Received: 15 May 2021 | Revised: 28 May 2022 | Accepted: 19 June 2022

DOI: 10.1002/rnc.6264

SPECIAL ISSUE ARTICLE

WILEY

Efficient secure state estimation against sparse integrity attack for regular linear system

Zishuo Li | Yilin Mo

Department of Automation and BNRist, Tsinghua University, Beijing, China

Correspondence
Yilin Mo, Department of Automation and BNRist, Tsinghua University, Beijing 100084, China.
Email: ylmo@mail.tsinghua.edu.cn

Funding information
National Key Research and Development Program of China, Grant/Award Number: 2018AAA0101601

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 the system is regular and system matrix A is non-singular, we propose a secure estimation scheme that is resilient to p-sparse attack as long as the system is $2p$-sparse detectable, which achieves 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. Two numerical examples including the IEEE 68 bus test system are provided to corroborate the results and illustrate the performance of the proposed estimator.

KEYWORDS

Kalman filter, secure estimation, sparse integrity attack, sparse observability/detectability

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, for example, the Stuxnet launched on Iran's nuclear facilities,[3] power blackouts in Ukraine,[4] North America and Europe[5] and so forth. 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 Reference 10. This article 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 switching estimation approach based on fault identification. The error correction approach usually takes measurements in a finite time window and adapts 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 Reference 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] adapt the 2-norm batch optimization approach for state estimation and a customized gradient descent algorithm to solve it efficiently. These works provide fundamental limits for static estimation against integrity attack, that is, 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 switching method, where the system operator switches between multiple estimate candidates[15-17] or sensor subset combinations[18-21] based on the evaluation of their reliability by consistency checking or malicious detection algorithms. Pasqualetti et al. propose a detection-and-switching mechanism[22] for noise-free linear system secure estimation, and the switching method is further studied in scenarios of bounded noise[16] LTI systems, Gaussian random noise[19] LTI systems and multi-mode hybrid systems.[17,23] However, the combinatorial number of possible benign sensor sets poses severe challenges for storage and online computation. For example, if there are p corrupted sensors out of m sensors. The number of all possible benign sensor combinations is $\binom{m}{m-p}$, which proliferates rapidly as either m or p grows. In view of this problem, researchers proposed various methods to reduce the combinatorial complexity. As far as we know, the main attempts can be roughly classified into three categories, the sequential switching method, the set cover approach, and the satisfiability modulo theory approach. An et al.[20] propose a sequential approach where the estimator switches among an ordered list of all possible benign sensor combinations. They proved that the switching mechanism is guaranteed to stop at a benign combination within a finite time. Moreover, the algorithm only needs to maintain one combination at each time index, reducing the online computation complexity significantly, with the cost that the delay before hitting a benign set may be considerable. The set cover approach by Lu et al.[21] reduces the number of candidates by searching for smaller benign sensor sets with cardinality $m - 2p$, which is generated by solving the set cover problem. It is proved that the number of candidates is at least reduced by half. With the help of the satisfiability modulo theory, Mishra et al.[19] reduce the search space by pruning in the process of malicious sensor detection. It is proved that the number of iterations is at least reduced to $\binom{m}{m-2p+1}$. This reduction is significant when p is nearly half of m.

Different from the error correction or switching mechanism, our proposed method achieves secure estimation by adapting the decomposition-fusion scheme proposed by Liu et al.,[24] whose estimation scheme decomposes Kalman filters into local estimators and recovers the state estimate by securely fusing the local estimates using 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 detectability, which is the fundamental limit[16] for secure state estimation. Similar to Liu's result,[24] other results in the literature either impose $2p$-sparse observability conditions,[14,18,19] or more restrictive conditions than $2p$-sparse observability,[11,25] and the conditions are NP-hard to validate. In view of these problems, we pursue lower computational complexity by analyzing the observability structure of local estimators under the assumption that the system is regular, which is inspired by the work[26] of Mao et al. With this assumption, the system observability structure is simple and the sparse observability index can be calculated in polynomial time.

In summary, we focus on LTI systems with Gaussian noise and intend to propose an estimation scheme that provides the following two contributions: (1) Most of the secure estimators in the literature are designed to work under $2p$-sparse observability or more restrictive conditions, while for dynamic estimation problem, the fundamental limit is the less restrictive $2p$-sparse detectability.[16] Our proposed estimation scheme achieves this fundamental limit. (2) For general system, computing the sparse observability/detectability is an NP-hard problem[26] and thus validating the estimator security is also NP-hard. Nevertheless, with the assumption that the system is regular, we achieve polynomial computation complexity with respect to sensor number and system dimension.

Preliminary versions of some results have been presented in Reference 27. This article is significantly improved from the previous work and has the following merits:

- **Theoretical achievability:** The proposed estimator in this article is secure as long as the system is $2p$-sparse detectable, and this is proved to achieve the fundamental limit of secure dynamic estimation.
- **Off-line computational complexity:** Under the assumption that the system is regular and A is non-singular, the calculation of sparse detectability index can be done within polynomial time w.r.t. sensor number and system dimension.
- **On-line computational complexity:** The secure estimation can be obtained by solving a convex optimization problem, which can be done efficiently by developed optimization algorithms and off-the-shelf solvers.
- **Accuracy balancing between normal operation and under attack:** The proposed estimator has a tuning parameter γ. This article proves that larger γ implies a larger probability of recovering the Kalman estimation in the absence of attack (i.e., better performance without attack), and smaller γ implies smaller estimation error upper bound under attack (i.e., better performance under attack). By setting proper γ, the system can achieve a better trade-off between both scenarios.

The organization of this article is that, we introduce the problem formulation and preliminary results in Section 2, where the problems of previous results in the literature are analyzed. Then, the main results of this article are provided in Section 3 and collaborated by numerical simulations in Section 4. Section 5 finally concludes the article.

Notations: The set of natural numbers (non-negative integers) is denoted as $\mathbb{N}$. Cardinality of a set S is denoted as $|S|$. A' represents conjugate transpose of matrix A. Diagonal matrix with diagonal elements $A_1, \dots, A_k$ is denoted as $\mathrm{diag}(A_1, \dots, A_k)$. 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. $\mathbb{R}^{n\times 1}$ is also written as $\mathbb{R}^n$. $\mathrm{Cov}(\cdot)$ denotes the covariance of a random vector. The ith 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. Matrix inequality $A \prec B$ means that $B - A$ is positive define.

2 | PROBLEM FORMULATION AND LOCAL ESTIMATOR DESIGN

2.1 | Secure dynamic state estimation

In this article, 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 ith 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 $\{w(k)\}$. We further introduce an assumption that is common in estimation problems.

Assumption 1. The pair (A, C) is observable.

Notice that the above assumption is without loss of generality, as one can always perform a Kalman decomposition and only consider the observable space. 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[11,18] that the attacker can only compromise a fixed subset of sensors with known maximum cardinality. Denote the index set of all sensors as $\mathcal{I} \triangleq \{1, 2, \dots, m\}$. Define the support of vector $a \in \mathbb{R}^n$ as $\text{supp}(a) \triangleq \{i | 1 \leq i \leq n, a_i \neq 0\}$ where a_i is the ith entry of vector a. We have the following assumptions on the malicious adversary.

Definition 1 (Sparse attack). The attack $a(k)$ is a p-sparse attack if the vector sequence $\{a(k)\}_{k=0}^{\infty}$ satisfy that, there exists a time invariant index set $\mathcal{C} \subseteq \mathcal{I}$ with $|\mathcal{C}| = p$ such that $\bigcup_{k=0}^{\infty} \text{supp}\,\{a(k)\} = \mathcal{C}$. The set of all admissible p-sparse attack is defined as $\mathbb{A}(p)$.

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

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

Define $y(k_1 : k_2)$ as the sequence $\{y(k_1), y(k_1 + 1), \dots, y(k_2)\}$. Similar notation is also applied on $z(k), u(k)$. An estimator is a 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:

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

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

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

$$q_k \triangleq g_k\left(z, u\right) - g_k\left(y, u\right) = g_k\left(z, u\right) - g_k\left(z + a, u\right).$$

The estimator is said to be secure against p-sparse attack if there exists a constant scalar† M_{cov} such that the following holds:

$$\sup_{a \in \mathbb{A}(p), k \in \mathbb{N}} \{\text{Cov}\left(q_k\right)\} < M_{\text{cov}} I_n,$$

where I_n is the $n \times n$ identity matrix.

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

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

where

$$\begin{aligned}
&\hat{x}(k|k-1) = A\hat{x}(k-1) + Bu(k), P(k|k-1) = AP(k-1)A' + Q, \\
&K(k) = P(k|k-1)C'\left(CP(k|k-1)C' + R\right)^{-1},
\end{aligned}$$

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

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

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

†The constant may change under different noise covariances Q, R and different system parameter A, C.

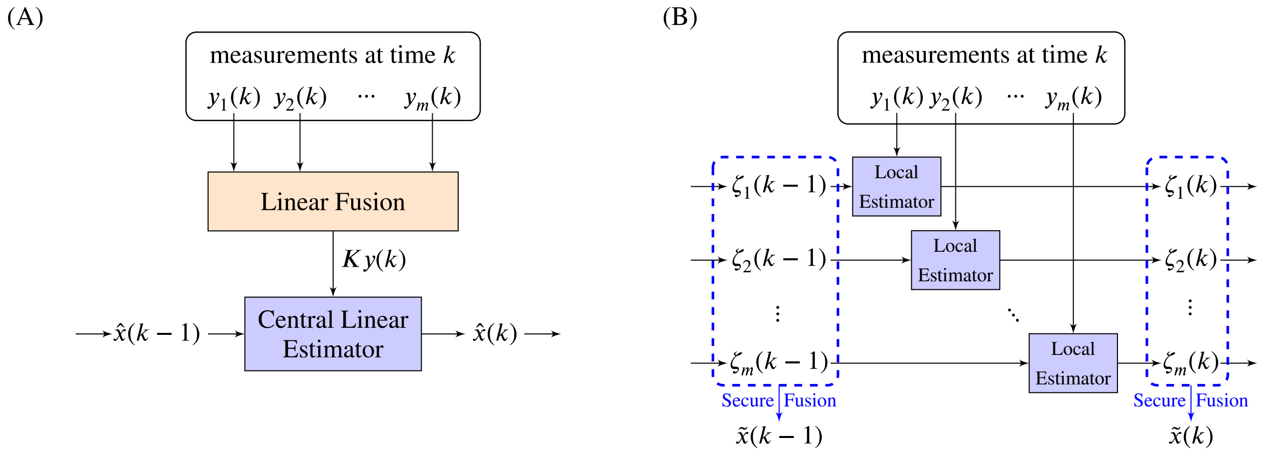

FIGURE 1 Information flow comparison between linear estimator and our design. (A) The information flow of a linear estimator (also known as Luenberger observer). $\hat{x}(k)$ is the estimated system state at time k; (B) The information flow of our proposed estimation scheme. $\tilde{x}(k)$ is the proposed secure estimation at time k

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:

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

In order to better convey our design idea of secure estimators, we illustrate the information flow of Kalman estimator (3) and our proposed secure estimator in Figure 1. As is shown in Figure 1A, for linear estimator (also known as Luenberger observer), all measurements $y(k) = \left[y_1'(k), y_2'(k), \dots, y_m'(k)\right]'$ are first multiplied by a gain matrix K to obtain $Ky(k)$. The result $Ky(k) = \sum_{i=1}^{m} K_i y_i(k)$ (K_i is ith column of K) can be seen as the linear fusion of measurements $y_1(k), y_2(k) \dots, y_m(k)$, and then the linearly fused measurements are used for the estimation update. In contrast, as shown in Figure 1B, our idea is to decompose the Kalman estimator into local linear estimators. Thus, the impact of the corrupted sensor i is isolated within the local estimator ζ_i, and other benign local estimators are not affected by the attack. Then, by a secure fusion scheme, the partly corrupted local estimations are fused to a secure estimation.

The following subsection introduces the details of local decomposition of Kalman estimator and previous results based on the local estimator design.

2.2 | Local decomposition of Kalman estimator

The following assumption is introduced to prevent system degradation.

Assumption 2. The matrix A is invertible; $A - KCA$ has n distinct eigenvalues. Moreover, $A - KCA$ and A do not share any eigenvalues.

Remark 1. From a theoretical perspective, the assumption of invertibility of matrix A is used to analyze the observability structure and guarantees that row span of G_i equals to observable space of sensor i (denoted as $\mathbb{O}_i$). If this assumption does not hold, it happens that rowspan $G_i \subsetneq \mathbb{O}_i$, which brings sophisticated discussion, but the secure estimation design may still exist. From a practical perspective, most discrete-time systems are derived from the discretization of a continuous-time system, thus $A_d = \exp(A_c \cdot \Delta t)$, where A_d and A_c are the system matrices for the discrete and continuous-time systems respectively and Δt is the sampling interval. In this case, the inversion of system matrix $A_d^{-1} = \exp(-A_c \cdot \Delta t)$ is always well-defined.

