sensors

MDPI

Article

Observability Decomposition-Based Decentralized Kalman Filter and Its Application to Resilient State Estimation under Sensor Attacks

Chanhwa Lee

School of Intelligent Mechatronics Engineering, Sejong University, Seoul 05006, Korea; chlee@sejong.ac.kr

Abstract: This paper considers a discrete-time linear time invariant system in the presence of Gaussian disturbances/noises and sparse sensor attacks. First, we propose an optimal decentralized multi-sensor information fusion Kalman filter based on the observability decomposition when there is no sensor attack. The proposed decentralized Kalman filter deploys a bank of local observers who utilize their own single sensor information and generate the state estimate for the observable subspace. In the absence of an attack, the state estimate achieves the minimum variance, and the computational process does not suffer from the divergent error covariance matrix. Second, the decentralized Kalman filter method is applied in the presence of sparse sensor attacks as well as Gaussian disturbances/noises. Based on the redundant observability, an attack detection scheme by the χ^2 test and a resilient state estimation algorithm by the maximum likelihood decision rule among multiple hypotheses, are presented. The secure state estimation algorithm finally produces a state estimate that is most likely to have minimum variance with an unbiased mean. Simulation results on a motor controlled multiple torsion system are provided to validate the effectiveness of the proposed algorithm.

Keywords: information fusion; decentralized Kalman filter; observability decomposition; attack resilience; secure state estimation; redundant observability; sparse sensor attack

Citation: Lee, C. Observability Decomposition-Based Decentralized Kalman Filter and Its Application to Resilient State Estimation under Sensor Attacks. *Sensors* **2022**, *22*, 6909. https://doi.org/10.3390/s22186909

Academic Editor: Fanglai Zhu

Received: 16 August 2022
Accepted: 8 September 2022
Published: 13 September 2022

1. Introduction

As control systems operate through network communication and become more complex due to increased connectivity, security against adversarial attacks is becoming more important and receiving attention [1–4]. In fact, attacks on control systems took place in reality [5–8], and many studies have been conducted on the security issues of systems whose measurements have been compromised by adversaries because sensors are one of the vulnerable points to malicious attackers in dynamical systems [9–15].

Among them, the state estimation problem when some of sensors are corrupted by attackers, often called a sparse sensor attack, has been investigated, and several solutions have been recently proposed [10–15]. The reference [10] introduces the basic concepts of the secure state estimation problem and formulates it as a non-convex combinatorial optimization problem. The problem is shown to be transformed into a convex optimization problem by using the results developed in the field of compressed sensing [16,17] under additional limiting assumptions. The relationship between this resilient state estimation problem and the notion of strong observability was revealed in [11]. A necessary and sufficient condition for the solvability of this problem is derived in [12,15] with the notion of redundant observability, more specifically, it requires the redundancy of observability twice as much as the sparsity of sensor attacks. A method to alleviate the computational complexity of the logic for finding a combination of non-attacked sensors, is proposed in [13,14]. In [15], the estimator is designed by a set of local observers with only a single sensor, and the decoder uses an error correction algorithm to generate a final state estimate based on the data collected from each local observer.

In addition to sparse sensor attacks, disturbances and noises are considered to enhance the robustness. First, bounded disturbances and noises are considered in [13,15,18], and in particular, the reference [15] explicitly derives the estimation error with the system parameters to provide an analysis of robustness. Second, zero-mean Gaussian white noises and disturbances rather than bounded ones were considered in [19–21], and Kalman filters were used to guarantee the state-estimation performance in a probabilistic manner. The reference [19] proposed an estimator with Kalman filters that searches a reliable subset of sensors and operates on the identified subset. A method of combining a secure state estimator and the standard Kalman filter by using the secure state estimator as a pre-filter for the Kalman filter when the set of attacked sensors changes over time, is proposed in [20]. It was shown in [21] that the optimal Kalman estimate can be decomposed into a weighted sum of local estimates, where each estimate uses only a single sensor measurement and that a secure state estimation can be achieved by a convex optimization under some additional assumptions.

This paper considers a general discrete-time linear dynamical system that is corrupted by sparse sensor attacks and Gaussian disturbances/noises. First, we construct local observers on each single sensor and design those local observers with Kalman filters using their own sensor data to cope with Gaussian disturbances/noises. The design of local observers is fully decentralized since it does not utilize any information including Kalman gains or error covariance matrices from other sensors as well as the sensor readings. Furthermore, the local observer's error covariance is guaranteed not to diverge since it is constructed in the observable subspace based on the observability decomposition, and thus, there is no numerical computational error in practice. Second, a novel information fusion scheme is developed to counteract sparse sensor attacks while maintaining the minimum variance properties. The information fusion center detects the presence of sensor attack in the selected subset of sensors by the χ^2 test, which is typically used in the area of fault detection [22,23]. If the χ^2 test concludes that there is an attack in the selected subset, a search algorithm is launched to choose a new index set of sensors that is most likely to be unattacked by the multiple hypothesis test. Each hypothesis produces a state estimate with minimum variance, assuming that the index set is attack-free so that each estimate is unbiased. Therefore, the information fusion scheme finally produces a state estimate that is most likely to have the minimum variance and to be unbiased.

Assuming that there exist only Gaussian disturbances/noises without any attacks, a basic information fusion Kalman filter scheme was proposed in [24,25]. The local observers in [24,25] were designed using a Kalman filter for the entire state variable with a single sensor, and a fusion algorithm generates the optimal state estimate with the minimum variance. However, as highlighted in [26], some components of the error covariance may diverge if a single-sensor system is not observable, and this can induce numerical computation problems in practice. This problem can be solved by reducing the target state space to an observable subspace and designing a Kalman filter for the reduced observable subsystem. The idea of decomposing a single-sensor system into the observable subsystem and the unobservable subsystem was proposed in [15] for the secure state estimator design under bounded disturbances/noises, and in [27] for the distributed Luenberger observer design of sensor networks. Hence, adopting this idea and designing the Kalman filter for the observable subsystem, the problem of divergent error covariance does not occur, and we derive the optimal information fusion algorithm even when the size of the local information is different each other.

The contributions of this paper can be summarized as follows:

(1) The proposed algorithm successfully estimates the state variable under sparse sensor attacks as well as Gaussian disturbances/noises. Our algorithm ensures the minimum variance, while [19] simply guarantees that its covariance is no worse than the worst case scenario with high probability;
(2) We only assume that the system is redundant observable, which is known as an equivalent condition for the secure state estimation to be solvable under sparse sensor

attacks. Note that [20] requires additional assumptions to reformulate the problem as a convex problem, and further, the combination of Kalman filter and the secure estimator implicitly supposes that the estimation error for the attack signal follows a zero-mean Gaussian distribution, which may not be true when the attack signal is intelligently designed in a coordinated way. The reference [21] needs the system matrix to be nonsingular, and both references [20] and [21] have additional assumptions about the closed-loop system;

(3) The construction of the local observer is completely decentralized, and the overall size of the observer is relatively small. As the combinatorial logic is embedded in the fusion center, we do not have to prepare all possible combinations of observers. Note that [19] does not utilize any decomposition, and thus, it asks for all combinations of observers. The local decomposition presented in [21] is not fully decentralized because the decomposition is performed using the global information of the output matrix and the Kalman gain;

(4) As a by-product obtained during the derivation process, the optimal decentralized information fusion Kalman filter scheme is developed based on the observability decomposition. Compared with the results in [24,25], the proposed scheme does not suffer from the numerical computational errors resulting from the diverging error covariance matrix. The algorithm in this paper guarantees that each error covariance matrix in the local observer converges by the observability decomposition, and this method can also be widely used for the multi-sensor information fusion Kalman filters that do not consider any attacks.

The rest of the paper is organized as follows. The remaining of this section introduces the notation used throughout the paper. The system model and problem formulation are given in Section 2. Section 3 presents the optimal multi-sensor information fusion Kalman filter based on the observability decomposition. We then give the attack detection algorithm by χ^2 test and the attack-resilient state estimation scheme by the multiple hypothesis test in Section 4. Finally, simulation results with a servo motor system are given in Section 5, and we provide our concluding remarks in Section 6. The preliminary results of this paper were studied in [28].

Notation: Throughout this paper, the following notations are adopted. For a set S, the number of elements in the set S is denoted by $|S|$. For a column vector $y \in \mathbb{R}^{\mathsf{p}}$ and its i-th element y_{i}, $\mathsf{supp}(y)$ denotes the number of nonzero elements of the vector y, that is, $\mathsf{supp}(y) := \{\mathsf{i} \in [\mathsf{p}] : y_{\mathsf{i}} \neq 0\}$ where the symbol $[\mathsf{p}]$ is used to represent the subset of natural numbers $\{1, 2, \cdots, \mathsf{p}\} \subset \mathbb{N}$. The number of nonzero elements of a vector y is defined by the ℓ_0 norm, and it is written as $\|y\|_0 := |\mathsf{supp}(y)|$. We say that the vector y is q-sparse if its ℓ_0 norm is less than or equal to q, that is, $\|y\|_0 \leq \mathsf{q}$.

For an index set $\mathcal{I} \subset [\mathsf{p}]$ and a vector $y \in \mathbb{R}^{\mathsf{p}}$ (or a matrix $C \in \mathbb{R}^{\mathsf{p}\times\mathsf{n}}$), $y_{\mathcal{I}} \in \mathbb{R}^{|\mathcal{I}|}$ (or $C_{\mathcal{I}} \in \mathbb{R}^{|\mathcal{I}|\times\mathsf{n}}$) denotes the vector (or the matrix) obtained from y (or C) by eliminating all i-th rows such that $\mathsf{i} \in \mathcal{I}^c$. Similarly, for two index sets $\mathcal{I}, \mathcal{J} \subset [\mathsf{p}]$ and a matrix $P \in \mathbb{R}^{\mathsf{p}\times\mathsf{p}}$, $P_{\mathcal{I},\mathcal{J}} \in \mathbb{R}^{|\mathcal{I}|\times|\mathcal{J}|}$ denotes the matrix obtained from P by eliminating all i-th rows and all j-th columns such that $\mathsf{i} \in \mathcal{I}^c$ and $\mathsf{j} \in \mathcal{J}^c$.

Let a finite sequence $\{\mu_{\mathsf{i}}\} = \{\mu_1, \mu_2, \cdots, \mu_{\mathsf{p}}\}$ with $\mu = \sum_{\mathsf{i}=1}^{\mathsf{p}} \mu_{\mathsf{i}}$ given. A stacked vector $z = \begin{bmatrix} z_1^\top & z_2^\top & \cdots & z_{\mathsf{p}}^\top \end{bmatrix}^\top \in \mathbb{R}^{\mu}$ is said to be partitioned by the sequence $\{\mu_{\mathsf{i}}\}$ if $z_{\mathsf{i}} \in \mathbb{R}^{\mu_{\mathsf{i}}}$ for all $\mathsf{i} \in [\mathsf{p}]$. For $\mathsf{j} \in [\mathsf{p}]$, an index set $\mathcal{I}_{\mathsf{j}}^{\{\mu_{\mathsf{i}}\}} := \left\{ (\sum_{\mathsf{i}=1}^{\mathsf{j}-1} \mu_{\mathsf{i}}) + 1, (\sum_{\mathsf{i}=1}^{\mathsf{j}-1} \mu_{\mathsf{i}}) + 2, \cdots, \sum_{\mathsf{i}=1}^{\mathsf{j}} \mu_{\mathsf{i}} \right\} \subset [\mu]$ represents the j-th partition among total p partitions when a vector $z \in \mathbb{R}^{\mu}$ is partitioned by the sequence $\{\mu_{\mathsf{i}}\}$. This notation is extended to a subset $\mathcal{J} \subset [\mathsf{p}]$ where $\mathcal{I}_{\mathcal{J}}^{\{\mu_{\mathsf{i}}\}}$ denotes $\bigcup_{\mathsf{j}\in\mathcal{J}} \mathcal{I}_{\mathsf{j}}^{\{\mu_{\mathsf{i}}\}}$. A vector $z \in \mathbb{R}^{\mu}$ partitioned by the sequence $\{\mu_{\mathsf{i}}\}$, is said to be ($\{\mu_{\mathsf{i}}\}$-stacked) q-sparse if $\left| \left\{ \mathsf{j} \in [\mathsf{p}] : z_{\mathcal{I}_{\mathsf{j}}^{\{\mu_{\mathsf{i}}\}}} \neq 0_{\mu_{\mathsf{j}}\times 1} \right\} \right| \leq \mathsf{q}$.

2. System Modeling and Problem Formulation

The plant and the attack model under consideration are presented, and the problem formulation is given in this section.

2.1. Plant Modeling with Gaussian Disturbances and Noises

A discrete-time linear time invariant (LTI) system under Gaussian disturbances and noises given by

$$\mathcal{P}: \begin{cases} x(k+1) = Ax(k) + Bu(k) + d(k) \\ y(k) = Cx(k) + n(k) \end{cases} \tag{1}$$

is considered. In the plant dynamics of (1), $x \in \mathbb{R}^{\mathsf{n}}$ is the state variable vector, $u \in \mathbb{R}^{\mathsf{m}}$ is the control input vector, and $y \in \mathbb{R}^{\mathsf{p}}$ is the sensor output vector. Furthermore, the dynamics is disrupted by the process disturbance $d \in \mathbb{R}^{\mathsf{n}}$, and the sensors are corrupted by the measurement noise $n \in \mathbb{R}^{\mathsf{p}}$. There are a total of p sensors that measure the system outputs, and the i-th sensor's measurement at time k is denoted by

$$y_{\mathsf{i}}(k) = c_{\mathsf{i}} x(k) + n_{\mathsf{i}}(k)$$

where c_{i} is the i-th row of the output matrix C, which implies that $C = [c_1^\top \ c_2^\top \ \cdots \ c_{\mathsf{p}}^\top]^\top$. Here, stochastic assumptions on the disturbance $d(k)$, the noise $n(k)$ and the initial state $x(0)$ of the system (1) are formally stated as follows.

