Automatica 136 (2022) 110083

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

On the computational complexity of the secure state-reconstruction problem☆

Yanwen Mao [a,*], Aritra Mitra [b], Shreyas Sundaram [c], Paulo Tabuada [a]

[a] Department of Electrical and Computer Engineering, University of California, Los Angeles, CA 90095, USA
[b] Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA
[c] School of Electrical and Computer Engineering at Purdue University, West Lafayette, IN 47907, USA

ARTICLE INFO

Article history:
Received 5 January 2021
Received in revised form 8 June 2021
Accepted 27 September 2021
Available online 11 December 2021

Keywords:
Resilient state secure state-reconstruction
Eigenvalue and sparse observability
Polynomial time algorithm

ABSTRACT

In this paper, we discuss the computational complexity of reconstructing the state of a linear system from sensor measurements that have been corrupted by an adversary. The first result establishes that the problem is, in general, NP-hard. We then introduce the notion of eigenvalue observability and show that the state can be reconstructed in polynomial time when each eigenvalue is observable by at least $2s+1$ sensors and at most s sensors are corrupted by an adversary. However, there is a gap between eigenvalue observability and the possibility of reconstructing the state despite attacks — this gap has been characterized in the literature by the notion of sparse observability. To better understand this, we show that when the **A** matrix of the linear system has unitary geometric multiplicity, the gap disappears, i.e., eigenvalue observability coincides with sparse observability, and there exists a polynomial time algorithm to reconstruct the state provided the state can be reconstructed.

1. Introduction

This paper is concerned with the detection of attacks on Cyber–Physical Systems (CPSs). The distributed nature of these large-scale systems often leads to increased vulnerabilities. Of particular concern are adversaries that exploit the distributed nature of CPSs to gain access to sensors and launch attacks by modifying their measurements (Cárdenas, Amin, & Sastry, 2008; Giraldo et al., 2018; Special issue on secure control of cyber physical systems, 2017). The most notorious example is the Stuxnet malware (Langner, 2011), which attacked numerous industrial control systems.

Over the last decade, a significant amount of research has focused on reconstructing the state in the presence of sensor attacks — we will refer to this as the Secure State-Reconstruction (SSR) problem throughout the paper. The first experimental demonstration of a stealthy attack on a control system was reported in Amin, Litrico, Sastry, and Bayen (2010) and it was followed by the first theoretical results developed for special classes of systems (Gupta, Langbort, & Başar, 2010; Sandberg, Teixeira, & Johansson, 2010). Stealthy attacks were then formalized in Smith (2011, 2015). An important step in the conceptual understanding of these attacks was given in Pasqualetti, Dörfler, and Bullo (2012, 2013), Sundaram and Hadjicostis (2010), where the existence of such attacks was characterized by the system theoretic notion of zero-dynamics.

In addition to detecting and identifying attacks, it is important to mitigate their effect by continuing to control the plant. Hence, researchers have invested a significant effort in developing algorithms to reconstruct the state since the papers (Fawzi, Tabuada, & Diggavi, 2011, 2014). However, the SSR problem is intrinsically an NP-hard problem (as we show in this paper). Based on how the NP-hardness is tackled, we classify the existing work in two classes: (1) brute force search (Chong, Wakaiki, & Hespanha, 2015; Lu & Yang, 2017), and (2) computationally efficient relaxations. The methods reported in the first class are better suited for small systems as the computational complexity grows combinatorially with the number of sensors. Noteworthy examples of the second class include: convex relaxations (Fawzi et al., 2014; Yong, Foo, & Frazzoli, 2016), distributed detection filters (Pasqualetti et al., 2013), specialized observers under sparsity constraints (Shoukry & Tabuada, 2015), satisfiability modulo theory techniques (Shoukry et al., 2018), and safety envelopes (Tiwari et al., 2014).

☆ This research was funded in part by the Army Research Laboratory, USA under Cooperative Agreement W911NF-17-2-0196, by the UC-NL, USA grant LFR-18-548554, by the NSF, USA award 1740047, and by the NSF CAREER, USA award 1653648. The material in this paper was partially presented at the 58th IEEE Conference on Decision and Control, December 11–13, 2019, Nice, France. This paper was recommended for publication in revised form by Associate Editor Tong Zhou under the direction of Editor Sophie Tarbouriech.

* Corresponding author.
E-mail addresses: yanwen.mao@ucla.edu (Y. Mao), amitra20@seas.upenn.edu (A. Mitra), sundara2@purdue.edu (S. Sundaram), tabuada@ucla.edu (P. Tabuada).

https://doi.org/10.1016/j.automatica.2021.110083
0005-1098/

The distributed version of the SSR problem has also attracted a substantial amount of interest given the distributed nature of CPSs. Several authors have studied the problem of estimating a static vector from a set of corrupted measurements, either over a distributed sensor network (Chen, Kar, & Moura, 2018a; Su & Shahrampour, 2019), or over a connected-on-average network (Chen, Kar, & Moura, 2018b). A control-theoretic approach to distributed function calculation was developed in Sundaram and Hadjicostis (2010). Follow-up works have analyzed the resilient consensus problem, both for discrete (LeBlanc, Zhang, Koutsoukos, & Sundaram, 2013), and continuous-time (LeBlanc, Zhang, Sundaram, & Koutsoukos, 2013) systems. The work in Tseng and Vaidya (2015) also evaluates this method in various network topologies. The problem of guaranteeing resilience in the context of distributed state estimation, when the state of the system evolves over time (based on potentially unstable dynamics) has been recently explored in Deghat, Ugrinovskii, Shames, and Langbort (2019), Mitra and Sundaram (2016), and Mitra and Sundaram (2019). In particular, the authors in Mitra and Sundaram (2019) develop a fully-distributed algorithm that reconstructs the evolving state despite attacks on certain sensors in the network.

Despite the wealth of literature on the security of CPSs, to the best of the authors' knowledge, a detailed characterization of the complexity of the SSR problem is still lacking. On the one hand, the papers (Fawzi et al., 2014; Pasqualetti et al., 2013; Shoukry et al., 2018; Shoukry & Tabuada, 2015; Tiwari et al., 2014; Yong et al., 2016) suggest that the SSR problem is computationally hard since they propose efficient relaxations to the problem. On the other hand, the paper (Mitra & Sundaram, 2019) implicitly proposes a polynomial-time solution to the SSR problem for certain cases. These observations naturally call for a better understanding of the complexity of the SSR problem, which is precisely the goal of this paper.

As we shall soon see, two alternate notions of observability, namely "sparse observability" introduced in Fawzi et al. (2014), Shoukry and Tabuada (2015) (see also (Sundaram & Hadjicostis, 2010) for an equivalent notion in continuous time), and "eigenvalue observability" (Chen, 1998), Mitra and Sundaram (2018), will play key roles in our characterization of the SSR problem complexity. Our contributions are the following:

(1) We show that the SSR problem is NP-hard.
(2) We provide a decomposition that identifies portions of the state that can be reconstructed in polynomial time and portions that are NP-hard to reconstruct.
(3) We offer a polynomial-time solution for the SSR problem under an eigenvalue observability assumption.
(4) We show that checking sparse observability is coNP-complete.
(5) We show that the notions of sparse observability and eigenvalue observability are equivalent when the geometric multiplicity of each eigenvalue of the system matrix $\mathbf{A}$ is 1.

These results can be understood as follows. Although the SSR problem is NP-hard, in general, there may be portions of the state that can be reconstructed in polynomial time. We perform a system decomposition to identify these different portions of the state. In particular, when all the eigenvalues of the system matrix $\mathbf{A}$ have unitary geometric multiplicity, the decomposition results in scalar SSR problems. This establishes the equivalence between sparse observability, a necessary and sufficient condition for the SSR problem to be solvable, and eigenvalue observability, a sufficient condition for the existence of a polynomial time algorithm. Interestingly, even if the unitary geometric multiplicity condition is not satisfied, we may still check eigenvalue observability and, if successful, solve the SSR problem in polynomial time. When the system does not satisfy the eigenvalue observability condition, we conjecture that the SSR problem is intractable since even checking sparse observability is coNP-complete. This paper improves upon the preliminary results in Mao, Mitra, Sundaram, and Tabuada (2019) by introducing a decomposition technique that is key to the aforementioned contributions 1 and 2.

The rest of the paper is organized as follows. In Section 2, we define the notation used throughout the paper. In Section 3, we introduce the system model and give a formal definition of the SSR problem, sparse observability, and eigenvalue observability. We prove that the SSR problem is NP-hard in Section 4. This is then followed by a result on breaking the overall SSR problem into several smaller independent SSR problems. As a special case, we show in Section 6 that under an eigenvalue observability assumption, the SSR problem can be solved in polynomial time. While checking eigenvalue observability can be done in polynomial time, in Section 7 we show that checking sparse observability is coNP-complete. We connect these two notions in Section 8 by showing that they are equivalent when the geometric multiplicity of each eigenvalue of the system matrix $\mathbf{A}$ is 1. Finally, we conclude the paper in Section 9.

2. Preliminaries and notations

The cardinality of a finite set $\mathcal{I} = \{\mathbf{i}_1, \ldots, \mathbf{i}_p\}$ is denoted by $|\mathcal{I}| = p$. For matrices $\mathbf{Q}_{i_1}, \ldots, \mathbf{Q}_{i_p}$ over the same field and with the same number of columns, we define the matrix $\mathbf{Q}_{\mathcal{I}} = \left[\mathbf{Q}_{i_1}^T | \mathbf{Q}_{i_2}^T | \ldots | \mathbf{Q}_{i_p}^T\right]^T$ by stacking the individual matrices vertically.

