{n_s \times n_s}$.

$$x = \begin{bmatrix} x_u \\ x_s \end{bmatrix}, \ H_i = \begin{bmatrix} H_{uu,i} & H_{us,i} \\ \mathbf{0}_{n_s \times n_u} & H_{ss,i} \end{bmatrix}.$$

Define $\eta_i \triangleq P_i \zeta_i$. Similar to x, $\tilde{x}$ and η_i are also divided to $\tilde{x}_u, \tilde{x}_s, \eta_{i,u}, \eta_{i,s}$ in the same way.

Proof of Theorem 4. Consider the KKT condition of problem (20) and denote the dual variables for equation constraints as $\lambda = [\lambda_1', \cdots, \lambda_m']' \in \mathbb{C}^{mn \times 1}$:

$$(\tilde{M}^{-1} + N'N)\mu + N'NH\tilde{x} - \lambda = \mathbf{0}, \tag{32}$$

$$H'N'N\mu + H'N'NH\tilde{x} - H'\lambda = \mathbf{0}, \tag{33}$$

$$\gamma \cdot \tilde{\nu} - \lambda = \mathbf{0}, \tag{34}$$

$$Y - H\tilde{x} - \mu - \nu = \mathbf{0}, \tag{35}$$

where $\tilde{\nu}$ belongs to the subgradient of $\|\nu\|_1$, i.e., for the i-th entry:

$$\begin{cases} [\tilde{\nu}]_i = -1, & [\nu]_i < 0 \\ [\tilde{\nu}]_i = 1, & [\nu]_i > 0 \\ [\tilde{\nu}]_i \in [-1, 1], & [\nu]_i = 0 \end{cases}.$$

Combining (32) and (33) leads to:

$$\begin{bmatrix} \tilde{M}^{-1} + N'N & N'NH \\ H'N'N & H'N'NH \end{bmatrix} \begin{bmatrix} \mu \\ \tilde{x} \end{bmatrix} = \begin{bmatrix} \lambda \\ H'\lambda \end{bmatrix}. \tag{36}$$

According to the definition of N, the first n_u rows of $H'N'N$ are zeros. Therefore, we extract the non-zeros part of equation (36) in the following:

$$\begin{bmatrix} \tilde{M}^{-1} + N'N & N'NH\mathcal{L}' \\ \mathcal{L}H'N'N & \mathcal{L}H'N'NH\mathcal{L}' \end{bmatrix} \begin{bmatrix} \mu \\ \tilde{x}_s \end{bmatrix} = \begin{bmatrix} \lambda \\ \mathcal{L}H'\lambda \end{bmatrix}. \tag{37}$$

where $\mathcal{L} \triangleq \begin{bmatrix} \mathbf{0}_{n_s \times n_u} & I_{n_s} \end{bmatrix}$. Rewrite (37) as:

$$\left(\begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & \mathcal{L}H'N' \end{bmatrix} \mathcal{W} \begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & NH\mathcal{L}' \end{bmatrix} \right) \begin{bmatrix} \mu \\ \tilde{x}_s \end{bmatrix} = \begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & \mathcal{L}H' \end{bmatrix} \lambda. \tag{38}$$

Notice that $\mathcal{W}$ is positive definite and $\begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & \mathcal{L}H'N' \end{bmatrix}$ is full row-rank, due to the Frobenius rank inequality, the matrix on the left of (38) is also invertible, and thus the following matrix is well-defined:

$$\mathcal{F} \triangleq \left(\begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & \mathcal{L}H'N' \end{bmatrix} \mathcal{W} \begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & NH\mathcal{L}' \end{bmatrix} \right)^{-1} \begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & \mathcal{L}H' \end{bmatrix}. \tag{39}$$

According to (34), $\|\lambda\|_\infty \le \gamma$. Therefore we have the following from (38)

$$\left\| \begin{bmatrix} \mu \\ \tilde{x}_s \end{bmatrix} \right\|_\infty \le \gamma \cdot \|\mathcal{F}\|_\infty. \tag{40}$$