Remark 2. From a practical perspective, we can freely assign the poles of $A-KCA$ by choosing a proper gain K since (A, CA) is observable[‡]. Hence, $A-KCA$ can satisfy Assumption 2 with a small estimation performance loss compared to Kalman estimator. From a theoretical perspective, a potential method to remove the assumption that $A-KCA$ has n distinct eigenvalues is that we can use the Jordan canonical form of $A-KCA$ instead of a diagonal matrix.[24] However, for notation simplicity and conclusion conciseness of this article, we leave it for future work.

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

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

where Π is a diagonal matrix with the eigenvalues of $A-KCA$ as diagonal entries. Denote these n eigenvalues (diagonal entries) as $\pi_1, \dots, \pi_n$. We design a local estimator that only takes the observations from sensor i. The local estimation $\zeta_i(k)$ is a n-dimensional vector initialized as $\zeta_i(0)=\mathbf{0}$ and satisfies the following dynamic:

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

where C_i is ith row of matrix C, and G_i is defined as the following $n \times n$ matrix:

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

Remark 3. Equation (5) is our design of the local estimator dynamics and also the definition of local estimation $\zeta_i(k)$. It is not a straightforward derivation from (4). The proof that this design can recover Kalman estimation $\hat{x}(k)$ by a linear combination of $\zeta_i(k)$ is shown in the previous work[24] and thus not presented in this article due to space limits.

In the following, we claim that the local state $\zeta_i(k)$ is actually estimating a linear transform of system state, that is, $\zeta_i(k)$ is estimating $G_ix(k)$ and their difference $\epsilon_i(k) \triangleq \zeta_i(k) - G_ix(k)$ has bounded covariance.

Lemma 1. *Define $\epsilon_i(k) \triangleq \zeta_i(k) - G_ix(k)$, then $\epsilon_i(k)$ satisfies the following dynamics:*

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

Since Π is a strictly stable matrix[§] and $w(k), v_i(k)$ are zero mean Gaussian random variables, when the attack is absence, that is, $a_i(k)=0$, the residue $\epsilon_i(k)$ is a Gaussian process with zero mean and bounded covariance. Thus, $\zeta_i(k)$ is a stable estimate of $G_ix(k)$.

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

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

Since it has been proved in Reference 24 Corollary 1 that $G_iA - \mathbf{1}_nC_iA = \Pi G_i$, one can verify that Equation (7) holds. ■

Define $\tilde{Q} \in \mathbb{R}^{mn\times mn}$ as the covariance of noise term $(G_i - \mathbf{1}_nC_i)w(k) - \mathbf{1}_nv_i(k+1)$ for all i, that is,

$$\tilde{Q} \triangleq \mathrm{Cov}\left(\begin{bmatrix} G_1 - \mathbf{1}_nC_1 \\ \vdots \\ G_m - \mathbf{1}_nC_m \end{bmatrix} w(k)\right) + \mathrm{Cov}\left(\begin{bmatrix} \mathbf{1}_nv_1(k+1) \\ \vdots \\ \mathbf{1}_nv_m(k+1) \end{bmatrix}\right) = \begin{bmatrix} G_1 - \mathbf{1}_nC_1 \\ \vdots \\ G_m - \mathbf{1}_nC_m \end{bmatrix} Q \begin{bmatrix} G_1 - \mathbf{1}_nC_1 \\ \vdots \\ G_m - \mathbf{1}_nC_m \end{bmatrix}' + R \otimes \mathbf{1}_{n\times n},$$

[‡] This comes from Assumption 1 and the invertibility of A.

[§] "Strictly stable" means all eigenvalues are within the open unit disk. Π is strictly stable since it shares same eigenvalues with $A-KCA$ and the latter is strictly stable.

where $\otimes$ is the Kronecker product. Define $\tilde{\Pi}$ by stacking m blocks of Π on the diagonal:

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

Therefore, in the absence of attack, the covariance of $\epsilon(k) \triangleq \left[\epsilon_1(k)', \dots, \epsilon_m(k)'\right]'$ satisfies

$$\mathrm{Cov}(\epsilon(k+1)) = \tilde{\Pi} \cdot \mathrm{Cov}(\epsilon(k)) \cdot \tilde{\Pi}' + \tilde{Q}.$$

Since $\tilde{\Pi}$ is a strictly stable matrix, the covariance of $\epsilon(k)$ converges to the solution $\tilde{W}$ of the following Lyapunov equation:

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

The secure estimation can be recovered by the solution of the following optimization problem where $\zeta(k) \triangleq \left[\zeta_1(k)', \dots, \zeta_m(k)'\right]'$ and $G \triangleq \left[G_1', \dots, G_m'\right]'$.

$$\underset{\check{x}(k),\mu(k),\nu(k)}{\text{minimize}} \quad \frac{1}{2}\mu(k)'\tilde{W}^{-1}\mu(k) + \gamma\|\nu(k)\|_1 \tag{8a}$$

$$\text{subject to} \quad \zeta(k) = G\check{x}(k) + \mu(k) + \nu(k). \tag{8b}$$

The parameter γ is a non-negative constant chosen by the system operator. The following theorem from Liu et al.[24] proves that the solution $\check{x}(k)$ to problem (8) is a secure estimation under specific condition.

Theorem 1 (24). *In the presence of p-sparse attack, the state estimation $\check{x}(k)$ is secure if the following inequality holds for all $x \neq \mathbf{0}$, $x \in \mathbb{R}^n$:*

$$\sum_{i\in C} \|G_i x\|_1 < \sum_{i\in \mathcal{I}\setminus C} \|G_i x\|_1, \quad \forall\, C \subset \mathcal{I}, |C| \leq p. \tag{9}$$

2.3 | Improvement upon previous results

In this subsection, we analyze condition (9) and present our improvement upon the results in Theorem 1. 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 (9) is **computationally hard**. The computational complexity can be significantly reduced by introducing the regularity assumption and further analysis on matrix G_i.
(2) Condition (9) **does not achieve the fundamental limit**. It is more restrictive than $2p$-sparse observability and has a gap from it, as proved in Lemma 2 in the following.

Lemma 2. *For $p > 0$, condition (9) implies that system (A, C) is $2p$-sparse observable and the reverse does not hold.*

Proof. Define the observability matrix of system (A, C_i) as

$$O_i \triangleq \left[C_i' \;\middle|\; (C_iA)' \;\middle|\; \cdots \;\middle|\; \left(C_iA^{n-1}\right)'\right]'.$$

Define the observable subspace of sensor i as

$$\mathbb{O}_i \triangleq \mathrm{rowspan}(O_i) = \mathrm{span}\left(C_i', (C_iA)', \dots, \left(C_iA^{n-1}\right)'\right).$$

According to Cayley–Hamilton theorem, $(A - \pi_j I)^{-1}$ can be written as a polynomial of A. As a result, according to the definition of G_i in (6), $\text{rowspan}(G_i) \subset \mathbb{O}_i$. Define the following sensor index set with respect to vector $x \in \mathbb{R}^n$ as

$$\mathcal{K}_O(x) \triangleq \{i \in \mathcal{I} | O_i x \neq \mathbf{0}\},\ \mathcal{K}_G(x) \triangleq \{i \in \mathcal{I} | G_i x \neq \mathbf{0}\}.$$

Then we have for arbitrary $x \in \mathbb{R}^n$, $\mathcal{K}_G(x) \subset \mathcal{K}_O(x)$.

We first prove condition (9)$\Rightarrow$ $2p$-sparse observability by contradiction. If the system is not $2p$-sparse observable, according to the definition, there exists $x_0 \neq \mathbf{0}$ such that $|\mathcal{K}_G(x_0)| \leq |\mathcal{K}_O(x_0)| \leq 2p$. Define compromised sensor index set C_0 as

$$C_0 \triangleq \underset{|C|=p}{\arg\max} \sum_{i \in C} \|G_i x_0\|_1 . \tag{10}$$

As a result,

$$\sum_{i \in C_0} \|G_i x_0\|_1 \geq \frac{1}{2} \sum_{i \in \mathcal{I}} \|G_i x_0\|_1 \geq \sum_{i \in \mathcal{I} \setminus C_0} \|G_i x_0\|_1 . \tag{11}$$

This contradicts to condition (9).

We further prove that the reverse does not hold by giving a counter example. Suppose the system sparse observability index is $2p$. Then there exists $x_0 \in \mathbb{R}^n, x_0 \neq \mathbf{0}$, such that $|\mathcal{K}_O(x_0)| = 2p + 1$. Consider the case where $|\mathcal{K}_G(x_0)| \leq 2p$ due to A is not full-rank (this can be seen in the proof of Theorem 3 (2) in the Appendix). Define C_0 same as in (10). As a result, one obtains $\sum_{i \in C_0} \|G_i x_0\|_1 \geq \sum_{i \in \mathcal{I} \setminus C_0} \|G_i x_0\|_1$. This means condition (9) is violated and $2p$-sparse observability does not necessarily imply condition (9). ■

Lemma 2 proves that condition (9) is more restrictive than $2p$-sparse observability and has a gap from it. Moreover, $2p$-sparse observability implies $2p$-sparse detectability while the reverse does not hold according to Definition 2. Our proposed scheme in this article fills these gaps (seen in the following illustration). Moreover, the condition of $2p$-sparse detectability is the fundamental limit of this problem and will be proved in Theorem 7. The relationships between secure conditions in related works are illustrated in the following.

$$\begin{tikzcd}
 & & \text{Fundamental limit} \arrow[d] \\
\text{Condition (9)} \arrow[r, phantom, "\overset{\Rightarrow}{\nLeftarrow}"] & 2p\text{-sparse observability} \arrow[r, phantom, "\overset{\Rightarrow}{\nLeftarrow}"] & 2p\text{-sparse detectability} \\
\text{Result in Liu et al.}^{24} \arrow[u] & \text{Result in our previous work}^{27} \arrow[u] & \text{Result in this article} \arrow[u]
\end{tikzcd}$$

In the following section, we propose a secure estimation scheme that improves the aforementioned points. Under Assumption 2 and that the system is regular, 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, we first present two facts about the observable subspace of sensor i. These results facilitate the construction of the canonical form of matrix G_i. On the basis of this canonical form, we propose our secure estimator based on the separated design of unstable modes and stable modes. We need the following notations throughout the article.

Define the observability matrix of system (A, C_i) as

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

Define the observable subspace with respect to sensor i as

$$\mathbb{O}_i \triangleq \text{rowspan}(O_i) = \text{span}\left(C_i', (C_iA)', \ \dots\ , \left(C_iA^{n-1}\right)'\right),$$

where span$(\mathcal{S})$ is the linear span of the vectors in set $\mathcal{S}$, and rowspan(X) is the linear span of rows of matrix X.

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

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

where $\bar{A} \triangleq TAT^{-1}$ is similar to A and $\overline{x} = Tx$, we can assume that A is in Jodan canonical form without loss of generality. In the remaining of this section, we assume that A is in the Jordan canonical form and the eigenvalues are sorted (according to its magnitude) in a descendant order from upper left to lower right on the diagonal.

3.1 | Canonical form of G_i

Define set $\mathcal{J} \triangleq \{1, 2, \ \dots\ , n\}$ as the index of all states and recall that $\mathcal{I} \triangleq \{1, 2, \ \dots\ , m\}$ is the index of sensors. Define state index set $\mathcal{E}_i \subset \mathcal{J}$ and sensor index set $\mathcal{F}_j \subset \mathcal{I}$ as:

$$\mathcal{E}_i \triangleq \left\{j \in \mathcal{J} \mid O_i^{(j)} \neq \mathbf{0}\right\}, \ \mathcal{F}_j \triangleq \left\{i \in \mathcal{I} \mid O_i^{(j)} \neq \mathbf{0}\right\}, \tag{12}$$

where $O_i^{(j)}$ is the jth column of matrix O_i. Define e_j as the n-dimensional canonical basis vector with 1 on the jth entry and 0 on the other entries. We have the following theorem revealing the structure of observable subspace. The proof is provided in Appendix A for legibility.

Theorem 2. *The following two statements hold true.*

(1) *When all eigenvalues of A has geometric multiplicity 1, $\left\{e_j, j \in \mathcal{E}_i\right\}$ is a group of canonical basis vector of linear subspace $\mathbb{O}_i$, that is, $\mathbb{O}_i = \text{span}\left(e_j, j \in \mathcal{E}_i\right)$.*

(2) *If A is invertible, $\mathbb{O}_i = \text{rowspan}(G_i)$, where G_i is defined in (6).*

In the following, we propose the canonical form of G_i under the following assumption and based on the results in Theorem 2.

Assumption 3. All the **unstable** eigenvalues of A have geometric multiplicity 1, that is, the linear system is regular.