Assumption 1. *The disturbance $d(k)$ and measurement noise $n(k)$ are independent and identically distributed (i.i.d.) white Gaussian process with zero-mean and covariance matrices Q and R, respectively. More specifically,*

$$\begin{aligned} d(k) &\sim N(0_{\mathsf{n}\times 1},\ Q), \\ n(k) &\sim N(0_{\mathsf{p}\times 1},\ R), \\ \mathbf{E}[d(k)] = 0_{\mathsf{n}\times 1}, &\quad \mathbf{E}[d(k)d^\top(t)] = Q\delta_{kt}, \\ \mathbf{E}[n(k)] = 0_{\mathsf{p}\times 1}, &\quad \mathbf{E}[n(k)n^\top(t)] = R\delta_{kt}, \\ &\quad \mathbf{E}[n(k)d^\top(t)] = O_{\mathsf{p}\times \mathsf{n}}, \end{aligned}$$

where the symbol $\mathbf{E}[\cdot]$ represents the expected value of a random variable and δ_{kt} is the Kronecker delta function. Furthermore, the initial state $x(0)$ is a Gaussian distributed random variable with the mean $\bar{x}_0$ and covariance matrix P_0,

$$\begin{aligned} x(0) &\sim N(\bar{x}_0, P_0), \\ \mathbf{E}[x(0)] = \bar{x}_0, &\quad \mathbf{E}[(x(0) - \bar{x}_0)(x(0) - \bar{x}_0)^\top] = P_0, \end{aligned}$$

and is independent of $d(k)$ and $n(k)$.

2.2. Attack Modeling with Sparse Sensor Attacks

Among various attack scenarios [3], we consider false data injection attacks on sensors. Adversarial attackers can inject arbitrary inputs to some (not all) sensors so that a part of the measurements is compromised. Some additive inputs may be induced by cyber or physical tampering with the sensors, or adversaries may penetrate into the communication network on the output side of the plant because those communication links are not secure. In both cases, the attack is characterized by the attack vector $a \in \mathbb{R}^{\mathsf{p}}$ as in

$$\begin{aligned} y^a(k) &= y(k) + a(k) \\ &= Cx(k) + n(k) + a(k) \\ &= Cx(k) + n^a(k) \end{aligned} \tag{2}$$

where $y^a \in \mathbb{R}^{\mathsf{p}}$ denotes sensor readings with a potential attack, while $y \in \mathbb{R}^{\mathsf{p}}$ is the original healthy sensor data affected by the measurement noise only. Similarly, $n^a \in \mathbb{R}^{\mathsf{p}}$ represents the total sensor contamination signal including both the noise n and the attack a.

Here, it is assumed that the adversaries can compromise only a part of the sensors, not all of them. Assuming that the attacker's resources are limited, we suppose that the attacker can contaminate up to q out of p measurement outputs. Therefore, a formal condition on the sparsity of the attack vector a can be given as follows.

Assumption 2. *The sensor attack vector $a(k)$ is q-sparse for all $k \geq 0$, that is, $\|a(k)\|_0 \leq \mathsf{q}$, $^{\forall}k \geq 0$. Moreover, it holds that*

$$|\{\mathsf{i} \in [\mathsf{p}] : a_{\mathsf{i}}(k) \neq 0 \text{ for some } k \geq 0\}| \leq \mathsf{q}.$$

This assumption tells more than $\|a(k)\|_0 \leq \mathsf{q}$ for all $k \geq 0$, in the sense that the compromised sensor channels are not altered for all time. In practice, this may be the case because it takes quite a long time and much effort to infiltrate into a new sensor from a malicious attacker's point of view. Thus, without loss of generality, it can be assumed that the attack channels remain the same in the long term although it is not revealed to the controller which channels are attacked. However, if the attacked sensor channel changes but does not change frequently, the resilient state estimation scheme to be presented is still applicable. We will simply refer to this assumption as a "q-sparse sensor attack".

2.3. Problem Formulation

For the given discrete-time LTI system (1) under Assumptions 1 and 2, this paper investigates how to design an estimator that can recover the state variable x correctly. First, the Gaussian distributed disturbances/noises are handled appropriately, and the optimality in the sense of minimum variance should be recovered. Second, the security against the sparse sensor attack is enhanced, and the attack-resilient estimation with the unbiased state estimate should be achieved. More specifically, this paper considers the problem of proposing a secure and robust state estimation algorithm that generates the estimate that is most likely to have the minimum variance and to be unbiased. In this process, the concept of "redundant observability", which characterizes the ability of coping with the sparse sensor attack, is utilized to ensure successful state estimation.

The basic condition for the observability of the system (1) with the attack model (2) satisfying Assumption 2, is given in the following assumption. Note that the assumption of "$2\mathsf{q}$ redundant observability" is an equivalent condition for the system to be observable under q-sparse sensor attacks ([15], Proposition 2,3,6). Here, the state estimation problem becomes challenging because this redundant observability does not guarantee for the entire states to be recovered with only a single sensor.

Assumption 3. *The system (1), or the pair (A, C), is $2\mathsf{q}$ redundant observable. In other words, each pair $(A, C_{\mathcal{I}})$ is observable for any $\mathcal{I} \subset [\mathsf{p}]$ satisfying $|\mathcal{I}| \geq \mathsf{p} - 2\mathsf{q}$.*

3. Optimal Information Fusion Kalman Filter Based on Observability Decomposition

3.1. Kalman Observability Decomposition with Single Sensor

Since conventional Luenburger observers or Kalman filters typically have the form of

$$\hat{x}(k+1) = (A - KC)\hat{x}(k) + Bu(k) + Ky^a(k),$$

the whole state estimates $\hat{x}$ are affected by the single sensor attack signal due to the observer gain K. In other words, any single non-zero component of a can alter all components of the state estimate $\hat{x}$. Hence, we design a collection of observers where each local observer utilizes only a single sensor information so that an attack signal for one sensor channel only interferes with the corresponding local observer and leaves other local observers unaffected.

Consider a single-output system

$$\mathcal{P}_{\mathsf{i}} : \begin{cases} x(k+1) = Ax(k) + Bu(k) + d(k) \\ y_{\mathsf{i}}^{a}(k) = c_{\mathsf{i}} x(k) + n_{\mathsf{i}}^{a}(k). \end{cases} \tag{3}$$

where the i-th component of $y^a(k)$ in (2), $y_{\mathsf{i}}^a(k)$, is the output and the dynamics is given by (1). Since the pair (A, c_{i}) is not necessarily observable, an estimator of the system (3) generally recovers only an (observable) portion of the full state x. The Kalman observability decomposition, which clearly describes the observable portion of the system, is now briefly introduced. For the single-output system (3), the observability matrix is written as

$$G_{\mathsf{i}} := \begin{bmatrix} c_{\mathsf{i}} \\ c_{\mathsf{i}} A \\ c_{\mathsf{i}} A^2 \\ \vdots \\ c_{\mathsf{i}} A^{\mathsf{n}-1} \end{bmatrix}, \tag{4}$$

and we denote μ_{i} as the rank of the observability matrix G_{i}. The null space of G_{i}, $\mathcal{N}(G_{\mathsf{i}})$, is the so-called unobservable subspace, and the column range space of $G_{\mathsf{i}}^{\top}$, $\mathcal{R}(G_{\mathsf{i}}^{\top})$, is often called the observable subspace.

One can define the similarity transformation as

$$\begin{bmatrix} z_{\mathsf{i}} \\ w_{\mathsf{i}} \end{bmatrix} = \begin{bmatrix} Z_{\mathsf{i}}^{\top} \\ W_{\mathsf{i}}^{\top} \end{bmatrix} x \tag{5}$$

where $Z_{\mathsf{i}} \in \mathbb{R}^{\mathsf{n} \times \mu_{\mathsf{i}}}$ is the matrix whose columns are th orthonormal basis of $\mathcal{R}(G_{\mathsf{i}}^{\top})$ and $W_{\mathsf{i}} \in \mathbb{R}^{\mathsf{n} \times (\mathsf{n}-\mu_{\mathsf{i}})}$ is the matrix whose columns are the orthonormal basis of $\mathcal{N}(G_{\mathsf{i}})$. Here, the size of those matrices is determined by

$$\mu_{\mathsf{i}} = \mathsf{rank}(G_{\mathsf{i}}) = \mathsf{dim}(\mathcal{R}(G_{\mathsf{i}}^{\top})) \quad \text{and} \quad \mathsf{n} - \mu_{\mathsf{i}} = \mathsf{nullity}(G_{\mathsf{i}}) = \mathsf{dim}(\mathcal{N}(G_{\mathsf{i}})).$$

Note that the observable subspace $\mathcal{R}(G_{\mathsf{i}}^{\top})$ is the span of column vectors in Z_{i} and the unobservable subspace $\mathcal{N}(G_{\mathsf{i}})$ is the span of column vectors in W_{i}. Since the matrix $[Z_{\mathsf{i}} \;\; W_{\mathsf{i}}]$ is orthogonal, we have

$$\begin{bmatrix} Z_{\mathsf{i}}^{\top} \\ W_{\mathsf{i}}^{\top} \end{bmatrix} \begin{bmatrix} Z_{\mathsf{i}} & W_{\mathsf{i}} \end{bmatrix} = \begin{bmatrix} Z_{\mathsf{i}}^{\top} Z_{\mathsf{i}} & Z_{\mathsf{i}}^{\top} W_{\mathsf{i}} \\ W_{\mathsf{i}}^{\top} Z_{\mathsf{i}} & W_{\mathsf{i}}^{\top} W_{\mathsf{i}} \end{bmatrix} = \begin{bmatrix} I_{\mu_{\mathsf{i}} \times \mu_{\mathsf{i}}} & O_{\mu_{\mathsf{i}} \times (\mathsf{n}-\mu_{\mathsf{i}})} \\ O_{(\mathsf{n}-\mu_{\mathsf{i}}) \times \mu_{\mathsf{i}}} & I_{(\mathsf{n}-\mu_{\mathsf{i}}) \times (\mathsf{n}-\mu_{\mathsf{i}})} \end{bmatrix}.$$

Moreover, because the unobservable subspace is A-invariant, any columns of AW_{i} belong to $\mathcal{N}(G_{\mathsf{i}}) = \mathcal{R}(W_{\mathsf{i}})$. Therefore, the Kalman observability decomposition of the system (3) is obtained by the transformation (5) as

$$\mathcal{P}_{\mathsf{i}}' : \begin{cases} \begin{bmatrix} z_{\mathsf{i}}(k+1) \\ w_{\mathsf{i}}(k+1) \end{bmatrix} = \begin{bmatrix} Z_{\mathsf{i}}^{\top} A Z_{\mathsf{i}} & O_{\mu_{\mathsf{i}} \times (\mathsf{n}-\mu_{\mathsf{i}})} \\ W_{\mathsf{i}}^{\top} A Z_{\mathsf{i}} & W_{\mathsf{i}}^{\top} A W_{\mathsf{i}} \end{bmatrix} \begin{bmatrix} z_{\mathsf{i}}(k) \\ w_{\mathsf{i}}(k) \end{bmatrix} + \begin{bmatrix} Z_{\mathsf{i}}^{\top} B \\ W_{\mathsf{i}}^{\top} B \end{bmatrix} u(k) + \begin{bmatrix} Z_{\mathsf{i}}^{\top} \\ W_{\mathsf{i}}^{\top} \end{bmatrix} d(k) \\ y_{\mathsf{i}}^{a}(k) = \begin{bmatrix} c_{\mathsf{i}} Z_{\mathsf{i}} & 0_{1 \times (\mathsf{n}-\mu_{\mathsf{i}})} \end{bmatrix} \begin{bmatrix} z_{\mathsf{i}}(k) \\ w_{\mathsf{i}}(k) \end{bmatrix} + n_{\mathsf{i}}^{a}(k). \end{cases} \tag{6}$$

Finally, the state $x \in \mathbb{R}^{\mathsf{n}}$ is decomposed into the observable sub-state $z_{\mathsf{i}} \in \mathbb{R}^{\mu_{\mathsf{i}}}$ and the unobservable sub-state $w_{\mathsf{i}} \in \mathbb{R}^{\mathsf{n}-\mu_{\mathsf{i}}}$. Further, the observable part of (6) can simply be written as

$$\mathcal{P}_{\mathsf{i}}^{o} : \begin{cases} z_{\mathsf{i}}(k+1) = S_{\mathsf{i}} z_{\mathsf{i}}(k) + Z_{\mathsf{i}}^{\top} B u(k) + Z_{\mathsf{i}}^{\top} d(k) \\ y_{\mathsf{i}}^{a}(k) = t_{\mathsf{i}} z_{\mathsf{i}}(k) + n_{\mathsf{i}}^{a}(k) \end{cases} \tag{7}$$

where $S_{\mathsf{i}} := Z_{\mathsf{i}}^{\top} A Z_{\mathsf{i}}$ and $t_{\mathsf{i}} := c_{\mathsf{i}} Z_{\mathsf{i}}$.

3.2. Decentralized Multi-Sensor Kalman Filter

Even though the Kalman filter can be applied to unobservable linear systems, the error covariance matrix may not converge in that case. According to ([29], Theorem 26), the detectability of the system is a sufficient condition for the convergence of the error covariance matrix in Kalman filtering. Since detectability is a slightly weaker concept than observability, the results in this paper dealing with observability can be generalized to the concept of detectability with slight modifications. The design of local state estimators for the observable subsystem (7) in the form of Kalman filters using only single sensor information, is derived in this subsection. By its construction, the pair $(Z_\mathsf{i}^\top A Z_\mathsf{i}, c_\mathsf{i} Z_\mathsf{i})$, or simply denoted as $(S_\mathsf{i}, t_\mathsf{i})$, is observable, and thus, the error covariance matrix of the Kalman filter designed for the system (7) converges to a positive semidefinite matrix ([29], Theorem 26).

Now, we design a decentralized Kalman filter with each single sensor output, which constitutes the local observer. Then, the design of an information fusion scheme, which collects all the information on state estimates and error covariance matrices from the decentralized Kalman filters, will be discussed in the next subsection. For the simplicity of the derivation, we assume that there are no attacks at this time, that is, $a(k) \equiv 0$. Thus, $n^a(k)$ and $y^a(k)$ are interpreted as $n(k)$ and $y(k)$, respectively, in this section.