We use $\mathbb{R}$ to denote the field of real numbers, $\mathbb{Q}$ to denote the field of rational numbers, and $\mathbb{C}$ to denote the field of complex numbers. For a matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, we use $\ker \mathbf{A}$ to denote the kernel of $\mathbf{A}$, $\mathrm{Im}(\mathbf{A})$ to denote the image of $\mathbf{A}$ and $\mathbf{A}|_V$ to denote the restriction of the linear map defined by $\mathbf{A}$ to the subspace V. We also denote by $\mathbf{A}(V)$ the set $\{y \in \mathbb{R}^n | y = \mathbf{A}x, x \in V\}$.

Let V be a vector space. The collection of vector spaces $\{V^j\}_{j=1,\ldots,r}$, with $V^j \subseteq V$, is said to be an internal direct sum of V, denoted by $V = \bigoplus_{j=1,\ldots,r} V^j$, if any vector $v \in V$ can be uniquely written as $v = v_1 + \ldots + v_r$ with $v_j \in V^j$. The direct sum comes equipped with canonical inclusions $\imath_j : V^j \to V$ taking $v_j \in V^j$ to $\imath_j(v_j) = v_j \in V$, and canonical projections $\pi_j : V \to V^j$ taking $v \in V$ to $\pi_j(v) = v_j \in V^j$.

As an example, consider $V = \mathbb{R}^4$ and let $V^1 = \mathrm{Im}(\mathbf{M}_1)$, $V^2 = \mathrm{Im}(\mathbf{M}_2)$, and $V^3 = \mathrm{Im}(\mathbf{M}_3)$ where $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{M}_3$ are the following linear transformations:

$$\mathbf{M}_1 = \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{M}_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \quad \mathbf{M}_3 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \end{bmatrix}. \tag{1}$$

The collection $\{V^1, V^2, V^3\}$ is an internal direct sum of V since all the column vectors are linearly independent. The canonical inclusions $\imath_j$ can be represented by $\mathbf{I}_4|_{V^j}$, the identity matrix $\mathbf{I}_4$ of order 4 restricted to the subspace V^j, since $\imath_j$ maps any vector $v \in V^j$ to $v \in V$. Conversely, the canonical projections π_j are represented by the matrices $\mathbf{P}_j = \mathbf{M}_i\mathbf{U}_j\mathbf{M}^{-1}$, where $\mathbf{U}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$, $\mathbf{U}_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}$, $\mathbf{U}_3 = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$, as well as $\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}$.

Let $V = \bigoplus_{j=1,\ldots,r} V^j$, $W = \bigoplus_{j=1,\ldots,r} W^j$, and consider a linear map $F : V \to W$ satisfying $F(V^j) \subseteq W^j$. Then, the linear map $F^{(j)} : V^j \to W^j$ defined by $F^{(j)} = \pi_j \circ F \circ \imath_j$ satisfies:

$$F^{(j)} \circ \pi_j = \pi_j \circ F \tag{2}$$

$$\imath_j \circ F^{(j)} = F \circ \imath_j, \tag{3}$$

where $\circ$ denotes function composition.

Continuing with our example, let $\mathbf{F}$ be represented by the matrix:

$$\mathbf{F} = \frac{1}{2}\begin{bmatrix} 2 & 0 & 0 & -4 \\ 1 & 3 & -1 & 4 \\ -1 & -1 & 3 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix}, \tag{4}$$

and note that $\mathbf{F}(V^j) \subseteq V^j$. The maps $\mathbf{F}^{(j)}$ are then given by $\mathbf{F}^{(1)} = \mathbf{P}_1\mathbf{F} \circ \imath_1 = \mathbf{P}_1\mathbf{F}|_{V^1} = \mathbf{I}_4|_{V^1}$, $\mathbf{F}^{(2)} = \mathbf{P}_2\mathbf{F} \circ \imath_2 = \mathbf{P}_2\mathbf{F}|_{V^2} = 2\mathbf{I}_4|_{V^2}$, as well as $\mathbf{F}^{(3)} = \mathbf{P}_3\mathbf{F} \circ \imath_3 = \mathbf{P}_3\mathbf{F}|_{V^3} = 3\mathbf{I}_4|_{V^3}$. Since the vector subspaces V^j are the generalized eigenspaces of $\mathbf{F}$ corresponding to each different eigenvalue, the matrices $\mathbf{F}^{(j)}$ are simply the identity matrix restricted to V^j multiplied by the corresponding eigenvalue.

We denote by $\lambda_1, \ldots, \lambda_r \in \mathbb{C}$ the (counted without repetition) eigenvalues of $\mathbf{A}$ and by $sp(\mathbf{A}) = \{\lambda_1, \ldots, \lambda_r\}$ its spectrum. The algebraic multiplicity of an eigenvalue λ_j, denoted by $\alpha(\lambda_j)$, is the number of times (counted with repetition) that λ_j is a solution of $\det(\mathbf{A} - \lambda_j\mathbf{I}_n) = 0$. The geometric multiplicity of an eigenvalue λ_j, denoted by $\gamma(\lambda_j)$, is the dimension of the vector space $\ker(\mathbf{A} - \lambda_j\mathbf{I}_n)$. We denote the space of generalized eigenvectors associated with λ_j, $\ker(\mathbf{A} - \lambda_j\mathbf{I}_n)^{\alpha(\lambda_j)}$, by V_j. Note that V_j has dimension $\alpha(\lambda_j)$ and $\gamma(\lambda_j)$ Jordan chains.

Given a vector $\mathbf{b} \in \mathbb{R}^n$, we denote by $\|\mathbf{b}\|_0$ the number of non-zero entries in $\mathbf{b}$.

3. Problem formulation

3.1. System model

Consider a discrete-time linear time-invariant system under sensor attacks of the following form:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) \tag{5}$$

$$\mathbf{y}_i(k) = \mathbf{C}_i\mathbf{x}(k) + \mathbf{e}_i(k), \tag{6}$$

where $\mathbf{x}(k) \in \mathbb{R}^n$ and $\mathbf{y}_i(k) \in \mathbb{R}^{p_i}$ represent the state of the system and the measurement acquired by sensor i respectively. The vector $\mathbf{e}_i(k) \in \mathbb{R}^{p_i}$ models the attack on sensor i. If sensor i is attacked by an adversary, then $\mathbf{e}_i(k)$ can be arbitrary, otherwise, $\mathbf{e}_i(k)$ remains zero for any k. Let $\mathcal{V}$ denote the set of sensors, and let $N = |\mathcal{V}|$. We use $\mathbf{C} = \left[\mathbf{C}_1^T|\mathbf{C}_2^T|\cdots|\mathbf{C}_N^T\right]^T$ to denote the collection of the sensor observation matrices, $\mathbf{y}(k) = \begin{bmatrix}\mathbf{y}_1^T(k) & \cdots & \mathbf{y}_N^T(k)\end{bmatrix}^T$ and $\mathbf{e}(k) = \begin{bmatrix}\mathbf{e}_1^T(k) & \cdots & \mathbf{e}_N^T(k)\end{bmatrix}^T$ to represent the collective measurement vector and the collective attack vector, respectively.

We define $\mathcal{O}_i = \left[\mathbf{C}_i^T|(\mathbf{C}_i\mathbf{A})^T|\ldots|(\mathbf{C}_i\mathbf{A}^{\tau_i-1})^T\right]^T$ to be the observability matrix of sensor i with τ_i being the observability index of the pair $(\mathbf{A}, \mathbf{C}_i)$. We also define two more vectors $\mathbf{Y}_i = \begin{bmatrix}\mathbf{y}_i^T(0) & \ldots & \mathbf{y}_i^T(\tau_i-1)\end{bmatrix}^T$ and $\mathbf{E}_i = \begin{bmatrix}\mathbf{e}_i^T(0) & \ldots & \mathbf{e}_i^T(\tau_i-1)\end{bmatrix}^T$ to be the collection of measurements and attacks of sensor i over the time horizon $[0, \tau_i - 1]$, respectively. An equivalent expression for the measurements is:

$$\mathbf{Y}_i = \mathcal{O}_i\mathbf{x}(0) + \mathbf{E}_i. \tag{7}$$

In the remainder of the paper, we drop the time indices to simplify notation.

3.2. The Secure State-Reconstruction problem

Problem 1 (*Secure state-reconstruction*)**.**

Input: Matrices $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{C}_i \in \mathbb{R}^{p_i\times n}$, $i = 1, \ldots, N$, and a set of vectors $\mathbf{Y}_i \in \mathbb{R}^{p_i\tau_i}$, $i = 1, \ldots, N$.

Question: Find a vector $\mathbf{x} \in \mathbb{R}^n$ and a set $\mathcal{I}$ of minimal cardinality such that $\mathbf{Y}_j = \mathcal{O}_j\mathbf{x}$ for all $j \notin \mathcal{I}$.

In other words, the SSR problem requires the reconstruction of a state $\mathbf{x}$ and the simplest attack explanation in the form of the least number of attacked sensors. Note that when the solution $\mathbf{x}$ is unique, we have found the state of the linear system. Although uniqueness of solutions is essential when handling attacks, we can study the complexity of the SSR problem independently of the number of solutions. To make this clear, we will explicitly state the uniqueness requirements when needed.

3.3. Sparse observability and eigenvalue observability

The notions of sparse observability and eigenvalue observability are instrumental to the results in this paper.

Definition 1 (*Sparse Observability Index*)**.** The sparse observability index of the pair $(\mathbf{A}, \mathbf{C})$ in system (5)–(6) is the largest integer k such that $\ker \mathcal{O}_{\mathcal{V}\setminus\mathcal{K}} = \{0\}$ for any $\mathcal{K} \subseteq \mathcal{V}$, $|\mathcal{K}| \leq k$. When the sparse observability index is r, we say that system (5)–(6) is $r-$sparse observable.