Now we continue to prove that the estimation of unstable states $\tilde{x}_u$ are resilient. Rewrite the optimization problem (20) as

$$\underset{\tilde{x},\ \mu}{\text{minimize}} \quad \frac{1}{2} \begin{bmatrix} \mu \\ NH\tilde{x} \end{bmatrix}' \mathcal{W} \begin{bmatrix} \mu \\ NH\tilde{x} \end{bmatrix} + \gamma \|Y - \mu - H\tilde{x}\|_1$$

where the time index is omitted for notation simplicity. Consider the 1-norm term in the objective function:

$$\|Y - \mu - H\tilde{x}\|_1 = \sum_{i=1}^m \|\eta_{i,u} - \mu_{i,u} - (H_{uu,i}\tilde{x}_u + H_{us,i}\tilde{x}_s)\|_1 + \sum_{i=1}^m \|\eta_{i,s} - \mu_{i,s} - H_{ss,i}\tilde{x}_s\|_1$$

where $\eta_{i,u}, \mu_{i,u}$ is the vector composed of first n_u element of η_i, μ_i and $\eta_{i,s}, \mu_{i,s}$ is the vector composed of last n_s element of η_i, μ_i. Suppose that μ and $\tilde{x}_s$ have taken the value of optimal solution $\mu^*, \tilde{x}_s^*$, it is sufficient to minimize the following :

$$\min_{\tilde{x}_u} \sum_{i=1}^m \left\| \eta_{i,u} - \mu_{i,u}^* - H_{us,i}\tilde{x}_s^* - H_{uu,i}\tilde{x}_u \right\|_1. \tag{41}$$

Define $\xi_i \triangleq \eta_{i,u} - \mu_{i,u}^* - H_{us,i}\tilde{x}_s^*$ and recall $[\cdot]_j$ is the j-th entry of a vector. The objective function in (41) can be written as

$$\sum_{i=1}^m \sum_{j=1}^{n_u} |[\xi_i]_j - [H_{uu,i}\tilde{x}_u]_j| = \sum_{j=1}^{n_u} \sum_{i \in \mathcal{E}_j} |[\xi_i]_j - \tilde{x}_j|. \tag{42}$$

where $\mathcal{E}_j$ is the index set of sensors that can observe state j that is defined in (12). For each unstable state $j \in \mathcal{U}$, the minimizer $\tilde{x}_j$ of objective (42) could be explicitly written as the median of all $[\xi_i]_j$ among $i \in \mathcal{E}_j$.

Before proving that $\tilde{x}_j$ is bounded, let us define the following operator: $f_i : R \times R \times \cdots \times R \to R$, such that $f_i (\alpha_l, l \in \{1, \cdots, L\})$ equals to the i-th smallest element in the set $\{\alpha_1, \ldots, \alpha_L\}$. For even number i, we further define

$$f_{\frac{i+1}{2}} = \left(f_{\frac{i}{2}} + f_{\frac{i}{2}+1} \right) / 2.$$

Thus, $f_{(L+1)/2} (\alpha_l, l \in \{1, \cdots, L\})$ is the median number of set $\{\alpha_1, \ldots, \alpha_L\}$ and the solution to problem (41) is

$$\tilde{x}_j = f_{(|\mathcal{E}_j|+1)/2} ([\xi_i]_j, i \in \mathcal{E}_j), j \in \mathcal{U}.$$

Define the uncorrupted data corresponding to sensor i as $\eta_i^o = P_i \zeta_i^o$. Define ξ_i^o correspondingly as $\xi_i^o \triangleq \eta_{i,u}^o - \mu_{i,u}^* - H_{us,i}x_s^*$. Recalling that the number of honest sensors and compromised sensors that can observe unstable state $j \in \mathcal{U}$ are h_j and c_j, we have

$$f_{(h_j - c_j)} ([\xi_i^o]_j, i \in \mathcal{E}_j) \le f_{(m+1)/2} ([\xi_i]_j, i \in \mathcal{E}_j), \tag{43}$$
$$f_{(m+1)/2} ([\xi_i]_j, i \in \mathcal{E}_j) \le f_{2c_j} ([\xi_i^o]_j, i \in \mathcal{E}_j). \tag{44}$$

According to Lemma 3, $h_j - c_j > 0$ and $2c_j < h_j + c_j = |\mathcal{E}_j|$. As a result, according to (43) and (44), one obtains

$$\min\left\{[\xi_i^o]_j, i \in \mathcal{E}_j\right\} \le \tilde{x}_j \le \max\left\{[\xi_i^o]_j, i \in \mathcal{E}_j\right\}, \ j \in \mathcal{U}. \tag{45}$$