Stochastic assumptions on the disturbance $d(k)$ and the noise $n(k)$ of the system (1) are formally stated in Assumption 1 where the covariance matrix R of the measurement noise $n(k)$ is partitioned as

$$R = \begin{bmatrix} R_1 & R_{12} & \cdots & R_{1\mathsf{p}} \\ R_{21} & R_2 & \cdots & R_{2\mathsf{p}} \\ \vdots & \vdots & \ddots & \vdots \\ R_{\mathsf{p}1} & R_{\mathsf{p}2} & \cdots & R_\mathsf{p} \end{bmatrix}.$$

Finally, the assumption for each measurement noise $n_\mathsf{i}(k)$ (which is the same as $n_\mathsf{i}^a(k)$ in this section) of the system (3) can be written as follows:

$$\begin{aligned} n_\mathsf{i}(k) &\sim N(0, R_\mathsf{i}), \\ \mathbf{E}[n_\mathsf{i}(k)] = 0, \quad \mathbf{E}[n_\mathsf{i}(k) n_\mathsf{i}^\top(t)] &= R_\mathsf{i} \delta_{kt}, \\ \mathbf{E}[n_\mathsf{i}(k) n_\mathsf{j}^\top(t)] &= R_{\mathsf{ij}} \delta_{kt}, \quad \text{if } \mathsf{i} \neq \mathsf{j}, \\ \mathbf{E}[n_\mathsf{i}(k) d^\top(t)] &= 0_{1\times \mathsf{n}}. \end{aligned}$$

The local observer is designed by a Kalman filter for the observable subsystem (7). To this end, let $\hat{z}_\mathsf{i}(k|k-1)$ be the estimate of $z_\mathsf{i}(k)$ based on observations from $y^a(0)$ to $y^a(k-1)$. Similarly, $\hat{z}_\mathsf{i}(k|k)$ is the estimate of $z_\mathsf{i}(k)$ after we process the measurement $y^a(k)$ at time k. Following the conventional notations in a Kalman filter, we use the terms $P_\mathsf{i}(k|k-1)$ and $P_\mathsf{i}(k|k)$ to denote the estimation error covariance of $\hat{z}_\mathsf{i}(k|k-1)$ and $\hat{z}_\mathsf{i}(k|k)$, respectively. Thus, We have

$$\begin{aligned} P_\mathsf{i}(k|k-1) &= \mathbf{E}[(\hat{z}_\mathsf{i}(k|k-1) - z_\mathsf{i}(k))(\hat{z}_\mathsf{i}(k|k-1) - z_\mathsf{i}(k))^\top], \\ P_\mathsf{i}(k|k) &= \mathbf{E}[(\hat{z}_\mathsf{i}(k|k) - z_\mathsf{i}(k))(\hat{z}_\mathsf{i}(k|k) - z_\mathsf{i}(k))^\top]. \end{aligned} \tag{8}$$

Then, the Kalman filter has the following form of

$$\begin{aligned} &\mathcal{O}_\mathsf{i}: \ \hat{z}_\mathsf{i}(k+1|k+1) \\ &= S_\mathsf{i}\hat{z}_\mathsf{i}(k|k) + Z_\mathsf{i}^\top B u(k) + K_\mathsf{i}(k+1)\Big(y_\mathsf{i}^a(k+1) - t_\mathsf{i}\Big(S_\mathsf{i}\hat{z}_\mathsf{i}(k|k) + Z_\mathsf{i}^\top B u(k)\Big)\Big) \\ &= (I - K_\mathsf{i}(k+1)t_\mathsf{i})\Big(S_\mathsf{i}\hat{z}_\mathsf{i}(k|k) + Z_\mathsf{i}^\top B u(k)\Big) + K_\mathsf{i}(k+1) y_\mathsf{i}^a(k+1), \end{aligned} \tag{9}$$

where

$$\hat{z}_i(k+1|k+1) = \hat{z}_i(k+1|k) + K_i(k+1)(y_i^a(k+1) - t_i\hat{z}_i(k+1|k)) \tag{10a}$$

$$\hat{z}_i(k+1|k) = S_i\hat{z}_i(k|k) + Z_i^\top Bu(k) \tag{10b}$$

$$K_i(k+1) = P_i(k+1|k)t_i^\top \left(t_i P_i(k+1|k)t_i^\top + R_i\right)^{-1} \tag{10c}$$

$$P_i(k+1|k) = S_i P_i(k|k)S_i^\top + Z_i^\top Q Z_i \tag{10d}$$

$$P_i(k+1|k+1) = (I - K_i(k+1)t_i)P_i(k+1|k) \tag{10e}$$

with initial value of

$$\hat{z}_i(0|0) = Z_i^\top \bar{x}_0, \quad P_i(0|0) = Z_i^\top P_0 Z_i.$$

The above Equations (10) describe the recursive form of how the state estimate $\hat{z}_i$, the Kalman gain K_i, and the error covariance matrix P_i evolve. The error covariance P_i of the i-th local observer defined in (8), is governed by Equations (10d) and (10e), which ensure that the covariance matrix $P_i(k|k)$ can be calculated by the following recursive form:

$$\mathcal{L}_i: \; P_i(k+1|k+1) = (I - K_i(k+1)t_i)(S_i P_i(k|k)S_i^\top + Z_i^\top Q Z_i) \tag{11}$$

with the initial value of

$$P_i(0|0) = Z_i^\top P_0 Z_i.$$

Similarly, the error cross covariance P_{ij} of the i-th and j-th local observers can be defined by

$$\begin{aligned} P_{ij}(k|k-1) &= \mathbf{E}[(\hat{z}_i(k|k-1) - z_i(k))(\hat{z}_j(k|k-1) - z_j(k))^\top], \\ P_{ij}(k|k) &= \mathbf{E}[(\hat{z}_i(k|k) - z_i(k))(\hat{z}_j(k|k) - z_j(k))^\top], \end{aligned} \tag{12}$$

and the recursive formula for P_{ij} is derived here. To this end, define the estimation error

$$\begin{aligned} \tilde{z}_i(k+1|k) &:= \hat{z}_i(k+1|k) - z_i(k+1) \\ \tilde{z}_i(k+1|k+1) &:= \hat{z}_i(k+1|k+1) - z_i(k+1), \end{aligned} \tag{13}$$

and we have that

$$\begin{aligned} \tilde{z}_i(k+1|k) &= \left(S_i\hat{z}_i(k|k) + Z_i^\top Bu(k)\right) - \left(S_i z_i(k) + Z_i^\top Bu(k) + Z_i^\top d(k)\right) \\ &= S_i\tilde{z}_i(k|k) - Z_i^\top d(k) \end{aligned} \tag{14a}$$

$$\begin{aligned} \tilde{z}_i(k+1|k+1) &= \left(\hat{z}_i(k+1|k) + K_i(k+1)(y_i^a(k+1) - t_i\hat{z}_i(k+1|k))\right) - z_i(k+1) \\ &= (I - K_i(k+1)t_i)\tilde{z}_i(k+1|k) + K_i(k+1)n_i^a(k+1). \end{aligned} \tag{14b}$$

By substituting (14a) into (14b), the dynamics of the error $\tilde{z}_i(k|k)$ is obtained as

$$\begin{aligned} \mathcal{F}_i: \; \tilde{z}_i(k+1|k+1) = (I - K_i(k+1)t_i)S_i\tilde{z}_i(k|k) - (I - K_i(k+1)t_i)Z_i^\top d(k) \\ + K_i(k+1)n_i^a(k+1). \end{aligned} \tag{15}$$

The errors $\tilde{z}_i(k|k)$ and $\tilde{z}_j(k|k)$ for $i \neq j$ may be correlated; thus, by using (15), the error cross covariance between $\tilde{z}_i(k|k)$ and $\tilde{z}_j(k|k)$ can be computed recursively. From the recursive form of (15), note that $\tilde{z}_i(k|k)$ is a linear combination of elements in

$$\{\tilde{z}_i(0|0), d(0), \cdots, d(k-1), n_i^a(0), \cdots, n_i^a(k)\}. \tag{16}$$

Therefore, by Assumption 1, we have (i) $n_i^a(k+1)$ and $d(k)$ are orthogonal, (ii) $\tilde{z}_i(k|k)$ and $d(k)$ are orthogonal, and (iii) $\tilde{z}_i(k|k)$ and $n_j^a(k+1)$ are orthogonal. Using these facts, one

can derive the recursive form of the error cross covariance between $\tilde{z}_i(k|k)$ and $\tilde{z}_j(k|k)$ as follows:

$$\begin{aligned}\mathcal{L}_{ij}: \; P_{ij}(k+1|k+1) &= \mathbf{E}[\tilde{z}_i(k+1|k+1)\tilde{z}_j^\top(k+1|k+1)] \\ &= (I - K_i(k+1)t_i)\Big(S_i\mathbf{E}[\tilde{z}_i(k|k)\tilde{z}_j^\top(k|k)]S_j^\top + Z_i^\top Q Z_j\Big)(I - K_j(k+1)t_j)^\top \\ &\quad + K_i(k+1)\mathbf{E}[n_i^a(k+1){n_j^a}^\top(k+1)]K_j^\top(k+1) \\ &= (I - K_i(k+1)t_i)\Big(S_iP_{ij}(k|k)S_j^\top + Z_i^\top Q Z_j\Big)(I - K_j(k+1)t_j)^\top \\ &\quad + K_i(k+1)R_{ij}K_j^\top(k+1),\end{aligned} \tag{17}$$

with the initial value of

$$P_{ij}(0|0) = Z_i^\top P_0 Z_j.$$

3.3. Optimal Information Fusion Based on Observability Decomposition

Based on the equivalence $Z_i^\top x = z_i$ in (5) and the definition $\tilde{z}_i = \hat{z}_i - z_i$ in (13), we have

$$\hat{z}_i = z_i + \tilde{z}_i = Z_i^\top x + \tilde{z}_i. \tag{18}$$

Stacking Equations (18) for all $i \in [p]$ leads to the following equation of

$$\begin{bmatrix}\hat{z}_1(k|k)\\ \vdots \\ \hat{z}_p(k|k)\end{bmatrix} = \begin{bmatrix}z_1(k)\\ \vdots \\ z_p(k)\end{bmatrix} + \begin{bmatrix}\tilde{z}_1(k|k)\\ \vdots \\ \tilde{z}_p(k|k)\end{bmatrix} = \begin{bmatrix}Z_1{}^\top\\ \vdots \\ Z_p{}^\top\end{bmatrix} x(k) + \begin{bmatrix}\tilde{z}_1(k|k)\\ \vdots \\ \tilde{z}_p(k|k)\end{bmatrix}. \tag{19}$$

Finally, (19) is written in a compact form as

$$\hat{z}(k|k) = \Phi x(k) + \tilde{z}(k|k) = \Phi x(k) + v^a(k) \in \mathbb{R}^\mu, \tag{20}$$

where the matrix

$$\Phi := \begin{bmatrix}Z_1{}^\top\\ \vdots \\ Z_p{}^\top\end{bmatrix} \in \mathbb{R}^{\mu \times n} \tag{21}$$

is composed of the similarity transformation matrices Z_i's and $v^a(k)$ is used for a simple notation of $\tilde{z}(k|k)$. In Equation (20),

$$\mu := \sum_{i=1}^{p} \mu_i$$

denotes the size of the stacked vector.

It should be noted that all the information in (20) except the actual state $x(k)$, are known or accessible to us. In Section 3.1, the matrix Φ is generated from the orthonormal basis of the observable subspace $\mathcal{R}(G_i^\top)$ where G_i is the observability matrix given by (4). In Section 3.2, each local observer $\mathcal{O}_i$ in (9) provides the state estimate $\hat{z}_i$ for the observable sub-state z_i. Now, the stochastic properties of the last term

$$v^a(k) = \tilde{z}(k|k) = \begin{bmatrix}\tilde{z}_1(k|k)\\ \tilde{z}_2(k|k)\\ \vdots \\ \tilde{z}_p(k|k)\end{bmatrix}$$

are analyzed. First, its mean is zero because $\tilde{z}_i(k|k)$ is a linear combination of elements in (16) by the Formula (15), and Assumption 1 ensures that every component in (16) has a

zero mean. Second, the covariance matrix of $v^a(k)$ can be obtained since the error covariance matrix P_{i} is computed by each local observer $\mathcal{L}_{\mathsf{i}}$ in (11), and the error cross covariance matrix P_{ij} is generated by the second layer of the multi-sensor Kalman filter $\mathcal{L}_{\mathsf{ij}}$ in (17) with collected information from local observers (see Figure 1 for the structure of the proposed Kalman filter). In summary, we have

$$v^a(k) \sim N\big(0_{\mu\times 1}, P(k|k)\big), \tag{22}$$

where

$$P(k|k) = \begin{bmatrix} P_1(k|k) & P_{12}(k|k) & \cdots & P_{1\mathsf{p}}(k|k) \\ P_{21}(k|k) & P_2(k|k) & \cdots & P_{2\mathsf{p}}(k|k) \\ \vdots & \vdots & \ddots & \vdots \\ P_{\mathsf{p}1}(k|k) & P_{\mathsf{p}2}(k|k) & \cdots & P_{\mathsf{p}}(k|k) \end{bmatrix}, \tag{23}$$

which can be recursively computed by (11) and (17). Finally, Equation (20) depicts a linear model with the measured data vector $\hat{z}$, the known matrix Φ, the noise vector v^a with a zero-mean Gaussian distribution, and the unknown vector x to be estimated.

Based on the statistical estimation and detection theory [30,31], an elaborate derivation process to recover the optimal estimate of x in (20), is now presented. The minimum variance unbiased estimator (MVUE) for the data model (20) with v^a satisfying $v^a \sim N(0_{\mu\times 1}, P)$ is introduced as follows.