It is proved in Fawzi et al. (2014), Shoukry and Tabuada (2015) (see also Chong et al. (2015) for a similar notion in continuous time) that the possibility of uniquely reconstructing the state $\mathbf{x}(k)$ is characterized by the sparse observability index.

Theorem 1 (*Chong et al., 2015; Fawzi et al., 2014; Shoukry & Tabuada, 2015*)**.** *Consider the linear system (5)–(6) where at most s sensors are subject to attacks. The state $\mathbf{x}(k)$ can be uniquely reconstructed if and only if the sparse observability index of the pair $(\mathbf{A}, \mathbf{C})$ is at least $2s$.*

In view of this result, computing the sparse observability index of a system is of great interest since it characterizes the maximum number of arbitrary sensor attacks that can be tolerated without compromising the ability to uniquely reconstruct the state.

In addition to sparse observability, we will require the notion of eigenvalue observability (Chen, 1998; Mitra & Sundaram, 2018).

Definition 2 (*Eigenvalue Observability Index*)**.** We say that an eigenvalue $\lambda \in sp(\mathbf{A})$ is observable w.r.t. sensor i if the linear map defined by $\begin{bmatrix}\mathbf{A} - \lambda\mathbf{I}_n \\ \mathbf{C}_i\end{bmatrix}$ is injective.

If the above condition is satisfied, we say that "sensor i can observe the states in the generalized eigenspace corresponding to λ", or briefly, we say "sensor i can observe eigenvalue λ". Let the set of all sensors that can observe an eigenvalue λ be denoted $\mathcal{S}_\lambda$. The eigenvalue observability index of system (5)–(6) is the largest integer k such that each eigenvalue of the matrix $\mathbf{A}$ is observable by at least $k+1$ distinct sensors. When the eigenvalue observability index is k, we say that system (5)–(6) is k-eigenvalue observable.

We study the SSR problem under the following assumptions.

Assumption 1. For each sensor $i \in \{1, \ldots, N\}$ under attack, the adversary can only manipulate sensor i's measurements through the signal $\mathbf{e}_i(k)$ in (6).

Assumption 2. The adversary is omniscient, i.e., we assume the adversary has full knowledge of the system state, measurements, and plant model. Moreover, all the attacked sensors are allowed to work cooperatively.

4. SSR is hard

Fawzi et al. established in Fawzi et al. (2014) a connection between the SSR problem and compressed sensing by drawing inspiration from the ideas of Candes and Tao in Candes and Tao (2005). We take this approach further by also using the ideas in Candes and Tao (2005) to establish that the SSR problem is NP-hard. To do so, we first define the compressed sensing problem.

Problem 2 (*Compressed Sensing*)**.**
Input: A full row rank matrix $\mathbf{F} \in \mathbb{Q}^{m\times n}$, a vector $\mathbf{b} \in \mathbb{Q}^m$.
Question: Find the sparsest solution of $\mathbf{Fx} = \mathbf{b}$.

The compressed sensing problem yields the solution to the minimization problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{x}\|_0 \\ \text{s.t.} \quad & \mathbf{Fx} = \mathbf{b}. \end{aligned} \tag{8}$$

Theorem 2 (*Fawzi et al., 2014*)**.** *The SSR problem is NP-hard.*

Proof. Given an instance of the compressed sensing problem, we generate an instance of the SSR problem as follows. Let the system matrix be of the form $\mathbf{A} = \mathbf{I}_n$, and the collective observation matrix $\mathbf{C}$ satisfy $\text{Im}\mathbf{C} = \ker \mathbf{F}$. Let the measurements of the sensors be scalar-valued, i.e., let $\mathbf{C}_i$ be the ith row of $\mathbf{C}$. Note that based on the above $\mathbf{A}$ matrix, the observability index for each sensor $i \in \{1, \ldots, N\}$ is given by $\tau_i = 1$, and thus $\mathcal{O}_i = \mathbf{C}_i$. Finally, let $\mathbf{Y}$ be any solution to the equation $\mathbf{FY} = \mathbf{b}$. Since the linear equation $\mathbf{FY} = \mathbf{b}$ is underdetermined, finding a solution $\mathbf{Y}$ can be done in polynomial time (Laub, 2004). For each $i \in \{1, \ldots, N\}$, set $\mathbf{Y}_i$ to be the ith row of $\mathbf{Y}$. Thus, given an instance of the compressed sensing problem, the instance of the SSR problem described above can be constructed in polynomial time.

The SSR problem for the constructed instance degenerates to:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{e}} \quad & \|\mathbf{e}\|_0 \\ \text{s.t.} \quad & \mathbf{Cx} + \mathbf{e} = \mathbf{Y}. \end{aligned} \tag{9}$$

We now show these two problems have the same solution. It is simple to see that any solution $(\mathbf{x}, \mathbf{e})$ of $\mathbf{Cx} + \mathbf{e} = \mathbf{Y}$ provides a solution to $\mathbf{Fe} = \mathbf{b}$, since by applying $\mathbf{F}$ we obtain:

$$\begin{aligned} &\mathbf{F}(\mathbf{Cx} + \mathbf{e}) = \mathbf{FY} \\ &\quad \Leftrightarrow \mathbf{Fe} = \mathbf{b}. \end{aligned} \tag{10}$$

To prove the converse, we show that for every $\mathbf{e}$ such that $\mathbf{Fe} = \mathbf{b}$, there exists some $\mathbf{x}$ satisfying $\mathbf{Cx} + \mathbf{e} = \mathbf{Y}$. Recalling that $\mathbf{FY} = \mathbf{b}$, we obtain $\mathbf{F}(\mathbf{Y} - \mathbf{e}) = \mathbf{0}$, i.e., $\mathbf{Y} - \mathbf{e} \in \ker \mathbf{F}$. Since $\ker \mathbf{F} = \text{Im}\mathbf{C}$, there exists an $\mathbf{x}$ such that $\mathbf{Cx} = \mathbf{Y} - \mathbf{e}$, as desired.

Noticing that the equations $\mathbf{Fe} = \mathbf{b}$ and $\mathbf{Cx} + \mathbf{e} = \mathbf{Y}$ have the same solutions for $\mathbf{e}$, we conclude that they also have the same sparsest solution. In other words, if there exists an algorithm $\mathcal{A}$ that solves the SSR problem for the specific instance constructed by us, such an algorithm will also yield a solution to the given instance of the compressed sensing problem. It then follows that since the compressed sensing problem is NP-hard (Natarajan, 1995), the secure state reconstruction problem is also NP-hard. □

5. System decomposition

In the previous section, we proved that the SSR problem is in general NP-hard. This means there does not exist a polynomial-time solution unless $P = NP$. Despite this fact, we show in this section how to decompose the SSR problem into smaller instances. In the next section, we identify which of these smaller instances are NP-hard, and which ones are solvable in polynomial time.

Lemma 1. *Assume the existence of a collection of vector spaces* $\{X^j\}_{j=1,\ldots,r}$ *satisfying:*

(1) $\mathbb{C}^n = \bigoplus_{j=1,\ldots,r} X^j$;
(2) $\mathbf{A}(X^j) \subseteq X^j$ *for* $j = 1, \ldots, r$;
(3) $\mathcal{O}_i(\mathbb{C}^n) = \bigoplus_{j=1,\ldots,r} \mathcal{O}_i^j(X^j)$ *for* $i = 1, \ldots, p$,

then for any $\mathbf{Y}_i$*, a solution* $\mathbf{x}$ *of the equation:*

$$\mathbf{Y}_i = \mathcal{O}_i\mathbf{x}, \tag{11}$$

whenever it exists, can be written as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 + \ldots + \mathbf{x}_r$ *with* $\mathbf{x}_j = \pi_j(\mathbf{x}) \in X^j$ *given by the solution of:*

$$\mathbf{Y}_i^j = \mathcal{O}_i^j\mathbf{x}_j, \tag{12}$$

for $\mathbf{Y}_i^j = \pi_j(\mathbf{Y}_i) \in \mathcal{O}_i^j(X^j)$ *and* $\mathcal{O}_i^j = \pi_j \circ \mathcal{O}_i \circ \iota_j$.

Proof. Let $\mathbf{x}_j$ be the solution of (12) and note that:

$$\mathbf{Y}_i^j = \mathcal{O}_i^j\mathbf{x}_j \Rightarrow \iota_j(\mathbf{Y}_i^j) = \iota_j \circ \mathcal{O}_i^j(\mathbf{x}_j) = \mathcal{O}_i \circ \iota_j(\mathbf{x}_j) = \mathcal{O}_i\mathbf{x}_j, \tag{13}$$

where the third equality follows from (3). By summing over j we obtain:

$$\mathbf{Y}_i = \sum_{j=1}^{r} \iota_j(\mathbf{Y}_i^j) = \sum_{j=1}^{r} \mathcal{O}_i\mathbf{x}_j = \mathcal{O}_i \sum_{j=1}^{r} \mathbf{x}_j = \mathcal{O}_i\mathbf{x}. \tag{14}$$

Hence, the solutions to (12) provide a solution to (11). Consider now (11):

$$\begin{aligned} \mathbf{Y}_i = &\mathcal{O}_i\mathbf{x} \Rightarrow \pi_j(\mathbf{Y}_i) = \pi_j \circ \mathcal{O}_i(\mathbf{x}) \\ &\Rightarrow \mathbf{Y}_i^j = \mathcal{O}_i^j \circ \pi_j(\mathbf{x}) = \mathcal{O}_i^j\mathbf{x}_j. \end{aligned} \tag{15}$$

where the third equality follows from (2). Hence, if $\mathbf{x}$ is a solution to (11), then $\mathbf{x}_i$ is a solution to (12). □

Intuitively, we treat the state–space $\mathbb{R}^n$ as the direct sum of multiple subspaces. If the images of these subspaces under the linear map $\mathcal{O}_i$ are pairwise non-overlapping, we are able to project the state vector $\mathbf{x}$ onto these subspaces, project the measurement $\mathbf{Y}_i$ onto the image under the linear map $\mathcal{O}_i$ of these subspaces, and then establish a one-to-one correspondence between the projected state vector and the projected measurement. This effectively decomposes the original problem into r subproblems, each of dimension $\dim(X^j)$. As formalized in the next result, the spaces X^j can always be taken to be the generalized eigenspaces of $\mathbf{A}$.