Consider the following optimization problem where observation are not influenced by attack:

$$\underset{\tilde{x}^o, \mu^o}{\text{minimize}} \quad \frac{1}{2}\begin{bmatrix} \mu^o \\ NH\tilde{x}^o \end{bmatrix}' \mathcal{W} \begin{bmatrix} \mu^o \\ NH\tilde{x}^o \end{bmatrix} + \gamma^o \left\| Y^o - \mu^o - H\tilde{x}^o \right\|_1,$$

where Y^o is composed of $P_i \zeta_i^o$. Denote the solution to this problem as $\tilde{x}^o, \mu^o$. According to Theorem 3, by choosing

$$\gamma^o = \left\| \mathcal{W}\begin{bmatrix} (I - GF)\,\epsilon^o(k) \\ NH\hat{x}^o(k) \end{bmatrix} \right\|_\infty,$$

the solution coincides with Kalman estimation, i.e., $\tilde{x}^o(k) = \hat{x}^o(k)$. Similar to previous analysis, the solution $\tilde{x}^o$ satisfies

$$[\tilde{x}^o]_j = f_{(|\mathcal{E}_j|+1)/2}\left([P_i \zeta_i^o - \mu_i^o]_j, i \in \mathcal{E}_j\right), \forall j \in \mathcal{U} \cup \mathcal{S}. \tag{46}$$

Combining (45) and (46) leads to that, for every $j \in \mathcal{U}$,

$$\left|[\tilde{x}]_j - [\tilde{x}^o]_j\right| = \left|[\tilde{x}]_j - [\hat{x}^o]_j\right| \le \max_{i_1, i_2 \in \mathcal{E}_j} \left|\left[P_{i_1}\zeta_{i_1}^o(k)\right]_j - \left[P_{i_2}\zeta_{i_2}^o(k)\right]_j\right| + \|\mu^*\|_\infty + \|\mu^o\|_\infty$$

$$\le \max_{i_1, i_2 \in \mathcal{E}_j} \left|\left[P_{i_1}\zeta_{i_1}^o(k)\right]_j - \left[P_{i_2}\zeta_{i_2}^o(k)\right]_j\right| + (\gamma + \gamma^o)\|\mathcal{F}\|_\infty.$$

Recall that $P_i \epsilon_i(k) = P_i \zeta_i(k) - H_i x(k)$. Since for all $i_1, i_2 \in \mathcal{E}_j$, one obtains $[H_{i_1} x(k)]_j = [x(k)]_j = [H_{i_2} x(k)]_j, \forall k \in \mathbb{Z}^+$. Thus, we have

$$\max_{i_1, i_2 \in \mathcal{E}_j} \left|\left[P_{i_1}\zeta_{i_1}^o(k)\right]_j - \left[P_{i_2}\zeta_{i_2}^o(k)\right]_j\right| = \max_{i_1, i_2 \in \mathcal{E}_j} \left|\left[P_{i_1}\epsilon_{i_1}^o(k)\right]_j - \left[P_{i_2}\epsilon_{i_2}^o(k)\right]_j\right|,$$

whose variance is uniformly bounded for all k according to Lemma 1. Similarly, γ^o is also uniformly bounded for all k. Recalling that for $j \in \mathcal{S}$, we have $|\tilde{x}_j(k)| \le \gamma \cdot \|\mathcal{F}\|_\infty$ from (40). Since $\hat{x}_j^o(k)$ is the oracle Kalman estimation of a stable state, its variance is always bounded. As a result, our estimation $\tilde{x}(k)$ is secure according to Definition 3. $\square$

References

[1] U.S. Department of Homeland Security (DHS) S and T Cyber Security Division. Cyber security division technology guide 2018, 2018.

[2] Alvaro Cárdenas, Saurabh Amin, Bruno Sinopoli, Annarita Giani, Adrian Perrig, Shankar Sastry, et al. Challenges for securing cyber physical systems. In *Workshop on future directions in cyber-physical systems security*, volume 5, 2009.

[3] R. Langner. Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security Privacy*, 9(3):49–51, May 2011.

[4] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong. The 2015 ukraine blackout: Implications for false data injection attacks. *IEEE Transactions on Power Systems*, 32(4):3317–3318.

[5] G. Andersson, P. Donalek, R. Farmer, N. Hatziargyriou, I. Kamwa, P. Kundur, N. Martins, J. Paserba, P. Pourbeik, J. Sanchez-Gasca, R. Schulz, A. Stankovic, C. Taylor, and V. Vittal. Causes of the 2003 major grid blackouts in north america and europe, and recommended means to improve system dynamic performance. *IEEE Transactions on Power Systems*, 20(4):1922–1928, 2005.