Recalling that A is in the Jordan canonical form and eigenvalues are sorted, then matrix A can be written as

$$A = \begin{pmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & A_2 \end{pmatrix}, \tag{13}$$

where block A_1 is composed of the Jordan blocks with unstable eigenvalues ($|\lambda| \geq 1$) 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 state entries of state as $\mathcal{U} \triangleq \{1, 2, \ \dots\ , n_u\}$ and index set of stable state entries as $\mathcal{S} \triangleq \{n_u + 1, \ \dots\ , n\}$. Notice that $\mathcal{U} \cup \mathcal{S} = \mathcal{J}$. Furthermore, a matrix X can be divided vertically into 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 (corresponding to unstable part) and $X^{\mathcal{S}}$ is composed of last n_s columns of X (corresponding to stable part).

Based on these notations, the following theorem follows from Theorem 2.

Theorem 3. *Assume system matrix A satisfies Assumption 3, then the following equation holds:*

$$\operatorname{rowspan}\left(G_i^{\mathcal{U}}\right) = \operatorname{rowspan}\left(O_i^{\mathcal{U}}\right) = \operatorname{rowspan}\left(H_i^{\mathcal{U}}\right), \tag{14}$$

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

$$H_i^{\mathcal{U}} \triangleq \left[\begin{array}{ccc} \mathbb{I}\{1 \in \mathcal{E}_i\} & & \\ & \ddots & \\ & & \mathbb{I}\{n_u \in \mathcal{E}_i\} \\ \hline & \mathbf{0}_{(n-n_u)\times n_u} & \end{array}\right], \tag{15}$$

and $\mathbb{I}\{j \in \mathcal{E}\}$ is the indicator function that takes the value 1 when $j \in \mathcal{E}$ is true and value 0 when $j \notin \mathcal{E}$. Therefore, there exists an invertible $n \times n$ matrix P_i such that $P_i G_i^{\mathcal{U}} = H_i^{\mathcal{U}}$.

Proof. Define transformation matrix

$$S \triangleq \begin{bmatrix} I_{n_u} & \mathbf{0}_{n_u \times n_s} \end{bmatrix} \in \mathbb{R}^{n_u \times n}. \tag{16}$$

Consider a linear system with system matrix SAS' and output matrix CS'. As a result, the observability matrix of this system with respect to sensor i is $\mathcal{O}_i^{\mathcal{U}}$. According to Assumption 2, SAS' is invertible. Plugging this unstable subsystem into Theorem 2 (2) leads to $\operatorname{rowspan}\left(G_i^{\mathcal{U}}\right) = \operatorname{rowspan}\left(O_i^{\mathcal{U}}\right)$. We proceed to prove $\operatorname{rowspan}\left(O_i^{\mathcal{U}}\right) = \operatorname{rowspan}\left(H_i^{\mathcal{U}}\right)$.

According to Assumption 3, all the eigenvalues of A_1 have geometric multiplicity one. Plugging A_1 into Theorem 2 (1), we have the following equation where e_j^u is the n_u-dimensional canonical basis vector with 1 on the jth entry and 0 on the other entries.

$$\operatorname{rowspan}\left(O_i^{\mathcal{U}}\right) = \operatorname{span}\left(e_j^u, j \in \mathcal{E}_i \cap \mathcal{U}\right) = \operatorname{span}\left(\begin{bmatrix} \mathbb{I}\{1 \in \mathcal{E}_i\} & & \\ & \ddots & \\ & & \mathbb{I}\{n_u \in \mathcal{E}_i\} \end{bmatrix}\right) = \operatorname{rowspan}\left(H_i^{\mathcal{U}}\right).$$

As a result, $\operatorname{rowspan}\left(O_i^{\mathcal{U}}\right) = \operatorname{rowspan}\left(H_i^{\mathcal{U}}\right)$ is proved. ■

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 detectability of sensor i. Therefore, the sparse detectability index can be directly obtained from matrix $H_i^{\mathcal{U}}$, or equivalently from set $\mathcal{E}_i$ or $\mathcal{F}_j$. We have the following theorem providing a efficient way to calculate the sparse detectability index.

Theorem 4. *Under Assumption 3, if $\mathcal{U} \neq \emptyset$, the sparse detectability index s of system (A, C) can be calculated as:*

$$s = \min_{j \in \mathcal{U}} |\mathcal{F}_j| - 1, \tag{17}$$

where $\mathcal{F}_j$ is defined in (12).

Remark 4. If $\mathcal{U} = \emptyset$, that is, the system matrix A is stable, we can always have trivial secure estimation (e.g., let the estimation be zero). Thus, we do not consider the sparse detectability index of stable system. If $\mathcal{U} \neq \emptyset$, the sparse detectability index is non-negative integer for detectable system.

Proof of Theorem 4. Define s as in (17). For arbitrary $\bar{s}$ that satisfies $\bar{s} \geq s + 1$, there exists a state index $j^* \in \mathcal{U}$ and a sensor index set C^* with $|C^*| = \bar{s}$ such that $C \supset \mathcal{F}_{j^*}$. As a result, state j^* cannot be observed by any sensor in $\mathcal{I} \setminus C^*$, that is,

$$e_{j^*}^u \notin \operatorname{rowspan}\left(O_i^{\mathcal{U}}\right), \ \forall i \in \mathcal{I} \setminus C^*,$$

and thus system $(A, C_{\mathcal{I}\setminus\mathcal{C}^*})$ is not detectable. As a result, the sparse detectability index cannot be larger than s defined in (17).

For arbitrary $\underline{s}$ that satisfies $\underline{s} \leq s$, arbitrary $\mathcal{C}$ with $|\mathcal{C}| = \underline{s}$, one obtains $\left(\cup_{j\in\mathcal{U}}\mathcal{F}_j\right) \setminus \mathcal{C} \neq \emptyset$, which means for all j, there exists $i^* \in \mathcal{I} \setminus \mathcal{C}$ such that: $e_j^u \in \text{rowspan}\left(O_{i^*}^{\mathcal{U}}\right)$. Therefore, system $(A, C_{\mathcal{I}\setminus\mathcal{C}})$ is detectable. According to Definition 2, the system is s-sparse detectable. ■

In conclusion, under Assumption 3, the matrix $G_i^{\mathcal{U}}$ has a canonical form $H_i^{\mathcal{U}}$. The canonical form is determined by sensor detectability, which can be checked easily by counting the number of non-zeros columns of matrix O_i. The index of non-zero columns of matrix O_i is recorded in set $\mathcal{E}_i$ and $\mathcal{F}_i$. Since the computation of O_i and checking non-zero columns only involve polynomial-time complexity w.r.t. sensor number m and system dimension n, the computation of sparse detectability index in (17) based on these sets can also be done in polynomial-time. Thus, the evaluation of system vulnerability can be done computational efficiently. The following subsection illustrates the theory introduced above by a numerical example.

3.2 | Illustrative example of detectability analysis

We provide an example to illustrate the state decomposition and canonical form H_i. Suppose

$$A = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}, \; C = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}.$$

The unstable eigenvalue is $\{2\}$ and the corresponding geometric multiplicity is 1. The stable eigenvalue is $\{1/2\}$ while the geometric multiplicity is 2. The unstable and stable index sets are respectively $\mathcal{U} = \{1, 2\}, \mathcal{S} = \{3, 4\}$. The observability matrix is in the following:

$$O_1 = \left[\underbrace{\begin{matrix} 1 & 0 \\ 1/2 & 1 \\ 1/4 & 1 \\ 1/8 & 3/4 \end{matrix}}_{O_1^{\mathcal{U}}} \; \underbrace{\begin{matrix} 1 & 0 \\ 2 & 0 \\ 4 & 0 \\ 8 & 0 \end{matrix}}_{O_1^{\mathcal{S}}}\right], O_2 = \left[\underbrace{\begin{matrix} 0 & 1 \\ 0 & 1/2 \\ 0 & 1/4 \\ 0 & 1/8 \end{matrix}}_{O_2^{\mathcal{U}}} \; \underbrace{\begin{matrix} 0 & 1 \\ 0 & 2 \\ 0 & 4 \\ 0 & 8 \end{matrix}}_{O_2^{\mathcal{S}}}\right], O_3 = \left[\underbrace{\begin{matrix} 1 & 0 \\ 1/2 & 1 \\ 1/4 & 1 \\ 1/8 & 3/4 \end{matrix}}_{O_3^{\mathcal{U}}} \; \underbrace{\begin{matrix} 1 & 1 \\ 2 & 2 \\ 4 & 4 \\ 8 & 8 \end{matrix}}_{O_3^{\mathcal{S}}}\right],$$

According to the definition, $\mathcal{E}_i \cap \mathcal{U}$ is the non-zero column index of matrix $O_i^{\mathcal{U}}$, that is, $\mathcal{E}_1 \cap \mathcal{U} = \{1, 2\}, \mathcal{E}_2 \cap \mathcal{U} = \{2\}, \mathcal{E}_3 \cap \mathcal{U} = \{1, 2\}$. Moreover, $\mathcal{F}_1 = \{1, 3\}, \mathcal{F}_2 = \{1, 2, 3\}$. As a result, $H_i^{\mathcal{U}}$ can be directly constructed from $\mathcal{E}_i$ rather than transformed from G_i:

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

One can obtain that,

$$\text{rowspan}\left(O_1^{\mathcal{U}}\right) = \text{span}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \text{rowspan}\left(H_1^{\mathcal{U}}\right),$$

$$\text{rowspan}\left(O_2^{\mathcal{U}}\right) = \text{span}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \text{rowspan}\left(H_2^{\mathcal{U}}\right),$$

$$\text{rowspan}\left(O_3^{\mathcal{U}}\right) = \text{span}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \text{rowspan}\left(H_3^{\mathcal{U}}\right).$$

The sparse detectability index s can be calculated according to (17): $s = \min\{2, 3\} - 1 = 1$.

3.3 Secure estimation design

Recalling the transformation P_i introduced in Theorem 3, define $\tilde{P} \triangleq \operatorname{diag}(P_1, \dots, P_m)$, $\tilde{M} \triangleq \tilde{P}\tilde{W}\tilde{P}'$ and

$$Y(k) \triangleq \begin{bmatrix} P_1\zeta_1(k) \\ \vdots \\ P_m\zeta_m(k) \end{bmatrix}, \ H \triangleq \begin{bmatrix} H_1 \\ \vdots \\ H_m \end{bmatrix}, \ \text{where } \ H_i = P_iG_i. \tag{18}$$

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.

$$\underset{\tilde{x}_{\mathrm{ls}}(k),\varphi(k)}{\text{minimize}} \quad \frac{1}{2} \begin{bmatrix} \varphi(k) \\ NH\tilde{x}_{\mathrm{ls}}(k) \end{bmatrix}' \mathcal{W} \begin{bmatrix} \varphi(k) \\ NH\tilde{x}_{\mathrm{ls}}(k) \end{bmatrix} \tag{19a}$$

$$\text{subject to} \quad Y(k) = H\tilde{x}_{\mathrm{ls}}(k) + \varphi(k), \tag{19b}$$

where

$$\mathcal{W} \triangleq \begin{bmatrix} \tilde{M}^{-1} + N'N & N' \\ N & I \end{bmatrix}. \tag{20}$$

Notice that $\mathcal{W}$ is strictly positive definite since $\tilde{M}$ is strictly positive definite. Define[¶]

$$F = \begin{bmatrix} F_1 & F_2 & \cdots & F_m \end{bmatrix}, \ \text{where } \ F_i \triangleq V\operatorname{diag}(V^{-1}K_i), \tag{21}$$

and V is defined in (4). Recall that $\epsilon(k) \triangleq \begin{bmatrix} \epsilon_1(k)', \dots, \epsilon_m(k)' \end{bmatrix}'$ and $\epsilon_i(k) = \zeta_i(k) - G_ix(k)$ from Lemma 1. We have the following lemma demonstrating that solution $\tilde{x}_{\mathrm{ls}}(k)$ to problem (19) equals to Kalman estimation $\hat{x}(k)$.

Lemma 3. *In the absence of attack, the solution to least square problem (19) coincides with the Kalman estimation, that is, the following holds where $\tilde{x}_{\mathrm{ls}}(k)$ and $\varphi(k)$ are the solutions of problem (19):*

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

Proof. Consider the following least square problem

$$\underset{\tilde{x}_{\mathrm{ls}}(k)}{\text{minimize}} \quad \frac{1}{2}(Y(k) - H\tilde{x}_{\mathrm{ls}}(k))'\tilde{M}^{-1}(Y(k) - H\tilde{x}_{\mathrm{ls}}(k)). \tag{22}$$

Based on Theorem 2 in Reference 24, the solution to problem (22) is equivalent to Kalman estimation. It is sufficient to prove that problems (19) and (22) are equivalent. Define

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

[¶] $\operatorname{diag}(V^{-1}K_i)$ is a $n \times n$ diagonal matrix with the jth diagonal entry equals to jth element of vector $V^{-1}K_i$.