Proposition 1. *The generalized eigenspaces* $V^1, \ldots, V^r$ *of* $\mathbf{A}$ *satisfy properties (1)-(3) in Lemma 1.*

Proof. Properties (1) and (2) in Lemma 1 follow directly from the definition of generalized eigenspace. To simplify notation, we will drop the sensor index i in this proof.

It also follows from the definition of generalized eigenspace that $\cup_{j=1,\ldots,r} V^j$ spans $\mathbb{C}^n$. Therefore, the set $\cup_{j=1,\ldots,r} \mathcal{O}(V^j)$ spans $\mathcal{O}(\mathbb{C}^n)$. Given this, to conclude property (3) we only need to show:

$$\mathcal{O}(V^j) \cap \mathcal{O}(V^k) = \{0\}, \qquad \forall j \neq k.$$

Moreover, it suffices to show that for any $\mathbf{x}_j \in V^j$ and $\mathbf{x}_k \in V^k$, with $j \neq k$, the equality $\mathcal{O}(\mathbf{x}_j + \mathbf{x}_k) = 0$ can only be satisfied if $\mathcal{O}\mathbf{x}_j = 0 = \mathcal{O}\mathbf{x}_k$.

We have the following sequence of equalities that is explained thereafter:

$$0 = \mathcal{O}(\mathbf{x}_j + \mathbf{x}_k) \tag{16}$$
$$= \mathcal{O}(\mathbf{A} - \lambda_k \mathbf{I}_n)^{\alpha(\lambda_k)}(\mathbf{x}_j + \mathbf{x}_k) \tag{17}$$
$$= \mathcal{O}(\mathbf{A} - \lambda_k \mathbf{I}_n)^{\alpha(\lambda_k)}(\mathbf{x}_j) \tag{18}$$
$$= \mathcal{O}\mathbf{x}_j. \tag{19}$$

The second step follows from $\ker \mathcal{O} \subseteq \ker \mathcal{O}(\mathbf{A} - \lambda_k \mathbf{I}_n)^{\alpha(\lambda_k)}$, the third step follows from $\mathbf{x}_k \in V^k = \ker(\mathbf{A} - \lambda_k \mathbf{I}_n)^{\alpha(\lambda_k)}$, and the fourth from the following sequence of steps:

$$\dim \ker \mathcal{O}\big|_{V^j} \leq \dim \ker \mathcal{O}(\mathbf{A} - \lambda_k \mathbf{I}_n)^{\alpha(\lambda_k)}\big|_{V^j} \tag{20}$$
$$= \dim \ker (\mathbf{A} - \lambda_k \mathbf{I}_n)^{\alpha(\lambda_k)}\big|_{V^j} \tag{21}$$
$$+ \dim \ker \mathcal{O}\big|_{(\mathbf{A}-\lambda_k \mathbf{I}_n)^{\alpha(\lambda_k)} V^j} \tag{22}$$
$$= \dim \ker \mathcal{O}\big|_{(\mathbf{A}-\lambda_k \mathbf{I}_n)^{\alpha(\lambda_k)} V^j} \tag{23}$$
$$\leq \dim \ker \mathcal{O}\big|_{V^j}. \tag{24}$$

The first step comes from $\ker \mathcal{O} \subseteq \ker \mathcal{O}(\mathbf{A} - \lambda_k \mathbf{I}_n)$. To show that the second step is true, we observe that $\dim \ker \mathbf{MN} = \dim \ker \mathbf{N} + \dim \ker(\mathbf{M}\big|_{\mathbf{N}(\mathbb{C}^n)})$ for any matrices $\mathbf{M}, \mathbf{N} \in \mathbb{C}^{n\times n}$. The third step comes from the map $(\mathbf{A}-\lambda_j \mathbf{I}_n)^{\alpha(\lambda_j)}\big|_{V^j}$ being injective if $j \neq k$, as the generalized eigenspaces V^j and V^k intersect only at the origin, and $\ker(\mathbf{A} - \lambda_j \mathbf{I}_n)^{\alpha(\lambda_j)} = V^j$. The fourth step follows by the $\mathbf{A}-$invariant nature of eigenspace V^j. This shows $\dim \ker \mathcal{O}\big|_{V^j} = \dim \ker \mathcal{O}(\mathbf{A} - \lambda_k \mathbf{I}_n)^{\alpha(\lambda_k)}\big|_{V^j}$ which, combined with $\ker \mathcal{O}\big|_{V^j} \subseteq \ker \mathcal{O}(\mathbf{A} - \lambda_k \mathbf{I}_n)^{\alpha(\lambda_k)}\big|_{V^j}$, can only hold when $\ker \mathcal{O}\big|_{V^j} = \ker \mathcal{O}(\mathbf{A}-\lambda_k \mathbf{I}_n)^{\alpha(\lambda_k)}\big|_{V^j}$. A symmetric argument can be used to show that $\mathcal{O}\mathbf{x}_k = 0$ and the claim is thus proved. □

Combining Lemma 1 and Proposition 1 results in a decomposition of the sensor measurements in (7):

$$\mathbf{Y}_i^j = \mathcal{O}_i^j \mathbf{x}_j, \quad j = 1, 2, \ldots, r, \tag{25}$$

where $\mathbf{Y}_i^j = \pi_j(Y_i)$ is the projection of measurement $\mathbf{Y}_i$ onto the vector space $\mathcal{O}_i(V^j)$, the linear transformation $\mathcal{O}_i^j$ is defined by $\mathcal{O}_i^j = \pi_j \circ \mathcal{O}_i \circ \imath_j$, $\mathbf{x}_j$ is given by $\mathbf{x}_j = \pi_j(\mathbf{x})$, $\pi_j : \mathbb{R}^n \to V^j$ is the canonical projection and $\imath_j : V^j \to \mathbb{R}^n$ is the canonical inclusion.

Theorem 3. *A solution* $\mathbf{x}$ *of the SSR problem with inputs* $\mathbf{A}, \mathbf{C}_i, \mathbf{Y}_i$ *is given by* $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_j$ *where* $\mathbf{x}_i$ *is the solution to the SSR problem with inputs* $\mathbf{A}^{(j)} = \pi_j \circ \mathbf{A} \circ \imath_j$, $\mathbf{C}_i^j = \mathbf{C}_i \circ \imath_j$, $\mathbf{Y}_i^j$.

Proof. Follows directly from Lemma 1, Proposition 1, and the properties of generalized eigenspaces. □

Theorem 3 lays the theoretical foundation for decomposing the SSR problem with n states into r sub-problems of the form:

$$\begin{aligned} \mathbf{x}_j(k+1) &= \mathbf{A}^{(j)}\mathbf{x}_j(k), \\ \mathbf{Y}_i^j(k) &= \mathcal{O}_i^j \mathbf{x}_j(k) + \mathbf{E}_i^j(k), \end{aligned} \tag{26}$$

each with $\alpha(\lambda_1), \alpha(\lambda_2), \ldots, \alpha(\lambda_r)$ states. The attack vector $\mathbf{E}_i^j$ is identically zero when sensor i is not under attack. The state of the original problem can be reconstructed by summing up the state reconstructions of each sub-problem.

We now illustrate the decomposition of (5)–(6) into (26) through an example. The matrix $\mathbf{A}$ is the same as the matrix $\mathbf{F}$ defined in (4) and the matrices $\mathbf{C}_i$ are given by:

$$\mathbf{C}_1 = \begin{bmatrix} 3 & 2 & 0 & 2 \end{bmatrix}, \quad \mathbf{C}_2 = \begin{bmatrix} 2 & 3 & 1 & -1 \end{bmatrix},$$
$$\mathbf{C}_3 = \begin{bmatrix} 2 & 2 & 0 & 0 \end{bmatrix}, \quad \mathbf{C}_4 = \begin{bmatrix} 2 & 3 & -1 & 0 \end{bmatrix}.$$

As we discussed below (4), the generalized eigenspaces of $\mathbf{A}$ are $V^1 = \mathrm{Im}(\mathbf{M}_1)$, $V^2 = \mathrm{Im}(\mathbf{M}_2)$, and $V^3 = \mathrm{Im}(\mathbf{M}_3)$ corresponding to eigenvalues 1, 2, and 3 respectively, where $\mathbf{M}_j$ are defined in (1) for $j = 1, 2, 3$. Also, recall that the projections π_1, π_2, and π_3 are $\mathbf{P}_j = \mathbf{M}_j(\mathbf{M}_j^T\mathbf{M}_j)^{-1}\mathbf{M}_j^T$ for $j = 1, 2, 3$. By definition, we have $\mathbf{x}_1 = \mathbf{P}_1\mathbf{x}$, $\mathbf{x}_2 = \mathbf{P}_2\mathbf{x}$, $\mathbf{x}_3 = \mathbf{P}_3\mathbf{x}$, and $\mathbf{A}^{(1)} = \mathbf{P}_1\mathbf{A}|_{V^1}$, $\mathbf{A}^{(2)} = \mathbf{P}_2\mathbf{A}|_{V^2}$, $\mathbf{A}^{(3)} = \mathbf{P}_3\mathbf{A}|_{V^3}$. Hence the decomposition of $\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k)$ is given by:

$$\mathbf{P}_j\mathbf{x}(k+1) = (\mathbf{P}_j\mathbf{A}|_{Vj})(\mathbf{P}_j\mathbf{x}(k)), \quad j = 1, 2, 3.$$