Theorem 1 ([30], Theorem 4.2)**.** *For the measurement $\hat{z} = \Phi x + v^a \in \mathbb{R}^{\mu}$ with $x \in \mathbb{R}^{\mathsf{n}}$ and $v^a \in \mathbb{R}^{\mu}$ such that $v^a \sim N(0_{\mu\times 1}, P)$ for some $P > 0$, the minimum variance unbiased estimator (MVUE) of x is*

$$\mathcal{D}: \ \hat{x}_{\text{MVUE}} = \left(\Phi^\top P^{-1}\Phi\right)^{-1}\Phi^\top P^{-1}\hat{z} \tag{24}$$

and the corresponding covariance matrix of $\hat{x}_{\text{MVUE}}$ is

$$P_{\hat{x}_{\text{MVUE}}} = \left(\Phi^\top P^{-1}\Phi\right)^{-1}, \tag{25}$$

which achieves the minimum covariance in the sense that $P_{\hat{x}_{\text{MVUE}}} \leq P_{\hat{x}}$ for any type of estimator $\hat{x}$.

Proof. The results directly follows from the Gauss–Markov Theorem ([30], Theorem 6.1). However, we provide a direct proof for the readers convenience, and it follows the procedure in the proof of ([24], Theorem 1) or ([25], Theorem 1). We introduce a linear unbiased estimator

$$\hat{x} = \Omega\hat{z}$$

and, from the unbiased assumption, it follows that

$$\mathbf{E}[\hat{x}] = \mathbf{E}[\Omega\hat{z}] = \Omega\mathbf{E}[\Phi x + v^a] = \Omega\Phi\mathbf{E}[x] = \mathbf{E}[x].$$

Thus, we have

$$\Omega\Phi = I_{\mathsf{n}\times\mathsf{n}}. \tag{26}$$

Let the covariance matrix of the estimation error $\tilde{x} := \hat{x} - x$ be P_x. Then, the estimation error $\tilde{x}$ is obtained that

$$\tilde{x} = \hat{x} - x = \Omega\hat{z} - x = \Omega\hat{z} - \Omega\Phi x = \Omega(\hat{z} - \Phi x) = \Omega v^a,$$

and the covariance matrix P_x can be computed as

$$P_x = \mathbf{E}\left[\tilde{x}\tilde{x}^\top\right] = \mathbf{E}\left[\Omega v^a {v^a}^\top \Omega^\top\right] = \Omega\mathbf{E}\left[v^a {v^a}^\top\right]\Omega^\top = \Omega P\Omega^\top.$$

In order to find the minimum variance estimator, set the trace of the covariance matrix P_x as the performance index

$$J := \mathrm{tr}(P_x) = \mathrm{tr}\left(\Omega P \Omega^\top\right).$$

The Lagrangian [32] associated with J becomes

$$L = J + 2\mathrm{tr}(\Lambda(\Omega\Phi - I_{\mathsf{n}\times\mathsf{n}}))$$

where $\Lambda \in \mathbb{R}^{\mathsf{n}\times\mathsf{n}}$ is a matrix representing the Lagrange multipliers. By solving

$$\frac{\partial L}{\partial \Omega} = O_{\mathsf{n}\times\mu},$$

we have

$$\Omega P + \Lambda^\top \Phi^\top = O_{\mathsf{n}\times\mu}. \tag{27}$$

Combining (26) and (27) results in the following equation of

$$\begin{bmatrix}\Omega & \Lambda^\top\end{bmatrix}\begin{bmatrix}P & \Phi \\ \Phi^\top & O_{\mathsf{n}\times\mathsf{n}}\end{bmatrix} = \begin{bmatrix}O_{\mathsf{n}\times\mu} & I_{\mathsf{n}\times\mathsf{n}}\end{bmatrix}.$$

Therefore, the matrix inversion lemma ([33], Section 2.3) yields the solution as

$$\begin{bmatrix}\Omega & \Lambda^\top\end{bmatrix} = \begin{bmatrix}O_{\mathsf{n}\times\mu} & I_{\mathsf{n}\times\mathsf{n}}\end{bmatrix}\begin{bmatrix}P & \Phi \\ \Phi^\top & O_{\mathsf{n}\times\mathsf{n}}\end{bmatrix}^{-1} = \begin{bmatrix}\left(\Phi^\top P^{-1}\Phi\right)^{-1}\Phi^\top P^{-1} & -\left(\Phi^\top P^{-1}\Phi\right)^{-1}\end{bmatrix}.$$

Thus, we have $\Omega = \left(\Phi^\top P^{-1}\Phi\right)^{-1}\Phi^\top P^{-1}$. Finally, the MVUE of x in (24), is obtained from $\hat{x}_{\mathrm{MVUE}} = \Omega\hat{z} = \left(\Phi^\top P^{-1}\Phi\right)^{-1}\Phi^\top P^{-1}\hat{z}$, and the corresponding covariance matrix in (25) is computed by $P_{\hat{x}_{\mathrm{MVUE}}} = \Omega P \Omega^\top = \left(\Phi^\top P^{-1}\Phi\right)^{-1}$. □

Theorem 1 explains how the optimal estimate is computed. The information fusion center $\mathcal{D}$ calculates the MVUE by (24) and its covariance by (25). In summary, the whole structure of the decentralized multi-sensor information fusion Kalman filter is shown in Figure 1. The first layer is composed of the local observer $\mathcal{O}_{\mathsf{i}}$, which generates the estimate $\hat{z}_{\mathsf{i}}$ and the Kalman gains K_{i} as given in (9) and (10). A part of the local observer $\mathcal{O}_{\mathsf{i}}$, denoted as $\mathcal{L}_{\mathsf{i}}$, provides the error covariance matrix P_{i}. The second layer $\mathcal{L}_{\mathsf{ij}}$ collects the Kalman gain K_{i}'s from the first layer and gives the error cross covariance matrix P_{ij} by (17). Finally, the third layer operates as an optimal information fusion center $\mathcal{D}$ as described in Theorem 1 and computes the optimal estimate with the minimum covariance.

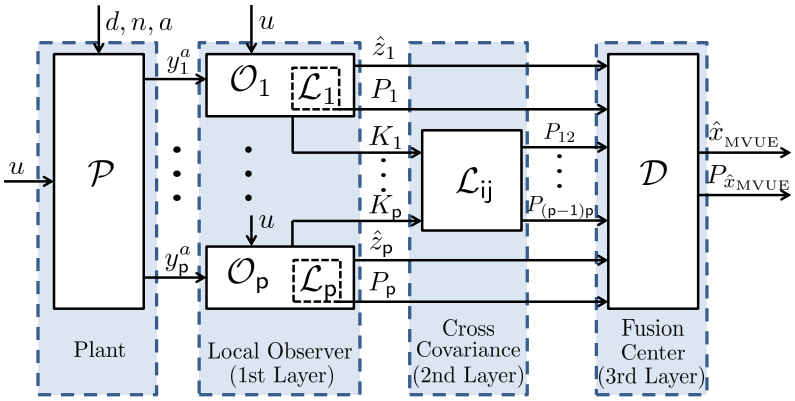


Figure 1. Structure of decentralized multi-sensor information fusion Kalman filter.

Remark 1. *Note that Gauss–Markov Theorem ([30], Theorem 6.1) gives the best linear unbiased estimator (BLUE) for the measurement* $\hat{z} = \Phi x + v^a$ *where* v^a *is a random variable, whose probability density function (PDF) is not restricted to a Gaussian distribution, with a zero mean and covariance P. Since the BLUE is also the MVUE for Gaussian data, the results of Theorem 1 also follow directly from the Gauss–Markov Theorem. The state estimate* $\hat{x}_{\mathrm{MVUE}}$ *given in Theorem 1 is the optimal estimate since it achieves the minimum variance with an unbiased mean. A special case of Theorem 1 is considered in ([24], Theorem 1) and ([25], Theorem 1) for an information fusion scheme; however, the scheme in [24,25] may not be successful for a system whose local systems with a single sensor are not observable because the covariance matrix P could diverge in that case, whereas the covariance matrix P does not diverge in our scheme due to the Kalman observability decomposition.*

4. Attack Resilient and Secure State Estimation by Decentralized Kalman Filter

4.1. Effect of Sparse Sensor Attack on Information Fusion Kalman FIlter

In the previous section, we assumed that all sensors were attack-free, that is, $a(k) \equiv 0$. Hence, $n_{\mathrm{i}}^a(k)$ and $y_{\mathrm{i}}^a(k)$ in (3) and (7) were regarded as non-attacked noise $n_{\mathrm{i}}(k)$ and output $y_{\mathrm{i}}(k)$, respectively. The effects of a sparse sensor attack satisfying Assumption 2 on the information fusion Kalman filter developed in Section 3 are investigated in this subsection.

By linearity, the Kalman filter in (10) can be divided into two parts with $\hat{z}_{\mathrm{i}} =: g_{\mathrm{i}} + e_{\mathrm{i}}$ as in

$$g_{\mathrm{i}}(k+1|k+1) := g_{\mathrm{i}}(k+1|k) + K_{\mathrm{i}}(k+1)(y_{\mathrm{i}}(k+1) - t_{\mathrm{i}} g_{\mathrm{i}}(k+1|k)), \tag{28a}$$
$$e_{\mathrm{i}}(k+1|k+1) := e_{\mathrm{i}}(k+1|k) + K_{\mathrm{i}}(k+1)(a_{\mathrm{i}}(k+1) - t_{\mathrm{i}} e_{\mathrm{i}}(k+1|k)), \tag{28b}$$
$$g_{\mathrm{i}}(k+1|k) := S_{\mathrm{i}} g_{\mathrm{i}}(k|k) + Z_{\mathrm{i}}^\top B u(k), \tag{28c}$$
$$e_{\mathrm{i}}(k+1|k) := S_{\mathrm{i}} e_{\mathrm{i}}(k|k). \tag{28d}$$

Note that $g_{\mathrm{i}}(k+1|k+1)$ and $e_{\mathrm{i}}(k+1|k+1)$ have the same dynamics with (10a), while the incoming signal $y_{\mathrm{i}}^a(k+1)$ is divided into two parts with $y_{\mathrm{i}}(k+1)$ and $a_{\mathrm{i}}(k+1)$ assigned to the dynamics of $g_{\mathrm{i}}(k+1|k+1)$ and $e_{\mathrm{i}}(k+1|k+1)$, respectively. Similarly, $g_{\mathrm{i}}(k+1|k)$ and $e_{\mathrm{i}}(k+1|k)$ have the same dynamics with (10b), whereas the incoming signal $u(k)$ is solely assigned to the dynamics of $g_{\mathrm{i}}(k+1|k)$. By setting the initial conditions as

$$g_{\mathrm{i}}(0|0) = \hat{z}_{\mathrm{i}}(0|0) = Z_{\mathrm{i}}^\top \bar{x}_0 \quad \text{and} \quad e_{\mathrm{i}}(0|0) = 0_{\mu_{\mathrm{i}} \times 1},$$

it easily follows from (10a) and (10b) that

$$\begin{aligned} \hat{z}_{\mathrm{i}}(k+1|k+1) &= g_{\mathrm{i}}(k+1|k+1) + e_{\mathrm{i}}(k+1|k+1), \\ \hat{z}_{\mathrm{i}}(k+1|k) &= g_{\mathrm{i}}(k+1|k) + e_{\mathrm{i}}(k+1|k). \end{aligned} \tag{29}$$

Finally, the local observer $\mathcal{O}_{\mathrm{i}}$ in (9) is divided into $\mathcal{O}_{\mathrm{i}}^y$ and $\mathcal{O}_{\mathrm{i}}^a$, as follows:

$$\mathcal{O}_{\mathrm{i}}^y : \; g_{\mathrm{i}}(k+1|k+1) = (I - K_{\mathrm{i}}(k+1)t_{\mathrm{i}})\Big(S_{\mathrm{i}} g_{\mathrm{i}}(k|k) + Z_{\mathrm{i}}^\top B u(k)\Big) + K_{\mathrm{i}}(k+1) y_{\mathrm{i}}(k+1), \tag{30a}$$
$$\mathcal{O}_{\mathrm{i}}^a : \; e_{\mathrm{i}}(k+1|k+1) = (I - K_{\mathrm{i}}(k+1)t_{\mathrm{i}}) S_{\mathrm{i}} e_{\mathrm{i}}(k|k) + K_{\mathrm{i}}(k+1) a_{\mathrm{i}}(k+1). \tag{30b}$$

Now, define the attack-free estimation error

$$\begin{aligned} v_{\mathrm{i}}(k+1|k+1) &:= g_{\mathrm{i}}(k+1|k+1) - z_{\mathrm{i}}(k+1), \\ v_{\mathrm{i}}(k+1|k) &:= g_{\mathrm{i}}(k+1|k) - z_{\mathrm{i}}(k+1), \end{aligned} \tag{31}$$

and we have that

$$\begin{aligned} v_i(k+1|k) &= \left(S_i g_i(k|k) + Z_i^\top B u(k)\right) - \left(S_i z_i(k) + Z_i^\top B u(k) + Z_i^\top d(k)\right) \\ &= S_i v_i(k|k) - Z_i^\top d(k) \qquad (32a) \\ v_i(k+1|k+1) &= \left(g_i(k+1|k) + K_i(k+1)(y_i(k+1) - t_i g_i(k+1|k))\right) - z_i(k+1) \\ &= (I - K_i(k+1)t_i)v_i(k+1|k) + K_i(k+1)n_i(k+1) \qquad (32b) \\ &= (I - K_i(k+1)t_i)S_i v_i(k|k) - (I - K_i(k+1)t_i)Z_i^\top d(k) \\ &\quad + K_i(k+1)n_i(k+1), \qquad (32c) \end{aligned}$$

which is the same as (14) and (15) with n_i^a replaced by n_i. By (29) and (31), the total state-estimation error defined in (13) satisfies

$$\tilde{z}_i(k+1|k+1) = v_i(k+1|k+1) + e_i(k+1|k+1), \qquad (33)$$

and, from (30b) and (32c), its dynamic equation is given as follows:

$$\begin{aligned} \mathcal{F}_i : \tilde{z}_i(k+1|k+1) &= (I - K_i(k+1)t_i)S_i\tilde{z}_i(k|k) - (I - K_i(k+1)t_i)Z_i^\top d(k) \\ &\quad + K_i(k+1)n_i(k+1) + K_i(k+1)a_i(k+1), \end{aligned} \qquad (34)$$

which is a rewrite of (15) using the fact $n_i^a = n_i + a_i$.