Consider the objective function of problem (22) added by a constant term[#]:

$$\frac{1}{2}(Y - H\tilde{x}_{\text{ls}})'\tilde{M}^{-1}(Y - H\tilde{x}_{\text{ls}}) + \frac{1}{2}Y'N'NY = \frac{1}{2}\begin{bmatrix} Y - H\tilde{x}_{\text{ls}} \\ NY \end{bmatrix}' \begin{bmatrix} \tilde{M}^{-1} & \mathbf{0} \\ \mathbf{0} & I \end{bmatrix} \begin{bmatrix} Y - H\tilde{x}_{\text{ls}} \\ NY \end{bmatrix}. \tag{23}$$

Notice that

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

and (23) can be written as

$$\frac{1}{2}\left(\mathcal{M}\begin{bmatrix} Y - H\tilde{x}_{\text{ls}} \\ NH\tilde{x}_{\text{ls}} \end{bmatrix}\right)' \begin{bmatrix} \tilde{M}^{-1} & \mathbf{0} \\ \mathbf{0} & I \end{bmatrix} \left(\mathcal{M}\begin{bmatrix} Y - H\tilde{x}_{\text{ls}} \\ NH\tilde{x}_{\text{ls}} \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} Y - H\tilde{x}_{\text{ls}} \\ NH\tilde{x}_{\text{ls}} \end{bmatrix}' \mathcal{W} \begin{bmatrix} Y - H\tilde{x}_{\text{ls}} \\ NH\tilde{x}_{\text{ls}} \end{bmatrix}. \tag{24}$$

Substituting φ in (19) with $Y - H\tilde{x}_{\text{ls}}$ leads to (24). Thus, optimizing objective function (22) is equivalent to optimizing (24), and the latter is equivalent to problem (19). ■

Based on least square problem (19), 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.

$$\underset{\tilde{x}(k),\mu(k),\nu(k)}{\text{minimize}} \quad \frac{1}{2}\begin{bmatrix} \mu(k) \\ NH\tilde{x}(k) \end{bmatrix}' \mathcal{W} \begin{bmatrix} \mu(k) \\ NH\tilde{x}(k) \end{bmatrix} + \gamma \left\|\nu(k)\right\|_1 \tag{25a}$$

$$\text{subject to} \quad Y(k) = H\tilde{x}(k) + \mu(k) + \nu(k). \tag{25b}$$

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

Theorem 5. *In the absence of attack, if the parameter γ in problem (25) satisfies*

$$\left\| \mathcal{W} \begin{bmatrix} (I - GF)\,\epsilon(k) \\ NH\hat{x}(k) \end{bmatrix} \right\|_\infty \le \gamma, \tag{26}$$

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

$$\tilde{x}(k) = \hat{x}(k). \tag{27}$$

Noticing that $\epsilon(k)$ is a stationary Gaussian process from Lemma 1, and $\hat{x}(k)$ is a Gaussian random variable ($NH\hat{x}(k)$ denotes the stable part of $\hat{x}(k)$), the probability that inequality (26) 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.

In order to quantify the estimation difference between the attack is absent and present, we consider the following local estimation $\zeta_i^o(k)$ and Kalman estimation $\hat{x}^o(k)$ that are **not affected by the attack**:

$$\zeta_i^o(k+1) = \Pi\zeta_i^o(k) + \mathbf{1}_n z_i(k+1) + (G_i - \mathbf{1}_n C_i)Bu(k), \tag{28}$$

$$\hat{x}^o(k+1) = (I - KC)\left(A\hat{x}^o(k) + Bu(k)\right) + Kz(k+1), \tag{29}$$

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

where $z(k) = Cx(k) + v(k)$ is the original (unmanipulated) measurement. Since the Kalman estimation $\hat{x}^o(k)$ has direct access to the unmanipulated measurements, we name it as **oracle** Kalman estimation. 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 6. *Under Assumption 2 and 3, in presence of arbitrary admissible p-sparse attack, if the system (A, C) is $2p$-sparse detectable, then the estimation difference between $\tilde{x}(k)$ solved from (25) and oracle Kalman estimation $\hat{x}^o(k)$ satisfies*

$$\left|[\tilde{x}(k)]_j - [\hat{x}^o(k)]_j\right| \leq \begin{cases} \max\limits_{i_1, i_2 \in \mathcal{F}_j} \left| \left[P_{i_1} \zeta_{i_1}^o(k)\right]_j - \left[P_{i_2} \zeta_{i_2}^o(k)\right]_j \right| + (\gamma + \gamma^o(k)) \left\|\mathcal{H}\right\|_\infty, \ j \in \mathcal{U} \\ \gamma \cdot \left\|\mathcal{H}\right\|_\infty + \left|[\hat{x}^o(k)]_j\right|, \ j \in \mathcal{S} \end{cases}, \tag{30}$$

where

$$\begin{aligned} \gamma^o(k) &\triangleq \left\| \mathcal{W} \begin{bmatrix} (I - GF)\, \epsilon^o(k) \\ NH\hat{x}^o(k) \end{bmatrix} \right\|_\infty, \\ \mathcal{H} &\triangleq \left(\begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & \mathcal{L}H'N' \end{bmatrix} \mathcal{W} \begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & NH\mathcal{L}' \end{bmatrix} \right)^{-1} \begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & \mathcal{L}H' \end{bmatrix}, \\ \mathcal{L} &\triangleq \begin{bmatrix} \mathbf{0}_{n_s \times n_u} & I_{n_s} \end{bmatrix}, \end{aligned} \tag{31}$$

with $\mathcal{E}_i$ defined in (12) and $[\cdot]_j$ being the jth 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 $k \in \mathbb{N}$, our proposed estimation $\tilde{x}(k)$ is secure.

Under Assumption 3, our proposed estimator is secure if the system is $2p$-sparse detectable. The maximum estimation difference from oracle Kalman estimation is shown in (30). Theorem 6 indicates smaller γ leads to lower estimation difference upper bound in the presence of attack, that is, smaller γ means better performance under attack. However, based on Theorem 5, smaller γ decreases the probability of recovering the optimal Kalman estimation in the absence of attack, that is, greater γ means better performance without attack. The choice of γ represents the accuracy trade-off between normal operation and under attack.

Moreover, since sparse detectability index only requires simple computation according to Theorem 4, our work reduces the complexity of evaluating system vulnerability significantly under Assumption 3. For general A that has eigenvalues with geometric multiplicity larger than 1 (A is derogatory), computing sparse observability is an NP-hard problem,[26] 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 by off-the-shelf solvers.

3.4 | Fundamental limit

This subsection 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 and confirms that we have achieved the fundamental limit.

Theorem 7. *If the system matrix A is strictly unstable and the system is not $2p$-sparse detectable, there always exists a p-sparse attack strategy that* **no** *estimator is secure.*

Proof. The proof of bounded noise scenario can be seen in Reference 16 Theorem 1 and we provide the proof of Gaussian noise here for paper self-consistency. If the system is not $2p$-sparse detectable, then there exists an eigenvector ξ of A that corresponds to an unstable eigenvalue λ such that $A\xi = \lambda\xi$, $C\xi = \mathbf{0}$. If ξ is a complex vector, then $A\xi = \lambda\xi, A\overline{\xi} = \overline{\lambda}\,\overline{\xi}$, $C\xi = \mathbf{0}, C\overline{\xi} = \mathbf{0}$, where $\overline{\xi}$ represents the conjugate of ξ. As a result, there exists a set $\mathcal{B}$ with $\mathcal{B} \subset \mathcal{I}$, $|\mathcal{B}| = 2p$, such that the linear transformation defined by $\mathcal{O}_t$ has a non-trivial kernel, where

$$\mathcal{O}_t \triangleq \left[C'_{\mathcal{I}\setminus\mathcal{B}} \mid \left(C_{\mathcal{I}\setminus\mathcal{B}} A\right)' \mid \cdots \mid \left(C_{\mathcal{I}\setminus\mathcal{B}} A^{t-1}\right)' \right]'.$$

In other words, either the following two statements is true.

$$(1)\text{ if } \xi \text{ is a real vector: } \mathcal{O}_t\xi = \mathbf{0}, \forall t \in \mathbb{N}. \quad (2)\text{ if } \xi \text{ is a complex vector: } \mathcal{O}_t\left(\xi + \bar{\xi}\right) = \mathbf{0}, \forall t \in \mathbb{N}.$$

Based on this result, we intend to prove that the following proposition is true.

Proposition 1. *There exists zero mean Gaussian disturbances $\{w(k)\}_{k=0}^{\infty}$, $\{v(k)\}_{k=0}^{\infty}$, two p-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) and (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 $\mathcal{B}$ into $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ such that $\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$ and $|\mathcal{B}_1| = |\mathcal{B}_2| = p$. Define the following two trajectories:

$$\begin{aligned} &\textbf{System 1:} \quad x^{(1)}(0) = \mathbf{0}, && \textbf{System 2:} \quad x^{(2)}(0) = \xi, \\ &a^{(1)}(k) = \begin{cases} C_iA^k\xi, i \in \mathcal{B}_1 \\ \mathbf{0}, i \in \mathcal{I} \setminus \mathcal{B}_1 \end{cases}, && a^{(2)}(k) = \begin{cases} -C_iA^k\xi, i \in \mathcal{B}_2 \\ \mathbf{0}, i \in \mathcal{I} \setminus \mathcal{B}_2 \end{cases}. \end{aligned} \tag{32}$$

Noticing that $A\xi = \lambda\xi$, $C\xi = \mathbf{0}$, one obtains for all $k \in \mathbb{N}$:

$$y_i^{(1)}(k) = y_i^{(2)}(k) = \begin{cases} C_i\left(A^k\xi + \sum_{t=0}^{k-1} A^{k-1-t}w(t)\right) + v_i(k), & i \in \mathcal{B}_1, \\ C_i\left(\sum_{t=0}^{k-1} A^{k-1-t}w(t)\right) + v_i(k), & i \in \mathcal{I} \setminus \mathcal{B}_1. \end{cases} \tag{33}$$

However, $\|x^{(1)}(k) - x^{(2)}(k)\|_2 = \|\lambda^k\xi\|_2$ is unbounded since $|\lambda| > 1$. If ξ is complex, replace ξ in (32) by $\xi + \bar{\xi}$ and one can also verify that $y^{(1)}(k) = y^{(2)}(k),\ \forall k \geq 0$ and $x^{(1)}(k) = \mathbf{0}$ while $x^{(2)}(k) = \lambda^k\left(\xi + \bar{\xi}\right)$. As a result, Proposition 1 is proved, and there exists no secure estimation since the system has identical output $y(k)$ but diverging state $x(k)$ for two trajectories. ■

In view of Theorem 7, our proposed estimator achieves the fundamental limit of the secure dynamic estimation problem, that is, provides a secure estimation whenever the system can be securely estimated. This result is stronger than other secure estimators in the literature which require $2p$-sparse observability,[11-13,18,19,24] since our requirement $2p$-sparse detectability is less restrictive. The performance of our proposed estimator is corroborated by the numerical simulation in the next section.

4 | NUMERICAL SIMULATION

4.1 | Inverted pendulum system (corroboration of algorithm security and accuracy)

We use an inverted pendulum for the numerical simulation to corroborate the performance of our proposed estimation scheme in the absence of attack and in the presence of different sparse attacks. The mass of the cart and the mass of the pendulum are both 1 kg. The length of pendulum is 1 m and the moment of inertia of the pendulum is $1/3$ kg m^2. The control input $u(k)$ is the force applied on the cart. The state x_1, x_2, x_3, x_4 represent cart position, cart velocity, pendulum angle from vertical and pendulum angle velocity respectively. We consider the system linearized at $x_3 = x_4 = 0$, and we sample the continuous-time linear system periodically with sampling interval $T_s = 0.02$ s. The first 3 sensors are monitoring state