We now illustrate how to decompose the measurement equation $\mathbf{Y}_1(k) = \mathcal{O}_1\mathbf{x}(k) + \mathbf{E}_1(k)$ for sensor 1. The observability matrix $\mathcal{O}_1$ of sensor 1 is given by:

$$\mathcal{O}_1 = \begin{bmatrix} 3 & 2 & 0 & 2 \\ 4 & 3 & -1 & 4 \\ 6 & 5 & -3 & 10 \\ 10 & 9 & -7 & 28 \end{bmatrix}.$$

We first compute the projections $\widetilde{\pi}_1^1$, $\widetilde{\pi}_1^2$ and $\widetilde{\pi}_1^3$ that map $\mathcal{O}_1(\mathbb{R}^4)$ to $\mathcal{O}_1(V^1)$, $\mathcal{O}_1(V^2)$, and $\mathcal{O}_1(V^3)$, respectively. To do this, we define the matrices:

$$\widetilde{\mathbf{M}}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \widetilde{\mathbf{M}}_2 = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 8 \end{bmatrix}, \text{ and } \widetilde{\mathbf{M}}_3 = \begin{bmatrix} 1 \\ 3 \\ 9 \\ 27 \end{bmatrix},$$

which satisfy $\mathcal{O}_1(V^1) = \mathrm{Im}(\widetilde{\mathbf{M}}_1)$, $\mathcal{O}_1(V^2) = \mathrm{Im}(\widetilde{\mathbf{M}}_2)$, and $\mathcal{O}_1(V^3) = \mathrm{Im}(\widetilde{\mathbf{M}}_3)$. We also remark that the collection $\{\mathcal{O}_1(V^1), \mathcal{O}_1(V^2), \mathcal{O}_1(V^3)\}$ is an internal direct sum of the vector space $\mathcal{O}_1(\mathbb{R}^4)$. Therefore, by defining $\widetilde{\mathbf{M}} = \begin{bmatrix} \widetilde{\mathbf{M}}_1 & \widetilde{\mathbf{M}}_2 & \widetilde{\mathbf{M}}_3 \end{bmatrix}$ and $\widetilde{\mathbf{U}}_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, $\widetilde{\mathbf{U}}_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$, $\widetilde{\mathbf{U}}_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$, each projection $\widetilde{\pi}_1^i$ can be represented by the projection matrix:

$$\widetilde{\mathbf{P}}_1^i = \widetilde{\mathbf{M}}_i\widetilde{\mathbf{U}}_i(\widetilde{\mathbf{M}}^T\widetilde{\mathbf{M}})^{-1}\widetilde{\mathbf{M}}^T, \ i = 1, 2, 3.$$

By definition, $\mathbf{Y}_1^j = \widetilde{\mathbf{P}}_1^j\mathbf{Y}_1$, $\mathbf{E}_1^j = \widetilde{\mathbf{P}}_1^j\mathbf{E}_1$ and $\mathcal{O}_1^j = \widetilde{\mathbf{P}}_1^j\mathcal{O}_1|_{V^j}$ for $j = 1, 2, 3$. In summary, the decomposition of measurement $\mathbf{Y}_1(k) = \mathcal{O}_1\mathbf{x}(k) + \mathbf{E}_1(k)$ is given by:

$$\widetilde{\mathbf{P}}_1^j\mathbf{Y}_1(k) = (\widetilde{\mathbf{P}}_1^j\mathcal{O}_1|_{V^j})(\mathbf{P}_1^j\mathbf{x}(k)) + \widetilde{\mathbf{P}}_1^j\mathbf{E}_1(k), \quad j = 1, 2, 3.$$

6. Classes of SSR problems solvable in polynomial time

While in the previous section we established that the SSR problem is NP-hard, in this section we leverage the results in Section 5 to answer a simple but important question: when can we solve the SSR problem in polynomial time? Our answer relies heavily on the system decomposition technique introduced in Section 5. The first result establishes that the decomposition can be done in polynomial time.

Proposition 2. *The computational complexity of decomposing the system (5)–(6) into sub-systems (26) is within* $O(pn^3)$.

Proof. To prove this result, we list all the steps involved in the decomposition from (5)–(6) to (26) and list the computational complexity of each step.

Offline preparation 1: compute the observability matrix of each sensor $\mathcal{O}_i$. The computational complexity of this step is $O(pn^2)$.

Offline preparation 2: find the eigenvalues of the matrix $\mathbf{A}$ as well as its generalized eigenspaces V^j. This can be done by finding the Jordan form of $\mathbf{A}$. The computational complexity of this step is $O(n^3)$.

Offline preparation 3: determine the image of each generalized eigenspace V^j under the observability matrix $\mathcal{O}_i$, i.e., $\mathcal{O}_i(V^j)$. In this step, we perform p times two $n \times n$ matrix multiplications and thus the complexity of this step is $O(pn^3)$.

Offline preparation 4: find the projection matrix for each generalized eigenspace and each sensor. The computational complexity of this step is $O(pn^3)$.

Online task: at each time instance, project the measurements $\mathbf{Y}_i(k)$ of each sensor i onto each generalized eigenspace. In this step, for each sensor we multiply a $n \times n$ matrix by a $n \times 1$ vector r times. This requires $O(pn^2 r)$ time.

We thus conclude that we can decompose the system (5)–(6) into sub-systems (26) within $O(pn^3)$ and finish the proof. □

Before giving an answer to the question we stated at the beginning of this section, we relate the sparse observability index defined for the system (5)–(6) and the sparse observability index for each subsystem (26) with j ranging from 1 to r in the following two results. Note that, since the state space of (26) is V^j, sparse observability is characterized by the injectivity of $\mathcal{O}_i^j|_{V^j}$ whereas eigenvalue observability is characterized by injectivity of the linear map $\begin{bmatrix} \mathbf{A}^{(j)} - \lambda_j \mathbf{I}_n^{(j)} \\ \mathbf{C}_i^j \end{bmatrix}$, where we define $\mathbf{I}_n^{(j)} = \pi_j \circ \mathbf{I}_n \circ \imath_j$. We now have the following results.

Theorem 4. *The system (5)–(6) is k-sparse observable if and only if for each $j \in \{1, 2, \ldots, r\}$, the system (26) is k-sparse observable.*

Proof. This result can be easily established by observing that $\ker \mathcal{O}_i = \oplus_{j=1}^r \ker \mathcal{O}_i^j$ holds for any sensor i. We omit the proof here in the interest of space. □

Similarly, to relate the eigenvalue observability index defined for the overall system and the eigenvalue observability index for each subsystem, we have the following result.

Theorem 5. *The system (5)–(6) is k-eigenvalue observable if and only if for each $j \in \{1, 2, \ldots, r\}$, the system (26) is k-eigenvalue observable.*

Proof. By the definition of eigenvalue observability, it suffices to show the matrix $\begin{bmatrix} \mathbf{A} - \lambda_j \mathbf{I}_n \\ \mathbf{C}_i \end{bmatrix}$ has full column rank if and only if each matrix $\begin{bmatrix} \mathbf{A}^{(j)} - \lambda_j \mathbf{I}_n^{(j)} \\ \mathbf{C}_i^j \end{bmatrix}$ defines an injective map with domain V^j, for j ranging from 1 to r.

Consider the map $F : V \to V \times \mathbb{R}^{p_i}$ defined by the matrix $\begin{bmatrix} \mathbf{A} - \lambda_j \mathbf{I}_n \\ \mathbf{C}_i \end{bmatrix}$ and note that F being injective is equivalent to $\ker F = \{0\}$. Note also that the result immediately follows if we establish that $\ker F \subseteq V^j$. This can be seen by noting that $F\mathbf{x} = 0$ for $\mathbf{x} \in \mathbb{R}^n$ degenerates to $F\mathbf{x} = 0$ for $\mathbf{x} \in V^j$ and (given $\mathbf{x} = \imath_j \mathbf{x}$) can be written as $F\imath_j \mathbf{x} = 0$:

$$\begin{bmatrix} \mathbf{A} \circ \imath_j - \lambda_j \imath_j \\ \mathbf{C}_i \circ \imath_j \end{bmatrix} \mathbf{x} = 0. \tag{27}$$

Moreover, since $(\mathbf{A} - \lambda_j \mathbf{I}_n)(V^j) \subseteq V^j$ we have the equality $\pi_j(\mathbf{A} - \lambda_j \mathbf{I}_n)\imath_j \mathbf{x} = (\mathbf{A} - \lambda_j \mathbf{I}_n)\imath_j \mathbf{x}$. Therefore, (27) degenerates into:

$$\begin{bmatrix} \pi_j \circ \mathbf{A} \circ \imath_j - \lambda_j \pi_j \circ \imath_j \\ \mathbf{C}_i \circ \imath_j \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{A}^{(j)} - \lambda_j \mathbf{I}_n^{(j)} \\ \mathbf{C}_i^j \end{bmatrix} \mathbf{x} = 0. \tag{28}$$