For notational simplicity, $\hat{z}_i(k|k)$, $v_i(k|k)$, and $e_i(k|k)$ are denoted by $\hat{z}_i(k)$, $v_i(k)$, and $e_i(k)$, respectively. Then, Equation (19) becomes

$$\begin{bmatrix} \hat{z}_1(k) \\ \vdots \\ \hat{z}_p(k) \end{bmatrix} = \begin{bmatrix} Z_1{}^\top \\ \vdots \\ Z_p{}^\top \end{bmatrix} x(k) + \begin{bmatrix} v_1(k) \\ \vdots \\ v_p(k) \end{bmatrix} + \begin{bmatrix} e_1(k) \\ \vdots \\ e_p(k) \end{bmatrix}, \qquad (35)$$

which can be written in a compact form as

$$\hat{z}(k) = \Phi x(k) + v(k) + e(k) \in \mathbb{R}^\mu. \qquad (36)$$

The above Equation (36) is nothing but (20) with v^a replaced by $v + e$. The properties of v are exactly identical with those of v^a in (22) because the derivation in (22) is under the assumption of $a \equiv 0$ meaning $e \equiv 0$ in this case. Thus, we have

$$v(k) \sim N(0_{\mu\times 1}, P(k)), \qquad (37)$$

where $P(k)$ simply denotes $P(k|k)$ in (23). The attack-induced signal $e(k) = [e_1^\top(k), \cdots, e_p^\top(k)]^\top$ evolves according to Equation (30b) (or equivalently (28b) and (28d)) with an initial value of $e_i(0) = e_i(0|0) = 0_{\mu_i\times 1}$. Therefore, we have $e_i \equiv 0_{\mu_i\times 1}$ for the healthy sensor with $a_i \equiv 0$, while $e_i \not\equiv 0_{\mu_i\times 1}$ generally holds for the attacked sensor with $a_i \not\equiv 0$. Finally, the stacked error vector $e \in \mathbb{R}^\mu$ partitioned by the sequence $\{\mu_i\}$, is ($\{\mu_i\}$-stacked) q-sparse by Assumption 2.

4.2. Detection of Sparse Sensor Attack

In the previous subsection, the measurement data have the form $\hat{z} = \Phi x + v + e \in \mathbb{R}^\mu$ with unknown signals x, v, and e where the noise-induced signal v can be considered as a random variable whose distribution is $N(0_{\mu\times 1}, P)$ and the attack-induced signal e is ($\{\mu_i\}$-stacked) q-sparse. To investigate the properties of the matrix Φ in the measurement data, we borrow the definition of ($\{\mu_i\}$-stacked) q-error detectability and its characterization from [15]. There is a slight modification in the following Definition 1 and Lemma 1 from [15].

They do not append any additional zeros, whereas [15] adds additional zeros to match the size of all partitioned vectors and matrices.

Definition 1 ([15], Definition 1). *[-15]For a finite sequence* $\{\mu_i\} = \{\mu_1, \mu_2, \cdots, \mu_p\}$ *with* $\mu = \sum_{i=1}^{p} \mu_i$*, a coding matrix* $\Phi \in \mathbb{R}^{\mu \times n}$ *is said to be* ($\{\mu_i\}$-stacked) q-error detectable *if, for all* $x, x' \in \mathbb{R}^n$ *and* ($\{\mu_i\}$*-stacked)* q*-sparse* $e \in \mathbb{R}^{\mu}$ *such that* $\Phi x + e = \Phi x'$*, it holds that* $x = x'$*.*

Accordingly, the matrix $\Phi \in \mathbb{R}^{\mu \times n}$ is not ($\{\mu_i\}$-stacked) q-error detectable if and only if there exist $x, x' \in \mathbb{R}^n$ satisfying $x \neq x'$, and ($\{\mu_i\}$-stacked) q-sparse $e \in \mathbb{R}^{\mu}$ such that $\Phi x + e = \Phi x'$. In other words, the matrix $\Phi \in \mathbb{R}^{\mu \times n}$ is ($\{\mu_i\}$-stacked) q-error undetectable if and only if there exist a non-zero $x_e \in \mathbb{R}^n$ and ($\{\mu_i\}$-stacked) q-sparse $e \in \mathbb{R}^{\mu}$ such that $\Phi x_e = e$. Typically, in terms of vectors, the vector $e \in \mathbb{R}^{\mu}$ is said to be undetectable with respect to $\Phi \in \mathbb{R}^{\mu \times n}$ if $e = \Phi x_e \in \mathbb{R}^{\mu}$ for some $x_e \in \mathbb{R}^n$.

Lemma 1 ([15], Proposition 1). *For a finite sequence* $\{\mu_i\} = \{\mu_1, \mu_2, \cdots, \mu_p\}$ *with* $\mu = \sum_{i=1}^{p} \mu_i$ *and a matrix* $\Phi \in \mathbb{R}^{\mu \times n}$*, the followings are equivalent:*
(i) The matrix $\Phi \in \mathbb{R}^{\mu \times n}$ *is* ($\{\mu_i\}$*-stacked)* q*-error detectable.*
(ii) For every set $\mathcal{J} \subset [p]$ *satisfying* $|\mathcal{J}| \geq p - q$*,* $\Phi_{\mathcal{I}_{\mathcal{J}}^{\{\mu_i\}}}$ *has full column rank.*
(iii) For any $x \in \mathbb{R}^n$ *where* $x \neq 0_{n \times 1}$*, the vector* $\Phi x \in \mathbb{R}^{\mu}$ *is not* ($\{\mu_i\}$*-stacked)* q*-sparse.*

With the estimate $\hat{x}$ of x obtained by MVUE of (24) in Theorem 1, we can calculate the estimated output $\Phi\hat{x}$ and generate a residual signal r, which is a difference between the real measurement and the estimated output, that is, $r := \hat{z} - \Phi\hat{x}$. Then, the residual r becomes another random variable whose distribution is also Gaussian. Finally, the mean and covariance of the Gaussian distributed random variable r is computed in the following theorem.

Theorem 2. *For the measurement* $\hat{z} = \Phi x + v + e \in \mathbb{R}^{\mu}$ *where* $\Phi \in \mathbb{R}^{\mu \times n}$ *has full column rank and* v *satisfies* $v \sim N(0_{\mu \times 1}, P)$ *with* $P > 0$*, let* $\hat{x} = \Psi\hat{z} = \left(\Phi^\top P^{-1}\Phi\right)^{-1}\Phi^\top P^{-1}\hat{z}$ *and*

$$r := \hat{z} - \Phi\hat{x} = (I_{\mu\times\mu} - \Phi\Psi)\hat{z} = (I_{\mu\times\mu} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})\hat{z}, \tag{38}$$

where $\Psi := (\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1}$*. Then, the residual* r *is Gaussian distributed with mean* $(I_{\mu\times\mu} - \Phi\Psi)e$ *and covariance* $(I_{\mu\times\mu} - \Phi\Psi)P$*,*

$$r \sim N\Big((I_{\mu\times\mu} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})e, P - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top\Big). \tag{39}$$

Furthermore, $e = \Phi x_e \in \mathbb{R}^{\mu}$ *for some* $x_e \in \mathbb{R}^n$ *if and only if the mean of* r*,* $\mathbf{E}[r] (= (I_{\mu\times\mu} - \Phi\Psi)e)$*, satisfies* $\mathbf{E}[r] = 0_{\mu\times 1}$*. In other words,* e *is undetectable with respect to* Φ *if and only if* $\mathbf{E}[r] = 0_{\mu\times 1}$*.*

Proof. First, the mean of r is computed as follows.

$$\begin{aligned}
\mathbf{E}[r] &= \mathbf{E}[(I_{\mu\times\mu} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})\hat{z}] \\
&= (I_{\mu\times\mu} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})\mathbf{E}[\Phi x + v + e] \\
&= (I_{\mu\times\mu} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})(\Phi x + e) \\
&= (I_{\mu\times\mu} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})e = (I_{\mu\times\mu} - \Phi\Psi)e
\end{aligned} \tag{40}$$

Second, because it easily follows that

$$\begin{aligned}
r - \mathbf{E}[r] &= (I_{\mu\times\mu} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})(\hat{z} - e) \\
&= (I_{\mu\times\mu} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})(\Phi x + v) \\
&= (I_{\mu\times\mu} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})v = (I_{\mu\times\mu} - \Phi\Psi)v,
\end{aligned}$$

the covariance matrix is calculated as

$$\begin{aligned}&\mathbf{E}[(r-\mathbf{E}[r])(r-\mathbf{E}[r])^\top] = \mathbf{E}[(I_{\mu\times\mu}-\Phi\Psi)vv^\top(I_{\mu\times\mu}-\Phi\Psi)^\top]\\ &= (I_{\mu\times\mu}-\Phi\Psi)\mathbf{E}[vv^\top](I_{\mu\times\mu}-\Phi\Psi)^\top = (I_{\mu\times\mu}-\Phi\Psi)P(I_{\mu\times\mu}-\Phi\Psi)^\top\\ &= (I_{\mu\times\mu}-\Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})P(I_{\mu\times\mu}-\Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})^\top\\ &= P-\Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top = (I_{\mu\times\mu}-\Phi\Psi)P.\end{aligned}$$

Moreover, note that

$$\mathbf{E}[r] = (I_{\mu\times\mu}-\Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})\mathbf{E}[\hat{z}]$$

because of (40), and

$$\mathbf{E}[\hat{z}] = \mathbf{E}[\Phi x+v+e] = \Phi x+e.$$

Since $\Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1}$ is a projection matrix and it projects $\mathbf{E}[\hat{z}]$ onto the range space of Φ, $\mathcal{R}(\Phi)$, we have $\mathbf{E}[\hat{z}] = \Phi x+e \notin \mathcal{R}(\Phi)$ if and only if $\mathbf{E}[\hat{z}] \neq \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1}\mathbf{E}[\hat{z}]$. This implies that $e \notin \mathcal{R}(\Phi)$ if and only if $(I_{\mu\times\mu}-\Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})\mathbf{E}[\hat{z}] \neq 0_{\mu\times 1}$. This completes the proof. □

Theorem 2 clarifies the mean and covariance of the Gaussian random variable r, and further, characterization of undetectable attacks with statistical analysis is also given. Now, one can derive a detection criterion of ($\{\mu_i\}$-stacked) q-sparse errors based on the property of the residual signal r, assuming that $\Phi \in \mathbb{R}^{\mu\times n}$ is ($\{\mu_i\}$-stacked) q-error detectable and that $e \in \mathbb{R}^\mu$ is actually ($\{\mu_i\}$-stacked) q-sparse. This detection strategy is summarized in the following theorem.

Theorem 3. *For a finite sequence $\{\mu_i\} = \{\mu_1, \mu_2, \cdots, \mu_p\}$ with $\mu = \sum_{i=1}^{p}\mu_i$ and the measurement $\hat{z} = \Phi x+v+e \in \mathbb{R}^\mu$ where $\Phi \in \mathbb{R}^{\mu\times n}$ is ($\{\mu_i\}$-stacked) q-error detectable, $e \in \mathbb{R}^\mu$ is ($\{\mu_i\}$-stacked) q-sparse, and $v \in \mathbb{R}^\mu$ satisfies $v \sim N(0_{\mu\times 1}, P)$ with $P > 0$, let*

$$r = \hat{z}-\Phi\hat{x} = \hat{z}-\Phi\Psi\hat{z} = (I_{\mu\times\mu}-\Phi\Psi)\hat{z} = (I_{\mu\times\mu}-\Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})\hat{z}$$

be given. Then, $e = 0_{\mu\times 1}$ if and only if $\mathbf{E}[r] = 0_{\mu\times 1}$. Moreover, when $e = 0_{\mu\times 1}$, the vector x is exactly recovered by the expectation value of $\hat{x} = \Psi\hat{z} = (\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1}\hat{z}$, that is, $x = \mathbf{E}[\hat{x}]$, which means that $\hat{x}$ is an unbiased estimate of x.

Proof. From Theorem 2, the ($\{\mu_i\}$-stacked) q-sparse e satisfies $e = \Phi x_e \in \mathbb{R}^\mu$ for some $x_e \in \mathbb{R}^n$ if and only if $\mathbf{E}[r] = 0_{\mu\times 1}$. However, any non-zero $e = \Phi x_e \in \mathbb{R}^\mu$ for some $x_e \in \mathbb{R}^n$ is not ($\{\mu_i\}$-stacked) q-sparse by Lemma 1. (iii) since $\Phi \in \mathbb{R}^{\mu\times n}$ is ($\{\mu_i\}$-stacked) q-error detectable. Therefore, the ($\{\mu_i\}$-stacked) q-sparse $e = \Phi x_e \in \mathbb{R}^\mu$ should be zero, and the result directly follows. Furthermore, the property of an unbiased estimate (with minimum variance) is easily obtained from Theorem 1. □

From the observation of Theorems 2 and 3, the problem of detecting a non-zero ($\{\mu_i\}$-stacked) q-sparse error signal e with a ($\{\mu_i\}$-stacked) q-error detectable coding matrix $\Phi \in \mathbb{R}^{\mu\times n}$ can be rephrased as: Given the residual signal r, which comes from the Gaussian distribution $N(\mathbf{E}[r], P-\Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top)$, determine if $\mathbf{E}[r] = 0_{\mu\times 1}$ or $\mathbf{E}[r] \neq 0_{\mu\times 1}$. Therefore, the statistical decision theory [31] is helpful in this situation. More precisely, the χ^2 test for fault detection [22,23], which is widely used to detect unwanted error signals, such as faults or attacks, can be applied.