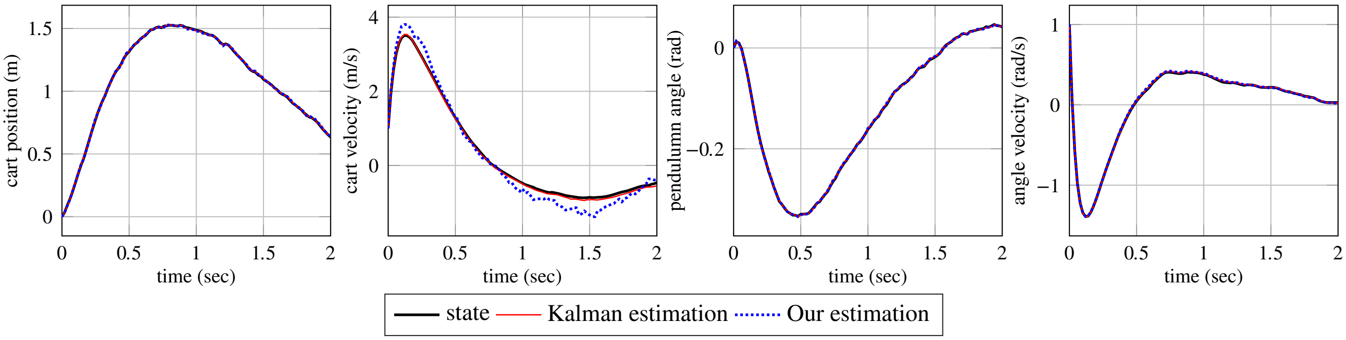

FIGURE 2 Estimation of states in the absence of attack

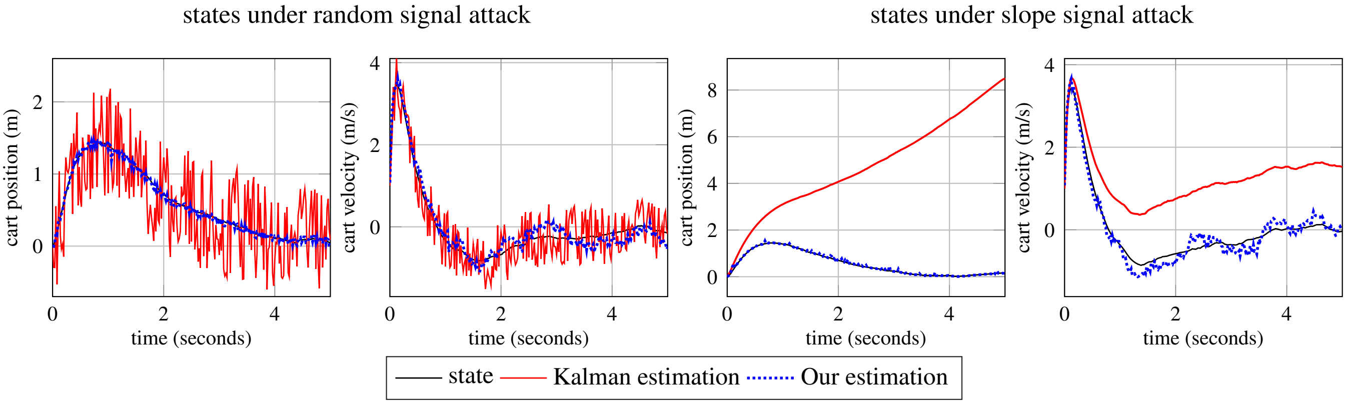

FIGURE 3 Estimation of states under random signal attack and slope signal attack on sensor 3

x_1 and sensor 4 is monitoring state x_3. The system matrix is:

$$A = \begin{bmatrix} 1 & 2.0 \cdot 10^{-2} & -2.0 \cdot 10^{-4} & 1.9 \cdot 10^{-5} \\ 0 & 1.0 \cdot 10^{0} & -2.0 \cdot 10^{-2} & 1.8 \cdot 10^{-3} \\ 0 & 1.0 \cdot 10^{-5} & 1.0 \cdot 10^{0} & 2.0 \cdot 10^{-2} \\ 0 & 1.0 \cdot 10^{-3} & 2.1 \cdot 10^{-1} & 9.8 \cdot 10^{-1} \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

One can calculate the sparse detectability index of the system according to Section 3 and obtain that the system is 2-sparse detectable and our proposed 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 $x(0) = [0,\ 1,\ 0,\ 1]'$ and 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 our proposed estimator in the absence of attack in Figure 2, where our proposed estimation substantially coincides with the Kalman estimation. The numerical difference attributes to large Gaussian noise that occurs occasionally which violates inequality (26) and error in numerical calculation.

In the following, we show the performance of our proposed estimator under random signal attack and slope signal attack launched on sensor 3 in Figure 3. The random signal attack is a time-independent random value uniformly distributed on interval $(-1, 1)$. The slope signal attack is a linearly increasing signal with a rate of 2 m/s, that is, $a(k) = 2kT_s$. As shown in the left two sub-figures of Figure 3, Kalman estimation (denoted as red line) is corrupted by the injected signal and has a larger estimation error than our proposed estimation. In the right two sub-figures, the Kalman estimation of cart position is driven away from its actual value, and the Kalman estimation of cart velocity has a stationary nonzero error. In contrast, our proposed estimator has smaller and bounded estimation error covariance under attack.

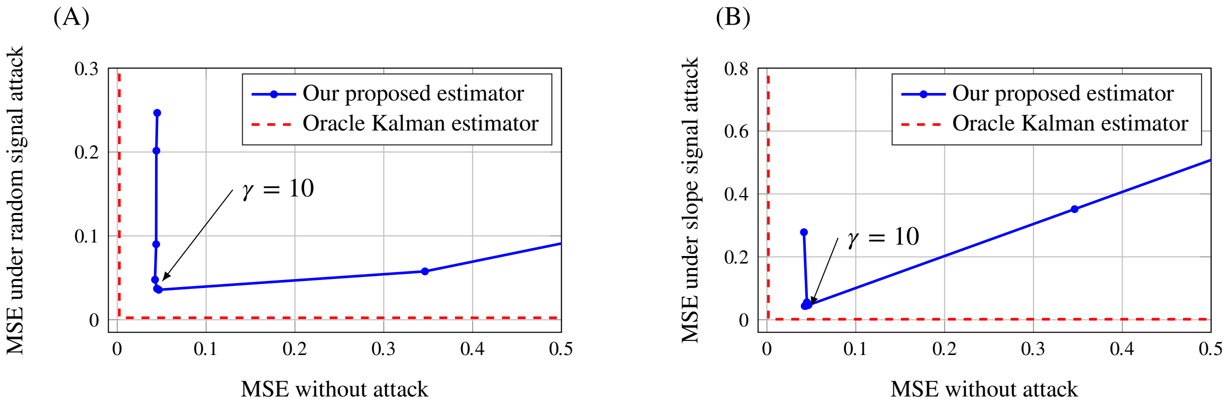

FIGURE 4 Estimation mean square error (MSE) with varying tuning parameter γ. (A) Performance trade-off under random signal attack and no attack. The MSE of Kalman filter under attack is 0.407; (B) Performance trade-off under slope signal attack and no attack. The MSE of Kalman filter under attack is 6.26

Our proposed estimator has a tuning parameter γ balancing the performance under and without attack. In the following, we show the trade-off between the estimation error under attack and without attack by tuning γ. Figure 4 illustrates the estimation mean square error MSE $\left(\text{MSE} = \frac{1}{N}\sum_{k=1}^{N} \|\tilde{x}(k) - x(k)\|_2^2\right)$ of our proposed estimator with varying tuning parameter γ with sensor 3 corrupted. In Figure 4A,B, the MSE of the **oracle Kalman** estimator is illustrated by the red dashed line, and that of our proposed estimator under attack is illustrated by the solid blue line. The point in the lower-left corner represents small estimation errors in both scenarios with attack and without attack. As shown in Figure 4A,B, by properly choosing γ, the MSE of our proposed estimator achieves good performance under both scenarios (with and without attack). The MSE of our proposed estimator without attack is 0.036, which is slightly larger than that of the Kalman estimator. Under both attacks, the MSE of our proposed estimator is roughly 0.040. In contrast, the MSE of Kalman estimation under random signal attack is 0.407 and is 6.26 under slope signal attack, which are significantly larger than that of the proposed secure estimator.

4.2 | IEEE 68-bus system (corroboration of algorithm low complexity)

We employ the IEEE 68-bus system, which is extensively used in the literature,[17] for simulation to further demonstrate the benefits of our scheme on low complexity and validate the performance. The IEEE 68-bus system is composed of 16 generator buses (indexed from 1 to 16) and 52 load buses (indexed from 17 to 68), as shown in Figure 5. The network topology is depicted as an undirected graph $(\mathcal{V}, \mathcal{E})$ with $\mathcal{V}$ representing the vertex set and $\mathcal{E}$ representing the edge set. We adapt the system dynamic as in Wood et al.[29] which is also seen in Yong et al..[17] The phase angle $\theta_i(t)$ and angular frequency $\omega_i(t)$ on each bus i satisfy:

$$\dot{\theta}_i(t) = \omega_i(t), \tag{34}$$

$$\dot{\omega}_i(t) = -\frac{1}{m_i}\left[D_i\omega_i(t) + \sum_{j\in\mathcal{N}_i} P_{tie}^{ij}(t) - P_i(t) + w_i(t)\right]. \tag{35}$$

The power flow between neighboring buses $(i,j) \in \mathcal{E}$ is given by $P_{tie}^{ij}(t) = -P_{tie}^{ji}(t) = t_{ij}\left(\theta_i(t) - \theta_j(t)\right)$. The power $P_i(t)$ denotes the mechanical power for the generator bus and the negative of power demand for the load bus. $P_i(t)$ is modeled as the system input and assumed to be known by the system operator. The noise $w_i(t)$ is a zero-mean Gaussian signal with covariance matrix $Q_i(t) = 10^{-6}I$, and the system parameters are $D_i = 1, t_{ij} = 15$ for all $i \in \mathcal{V}, j \in \mathcal{N}_i$, and $t_{ij} = 0$ otherwise. Angular momentums are $m_i = 10$ for generator buses and a larger value $m_i = 100$ for load buses. The system is sampled

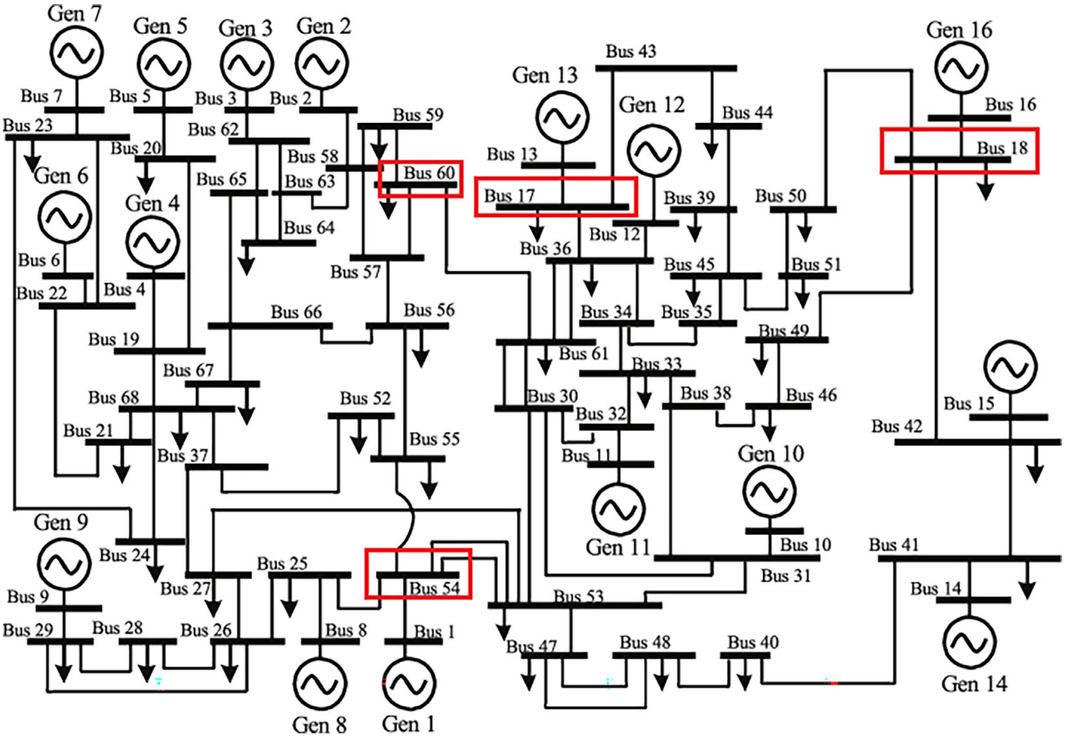

FIGURE 5 IEEE 68 bus test system with load bus 17,18,54,60 corrupted (framed by red box). *Source*: The figure is adapted from Ishizaki et al.[28]

at discrete times with sampling interval $T_s = 0.01$ s. The measurements are also discrete in time:

$$\begin{bmatrix} y_{i_1}(k) & y_{i_2}(k) & y_{i_3}(k) \end{bmatrix}^\top = \begin{bmatrix} P_{elec,i}(k) & \theta_i(k) & \omega_i(k) \end{bmatrix}^\top + v_i(k),$$

where $P_{\mathrm{elec},i}(k) = D_i\omega_i(k) + \mathbb{I}\{i \in \mathcal{V}_l\} \cdot P_i(k)$ is the electrical power output and $v_{i,k}$ is a zero-mean Gaussian noise signal with covariance matrix $R_i = T_s^4 I$.

The system has dimension $n = 136$ (each bus is associated with two scalar states ω_i and θ_i), and the sensor number is $m = 204$ (each bus has electrical power sensor, phase angle sensor, and angular frequency sensor). The complexity bounds of some methods in the literature are shown in Table 1. As shown in the table, if there is no *a priori* information about the sensor vulnerability, the number of possible combinations of corrupted sensor choices is huge, making maintaining all possible candidate filters[16] or sequentially switching among them until hitting a benign combination[20] computationally heavy or even intractable. Furthermore, for sliding window methods, the dimension of the convex optimization problem may be large ($m \times T$, and T is required to be no less than n). In view of the scalability issues, the Satisfiability Modulo Theory (SMT) is adapted by Shoukry et al.[18] and Mishra et al.[19] to harness the complexity. However, the proposed methods can reduce the computational complexity effectively only when the corrupted sensor number is relatively large compared to total sensor number (e.g., nearly half of the sensors are corrupted). If the corrupted sensor only takes a small portion (4 out of 204 sensors in this case), the reduction in computational complexity is not significant for the SMT-based method.