Therefore, we proceed by showing that $\ker F \subseteq V^j$. The equality $F\mathbf{x} = 0$ implies $(\mathbf{A} - \lambda_j \mathbf{I}_n)\mathbf{x} = 0$. If we write $\mathbf{x}$ as $\mathbf{x}_j + \mathbf{x}_{\bar{j}}$ with $\mathbf{x}_j = \pi_j(\mathbf{x})$ and $\mathbf{x}_{\bar{j}} = \sum_{k=1, k\neq j}^r \pi_k(\mathbf{x})$ we have $(\mathbf{A} - \lambda_j \mathbf{I}_n)(\mathbf{x}_j + \mathbf{x}_{\bar{j}}) = 0$. We now make two observations. The first is that $(\mathbf{A} - \lambda_j \mathbf{I}_n)\mathbf{x}_{\bar{j}} = 0$ implies $\mathbf{x}_{\bar{j}} = 0$ since $\mathbf{x}_{\bar{j}} \neq 0$ would imply that $\mathbf{x}_{\bar{j}} \in V^j$, by definition of V^j. The second observation is that $(\mathbf{A} - \lambda_j \mathbf{I}_n)(V^\ell) \subseteq V^\ell$, for $\ell \in \{1, \ldots, r\}$, implies that $(\mathbf{A} - \lambda_j \mathbf{I}_n)(\mathbf{x}_j + \mathbf{x}_{\bar{j}}) = 0$ iff $(\mathbf{A} - \lambda_j \mathbf{I}_n)\mathbf{x}_j = 0$ and $(\mathbf{A} - \lambda_j \mathbf{I}_n)\mathbf{x}_{\bar{j}} = 0$. Together with the first observation we have $\mathbf{x}_{\bar{j}} = 0$ which implies that $\mathbf{x} \in V^j$ and concludes the proof. □

Based on the above decomposition and the assumption that at most s sensors are attacked, we partition the set of eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_r\}$ as follows:

- We define $\mathcal{J}_1 \subseteq \{\lambda_1, \lambda_2, \ldots, \lambda_r\}$ to be the set of eigenvalues whose corresponding subsystems (26) are not $2s$-sparse observable.
- We define $\mathcal{J}_2 \subseteq \{\lambda_1, \lambda_2, \ldots, \lambda_r\} \setminus \mathcal{J}_1$ to be the set of eigenvalues whose corresponding subsystems (26) are $2s$-eigenvalue observable.
- We define $\mathcal{J}_3 = \{\lambda_1, \lambda_2, \ldots, \lambda_r\} \setminus \{\mathcal{J}_1 \cup \mathcal{J}_2\}$ to be the set of eigenvalues whose corresponding subsystems (26) are $2s$-sparse observable but not $2s-$eigenvalue observable.

6.1. Impossibility of reconstructing substates corresponding to eigenvalues in the set $\mathcal{J}_1$

It is established in Section 3 that the SSR problem does not admit a unique solution if it is not $2s-$sparse observable. Therefore, it is impossible to reconstruct the substates corresponding to eigenvalues in $\mathcal{J}_1$. Furthermore, by Theorem 4 if $\mathcal{J}_1$ is not empty, the overall system defined in (5)–(6) is not $2s-$sparse observable, which in turn means the solution is not unique.

6.2. Reconstructing the substates corresponding to eigenvalues in the set $\mathcal{J}_2$

We learned from Theorem 5 that if λ_j is observable w.r.t. sensor i, then after decomposing the system, λ_j is also observable w.r.t. to sensor i in the jth sub-system corresponding to this sensor. By the Popov–Belevitch–Hautus (PBH) test, the jth sub-system $(\mathbf{A}^{(j)}, \mathbf{C}_i^j)$ is observable, which shows that $\mathbf{x}_j$ can be reconstructed using only measurements from sensor i.

We now explain how to reconstruct the substates corresponding to eigenvalues in $\mathcal{J}_2$ based on majority voting. Consider any eigenvalue $\lambda_j \in \mathcal{J}_2$. Let S_{λ_j} represent the set of sensors w.r.t. which λ_j is observable. The result of the PBH test implies that $\mathbf{x}_j$ can be recovered using the measurements of each of the sensors in the set $\mathcal{S}_{\lambda_j}$. We denote by $x_j^{(l)}$ the lth component of $\mathbf{x}_j$. Based on the definition of the set $\mathcal{J}_2$, we have $|\mathcal{S}_{\lambda_j}| \geq (2s + 1)$. Consequently, since at most s sensors have been compromised, we are guaranteed at least $s + 1$ consistent copies of the state $x_j^{(l)}$. Thus, each component of the vector $x_j^{(l)}$ can be recovered via majority voting and therefore all the substates corresponding to eigenvalues in $\mathcal{J}_2$ can be reconstructed in polynomial time.

6.3. Computational complexity of reconstructing substates corresponding to eigenvalues in the set $\mathcal{J}_3$

The NP-hardness of solving the SSR problem has been established in Section 4. In this subsection, we argue that with the prescribed decomposition technique, the computational complexity of solving the SSR problem for substates corresponding to eigenvalues in $\mathcal{J}_3$ could be reduced whenever we only need to reconstruct substates whose dimension is smaller than n. Assuming s is the upper bound of the number of attacked sensors, we have the following theorem.

Theorem 6. *By applying the decomposition (26), the SSR problem can be solved in time $\sum_{\lambda_j \in \mathcal{J}_3} \mathcal{C}(p, n_j) + O(pn^3)$ if the system (5)–(6) is $2s-$sparse observable, where $\mathcal{C}(p, n)$ is the time complexity of solving an instance of the SSR problem with n states and p sensors whose corresponding system is $2s-$sparse observable.*

Before providing a proof we first discuss how this result may reduce the computational complexity of solving the SSR problem. For a large-scale CPS, it is not uncommon for the number of sensors to greatly exceed the number of states, i.e., $p \gg n$. We note that the computational complexity of brute force search grows exponentially with p. Also, the computational complexity of some brute force search algorithms (such as Chong et al. (2015)) to determine whether a set of sensors is attacked is at least $O(n^2)$. In other words, for such algorithms $\mathcal{C}(p,n) \geq O(p^2n^2)$. By assuming $p \gg n$ we make the following observations:

(1) $O(p^2n^2) \geq \sum_{j=1}^{r} O(p^2n_j^2)$, and equality holds only when $r=1$.
(2) $O(pn^3) \ll \sum_{j=1}^{r} O(p^2n_j^2)$.

The first observation shows that the computation required to solve all the sub-problems is smaller than what is required to solve the original problem. The second observation shows that, compared with the computational complexity of solving the SSR problem, the computation required for decomposition of the original system is negligible. These two facts indicate that by decomposing the SSR problem into simpler instances, we reduce the computational complexity of solving the SSR problem.

Proof of Theorem 6. We already established that reconstructing the state of each decomposed system is also an SSR problem and the solution $\mathbf{x}$ of the original problem is obtained by summing over all the projections, i.e., $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_r$. Therefore any algorithm that solves the SSR problem can be applied to solve each subproblem, i.e., we may solve each subproblem corresponding to $\lambda_j \in \mathcal{J}_3$ within time complexity $\mathcal{C}(p, n_j)$ since there are p sensors and n_j states. By the assumption that the system (5)–(6) is $2s-$sparse observable as well as Theorem 4, all subsystems are $2s-$sparse observable and hence $\mathcal{J}_1 = \{\phi\}$, and for each subproblem corresponding to $\lambda_j \in \mathcal{J}_2$ the time complexity of the majority voting algorithm is within $O(pn^2)$. In summary, the total computational complexity is:

$$\sum_{\lambda_j \in \mathcal{J}_2} O(pn_j^2) + \sum_{\lambda_j \in \mathcal{J}_3} \mathcal{C}(p, n_j) + O(pn^3) \tag{29}$$

$$= \sum_{\lambda_j \in \mathcal{J}_3} \mathcal{C}(p, n_j) + O(pn^3), \tag{30}$$

which finishes the proof. □

Remark 1. The actual complexity might be even smaller than $\sum_{\lambda_j \in \mathcal{J}_3} \mathcal{C}(p, n_j) + O(pn^3)$. This can be seen by noting that we solve each smaller SSR problem sequentially, and thus we can remove measurements from sensors that have been identified as being attacked when solving subsequent problems.

To conclude, we have the following result which answers the question at the beginning of this section by pointing out when the SSR problem can be solved in polynomial time, which actually is a corollary of Theorem 6.

Corollary 1. *Consider the system (5)–(6), and suppose at most s sensors are attacked. Let the eigenvalue observability index of system (5)–(6) be at least $2s$. Then, the SSR problem can be solved in polynomial time.*

Remark 2. Another understanding of this classification of eigenvalues into $\mathcal{J}_1$, $\mathcal{J}_2$, and $\mathcal{J}_3$ is provided by the vulnerability of the corresponding substates. Substates in $\mathcal{J}_1$ are the most vulnerable to attack since the defender may not even be able to identify the attacked set of sensors. Substates in $\mathcal{J}_2$ are robust against attacks since attacked sensors can be easily determined. For substates $\mathcal{J}_3$, the defender is able to identify the attacked sensors, but this task requires a substantially higher computational effort.

In other words, in the view of the adversary, a wise attacking strategy is to attack the substates corresponding to eigenvalues in $\mathcal{J}_1$, and it should avoid attacking states in $\mathcal{J}_2$ since majority voting will allow the defender to easily identify the compromised sensors.

6.4. Example — continued

In this subsection we continue the example in Sections 2 and 5 and show how to classify each subsystem under the assumption that the adversary can attack at most $s = 1$ sensor. We recall that V^1, V^2, V^3 are the eigenspaces corresponding to eigenvalues 1, 2, and 3, respectively. Also, after decomposition, we have $\mathbf{A}^{(j)} = \mathbf{P}_j\mathbf{A}|_{Vj}$ as well as $\mathcal{O}_i^j = \widetilde{\mathbf{P}}_i^j\mathcal{O}_i|_{Vj}$ for $i = 1, 2, 3, 4$ and $j = 1, 2, 3$.