One can simply apply the χ^2 test to detect the presence of error signals in the ($\{\mu_i\}$-stacked) measurement $\hat{z}$ given by (36), and its operating scheme is summarized in Algorithm 1. Initially, the attack detection alarm indicator f is set to 0, and then the residual r is computed according to Equation (38). Without any error signal (that is, $e = 0_{\mu\times 1}$), the residual r follows a Gaussian distribution $N(0, P-\Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top)$, which is shown

in (39). Now, define the standardized residual $\zeta := \left(P - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top)\right)^{-\frac{1}{2}} r$ whose distribution becomes $N(0_{\mu\times 1}, I_{\mu\times\mu})$. Thus, the 2-norm of ζ denoted by $g := \zeta^\top\zeta$ is an observation from a random variable $\mathbf{g}$, which satisfies a χ^2 distribution with μ degrees of freedom (DOF),

$$\mathbf{g} \sim \chi^2_\mu.$$

This means that g cannot be far away from zero. Finally, when g is greater than a threshold Δ_{TH}, the attack detection alarm is triggered by setting $f = 1$. Here, Δ_{TH} is the predetermined threshold value, and it decides the probability of false alarm and the probability of error detection. For example, when the threshold Δ_{TH} is chosen such that

$$\int_0^{\Delta_{TH}} p_{\mathbf{g}}(x)dx = 1 - \delta, \tag{41}$$

where $p_{\mathbf{g}}(x)$ denotes the PDF of the χ^2_μ distribution, the probability of false alarm becomes δ. As the probability of false alarm δ becomes smaller, the probability of error detection also decreases, which implies that there is a trade-off between the small false alarm and the high error detection ratio. Thus, one needs to choose Δ_{TH} as a good compromise between these two conflicting requirements.

Algorithm 1 Detection scheme based on the χ^2 test

Input: $\hat{z}$
Output: f
Initialization: $f = 0$
$\hat{x}_{\text{MVUE}} = (\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1}\hat{z}$
$r = \hat{z} - \Phi\hat{x}_{\text{MVUE}}$
$\zeta = \left(P - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top)\right)^{-\frac{1}{2}} r$
$g = \zeta^\top\zeta$
if $g \le \Delta_{TH}$ **then**
  $f = 0$
else if $g > \Delta_{TH}$ **then**
  $f = 1$
end if

4.3. Secure State Estimation under a Sparse Sensor Attack

In this subsection, an attack-resilient and secure state estimation scheme, which reconstructs the optimal estimate for the state x under Assumptions 1–3, is developed. First, characterization of the matrix Φ defined in (21) under Assumption 3 is given as follows.

Lemma 2 ([15], Proposition 1,2,3,6)**.** *For a finite sequence* $\{\mu_\mathsf{i}\} = \{\mu_1, \mu_2, \cdots, \mu_\mathsf{p}\}$ *with* $\mu_\mathsf{i} = \mathsf{rank}(G_\mathsf{i})$ *for* $\mathsf{i} \in [\mathsf{p}]$ *where* G_i *is the observability matrix given in* (4)*, the followings are equivalent:*
(i) The pair (A, C) *is* $2\mathsf{q}$ *redundant observable.*
(ii) The matrix Φ *is* $(\{\mu_\mathsf{i}\}$*-stacked)* $2\mathsf{q}$*-error detectable.*
(iii) For every set $\mathcal{J} \subset [\mathsf{p}]$ *satisfying* $|\mathcal{J}| \ge \mathsf{p} - 2\mathsf{q}$, $\Phi_{\mathcal{I}_\mathcal{J}^{\{\mu_\mathsf{i}\}}}$ *has full column rank.*
(iv) The pair (A, C) *is observable under* q*-sparse sensor attacks.*

Note that the redundancy for observability is $2\mathsf{q}$, which is twice the sparsity of the attack signal. This is the key point of constructing the state estimation algorithm. We can examine each subset $\mathcal{J}_\mathsf{k} \subset [\mathsf{p}]$ of sensors whose size is $\mathsf{p} - \mathsf{q}$. In other words, we have $\binom{\mathsf{p}}{\mathsf{q}}$ number of subsets $\mathcal{J}_1, \mathcal{J}_2, \cdots, \mathcal{J}_{\binom{\mathsf{p}}{\mathsf{q}}}$ where $\mathcal{J}_\mathsf{k} \subset [\mathsf{p}]$ and $|\mathcal{J}_\mathsf{k}| = \mathsf{p} - \mathsf{q}$ for $\mathsf{k} = 1, 2, \cdots, \binom{\mathsf{p}}{\mathsf{q}}$. Since Φ is $(\{\mu_\mathsf{i}\}$-stacked) $2\mathsf{q}$-error detectable by Assumption 3 and Lemma 2.(ii), it easily follows that $\Phi_{\mathcal{I}_{\mathcal{J}_\mathsf{k}}^{\{\mu_\mathsf{i}\}}}$ is q-error detectable for $\mathcal{J}_\mathsf{k}$ with $|\mathcal{J}_\mathsf{k}| = \mathsf{p} - \mathsf{q}$. This means that, even after

removing any q sensors, the remaining outputs still have q redundancy for observability. Therefore, the detection scheme of Theorem 3, which relies on the ($\{\mu_i\}$-stacked) q-error detectability of the coding matrix, can be applied for each subset $\mathcal{J}_k \subset [p]$ satisfying $|\mathcal{J}_k| = p - q$.

The configuration of the secure state estimator, which replaces the information fusion center $\mathcal{D}$ in Figure 1, is sketched in Figure 2, and its operation is described in Algorithm 2. Before explaining the operation, let Ψ denote $(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1}$ where Φ and P are given in (21) and (23), respectively. Furthermore, the notation for a sub-matrix is slightly abused for simplicity. For example, $P_{\mathcal{J}}$, $\Phi_{\mathcal{J}}$, and $\Psi_{\mathcal{J}}$ denote

$$P_{\mathcal{I}_{\mathcal{J}}^{\{\mu_i\}},\mathcal{I}_{\mathcal{J}}^{\{\mu_i\}}},\quad \Phi_{\mathcal{I}_{\mathcal{J}}^{\{\mu_i\}}},\quad \text{and}\quad \left(\Phi_{\mathcal{I}_{\mathcal{J}}^{\{\mu_i\}}}^\top \left(P_{\mathcal{I}_{\mathcal{J}}^{\{\mu_i\}},\mathcal{I}_{\mathcal{J}}^{\{\mu_i\}}}\right)^{-1}\Phi_{\mathcal{I}_{\mathcal{J}}^{\{\mu_i\}}}\right)^{-1}\Phi_{\mathcal{I}_{\mathcal{J}}^{\{\mu_i\}}}^\top \left(P_{\mathcal{I}_{\mathcal{J}}^{\{\mu_i\}},\mathcal{I}_{\mathcal{J}}^{\{\mu_i\}}}\right)^{-1},$$

respectively, where $\mathcal{I}_{\mathcal{J}}^{\{\mu_i\}} := \bigcup_{j\in\mathcal{J}}\left\{(\sum_{i=1}^{j-1}\mu_i)+1, (\sum_{i=1}^{j-1}\mu_i)+2, \cdots, \sum_{i=1}^{j}\mu_i\right\}$. Recall that $P_{\mathcal{I}_{\mathcal{J}}^{\{\mu_i\}},\mathcal{I}_{\mathcal{J}}^{\{\mu_i\}}}$ denotes the matrix obtained from P by eliminating all i-th rows and all j-th columns such that $i \notin \mathcal{I}_{\mathcal{J}}^{\{\mu_i\}}$ and $j \notin \mathcal{I}_{\mathcal{J}}^{\{\mu_i\}}$.

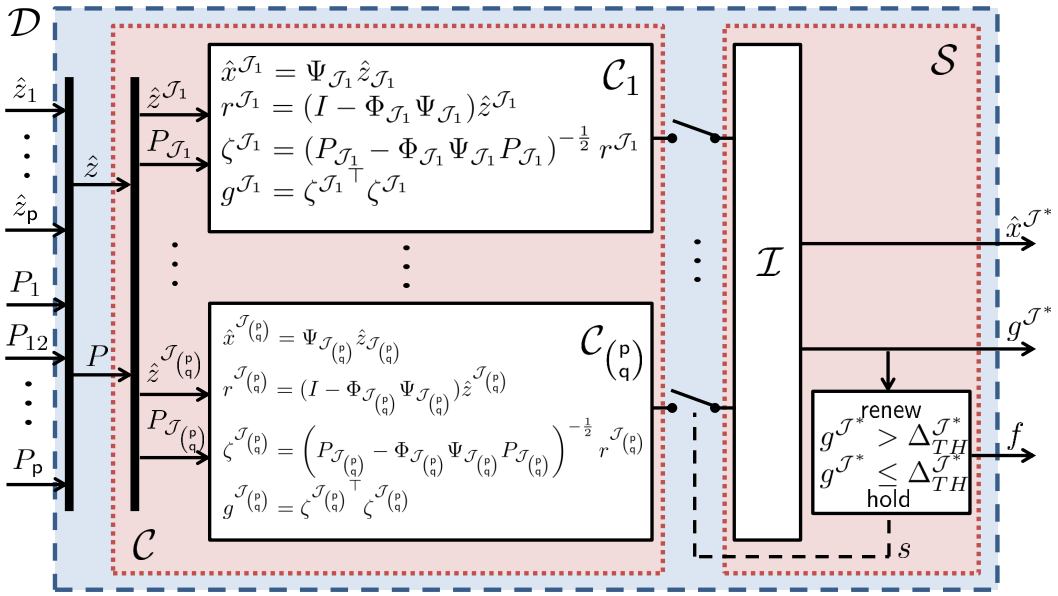


Figure 2. Configuration of the resilient estimation scheme with Gaussian disturbance/noise.

Initially, an attack-free index set $\mathcal{J}^*$, a state estimate $\hat{x}$, a standardized residual's norm g, and a fault alarm signal f, are set to $[p]$, $\Psi\hat{z}$, 0, and 0, respectively. The algorithm continually checks if there is any attack in the index set $\mathcal{J}^*$ based on Algorithm 1. For the given index set $\mathcal{J}^*$, the algorithm essentially calculates the MVUE $\hat{x} = \Psi_{\mathcal{J}^*}\hat{z}_{\mathcal{J}^*}$, the residual $r = \hat{z}_{\mathcal{J}^*} - \Phi_{\mathcal{J}^*}\hat{x}$, the standardized residual $\zeta = (P_{\mathcal{J}^*} - \Phi_{\mathcal{J}^*}\Psi_{\mathcal{J}^*}P_{\mathcal{J}^*})^{-\frac{1}{2}}r$, and its 2-norm $g = \zeta^\top\zeta$ only with the measurement and covariance data from the subset $\mathcal{J}^* \subset [p]$. Recall from Theorem 2 that, if $e_j = e_{\mathcal{I}_j^{\{\mu_i\}}} = 0_{\mu_j\times 1}$ for all $j \in \mathcal{J}^*$, we have $r \sim N(0_{\mu_{\mathcal{J}^*}\times 1}, P_{\mathcal{J}^*} - \Phi_{\mathcal{J}^*}\Psi_{\mathcal{J}^*}P_{\mathcal{J}^*})$ where $\mu_{\mathcal{J}^*} := \sum_{j\in\mathcal{J}^*}\mu_j = \left|\mathcal{I}_{\mathcal{J}^*}^{\{\mu_i\}}\right|$, and thus, $g = \zeta^\top\zeta$ is an observation from a random variable $\mathbf{g}_{\mathcal{J}^*}$, which satisfies a χ^2 distribution with $\mu_{\mathcal{J}^*}$ DOF,

$$\mathbf{g}_{\mathcal{J}^*} \sim \chi^2_{\mu_{\mathcal{J}^*}}. \tag{42}$$

Therefore, g is used to detect the presence of attack in the subset $\mathcal{J}^*$ by the χ^2 test. We compare g with the threshold $\Delta_{TH}^{\mathcal{J}^*}$, which is designed before running the algorithm

and determines the probability of false alarm and the probability of error detection. If $g \leq \Delta_{TH}^{\mathcal{J}^*}$, the index set $\mathcal{J}^*$ is declared to be attack-free by setting $f = 0$ and the algorithm simply maintains the selected optimal index set $\mathcal{J}^*$. Otherwise, when g is greater than the threshold $\Delta_{TH}^{\mathcal{J}^*}$, the attack detection alarm is triggered by setting $f = 1$, and the algorithm starts the process of searching new attack-free index set.