[6] Yilin Mo and Bruno Sinopoli. On the performance degradation of cyber-physical systems under stealthy integrity attacks. *IEEE Transactions on Automatic Control*, 61(9):2618–2624, 2016.

[7] Ilija Jovanov and Miroslav Pajic. Sporadic data integrity for secure state estimation. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 163–169, 2017.

[8] Kemi Ding, Xiaoqiang Ren, Daniel E. Quevedo, Subhrakanti Dey, and Ling Shi. Dos attacks on remote state estimation with asymmetric information. *IEEE Transactions on Control of Network Systems*, 6(2):653–666, 2019.

[9] Jing-Jing Yan and Guang-Hong Yang. Secure state estimation with switched compensation mechanism against dos attacks. *IEEE Transactions on Cybernetics*, pages 1–12, 2021.

[10] Derui Ding, Qing-Long Han, Xiaohua Ge, and Jun Wang. Secure state estimation and control of cyber-physical systems: A survey. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(1):176–190, 2021.

[11] H. Fawzi, P. Tabuada, and S. Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, 59(6):1454–1467, 2014.

[12] Miroslav Pajic, Insup Lee, and George J. Pappas. Attack-resilient state estimation for noisy dynamical systems. *IEEE Transactions on Control of Network Systems*, 4(1):82–92, 2017.

[13] Young Hwan Chang, Qie Hu, and Claire J. Tomlin. Secure estimation based kalman filter for cyber–physical systems against sensor attacks. *Automatica*, 95:399–412, 2018.

[14] Y. Shoukry and P. Tabuada. Event-triggered state observers for sparse sensor noise/attacks. *IEEE Transactions on Automatic Control*, 61(8):2079–2091, 2016.

[15] M. S. Chong, M. Wakaiki, and J. P. Hespanha. Observability of linear systems under adversarial attacks. In *2015 American Control Conference (ACC)*, pages 2439–2444, 2015.

[16] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada. Secure state estimation for cyber-physical systems under sensor attacks: a satisfiability modulo theory approach. *IEEE Transactions on Automatic Control*, 62(10):4917–4932, 2017.

[17] Shaunak Mishra, Yasser Shoukry, Nikhil Karamchandani, Suhas N. Diggavi, and Paulo Tabuada. Secure state estimation against sensor attacks in the presence of noise. *IEEE Transactions on Control of Network Systems*, 4(1):49–59, 2017.

[18] Y. Nakahira and Y. Mo. Attack-resilient $\mathcal{H}_2$, $\mathcal{H}_\infty$, and ℓ_1 state estimator. *IEEE Transactions on Automatic Control*, 63(12):4353–4360, 2018.

[19] Liwei An and Guang-Hong Yang. Secure state estimation against sparse sensor attacks with adaptive switching mechanism. *IEEE Transactions on Automatic Control*, 63(8):2596–2603, 2018.

[20] A. Lu and G. Yang. Secure switched observers for cyber-physical systems under sparse sensor attacks: A set cover approach. *IEEE Transactions on Automatic Control*, 64(9):3949–3955, 2019.

[21] X. Liu, Y. Mo, and E. Garone. Local decomposition of kalman filters and its application for secure state estimation. *IEEE Transactions on Automatic Control*, pages 1–1, 2020.

[22] J. M. Hendrickx, K. H. Johansson, R. M. Jungers, H. Sandberg, and K. C. Sou. Efficient computations of a security index for false data attacks in power networks. *IEEE Transactions on Automatic Control*, 59(12):3194–3208, 2014.

[23] Arpan Chattopadhyay and Urbashi Mitra. Attack detection and secure estimation under false data injection attack in cyber-physical systems. In *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6, 2018.

[24] Chengwei Wu, Zhongrui Hu, Jianxing Liu, and Ligang Wu. Secure estimation for cyber-physical systems via sliding mode. *IEEE Transactions on Cybernetics*, 48(12):3420–3431, 2018.

[25] Yanwen Mao, Aritra Mitra, Shreyas Sundaram, and Paulo Tabuada. On the computational complexity of the secure state-reconstruction problem, 2021.

[26] Zishuo Li and Yilin Mo. Low complexity secure state estimation design for linear system with non-derogatory a, 2021.

[27] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the royal statistical society series b-methodological*, 58:267–288, 1996.

[28] Yorie Nakahira and Yilin Mo. Attack-resilient h2, h-infinity, and l1 state estimator, 2018.