We first claim that $\lambda_3 = 3$ belongs to $\mathcal{J}_1$. To see why this is true, we remove $2s = 2$ sensors, sensor 1 and sensor 4, and explicitly compute $\mathcal{O}_2^3$ and $\mathcal{O}_3^3$. We have:

$$\mathcal{O}_2 = \begin{bmatrix} 2 & 3 & 1 & -1 \\ 3 & 4 & 0 & -1 \\ 5 & 6 & -2 & -1 \\ 9 & 10 & -6 & -1 \end{bmatrix}, \mathcal{O}_3 = \begin{bmatrix} 2 & 2 & 0 & 0 \\ 3 & 3 & -1 & 0 \\ 5 & 5 & -3 & -0 \\ 9 & 9 & -7 & 0 \end{bmatrix},$$

and $\mathcal{O}_2(V^3) = \mathcal{O}_3(V^3) = \{0\}$ which yields $(\widetilde{\mathbf{P}}_2^3\mathcal{O}_2)\mathbf{x}_3' = 0$ and $(\widetilde{\mathbf{P}}_3^3\mathcal{O}_3)\mathbf{x}_3'' = 0$ for any $\mathbf{x}_3'$ and $\mathbf{x}_3''$ in V^3. Therefore, we have $\mathcal{O}_2^3 = \mathcal{O}_3^3 = 0$. By the definition of sparse observability, we have $\ker \mathcal{O}_{\{2,3\}}^3 = V^3$ and hence the subsystems corresponding to eigenvalue 3 are not $2s-$sparse observable. Also, a similar analysis reveals that subsystems corresponding to eigenvalues λ_1 and λ_2 are both $2s-$sparse observable, hence $1 \notin \mathcal{J}_1$ and $2 \notin \mathcal{J}_1$.

Next we argue that $\lambda_2 = 2$ belongs to $\mathcal{J}_2$. To see why this is true, we first recall that $\mathbf{A}^{(2)} = \mathbf{P}_2\mathbf{A}|_{V^2}$, $\mathbf{I}_4^{(2)} = \mathbf{I}_4|_{V^2}$, $\mathbf{C}_i^2 = \mathbf{C}_i|_{V^2}$, and then check that for sensor 1, the matrix:

$$\begin{bmatrix} \mathbf{A}^{(2)} - 2\mathbf{I}_4^{(2)} \\ \mathbf{C}_1^2 \end{bmatrix} = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 \\ -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & -2 \\ 3 & 2 & 0 & 2 \end{bmatrix}\Bigg|_{V^2},$$

defines an injective map. We also run the same check on sensor 2, 3, and 4 to conclude that eigenvalue λ_2 is observable by all 4 sensors. Hence the subsystems corresponding to λ_2 are $2s-$eigenvalue observable. Proceeding in the same fashion we conclude that subsystems corresponding to eigenvalue λ_1 are not $2s-$eigenvalue observable. Therefore, the eigenvalue $\lambda_1 = 1$ belongs to $\mathcal{J}_3$.

In summary, the substates in V^3 cannot be securely reconstructed, the substates in V^1 can be securely reconstructed in the presence of at most 1 attacked sensor, and the substates in V^2 can be securely reconstructed and the reconstruction can be done efficiently.

7. Complexity of checking sparse observability

In the previous two sections, we studied the complexity of the SSR problem, and in particular, identified instances of the problem that can be solved in polynomial time. Recall that under at most s sensor attacks on the system (5)–(6), $2s$-sparse observability is necessary and sufficient for the SSR problem to yield a unique solution, namely the true initial state vector $\mathbf{x}(0)$. Given this result, we now take a step back and ask: what is the complexity of deciding whether a given system is $2s$-sparse observable? This question is highly relevant since it aims to identify the maximum number of sensor attacks that can be tolerated by a given system of the form (5)–(6). In what follows, we show that

determining the sparse-observability index (see Definition 1) of a system is computationally hard; we will focus on the case of scalar-valued sensors throughout, as it suffices to establish the computational complexity of the problem.

Problem 3 (*r-sparse Observability*)**.**
Input: A matrix $\mathbf{A} \in \mathbb{Q}^{n\times n}$, a matrix $\mathbf{C} \in \mathbb{Q}^{p\times n}$ and a positive integer r.
Question: Is the pair $(\mathbf{A}, \mathbf{C})$ r-sparse observable?

Note that if the answer to an instance of the r-sparse observability problem is "no", then there is a simple proof: one can provide a set of r rows of $\mathbf{C}$ that, if removed, result in a system that is no longer observable. However, it is not clear whether there is a similarly simple proof for "yes" instances. Thus, the r-sparse observability problem is in the class coNP.[1]

The complement of a decision problem is the problem obtained by switching the "yes" and "no" answers to all instances of that problem. If a problem is in the class coNP, then its complement is in the class NP, and vice versa.

We will show that the r-sparse observability problem is coNP-hard by showing that its complement is NP-hard. Specifically, we define the following complement problem to r-sparse observability.

Problem 4 (*r-sparse Unobservability*)**.**
Input: A matrix $\mathbf{A} \in \mathbb{Q}^{n\times n}$, a matrix $\mathbf{C} \in \mathbb{Q}^{p\times n}$ and a positive integer r.
Question: Is there a set of r rows that can be removed from $\mathbf{C}$ in order to yield a matrix $\bar{\mathbf{C}}$ such that $(\mathbf{A}, \bar{\mathbf{C}})$ is unobservable?

Note that the answer to an instance of r-sparse unobservability is "yes" if and only if the answer to the corresponding instance of r-sparse observability is "no" and vice versa. Further note that r-sparse unobservability is in the class NP.

We show that r-sparse unobservability is NP-complete by providing a reduction from the following *Linear Degeneracy* problem. This problem was shown to be NP-complete in Khachiyan (1995).

Problem 5 (*Linear Degeneracy (Khachiyan, 1995)*)**.**
Input: A full column rank matrix $\mathbf{F} \in \mathbb{Q}^{p\times n}$.
Question: Does $\mathbf{F}$ contain a degenerate (i.e., noninvertible) $n \times n$ submatrix?

In other words, the linear degeneracy problem asks whether it is possible to remove $p - n$ rows from matrix $\mathbf{F}$ so that the resulting (square) matrix is not full rank. We are now ready to prove the following result.

Theorem 7 (*Mao et al., 2019*)**.** *The r-sparse unobservability problem is NP-complete. Thus, the r-sparse observability problem is coNP-complete.*

Proof. Given an instance of the linear degeneracy problem (with matrix $\mathbf{F} \in \mathbb{Q}^{p\times n}$), we construct an instance of the r-sparse unobservability problem as follows: set $\mathbf{A} = \mathbf{I}_n$, $\mathbf{C} = \mathbf{F}$, and $r = p - n$.

We now show that the answer to the constructed instance of r-sparse unobservability is "yes" if and only if the answer to the given instance of linear degeneracy is "yes".

First, suppose that the answer to the constructed instance of r-sparse unobservability is "yes". Then there exists a set of r rows of $\mathbf{C}$ that can be removed such that the remaining rows are not sufficient to yield observability. However, since $\mathbf{A} = \mathbf{I}_n$, the above implies that there is a set of r rows of $\mathbf{C}$ that can be removed such that the remaining rows are not full column rank. Since $\mathbf{C} = \mathbf{F}$ and $r = p - n$, this means that there is an $n \times n$ submatrix of $\mathbf{F}$ that loses rank, and thus the answer to the linear degeneracy problem is "yes".

Next, we show that if the answer to the given instance of linear degeneracy is "yes", then the answer to the constructed instance of r-sparse unobservability is "yes". We will do this by showing the contrapositive: if the answer to the constructed instance of r-sparse unobservability is "no", then the answer to the given instance of linear degeneracy is "no". Suppose the answer to the constructed instance of r-sparse unobservability is "no". Then, by definition, the pair $(\mathbf{A}, \mathbf{C})$ is observable even after removing any arbitrary r rows from $\mathbf{C}$. However, since $\mathbf{A} = \mathbf{I}_n$, in order for the system to remain observable after removing r rows from $\mathbf{C}$, it must be the case that the remaining rows of $\mathbf{C}$ have full column rank. Thus, if the answer to the constructed instance of r-sparse unobservability is "no", then $\mathbf{C}$ has full column rank after removing any arbitrary $r = p - n$ rows. This means that every $n \times n$ submatrix of $\mathbf{C}$ is invertible. Since $\mathbf{C} = \mathbf{F}$, the answer to the given instance of linear degeneracy is "no" (i.e., there is no $n \times n$ submatrix of $\mathbf{F}$ that is degenerate).

Thus, we have shown that the answer to the constructed instance of r-sparse unobservability is "yes" if and only if the answer to the given instance of linear degeneracy is "yes". Since linear degeneracy is NP-complete, so is r-sparse unobservability.

Finally, since r-sparse observability is the complement of r-sparse unobservability, we have that r-sparse observability is coNP-complete. □

Remark 3. In Mitra and Sundaram (2019), certain necessary conditions were presented for estimating the state of a plant despite attacks in a distributed setting, i.e., where measurements of the plant are dispersed over a network of sensors. Specifically, these conditions impose certain requirements on the observation model (in addition to requirements on the communication structure), the complexity of checking which was left open. Interestingly, Theorem 7 resolves this question, and establishes that checking the necessary conditions in Mitra and Sundaram (2019) is computationally hard; since the focus of our paper is on centralized systems, we do not present details of this result here.

8. Connections between sparse observability and eigenvalue observability

In Sections 4 and 7, we showed that the SSR problem and the problem of determining the sparse observability index of a system are each computationally hard. At the same time, Section 6 gave us the positive result that certain instances of the SSR problem can be efficiently solved. In line with this finding, we are now motivated to ask: Can the sparse observability index of a system be computed in polynomial time for certain specific instances? In this section, we show that this is indeed the case by identifying instances of the problem where the notions of sparse observability and eigenvalue observability coincide. Given that the eigenvalue observability index of a system can always be computed in polynomial time based on simple rank tests, an equivalence between the two notions of observability immediately yields instances of the problem where the sparse observability index of the system can also be computed in polynomial time. With this in mind, in this section we will prove each of the implications indicated in Fig. 1. We begin with the following simple result.