Algorithm 2 Operation of the resilient estimation with Gaussian disturbance/noise

Input: $\hat{z}_1, \hat{z}_2, \cdots, \hat{z}_{\mathsf{p}}, P_1, P_{12}, \cdots, P_{\mathsf{p}(\mathsf{p}-1)}, P_{\mathsf{p}}$
Output: $\mathcal{J}^*, \hat{x}, g, f$
Initialization: $\mathcal{J}^* = [\mathsf{p}], \hat{x} = \Psi\hat{z}, g = 0, f = 0$
while system (1) is running **do**
  $\hat{x} = \Psi_{\mathcal{J}^*}\hat{z}_{\mathcal{J}^*}$
  $r = \hat{z}_{\mathcal{J}^*} - \Phi_{\mathcal{J}^*}\hat{x}$
  $\zeta = \left(P_{\mathcal{J}^*} - \Phi_{\mathcal{J}^*}\Psi_{\mathcal{J}^*}P_{\mathcal{J}^*}\right)^{-\frac{1}{2}} r$
  $g = \zeta^\top\zeta$
  if $g \leq \Delta_{TH}^{\mathcal{J}^*}$ **then**
    $f = 0$
  else if $g > \Delta_{TH}^{\mathcal{J}^*}$ **then**
    $f = 1$
    for $\mathcal{J} \subset [\mathsf{p}]$ satisfying $|\mathcal{J}| = \mathsf{p} - \mathsf{q}$ **do**
      $\hat{x}^{\mathcal{J}} = \Psi_{\mathcal{J}}\hat{z}_{\mathcal{J}}$
      $r^{\mathcal{J}} = \hat{z}_{\mathcal{J}} - \Phi_{\mathcal{J}}\hat{x}^{\mathcal{J}}$
      $\zeta^{\mathcal{J}} = \left(P_{\mathcal{J}} - \Phi_{\mathcal{J}}\Psi_{\mathcal{J}}P_{\mathcal{J}}\right)^{-\frac{1}{2}} r^{\mathcal{J}}$
      $g^{\mathcal{J}} = {\zeta^{\mathcal{J}}}^\top\zeta^{\mathcal{J}}$
    end for
    $\mathcal{J}^* = \underset{\substack{\mathcal{J}\subset[\mathsf{p}]\\|\mathcal{J}|=\mathsf{p}-\mathsf{q}}}{\arg\max}\ p_{\mathbf{g}_{\mathcal{J}}}\left(g^{\mathcal{J}}\right)$
  end if
end while

In order to find a new attack-free index set and, consequently, to recover the state x from the new index set, we search all subsets $\mathcal{J}_\mathsf{k}$'s in $[\mathsf{p}]$ whose size is $\mathsf{p} - \mathsf{q}$. For a detailed explanation, let

$$\left\{\mathcal{J}_1, \mathcal{J}_2, \cdots, \mathcal{J}_{\binom{\mathsf{p}}{\mathsf{q}}}\right\}$$

be the set $\{\mathcal{J} \subset [\mathsf{p}] : |\mathcal{J}| = \mathsf{p} - \mathsf{q}\}$. For each subset $\mathcal{J}_\mathsf{k}$ where $\mathsf{k} \in \left[\binom{\mathsf{p}}{\mathsf{q}}\right]$, the computing module $\mathcal{C}_\mathsf{k}$ calculates the MVUE $\hat{x}^{\mathcal{J}_\mathsf{k}} = \Psi_{\mathcal{J}_\mathsf{k}}\hat{z}_{\mathcal{J}_\mathsf{k}}$, the residual $r^{\mathcal{J}_\mathsf{k}} = \hat{z}_{\mathcal{J}_\mathsf{k}} - \Phi_{\mathcal{J}_\mathsf{k}}\hat{x}^{\mathcal{J}_\mathsf{k}}$, the standardized residual $\zeta^{\mathcal{J}_\mathsf{k}} = \left(P_{\mathcal{J}_\mathsf{k}} - \Phi_{\mathcal{J}_\mathsf{k}}\Psi_{\mathcal{J}_\mathsf{k}}P_{\mathcal{J}_\mathsf{k}}\right)^{-\frac{1}{2}} r^{\mathcal{J}_\mathsf{k}}$, and its 2-norm $g^{\mathcal{J}_\mathsf{k}} = {\zeta^{\mathcal{J}_\mathsf{k}}}^\top\zeta^{\mathcal{J}_\mathsf{k}}$ only with the measurement and covariance data from the subset $\mathcal{J}_\mathsf{k}$. Now, the new optimal subset $\mathcal{J}^*$ is decided by the maximum likelihood (ML) decision rule with the values of $g^{\mathcal{J}_\mathsf{k}}$'s, and the selector $\mathcal{S}$ chooses the optimal index set $\mathcal{J}^*$ by the ML decision rule. To this end, we wish to distinguish between $\binom{\mathsf{p}}{\mathsf{q}}$ hypotheses, $\mathcal{H}_1, \mathcal{H}_2, \cdots, \mathcal{H}_{\binom{\mathsf{p}}{\mathsf{q}}}$, which are given as follows:

$$\mathcal{H}_\mathsf{k}: \text{ the set } \mathcal{J}_\mathsf{k} \text{ is attack-free, i.e., } e_\mathsf{j} = e_{\mathcal{I}_\mathsf{j}^{\{\mu_\mathsf{i}\}}} = 0_{\mu_\mathsf{j}\times 1} \text{ for all } \mathsf{j} \in \mathcal{J}_\mathsf{k}.$$

Let us denote $\mathbf{g}_\mathsf{k}$ as a random variable such that $g^{\mathcal{J}_\mathsf{k}}$ is a single observation from $\mathbf{g}_\mathsf{k}$, whereas $\mathbf{g}_{\mathcal{J}_\mathsf{k}}$ denotes a random variable such that

$$\mathbf{g}_{\mathcal{J}_\mathsf{k}} \sim \chi^2_{\mu_{\mathcal{J}_\mathsf{k}}}$$

with $\mu_{\mathcal{J}_k} := \sum_{j\in\mathcal{J}_k} \mu_j = \left|\mathcal{I}_{\mathcal{J}_k}^{\{\mu_i\}}\right|$ and $p_{\mathbf{g}_{\mathcal{J}_k}}$ is the PDF of the $\chi^2_{\mu_{\mathcal{J}_k}}$ distribution. Note that, if the sensors indexed by $\mathcal{J}_k$ are attack-free, then the random variable $\mathbf{g}_k$ as well as $\mathbf{g}_{\mathcal{J}_k}$ follows the χ^2 distribution with $\mu_{\mathcal{J}_k}$ DOF. The ML decision rule choose the hypothesis $\mathcal{H}_{k^*}$ and the corresponding optimal index set $\mathcal{J}_{k^*}$ that maximize the likelihood $p_{\mathbf{g}_k}(g^{\mathcal{J}_k}; \mathcal{H}_k)$, which is the PDF of $\mathbf{g}_k$ being equal to the observation $g^{\mathcal{J}_k}$ under the hypothesis $\mathcal{H}_k$ (that is, under the condition that there is no attack signal in the measurements indexed by $\mathcal{J}_k$). Therefore, we have

$$\mathcal{J}^* = \mathcal{J}_{k^*} = \underset{k\in\left[\binom{p}{q}\right]}{\arg\max}\ p_{\mathbf{g}_k}\left(g^{\mathcal{J}_k}; \mathcal{H}_k\right) = \underset{\substack{\mathcal{J}\subset[p]\\ |\mathcal{J}|=p-q}}{\arg\max}\ p_{\mathbf{g}_{\mathcal{J}}}\left(g^{\mathcal{J}}\right),$$

where the last equality comes from the fact that $\mathbf{g}_k \sim \chi^2_{\mu_{\mathcal{J}_k}}$ under the hypothesis $\mathcal{H}_k$ so that it follows the PDF of the χ^2 distribution. Therefore, from the index set $\mathcal{J}_{k^*}$ corresponding to the ML hypothesis $\mathcal{H}_{k^*}$, the MVUE of the newly selected optimal index set $\mathcal{J}^*(=\mathcal{J}_{k^*})$, $\hat{x}^{\mathcal{J}^*}$, becomes the final suboptimal estimate of x.

Remark 2. *The proposed algorithm selects the subset of sensors $\mathcal{J}^* \subset [p]$, which is most likely to be attack-free with $|\mathcal{J}^*| = p - q$. Moreover, if the selected set $\mathcal{J}^*$ is actually attack-free, it gives the minimum variance with unbiased estimation. In short, Algorithm 2 generates a state estimate, which is most likely to have minimum variance with unbiased mean. However, we say that it is a suboptimal estimate of x instead of the optimal estimate because the decentralized multi-sensor information fusion Kalman filter cannot ensure to achieve the centralized optimal covariance even without attack as illustrated in ([24], Section 5).*

Remark 3. *Note that Algorithm 2 needs to prepare $\binom{p}{q}$ candidates and compare all those candidates. The time complexity of the error correction algorithm depends on the number of combinations $\binom{p}{q}$, and thus, it has the polynomial time complexity of $\mathcal{O}(p^{\min\{q,\, p-q\}})$. Therefore, the proposed algorithm may not be scalable for very large p with $q \approx p/2$ due to the combinatorial nature of the algorithm. The time complexity could be reduced by imposing additional restrictive assumptions as done in [20,21] which reformulate the problem into a convex optimization problem. However, in our scheme demanding minimal assumptions, the comibinatorial algorithm only needs to operate when an attack is detected. In addition, most of the time, only the attack detection algorithm requiring a small amount of computation, is executed. Another advantage of the proposed algorithm is that its space complexity is linear with the number of sensors p, that is, $\mathcal{O}(p)$. The total memory size required for local observers is $\sum_{i=1}^{p} \mu_i \le np$, whereas if all possible combinations of estimator candidates are configured as real observers, the observer's size becomes $n\binom{p}{q}$.*

5. Simulation Results

We consider a motor-controlled multi-DOF torsion system [34] as depicted in Figure 3. A continuous-time state-space model of the system when the control input is the torque τ (N·m) generated by the servo motor is given by

$$\mathcal{P}_c' : \begin{cases} \dot{x}(t) = A_c' x(t) + B_c' \tau(t) + d(t) \\ y(t) = C_c x(t) + n(t) \end{cases} \tag{43}$$

with the matrices

$$A_c' = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -\frac{k_1}{J_1} & -\frac{b_1}{J_1} & \frac{k_1}{J_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{k_1}{J_2} & 0 & -\frac{k_1+k_2}{J_2} & -\frac{b_2}{J_2} & \frac{k_2}{J_2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{k_2}{J_3} & 0 & -\frac{k_2}{J_3} & -\frac{b_3}{J_3} \end{bmatrix}, \quad B_c' = \begin{bmatrix} 0 \\ \frac{1}{J_1} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{44}$$

$$C_c = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \end{bmatrix},$$

where

$$x := \begin{bmatrix} \theta_1 \\ \dot{\theta}_1 \\ \theta_2 \\ \dot{\theta}_2 \\ \theta_3 \\ \dot{\theta}_3 \end{bmatrix} \quad \text{and} \quad y := \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_1 - \theta_2 \\ \theta_2 - \theta_3 \end{bmatrix}$$

are the state variable and the output measurement, respectively. Here, the unit for angular positions θ's and the unit for angular velocities $\dot{\theta}$'s are (rad) and (rad/s), respectively. The parameters are borrowed from [34], and we have that $J_1 = 0.0022$, $J_2 = J_3 = 0.000545$ (kg·m^2) for the moment of inertia, $b_1 = 0.015$, $b_2 = b_3 = 0.0015$ (N·m/(rad/s)) for the viscous damping ratio, and $k_1 = k_2 = 1$ (N·m/rad) for the flexible stiffness.

Note that the dynamics are the same as those of the three inertia system considered in [15]; however, Figure 3 additionally considers the servo motor system given as follows:

$$\tau = \frac{\eta_g K_g \eta_m k_t (u - K_g k_m \dot{\theta}_1)}{R_m}, \tag{45}$$

which generates the torque τ (N·m) from the input voltage of u (V). The parameters for the servo system are also borrowed from [34], and we have that $\eta_g = 0.9$ for the gearbox efficiency, $K_g = 70$ for the total gear ratio, $\eta_m = 0.69$ for the motor efficiency, $k_t = 0.00768$ (N·m/A) for the motor current torque constant, $k_m = 0.00768$ (V/(rad/s)) for the motor back electromotive force (EMF) constant, and $R_m = 2.6$ (Ω) for the motor armature resistance. Thus, the final continuous-time plant with the voltage u (V) as an input signal is obtained as

$$\mathcal{P}_c : \begin{cases} \dot{x}(t) = A_c x(t) + B_c u(t) + d(t) \\ y(t) = C_c x(t) + n(t) \end{cases} \tag{46}$$

with the matrices

$$A_c = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -\frac{k_1}{J_1} & -\frac{b_1}{J_1} - \frac{\eta_g K_g^2 \eta_m k_t k_m}{R_m} \frac{1}{J_1} & \frac{k_1}{J_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{k_1}{J_2} & 0 & -\frac{k_1+k_2}{J_2} & -\frac{b_2}{J_2} & \frac{k_2}{J_2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{k_2}{J_3} & 0 & -\frac{k_2}{J_3} & -\frac{b_3}{J_3} \end{bmatrix}, \; B_c = \begin{bmatrix} 0 \\ \frac{\eta_g K_g \eta_m k_t}{R_m} \frac{1}{J_1} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{47}$$

and the same C_c as in (44). Finally, the zero-order hold equivalent model of (46) is used for the discrete-time model $\mathcal{P}$ in (1), and the matrices are calculated by

$$A := e^{A_c T_s}, \quad B := \Big(\int_0^{T_s} e^{A_c \tau} d\tau \Big) B_c, \quad C := C_c \tag{48}$$

with the sampling time of $T_s = 0.002$ (s). By examining all possible combinations of sensors, it follows that the system $\mathcal{P}$ in (1) with A and C given in (48) is 2-redundant observable, and hence it is observable under 1-sparse sensor attack by Lemma 2.

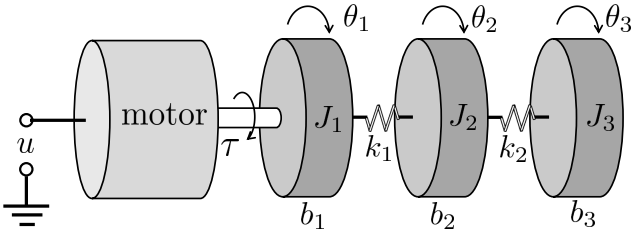


Figure 3. Motor control system of multi-DOF torsion modules.