For our proposed scheme, the dimension of the convex optimization problem is also large ($m \times n$). However, we can reduce the dimension significantly at the cost of a minor loss of accuracy. On the one hand, by only retaining the elements on the diagonal blocks of weighting matrix $\mathcal{W}$ and setting other elements to zero, we obtain that the optimization problem (25) can be decomposed to minimize the summation of m terms:

$$\text{minimize}_{v(k),\tilde{x}(k)} \sum_{i=1}^{m} \frac{1}{2} \begin{bmatrix} H_i\tilde{x}(k) + v_i(k) - P_i\zeta_i(k) \\ H_i^s\tilde{x}(k) \end{bmatrix}' \mathcal{W}_i \begin{bmatrix} H_i\tilde{x}(k) + v_i(k) - P_i\zeta_i(k) \\ H_i^s\tilde{x}(k) \end{bmatrix} + \gamma \left\| v_i(k) \right\|_1, \tag{36}$$

TABLE 1 Complexity comparison between different resilient estimation methods for m sensor system

	Switching		**Sliding window**		
Methods	**Subset selection**[16,17,20]	**Set cover approach**[21]	**Batched optimization**[11,14]	**SMT**[12,18]	**Decomposition-fusion**[27] **(this article)**
Complexity Criterion	Number of filters need to consider	Number of filters need to consider	Dimension of optimization problem	Number of subproblems need to solve/check	Dimension of optimization problem
p out of m sensors corrupted	$\binom{m}{p}$	At most $\frac{1}{2}\binom{m}{p}$ at least $\binom{m}{p}/\binom{2p}{p}$	1 problem with dimension $m \times T$ (T is window width)	At most $\binom{m}{m-2p+1}$	much lower than mn for most large systems
IEEE 68 bus case 4 sensors corrupted	7.006×10^7	At most 3.503×10^7 at least 1.001×10^6	1 problem with dimension 27,744	More than 7.006×10^7	204 problems with dimension 7

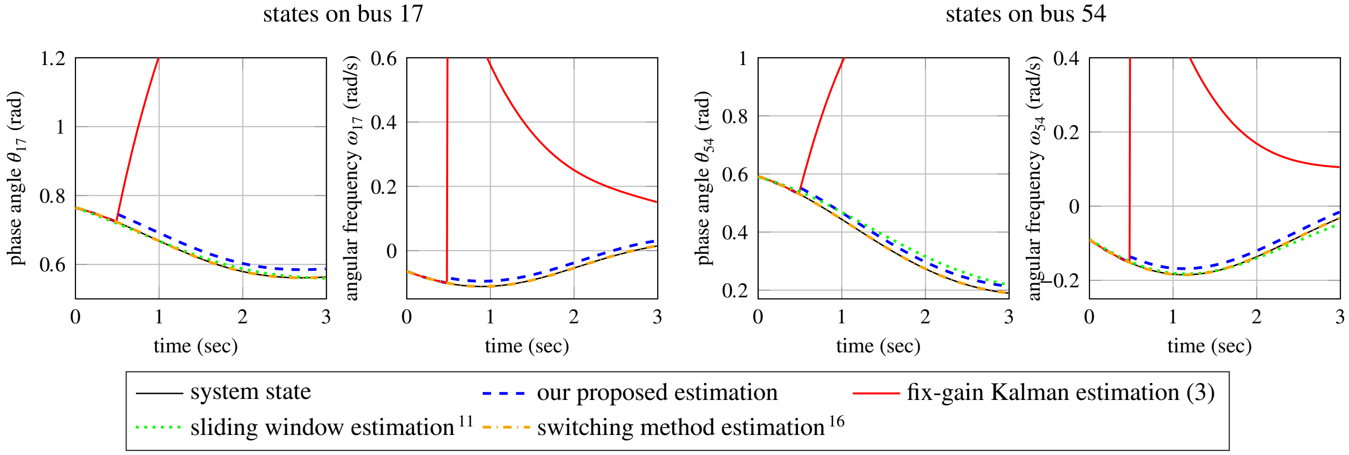

FIGURE 6 Estimation of states under **step signal attack** on IEEE 68 system. The left two sub-figures illustrate the phase angle and angular velocity of bus 17. The right two sub-figures illustrate those of bus 54

which can be transformed into minimizing m uncorrelated objective functions with the constraint that they have reached consensus on solution $\tilde{x}(k)$. On the other hand, due to the sparse observability structure (one sensor can only observe a few number of states), we know that the canonical matrix H_i is a low-rank diagonal matrix. By rearranging the rows of H_i, one can obtain a matrix with only first rank(H_i) rows non-zero, that is, $\begin{bmatrix} \tilde{H}_i \\ 0 \end{bmatrix}$, with $\tilde{H}_i$ of size rank$(H_i) \times n$. By substituting H_i by $\tilde{H}_i$ in optimization problem (36) (the corresponding rows and columns of matrix $\mathcal{W}_i$ and entries of vector v_i are also discarded), the dimension is further decreased. For this IEEE 68-bus system example, our proposed algorithm involves cooperatively solving $m = 204$ convex optimization problems whose optimization variable is of dimension 7 (much lower than system state dimension $n = 136$), and can be easily implemented in a distributed manner. However, this simplification will cause a minor loss of accuracy. It can be seen in Figure 6 that the estimation error is small and acceptable.

In the simulation, we launched two different attacks on the same four sensors, that is, the phase angle sensors at buses 17, 18, 54, 60. In Figure 6, sensors are corrupted by step signals with a magnitude of π. In Figure 7, sensors are corrupted by random signal that is uniformly distributed in interval $(-1, 1)$. The two attacks are both launched at $t = 0.5$ s. Figure 6 illustrates the estimation of states on buses 17 and 54, and Figure 6 illustrates the estimation of states on buses 18 and 60. Since the enormous number of possible corrupted sensor combinations that some methods require to consider

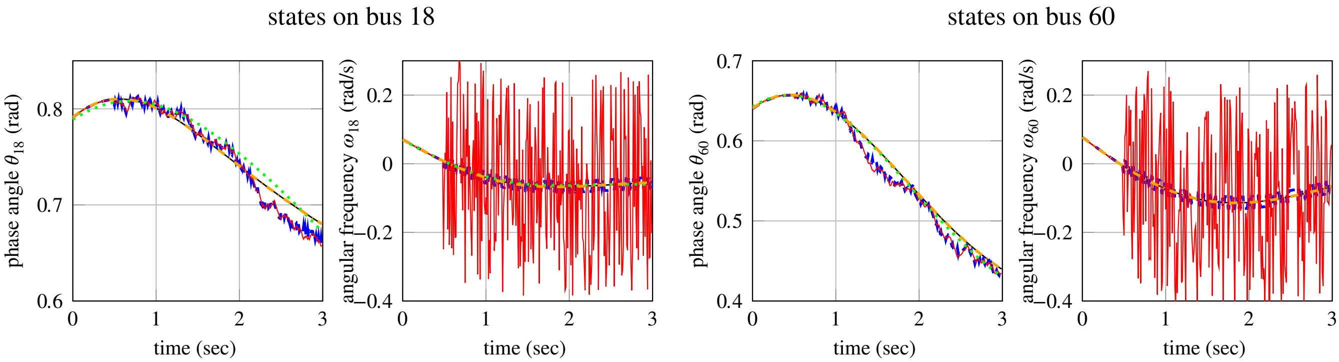

FIGURE 7 Estimation of states under **random signal attack** on IEEE 68 system. The left two sub-figures illustrate the phase angle and angular velocity of bus 18. The right two sub-figures illustrate those of bus 60

makes the numerical simulation exceedingly costly, we assume the switching method and the sliding window method know the corrupted sensor set in the simulation. For our proposed estimation and the fix-gain Kalman estimation, the corrupted sensor set is unknown. As shown in Figures 6 and 7, the fixed gain Kalman estimation (3) is not secure against sparse attack, and the estimated state (red line) deviates from real state (black line) under attack. Other secure estimation schemes can recover the system state with small errors. The estimation error of our proposed estimation scheme is acceptable considering the merit of low computational complexity.

5 | CONCLUSION

This article considers LTI systems with Gaussian noise against sparse integrity attack on an unknown subset of sensors. Under the assumption that the system is regular and A is non-singular, we propose an estimation scheme that is secure to p-sparse attack as long as the system is $2p$-sparse detectable. Our design first decomposes the Kalman estimator into local estimators and then fuses the local estimation in a secure manner. The secure fusion scheme is developed by the following two steps. (1) We first prove that the matrix G_i has a canonical form under Assumption 3, and the canonical form can be constructed according to the observability matrix O_i. The canonical form H_i constitutes the coefficient of the secure fusion optimization problem. (2) In order to guarantee that the stable states are always secure, the fusion scheme is carefully designed as a convex optimization problem where unstable and stable state entries play different roles. By this design, the estimation of stable states are always secured.

As a result, the proposed estimation scheme is proved to be secure against p-sparse attack as long as the system is $2p$-sparse detectable, and the system sparse detectability index can be calculated in polynomial time with respect to m and n by our design. 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 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 and low computation complexity. The proposed estimator is corroborated by two simulation examples, the inverted pendulum system, and IEEE 68 bus system. In the latter simulation case, we compared the estimation performance and computational complexity of our scheme with some existing methods in the literature. It is shown that our scheme has lower computation complexity at the cost of minor accuracy loss.

FUNDING INFORMATION

This work was supported by National Key Research and Development Program of China under Grant 2018AAA0101601.

CONFLICT OF INTEREST

We declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

ORCID

Zishuo Li https://orcid.org/0000-0003-4744-455X
Yilin Mo https://orcid.org/0000-0001-7937-6737

REFERENCES

1. Maughan D. Cyber security division technology guide 2018. U.S. Department of Homeland Security (DHS); 2018.
2. Cárdenas A, Amin S, Sinopoli B, et al. Challenges for securing cyber physical systems. Proceedings of the Workshop on Future Directions in Cyber-Physical Systems Security; Vol. 5, 2009:1.
3. Langner R. Stuxnet: dissecting a cyberwarfare weapon. *IEEE Secur Priv*. 2011;9(3):49-51. doi:10.1109/MSP.2011.67
4. Liang G, Weller SR, Zhao J, Luo F, Dong ZY. The 2015 Ukraine blackout: implications for false data injection attacks. *IEEE Trans Power Syst*. 2017;32(4):3317-3318. doi:10.1109/TPWRS.2016.2631891
5. Andersson G, Donalek P, Farmer R, et al. Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance. *IEEE Trans Power Syst*. 2005;20(4):1922-1928. doi:10.1109/TPWRS.2005.857942
6. Mo Y, Sinopoli B. On the performance degradation of cyber-physical systems under stealthy integrity attacks. *IEEE Trans Automat Contr*. 2016;61(9):2618-2624. doi:10.1109/TAC.2015.2498708
7. Jovanov I, Pajic M. Sporadic data integrity for secure state estimation. Proceedings of the 2017 IEEE 56th Annual Conference on Decision and Control (CDC); 2017:163-169. 10.1109/CDC.2017.8263660
8. Ding K, Ren X, Quevedo DE, Dey S, Shi L. DoS attacks on remote state estimation with asymmetric information. *IEEE Trans Control Netw Syst*. 2019;6(2):653-666. doi:10.1109/TCNS.2018.2867157
9. Yan J-J, Yang G-H. Secure state estimation with switched compensation mechanism against DoS attacks. *IEEE Trans Cybern*. 2021;1-12. [Early Access]. doi:10.1109/TCYB.2021.3060743
10. Ding D, Han Q-L, Ge X, Wang J. Secure state estimation and control of cyber-physical systems: a survey. *IEEE Trans Syst Man Cybern Syst*. 2021;51(1):176-190. doi:10.1109/TSMC.2020.3041121
11. Fawzi H, Tabuada P, Diggavi S. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Trans Automat Contr*. 2014;59(6):1454-1467. doi:10.1109/TAC.2014.2303233
12. Pajic M, Lee I, Pappas GJ. Attack-resilient state estimation for noisy dynamical systems. *IEEE Trans Control Netw Syst*. 2017;4(1):82-92. doi:10.1109/TCNS.2016.2607420
13. Chang YH, Hu Q, Tomlin CJ. Secure estimation based Kalman filter for cyber–physical systems against sensor attacks. *Automatica*. 2018;95:399-412. doi:10.1016/j.automatica.2018.06.010
14. Shoukry Y, Tabuada P. Event-triggered state observers for sparse sensor noise/attacks. *IEEE Trans Automat Contr*. 2016;61(8):2079-2091. doi:10.1109/TAC.2015.2492159
15. Chong MS, Wakaiki M, Hespanha JP. Observability of linear systems under adversarial attacks. *Am Control Conf*. 2015;2015:2439-2444. doi:10.1109/ACC.2015.7171098
16. Nakahira Y, Mo Y. Attack-resilient $\mathscr{H}_2$, $\mathscr{H}_\infty$, and ℓ_1 state estimator. *IEEE Trans Automat Contr*. 2018;63(12):4353-4360. doi:10.1109/TAC.2018.2819686
17. Yong SZ, Zhu M, Frazzoli E. Switching and data injection attacks on stochastic cyber-physical systems: modeling, resilient estimation, and attack mitigation. *ACM Trans Cyber-Phys Syst*. 2018;2(2):9:1-9:26. doi:10.1145/3204439
18. Shoukry Y, Nuzzo P, Puggelli A, Sangiovanni-Vincentelli AL, Seshia SA, Tabuada P. Secure state estimation for cyber-physical systems under sensor attacks: a satisfiability modulo theory approach. *IEEE Trans Automat Contr*. 2017;62(10):4917-4932. doi:10.1109/TAC.2017.2676679
19. Mishra S, Shoukry Y, Karamchandani N, Diggavi SN, Tabuada P. Secure state estimation against sensor attacks in the presence of noise. *IEEE Trans Control Netw Syst*. 2017;4(1):49-59. doi:10.1109/TCNS.2016.2606880
20. An L, Yang G-H. Secure state estimation against sparse sensor attacks with adaptive switching mechanism. *IEEE Trans Automat Contr*. 2018;63(8):2596-2603. doi:10.1109/TAC.2017.2766759
21. Lu A, Yang G. Secure switched observers for cyber-physical systems under sparse sensor attacks: a set cover approach. *IEEE Trans Automat Contr*. 2019;64(9):3949-3955. doi:10.1109/TAC.2019.2891405
22. Pasqualetti F, Dörfler F, Bullo F. Attack detection and identification in cyber-physical systems. *IEEE Trans Automat Contr*. 2013;58(11):2715-2729. doi:10.1109/TAC.2013.2266831
23. Yong SZ, Zhu M, Frazzoli E. Simultaneous mode, input and state estimation for switched linear stochastic systems. *Int J Robust Nonlinear Control*. 2021;31(2):640-661. doi:10.1002/rnc.5306
24. Liu X, Mo Y, Garone E. Local decomposition of kalman filters and its application for secure state estimation. *IEEE Trans Automat Contr*. 2021;66(10):5037-5044. doi:10.1109/TAC.2020.3044854
25. Hendrickx JM, Johansson KH, Jungers RM, Sandberg H, Sou KC. Efficient computations of a security index for false data attacks in power networks. *IEEE Trans Automat Contr*. 2014;59(12):3194-3208. doi:10.1109/TAC.2014.2351625
26. Mao Y, Mitra A, Sundaram S, Tabuada P. On the computational complexity of the secure state-reconstruction problem. *Automatica*. 2022;136:110083. doi:10.1016/j.automatica.2021.110083
27. Li Z, Mo Y. Low complexity secure state estimation design for linear system with non-derogatory dynamics. Proceedings of the 2021 60th IEEE Conference on Decision and Control (CDC); 2021:6591-6596. 10.1109/CDC45484.2021.9682830