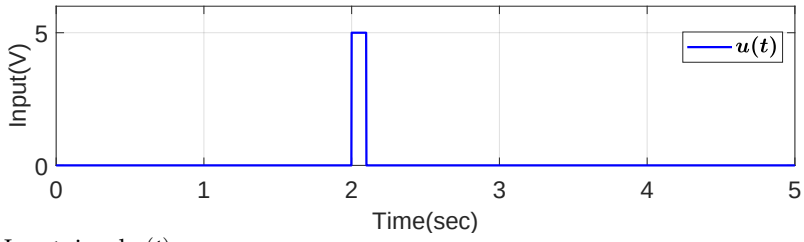
In addition, the disturbance d and the noise n are assumed to satisfy Assumption 1 with

$$Q = 0.001^2 \times \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad R = 0.001^2 \times \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 & -1 \\ 1 & -1 & 0 & 3 & -1 \\ 0 & 1 & -1 & -1 & 3 \end{bmatrix},$$

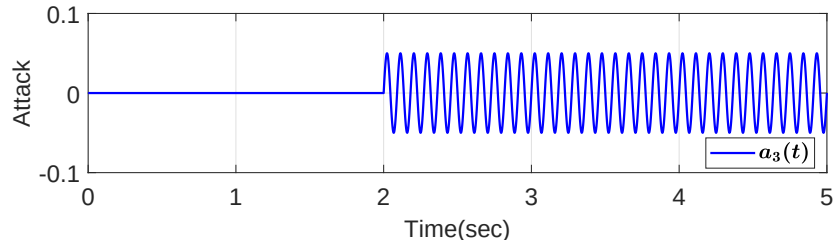
and the initial state $x(0)$ of the system (46) satisfies $x(0) \sim N(\bar{x}_0, P_0)$ as stated in Assumption 1 with the mean $\bar{x}_0$ and the covariance P_0 given by

$$\bar{x}_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad P_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

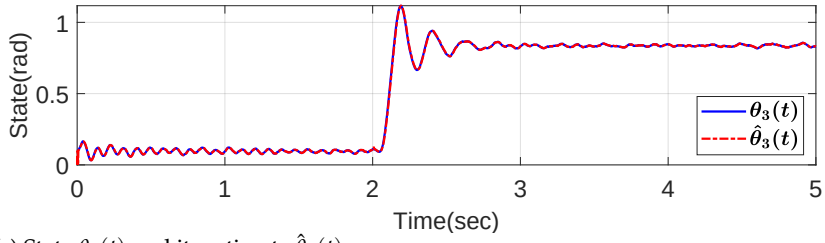
The simulation is performed under 1-sparse sensor attack on the third sensor with the signal shown in Figure 4b, which is made to mimic the motion pattern by the natural frequency as observed in Figure 4c,d. Moreover, the attack starts at 2 second, which is the same time when the square pulse input u is injected into the system as described in Figure 4a. Even under the attack signal, the resilient state estimation with multi-sensor information fusion Kalman filter based on the observability decomposition developed in Sections 3 and 4 works well. The states are recovered with a small error as demonstrated in Figure 4c,d, which are the state estimation results for θ_3 and $\dot{\theta}_3$, respectively.



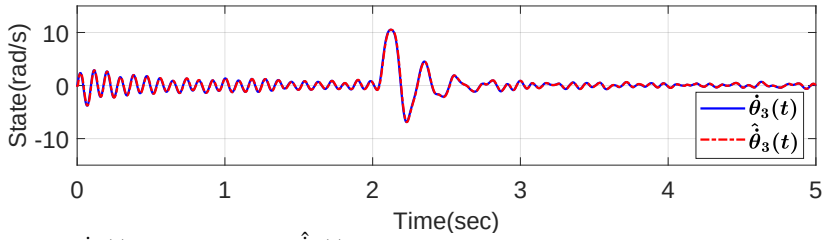
(**a**) Input signal $u(t)$



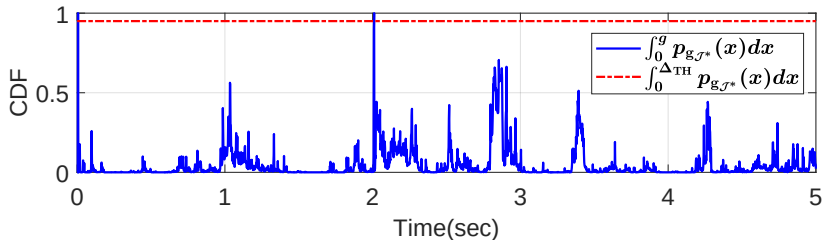
(**b**) Attack signal $a_3(t)$



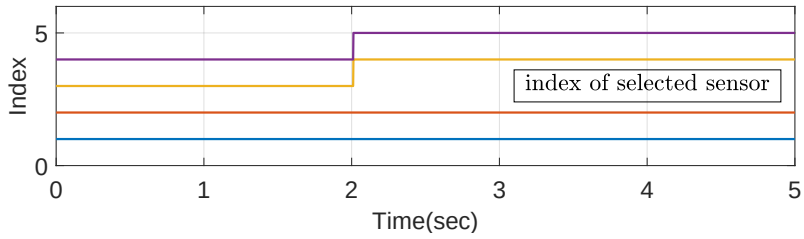
(**c**) State $\theta_3(t)$ and its estimate $\hat{\theta}_3(t)$



(**d**) State $\dot{\theta}_3(t)$ and its estimate $\hat{\dot{\theta}}_3(t)$



(**e**) Value of cumulative density function and its threshold for attack detection



(**f**) Index of selected sensors for resilient state estimation

Figure 4. Plot of signals in a multi-DOF torsion system.

In this simulation, the threshold Δ_{TH} for the attack detection is chosen by $\delta = 0.05$ in (41) so that the cumulative density function (CDF) satisfies $\int_0^{\Delta_{TH}} p_{\mathbf{g}_{\mathcal{J}^*}}(x)dx = 0.95$ where $p_{\mathbf{g}_{\mathcal{J}^*}}$ is the PDF of a random variable $\mathbf{g}_{\mathcal{J}^*}$, which satisfies a χ^2 distribution with $\mu_{\mathcal{J}^*}$ DOF, as stated in (42). Since Figure 4e shows that the 2-norm of the standardized residual, g, exceeds the threshold Δ_{TH} at the instant of 2 second, which is the initiation time of the attack, it is judged that there is an attack (the lines from 8 to 9 in Algorithm 2) and the estimation scheme begins to search the indices of attack-free sensors (the lines from 10 to 16 in Algorithm 2). As a result of the search algorithm, a new set of sensor indices is found by the ML decision rule (the line 16 in Algorithm 2), and the attacked third senor is excluded from 2 second as depicted in Figure 4f.

6. Conclusions

In this paper, the multi-sensor information fusion Kalman filter proposed in [24,25] was improved using the observability decomposition to ensure the convergence of the error covariance matrix of each local observer. The local observer of a decentralized Kalman filter with only a single sensor was designed for an observable subspace instead of the entire n-dimensional state vector without any global information. Then, the proposed decentralized information fusion Kalman filter was applied to the secure state estimation problem where some of sensors were compromised by a malicious attacker.

To cope with the zero-mean Gaussian distributed disturbances/noises, a local Kalman filter replaced the partial Luenberger observer designed in [15], where bounded disturbances/noises were considered in the state estimation problem under sparse sensor attacks. When there was no attack, the proposed algorithm guaranteed an optimal state estimate in the sense of minimum variance, and it generated a state estimate that was most likely to have the minimum variance with an unbiased mean in the presence of sparse sensor attacks.

The proposed algorithm can be applied to cyber-physical systems, including complex sensor networks operating based on linear dynamics under sparse sensor attacks as well as Gaussian disturbances/noises. We imposed the minimal assumption of the redundant observability, which is known to be the equivalent condition to solve the problem. Furthermore, the computational time was alleviated by running only a relatively light attack detection scheme for most of the execution time, and the memory size of the observer was reduced by constructing local observers only for observable subspaces.

One possible direction of future research is to develop a distributed attack-resilient state estimator. While this paper proposed a decentralized Kalman filter scheme, the fusion center collects all the data from each sensors. Although the construction of local Kalman filters is decentralized, the information fusion method is still centralized. Therefore, it is necessary to develop a fully distributed attack-resilient state estimation technique for a general sensor network without any central information fusion center.

Funding: This work was supported by the Materials & Components Technology Development Program (20017351, Development of Servo System Technology with a Current Response of 6.2 kHz and Power Regeneration for Automated Manufacturing Equipment Application) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LTI	Linear Time Invariant
i.i.d.	independent and identically distributed
MVUE	Minimum Variance Unbiased Estimator
BLUE	Best Linear Unbiased Estimator
PDF	Probability Density Function
DOF	Degrees Of Freedom
ML	Maximum Likelihood
EMF	ElectroMotive Force
CDF	Cumulative Density Function

References

1. Pasqualetti, F.; Dörfler, F.; Bullo, F. Attack detection and identification in cyber-physical systems. *IEEE Trans. Autom. Control* **2013**, *58*, 2715–2729. [CrossRef]
2. Sandberg, H.; Amin, S.; Johansson, K.H. Cyberphysical security in networked control systems: An introduction to the issue. *IEEE Control Syst. Mag.* **2015**, *35*, 20–23.
3. Teixeira, A.; Shames, I.; Sandberg, H.; Johansson, K.H. A secure control framework for resource-limited adversaries. *Automatica* **2015**, *51*, 135–148. [CrossRef]
4. Zhang, X.; Zhu, F.; Zhang, J.; Liu, T. Attack isolation and location for a complex network cyber-physical system via zonotope theory. *Neurocomputing* **2022**, *469*, 239–250. [CrossRef]
5. Langner, R. Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Secur. Priv.* **2011**, *9*, 49–51. [CrossRef]
6. Wright, A. Hacking cars. *Commun. ACM* **2011**, *54*, 18–19. [CrossRef]
7. Ten, C.-W.; Liu, C.-C.; Manimaran, G. Vulnerability assessment of cybersecurity for SCADA systems. *IEEE Trans. Power Syst.* **2008**, *23*, 1836–1846. [CrossRef]
8. Dutta, A.; Langbort, C. Confiscating flight control system by stealthy output injection attack. *J. Aerosp. Inf. Syst.* **2017**, *14*, 203–213. [CrossRef]
9. Liu, Y.; Ning, P.; Reiter, M.K. False data injection attacks against state estimation in electric power grids. *ACM Trans. Inf. Syst. Secur.* **2011**, *14*, 13:1–13:33. [CrossRef]
10. Fawzi, H.; Tabuada, P.; Diggavi, S. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Trans. Autom. Control* **2014**, *59*, 1454–1467. [CrossRef]
11. Chen, Y.; Kar, S.; Moura, J.M.F. Cyber-physical systems: Dynamic sensor attacks and strong observability. In Proceedings of the 40th IEEE International Conference on Acoustics, Speech and Signal Processing, Brisbane, Australia, 19–24 April 2015; pp. 1752–1756.
12. Shoukry, Y.; Tabuada, P. Event-triggered state observers for sparse sensor noise/attacks. *IEEE Trans. Autom. Control* **2016**, *61*, 2079–2091. [CrossRef]
13. Shoukry, Y.; Nuzzo, P.; Puggelli, A.; Sangiovanni-Vincentelli, A.L.; Seshiz, S.A.; Tabuada, P. Secure state estimation for cyber physical systems under sensor attacks: A satisfiability modulo theory approach. *IEEE Trans. Autom. Control* **2017**, *62*, 4917–4932. [CrossRef]
14. An, L.; Yang, G.-H. State estimation under sparse sensor attacks: A constrained set partitioning approach. *IEEE Trans. Autom. Control* **2019**, *64*, 3861–3868. [CrossRef]
15. Lee, C.; Shim, H.; Eun, Y. On redundant observability: From security index to attack detection and resilient state estimation. *IEEE Trans. Autom. Control* **2019**, *64*, 775–782. [CrossRef]
16. Candès, E.J.; Tao, T. Decoding by linear programming. *IEEE Trans. Inf. Theory* **2005**, *51*, 4203–4215. [CrossRef]
17. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [CrossRef]
18. Pajic, M.; Lee, I.; Pappas, G.J. Attack-resilient state estimation for noisy dynamical systems. *IEEE Trans. Control Netw. Syst.* **2017**, *4*, 82–92. [CrossRef]
19. Mishra, S.; Shoukry, Y.; Karamchandani, N.; Diggavi, S.; Tabuada, P. Secure state estimation against sensor attacks in the presence of noise. *IEEE Trans. Control Netw. Syst.* **2017**, *4*, 49–59. [CrossRef]
20. Chang, Y.H.; Hu, Q.; Tomlin, C.J. Secure estimation based Kalman filter for cyber-physical systems against sensor attacks. *Automatica* **2018**, *95*, 399–412. [CrossRef]
21. Liu, X.; Mo, Y.; Garone, E. Local decomposition of Kalman filters and its application for secure state estimation. *IEEE Trans. Autom. Control* **2021**, *66*, 5037–5044. [CrossRef]
22. Mehra, R.K.; Peschon, J. An innovations approach to fault detection and diagnosis in dynamic systems. *Automatica* **1971**, *7*, 637–640. [CrossRef]
23. Brumback, B.; Srinath, M. A chi-square test for fault-detection in Kalman filters. *IEEE Trans. Autom. Control* **1987**, *32*, 552–554. [CrossRef]
24. Sun, S.-L.; Deng, Z.-L. Multi-sensor optimal information fusion Kalman filter. *Automatica* **2004**, *40*, 1017–1023. [CrossRef]

25. Sun, S.-L. Multi-sensor optimal information fusion Kalman filters with applications. *Aerosp. Sci. Technol.* **2004**, *8*, 57–62. [CrossRef]
26. Kim, J.; Shim, H.; Wu, J. On distributed optimal Kalman-Bucy filtering by averaging dynamics of heterogeneous agents. In Proceedings of the 55th IEEE Conference on Decision and Control, Las Vegas, NV, USA, 12–14 December 2016; pp. 6309–6314.
27. Kim, T.; Lee, C.; Shim, H. Completely decentralized design of distributed observer for linear systems. *IEEE Trans. Autom. Control* **2020**, *65*, 4664–4678. [CrossRef]
28. Lee, C. Attack-Resilient Feedback Control Systems: Secure State Estimation under Sensor Attacks. Ph.D. Dissertation, Seoul National University, Seoul, Korea, 2018.
29. Simon, D. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*; Wiley-Interscience: Hoboken, NJ, USA, 2006.
30. Kay, S.M. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*; Prentice Hall PTR: Upper Saddle River, NJ, USA, 1993.
31. Kay, S.M. *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*; Prentice Hall PTR: Upper Saddle River, NJ, USA,1993.
32. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
33. Zhou, K.; Doyle, J.C. *Essentials of Robust Control*; Prentice Hall: Upper Saddle River, NJ, USA, 1998.
34. Quanser Inc. *Multi-DOF Torsion Experiment User Manual*; Quanser Inc.: Markham, ON, Canada, 2012.