28. Ishizaki T, Sasahara H, Inoue M, Kawaguchi T, Imura J-i. Modularity in design of dynamical network systems: retrofit control approach. *IEEE Trans Automat Contr*. 2021;66(11):5205-5220. doi:10.1109/TAC.2020.3035631
29. Allen JW, Gerald BS. *Power Generation, Operation, and Control*. 3rd ed. Springer; 2013.

How to cite this article: Li Z, Mo Y. Efficient secure state estimation against sparse integrity attack for regular linear system. *Int J Robust Nonlinear Control*. 2022;1-28. doi: 10.1002/rnc.6264

APPENDIX A. PROOF OF THEOREM 2

Before proving Theorem 2, we need the following lemma.

Lemma 4. *When all the eigenvalues of A has geometric multiplicity 1, non-zero columns of O_i are linearly independent.*

Proof of Lemma 4. Suppose A is in the Jordan canonical form, that is,

$$A = \text{diag}(J_1.J_2. \dots ,J_l),$$

where J_p is the Jordan block with respect to eigenvalue λ_p. Denote the algebraic multiplicity of λ_p as $\alpha(p)$. The integer l is the number of Jordan blocks and the size of block J_p equals to the algebraic multiplicity $\alpha(p)$ since the geometric multiplicity equals to one. Moreover, $\sum_{p=1}^{l}\alpha(p) = n$. Then $O_i = \left[C_i' \ (C_iA)' \ \cdots \ (C_iA^{n-1})'\right]'$ can be formulated into the following stack of vertical groups:

$$\left[\begin{array}{c|c|c|c} C_{i,1} & C_{i,2} & \cdots & C_{i,l} \\ C_{i,1}J_1 & C_{i,2}J_2 & \cdots & C_{i,l}J_l \\ \vdots & \vdots & \vdots & \vdots \\ C_{i,1}J_1^{n-1} & C_{i,2}J_2^{n-1} & \cdots & C_{i,l}J_l^{n-1} \end{array}\right],$$

where $C_{i,p}$ is the segment of C_i with entries corresponding to block J_p, that is, $C_{i,p}$ is composed of elements in row vector C_i with entry index from $\sum_{k=1}^{p-1}\alpha(k) + 1$ to $\sum_{k=1}^{p}\alpha(k)$. We concentrate on the vertical group corresponding to block J_p. With the help of Jordan canonical form, we can rewrite the vertical group as

$$\begin{bmatrix} C_{i,p} \\ C_{i,p}J_p \\ \vdots \\ C_{i,p}J_p^{n-1} \end{bmatrix} = \left[C_{i,p}(1)\beta_p(1) \ \middle| \ C_{i,p}(2)\beta_p(1) + C_{i,p}(1)\beta_p(2) \ \middle| \ \cdots \ \middle| \ \sum_{j=1}^{\alpha(p)} C_{i,p}(j)\beta_p\left(\alpha(p) + 1 - j\right)\right], \tag{A1}$$

where $C_{i,p}(k)$ is the kth entry of row vector $C_{i,p}$, and $\beta_p(k) \in$ is a $n \times 1$ column vector defined:

$$\beta_p(k) = \begin{bmatrix} I(1,k) \\ J_p(1,k) \\ J_p^2(1,k) \\ \vdots \\ J_p^{n-1}(1,k) \end{bmatrix}, \text{ where } J_p^q = \begin{bmatrix} \lambda_p & 1 & 0 & \cdots & 0 \\ 0 & \lambda_p & 1 & \cdots & 0 \\ 0 & 0 & \lambda_p & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & \lambda_p \end{bmatrix}^q.$$

$J_p^q(1,k)$ is the element at first row, kth column of matrix J_p^q, which is the qth power of matrix J_p. According to the Jordan canonical form, for a fixed p, vectors in $\{\beta_p(k), 1 \le k \le \alpha(p)\}$ are linearly independent. Moreover, since the geometric multiplicities of A are all one, $\lambda_j \neq \lambda_p$ when $j \neq p$. As a result, vectors in $\{\beta_p(k), 1 \leq p \leq l, 1 \leq k \leq \alpha(p)\}$ are linearly independent. According to Equation (A1), if the jth column is not zero vector, it is linearly independent to other columns. ■

Proof. **Proof of Theorem 2 (1)** According to the definition of $\mathcal{E}_i$, span $\left\{e_j | j \in \mathcal{E}_i\right\}$ = span $\left\{e_j | O_i e_j \neq \mathbf{0}, j \in \mathcal{J}\right\}$. We first prove the following two statements are equivalent:

(i) Vectors in set $\{O_i e_j | O_i e_j \neq \mathbf{0}, j \in \mathcal{J}\}$ are linearly independent.
(ii) $\mathbb{O}_i = \text{span}\left\{e_j | O_i e_j \neq \mathbf{0}, j \in \mathcal{J}\right\}$.

We first prove (ii)⇒(i) by proving its contrapositive. According to the definition of $\mathbb{O}_i$, we have $\mathbb{O}_i \triangleq \text{rowspan}(O_i)$. Since $O_i e_j$ is jth column of matrix O_i, (i) is stating that the non-zero columns of O_i are linear dependent. If (i) holds,

$$\dim(\mathbb{O}_i) = \text{rank}(O_i) < |\mathcal{E}_i| = \dim\left(\text{span}\left\{e_j | O_i e_j \neq \mathbf{0}, j \in \mathcal{J}\right\}\right),$$

and thus $\mathbb{O}_i \neq \text{span}\left\{e_j | O_i e_j \neq \mathbf{0}, j \in \mathcal{J}\right\}$.

We proceed to prove (i)⇒(ii). If the non-zero columns of O_i are linearly independent, by elementary row operations, O_i can be transformed into the following Smith standard form:

$$\tilde{O}_i = \begin{bmatrix} e'_{j_1} \\ e'_{j_2} \\ \vdots \\ e'_{j_{n(i)}} \\ \mathbf{0}_{(n-n(i))\times n} \end{bmatrix},$$

where $\{j_1, j_2, \ldots, j_{n(i)}\} = \{j \in \mathcal{J} | O_i e_j \neq \mathbf{0}\}$. As a result, $\mathbb{O}_i = \text{rowspan}(O_i) = \text{rowspan}(\tilde{O}_i) = \text{span}\left\{e_j | O_i e_j \neq \mathbf{0}, j \in \mathcal{J}\right\}$. At this point, it is proved that (i) and (ii) are equivalent.

Since (ii) is equivalent to $\mathbb{O}_i = \text{span}\left(e_j, j \in \mathcal{E}_i\right)$ and (i) holds true from Lemma 4, the proof of Theorem 2 (1) is completed.

Proof of Theorem 2 (2) 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 $q_\pi(x) \triangleq (x-\pi)^{-1}(p(x) - p(\pi))$ where $x \neq \pi$.

We have

$$\begin{aligned} q_\pi(x) &= (x-\pi)^{-1}\left(a_n(x^n - \pi^n) + a_{n-1}(x^{n-1} - \pi^{n-1}) + \cdots + a_1(x-\pi)\right) \\ &= a_n\left(x^{n-1} + \pi x^{n-2} + \pi^2 x^{n-3} + \cdots + \pi^{n-1}\right) + a_{n-1}\left(x^{n-2} + \pi x^{n-3} + \pi^2 x^{n-4} + \cdots + \pi^{n-2}\right) + \cdots + a_2(x+\pi) + a_1. \end{aligned} \tag{A2}$$

Equation (A2) is also valid when x is a square matrix. As a result, by rearranging the terms of A with the same power, one obtains

$$\begin{aligned} q_{\pi_j}(A) &= a_n\left(A^{n-1} + \pi_j A^{n-2} + \cdots + \pi_j^{n-1} I\right) + a_{n-1}\left(A^{n-2} + \pi_j A^{n-3} + \cdots + \pi_j^{n-2} I\right) + \cdots + a_2\left(A + \pi_j I\right) + \cdots + a_1 I \\ &= a_n A^{n-1} + \left(a_n \pi_j + a_{n-1}\right) A^{n-2} + \cdots + \left(a_n \pi_j^{n-1} + a_{n-1}\pi_j^{n-2} + \cdots + a_2 \pi_j + a_1\right) I, \end{aligned}$$

Define $b_{j,k}$ as the constant scalar coefficient of A^k in $q_{\pi_j}(A)$:

$$b_{j,k} \triangleq \sum_{i=0}^{n-k-1} a_{i+k+1} \pi_j^i. \tag{A3}$$

Thus, $q_{\pi_j}(A) = \sum_{k=0}^{n-1} b_{j,k} A^k$. According to the Cayley–Hamilton theorem, $p(A) = \mathbf{0}$. Thus, $q_{\pi_j}(A) = (A - \pi_j I)^{-1} \cdot \left(\mathbf{0} - p(\pi_j) I\right)$. The jth row of matrix G_i can be reformulated as

$$\begin{aligned} C_i A \left(A - \pi_j I\right)^{-1} &= -\frac{1}{p(\pi_j)} C_i A q_{\pi_j}(A) \\ &= -\frac{1}{p(\pi_j)} C_i \left(\sum_{k=0}^{n-1} b_{j,k} A^{k+1} \right) \\ &= -\frac{1}{p(\pi_j)} \begin{bmatrix} b_{j,0} & b_{j,1} & \cdots & b_{j,n-1} \end{bmatrix} \begin{bmatrix} C_i A \\ C_i A^2 \\ \vdots \\ C_i A^n \end{bmatrix} \\ &= -\frac{1}{p(\pi_j)} \begin{bmatrix} b_{j,0} & b_{j,1} & \cdots & b_{j,n-1} \end{bmatrix} O_i A. \end{aligned}$$

Therefore, G_i can be interpreted as follows

$$G_i = \mathcal{D}_1 \begin{bmatrix} b_{1,0} & b_{1,1} & \cdots & b_{1,n-1} \\ b_{2,0} & b_{2,1} & \cdots & b_{2,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n,0} & b_{n,1} & \cdots & b_{n,n-1} \end{bmatrix} O_i A = \mathcal{D}_1 \begin{bmatrix} \pi_1^{n-1} & \pi_1^{n-2} & \cdots & 1 \\ \pi_2^{n-1} & \pi_2^{n-2} & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ \pi_n^{n-1} & \pi_n^{n-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_n & 0 & \cdots & 0 \\ a_{n-1} & a_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \cdots & a_n \end{bmatrix} O_i A, \tag{A4}$$

where $\mathcal{D}_1 \triangleq \operatorname{diag}\left(-\frac{1}{p(\pi_1)}, -\frac{1}{p(\pi_2)}, \ldots, -\frac{1}{p(\pi_n)}\right)$. According to Assumption 2, all π_j are distinct eigenvalues and they are not the eigenvalues of A, that is, the diagonal matrix $\mathcal{D}_1$ and the Vandermonde matrix of π_i^j are invertible. Moreover, since $a_n \neq 0$, the lower triangular Toeplitz matrix of a_i is invertible and thus $\operatorname{rowspan}(G_i) = \operatorname{rowspan}(O_i A)$ from Equation (A4). We continue to prove $\operatorname{rowspan}(O_i) = \operatorname{rowspan}(O_i A)$. Considering that $A^n = -a_{n-1} A^{n-1} - \cdots - a_0 I$, one obtains the following Equation (A5).

$$O_i A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix} O_i. \tag{A5}$$

If A is invertible, we have $a_0 = (-1)^n \det(A) \neq 0$, which leads to the equation that $\operatorname{rowspan}(O_i) = \operatorname{rowspan}(O_i A)$. Therefore, one obtains $\mathbb{O}_i \triangleq \operatorname{rowspan}(O_i) = \operatorname{rowspan}(G_i)$. ■

APPENDIX B. PROOF OF THEOREM 3

Proof of Theorem 5. Considering the KKT condition of problem (25), 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 (25) and problem (19) are equivalent and the solution $\tilde{x}(k), \mu(k), \nu(k)$ satisfy

$$\tilde{x}(k) = \tilde{x}_{\mathrm{ls}}(k) = \hat{x}(k), \ \mu(k) = \varphi(k), \ \nu(k) = \mathbf{0}. \tag{B1}$$

According to Lemma 3, the solution $\varphi(k)$ of problem (19) satisfy the following equation:

$$\tilde{x}_{\text{ls}}(k) = \hat{x}(k),\ \ \varphi(k) = (I - GF)\epsilon(k), \tag{B2}$$

where $\hat{x}(k)$ is the fixed gain Kalman estimation defined in (3). Combining (B2) and (B1), result in Theorem 5 is obtained. ■

APPENDIX C. PROOF OF THEOREM 4

Before proving Theorem 6, we need the following lemma. Define the number of honest sensors and compromised sensors (w.r.t. compromised sensor index set $\mathcal{B}$) that can observe state j as:

$$h_j(\mathcal{B}) \triangleq |\mathcal{F}_j \cap \mathcal{I} \setminus \mathcal{B}|,\ \ c_j(\mathcal{B}) \triangleq |\mathcal{F}_j \cap \mathcal{B}|.$$

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

Lemma 5. *The following two propositions are equivalent.*

1. *The system is* $2p$*-sparse detectable.*
2. *For any* $\mathcal{B}$ *with* $|\mathcal{B}| = p$*, the inequality* $c_j(\mathcal{B}) < h_j(\mathcal{B})$ *holds for all* $j \in \mathcal{U}$*.*

Proof of Lemma 5. We prove the contrapositive of (1)⇒(2). Supposing that there exists j^* and $\mathcal{B}^*$ with $|\mathcal{B}^*| = p$ such that $c_{j^*}(\mathcal{B}^*) \geq h_{j^*}(\mathcal{B}^*)$, then $h_{j^*}(\mathcal{B}^*) \leq c_{j^*}(\mathcal{B}^*) \leq |\mathcal{B}^*| = p$. Noticing that $c_j(\mathcal{B}) + h_j(\mathcal{B}) = |\mathcal{F}_j|$ holds for all $\mathcal{B}$, we have $|\mathcal{F}_{j^*}| \leq 2p$. There exists sensor index set $\mathcal{B}$ that satisfy $\mathcal{B} \supseteq \mathcal{F}_{j^*}$ and $|\mathcal{B}| = 2p$. According to the definition of $\mathcal{F}_{j^*}$, there exists no sensor in set $\mathcal{I} \setminus \mathcal{B}$ who can observe state j^*, that is,

$$e^u_{j^*} \notin \text{rowspan}\left(O^{\mathcal{U}}_i\right),\ \forall i \in \mathcal{I} \setminus \mathcal{B}.$$

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

We proceed to prove (2)⇒(1). Since for any $\mathcal{B}$ with $|\mathcal{B}| = p$, $h_j(\mathcal{B}) > c_j(\mathcal{B}) \geq 0$, the system sparse detectability index is at least p. Therefore, for each $j \in \mathcal{U}$, there exists an $\mathcal{B}^*$ such that $c_j(\mathcal{B}^*) = p$, and thus $|\mathcal{F}_j| = h_j(\mathcal{B}^*) + c_j(\mathcal{B}^*) \geq 2p + 1$. According to the definition $\mathcal{F}_j$, there are at least $2p + 1$ sensors that can observe stable state j. Thus, the system is $2p$-sparse detectable. ■

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 3 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 6. Consider the KKT condition of problem (25) and denote the dual variables for equation constraints as $\lambda = [\lambda'_1, \ldots, \lambda'_m]' \in \mathbb{C}^{mn \times 1}$:

$$(\tilde{M}^{-1} + N'N)\mu + N'NH\tilde{x} - \lambda = \mathbf{0}, \tag{C1}$$

$$H'N'N\mu + H'N'NH\tilde{x} - H'\lambda = \mathbf{0}, \tag{C2}$$

$$\gamma \cdot \tilde{\nu} - \lambda = \mathbf{0}, \tag{C3}$$

$$Y - H\tilde{x} - \mu - \nu = \mathbf{0}, \tag{C4}$$

where $\tilde{\nu}$ belongs to the sub-gradient of $\|\nu\|_1$ and thus satisfies that $\|\tilde{\nu}\|_\infty \leq 1$.

Combining (C1) and (C2) leads to:

$$\begin{bmatrix} \tilde{M}^{-1} + N'N & N'NH \\ H'N'N & H'N'NH \end{bmatrix} \begin{bmatrix} \mu \\ \tilde{x} \end{bmatrix} = \begin{bmatrix} \lambda \\ H'\lambda \end{bmatrix}. \tag{C5}$$

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 (C5) in the following:

$$\begin{bmatrix} \tilde{M}^{-1} + N'N & N'NH\mathcal{L}' \\ \mathcal{L}H'N'N & \mathcal{L}H'N'NH\mathcal{L}' \end{bmatrix} \begin{bmatrix} \mu \\ \tilde{x}_s \end{bmatrix} = \begin{bmatrix} \lambda \\ \mathcal{L}H'\lambda \end{bmatrix}, \tag{C6}$$

where $\mathcal{L} \triangleq \begin{bmatrix} \mathbf{0}_{n_s \times n_u} & I_{n_s} \end{bmatrix}$. Rewrite (C6) as:

$$\left(\begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & \mathcal{L}H'N' \end{bmatrix} \mathcal{W} \begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & NH\mathcal{L}' \end{bmatrix} \right) \begin{bmatrix} \mu \\ \tilde{x}_s \end{bmatrix} = \begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & \mathcal{L}H' \end{bmatrix} \lambda. \tag{C7}$$

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 (C7) is also invertible, and thus the following matrix is well-defined:

$$\mathcal{H} \triangleq \left(\begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & \mathcal{L}H'N' \end{bmatrix} \mathcal{W} \begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & NH\mathcal{L}' \end{bmatrix} \right)^{-1} \begin{bmatrix} I_{mn} & \mathbf{0} \\ \mathbf{0} & \mathcal{L}H' \end{bmatrix}. \tag{C8}$$

According to (C3), $\|\lambda\|_\infty \leq \gamma$. Therefore we have the following from (C7):

$$\left\| \begin{bmatrix} \mu \\ \tilde{x}_s \end{bmatrix} \right\|_\infty \leq \gamma \cdot \|\mathcal{H}\|_\infty . \tag{C9}$$

Now we continue to prove that the estimation of unstable states $\tilde{x}_u$ are secure. Rewrite the optimization problem (25) 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:

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

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 :

$$\min_{\tilde{x}_u} \sum_{i=1}^{m} \left\| \eta_{i,u} - \mu_{i,u}^* - H_{us,i}\tilde{x}_s^* - H_{uu,i}\tilde{x}_u \right\|_1 . \tag{C10}$$

Define $\xi_i \triangleq \eta_{i,u} - \mu_{i,u}^* - H_{us,i}\tilde{x}_s^*$ and recall $[\cdot]_j$ is the jth entry of a vector. The objective function in (C10) can be written as

$$\sum_{i=1}^{m} \sum_{j=1}^{n_u} \left| [\xi_i]_j - [H_{uu,i}\tilde{x}_u]_j \right| = \sum_{j=1}^{n_u} \sum_{i \in \mathcal{F}_j} \left| [\xi_i]_j - \tilde{x}_j \right| . \tag{C11}$$

where $\mathcal{F}_j$ 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 (C11) could be explicitly written as the median of all $[\xi_i]_j$ among $i \in \mathcal{F}_j$.

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 $f_i\left(x_l, l \in \{1, \dots, L\}\right)$ equals to the ith smallest element in the set $\{x_1, \dots, x_L\}$. 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, $f_{(L+1)/2}\left(x_l, l \in \{1, \dots, L\}\right)$ is the median number of set $\{x_1, \dots, x_L\}$ and the solution to problem (C10) is

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

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

$$f_{(h_j - c_j)}\left([\xi_i^o]_j, i \in \mathcal{F}_j\right) \le f_{(m+1)/2}\left([\xi_i]_j, i \in \mathcal{F}_j\right), \tag{C12}$$

$$f_{(m+1)/2}\left([\xi_i]_j, i \in \mathcal{F}_j\right) \le f_{2c_j}\left([\xi_i^o]_j, i \in \mathcal{F}_j\right). \tag{C13}$$

According to Lemma 5, $h_j - c_j > 0$ and $2c_j < h_j + c_j = |\mathcal{F}_j|$. As a result, according to (C12) and (C13), one obtains

$$\min\left\{[\xi_i^o]_j, i \in \mathcal{F}_j\right\} \le \tilde{x}_j \le \max\left\{[\xi_i^o]_j, i \in \mathcal{F}_j\right\}, \ j \in \mathcal{U}. \tag{C14}$$

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 5, 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, that is, $\tilde{x}^o(k) = \hat{x}^o(k)$. Similar to previous analysis, the solution $\tilde{x}^o$ satisfies

$$[\tilde{x}^o]_j = f_{(|\mathcal{F}_j|+1)/2}\left([P_i\zeta_i^o - \mu_i^o]_j, i \in \mathcal{F}_j\right), \forall j \in \mathcal{U} \cup \mathcal{S}. \tag{C15}$$

Combining (C14) and (C15) leads to that, for every $j \in \mathcal{U}$,

$$\begin{aligned}
\left|[\tilde{x}]_j - [\tilde{x}^o]_j\right| = \left|[\tilde{x}]_j - [\hat{x}^o]_j\right| &\le \max_{i_1, i_2 \in \mathcal{F}_j}\left|\left[P_{i_1}\zeta_{i_1}^o(k)\right]_j - \left[P_{i_2}\zeta_{i_2}^o(k)\right]_j\right| + \|\mu^*\|_\infty + \|\mu^o\|_\infty \\
&\le \max_{i_1, i_2 \in \mathcal{F}_j}\left|\left[P_{i_1}\zeta_{i_1}^o(k)\right]_j - \left[P_{i_2}\zeta_{i_2}^o(k)\right]_j\right| + (\gamma + \gamma^o)\|\mathcal{H}\|_\infty.
\end{aligned}$$

Recall that $P_i\epsilon_i(k) = P_i\zeta_i(k) - H_i x(k)$. Since for all $i_1, i_2 \in \mathcal{F}_j$, 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

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

whose variance is uniformly bounded for all k according to Lemma 1. Similarly, γ^o is also uniformly bounded for all k.

For stable state,

$$\left|\tilde{x}_j - \hat{x}_j^o\right| \leq \left|\tilde{x}_j\right| + \left|\hat{x}_j^o\right| \leq \gamma \cdot \|\mathcal{H}\|_\infty + \left|\hat{x}_j^o\right|, \; j \in \mathcal{S},$$

where $\left|\tilde{x}_j\right| \leq \gamma \cdot \|\mathcal{H}\|_\infty, j \in \mathcal{S}$ comes from (C9). The oracle Kalman estimation $\hat{x}_j^o$ of stable state has bounded covariance. As a result, our estimation $\tilde{x}$ is secure according to Definition 3. ■