[1] See, e.g., Cormen, Leiserson, Rivest, and Stein (2009) for additional details on the complexity classes NP and coNP.

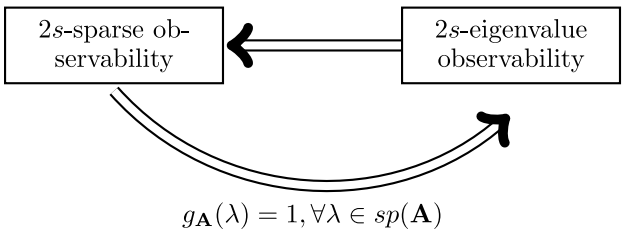


Fig. 1. Figure illustrating the hierarchy of relationships between different notions of observability.

Proposition 3 (*Mao et al., 2019*)**.** *Consider the linear system (5)–(6), and suppose its eigenvalue observability index is 2s. Then, the pair* $(\mathbf{A}, \mathbf{C})$ *is at least 2s-sparse observable.*

Proof. Consider any subset of sensors $\mathcal{F} \subset \mathcal{V}$, such that $|\mathcal{F}| \leq 2s$. To establish that the pair $(\mathbf{A}, \mathbf{C})$ is at least $2s$-sparse observable, we need to show that the pair $(\mathbf{A}, \mathbf{C}_{\mathcal{V}\setminus\mathcal{F}})$ is observable. Based on the PBH test, this amounts to checking that each eigenvalue $\lambda \in sp(\mathbf{A})$ is observable w.r.t. the observation matrix $\mathbf{C}_{\mathcal{V}\setminus\mathcal{F}}$. Let S_λ represent the set of sensors w.r.t. which λ is observable. A sufficient condition for this to happen is $|(\mathcal{V}\setminus\mathcal{F}) \cap \mathcal{S}_\lambda| \geq 1$, which is indeed true given that an eigenvalue observability index of $2s$ implies $|\mathcal{S}_\lambda| \geq (2f+1)$, $\forall \lambda \in sp(\mathbf{A})$, and the fact that $|\mathcal{F}| \leq 2s$. □

To see that the reverse implication does not hold in general, consider the following example.

Example 1. Consider an LTI system of the form (5)–(6) monitored by 6 sensors, with parameters as follows:

$$\mathbf{A} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}, \mathbf{C}_i = \begin{cases} \begin{bmatrix} 1 & 0 \end{bmatrix}, & \text{if } i \in \{1,2,3\}, \\ \begin{bmatrix} 0 & 1 \end{bmatrix}, & \text{if } i \in \{4,5,6\}. \end{cases} \tag{31}$$

Here $\lambda \in \mathbb{R}, |\lambda| \geq 1$. Suppose $s = 1$. Then, the removal of at most 2 sensors will ensure that at least one sensor from each of the sets $\{1,2,3\}$ and $\{4,5,6\}$ remains unattacked; given the measurement model in (31), this is sufficient to preserve observability w.r.t. the remaining sensors. In other words, the system is 2-sparse observable. However, it is easy to verify that the eigenvalue λ is not observable w.r.t. any sensor.

In view of Proposition 3 and Example 1, we conclude that $2s$-sparse observability of a system is in general less restrictive than the condition that the eigenvalue observability index of the system is $2s$. In what follows, we establish that the two aforementioned notions coincide when additional structure is imposed on the spectrum of $\mathbf{A}$.

Proposition 4 (*Mao et al., 2019*)**.** *Consider the linear system model given by (5)–(6), and suppose* $\lambda \in sp(\mathbf{A})$ *has geometric multiplicity* 1*. Consider any non-empty subset of sensors* $\mathcal{S} = \{i_1, i_2, \ldots, i_{|\mathcal{S}|}\} \subseteq \mathcal{V}$*. Then, the eigenvalue* λ *is observable w.r.t. the pair* $(\mathbf{A}, \mathbf{C}_\mathcal{S})$ *if and only if there exists a sensor* $i_p \in \mathcal{S}$ *such that* λ *is observable w.r.t. sensor* i_p*, i.e.,* λ *is observable w.r.t. the pair* $(\mathbf{A}, \mathbf{C}_{i_p})$*.*

Proof. Consider a similarity transformation that maps $\mathbf{A}$ to its Jordan canonical form $\mathbf{J}$. Let this transformation map $\mathbf{C}_\mathcal{S}$ to $\bar{\mathbf{C}}_\mathcal{S}$, and $\mathbf{C}_{i_j}$ to $\bar{\mathbf{C}}_{i_j}$, for each $i_j \in \mathcal{S}$. Since λ has geometric multiplicity 1, there exists a single Jordan block corresponding to λ in $\mathbf{J}$. Let this Jordan block be denoted $\mathbf{J}_\lambda$. Without loss of generality, suppose $\mathbf{J}$ is of the following form:

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_\lambda & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{J}} \end{bmatrix}, \tag{32}$$

where $\bar{\mathbf{J}}$ is the collection of the Jordan blocks corresponding to eigenvalues in $sp(\mathbf{A})\setminus\{\lambda\}$. Based on the PBH test, λ is observable w.r.t. the pair $(\mathbf{J}, \bar{\mathbf{C}}_\mathcal{S})$ if and only if the following condition holds:

$$\operatorname{rank} \begin{bmatrix} \mathbf{J} - \lambda \mathbf{I}_n \\ \bar{\mathbf{C}}_\mathcal{S} \end{bmatrix} = n. \tag{33}$$

Given the structure of $\mathbf{J}$ in (32), and the fact that λ has geometric multiplicity 1, it is easy to see that (33) holds if and only if there is at least one non-zero entry in the first column of $\bar{\mathbf{C}}_\mathcal{S}$. However, the preceding condition holds if and only if there exists some sensor $i_p \in \mathcal{S}$ with at least one non-zero entry in the first column of $\bar{\mathbf{C}}_{i_p}$; the latter is precisely the condition for observability of λ w.r.t. the sensor i_p, given that $g_\mathbf{A}(\lambda) = 1$. To complete the proof, it suffices to notice that a similarity transformation preserves the observability of an eigenvalue. □

We now make use of the previous result to establish an equivalence between sparse observability and eigenvalue observability.

Proposition 5. *Consider the linear system model (5)–(6), and suppose every eigenvalue of* $\mathbf{A}$ *has geometric multiplicity* 1*. Then, the pair* $(\mathbf{A}, \mathbf{C})$ *is* $2s$*-sparse observable if and only if the eigenvalue observability of the system is* $2s$*.*

Proof. For necessity, we proceed via contradiction. Suppose the pair $(\mathbf{A}, \mathbf{C})$ is $2s$-sparse observable, but there exists some $\lambda \in sp(\mathbf{A})$ that is observable w.r.t. at most $2s$ distinct sensors. Recall that the set of sensors w.r.t. which λ is observable is denoted $\mathcal{S}_\lambda$. Based on our hypothesis, $|\mathcal{S}_\lambda| \leq 2s$. Suppose $|\mathcal{S}_\lambda| = 2s$ (since an identical argument can be sketched when $|\mathcal{S}_\lambda| < 2s$). Since $(\mathbf{A}, \mathbf{C})$ is $2s$-sparse observable, the pair $(\mathbf{A}, \mathbf{C}_{\mathcal{V}\setminus\mathcal{S}_\lambda})$ is observable. However, based on Proposition 4, this requires λ to be observable w.r.t. at least one sensor in $\mathcal{V}\setminus\mathcal{S}_\lambda$, leading to the desired contradiction. This completes the proof of necessity. For sufficiency, note from Proposition 3 that the pair $(\mathbf{A}, \mathbf{C})$ is at least $2s$-sparse observable whenever its eigenvalue observability index is $2s$; the fact that the observability index is no more than $2s$ follows from the additional assumption on the geometric multiplicity of eigenvalues, and arguments similar to those used for establishing necessity. □

It directly follows from the definition of eigenvalue observability that the eigenvalue observability index of a system can be computed in polynomial time. Hence, we have the following corollaries of Proposition 5.

Corollary 2. *When all the eigenvalues of the matrix* $\mathbf{A}$ *have geometric multiplicity 1, the sparse observability index of the system can be computed in polynomial time.*

Corollary 3. *For a 2s-sparse observable system (5)–(6), when all the eigenvalues of the matrix* $\mathbf{A}$ *have geometric multiplicity 1, the SSR problem can be solved in polynomial time.*

Proof. It is shown in Proposition 5 that under the unitary geometric multiplicity assumption, a $2s$-sparse observable system is also $2s$-eigenvalue observable. Thus, such a system satisfies the hypotheses in the statement of Corollary 1, and we immediately obtain the existence of a polynomial-time solution for the SSR problem. □

9. Conclusion

In this paper, we showed that when the eigenvalues of the system matrix **A** have unitary geometric multiplicity, the SSR problem is tractable since both checking the sparse observability (see Corollary 2) as well as solving the SSR problem (see Corollary 1) can be performed in polynomial time. When at least one of the eigenvalues has geometric multiplicity greater than one, we can still compute the eigenvalue observability index and, if it is at least $2s$, solve the SSR problem in polynomial time if at most s sensors are attacked. However, in this case, eigenvalue observability is no longer necessary for the SSR problem to be solvable. Since even checking sparse observability is coNP-complete, we conjecture that the SSR problem may be intractable in this case. The authors are currently investigating this conjecture. However, even in this case, the computational complexity of solving the SSR problem can be reduced, when the system matrix **A** has at least 2 distinct eigenvalues.

Acknowledgment

Shreyas Sundaram thanks Lintao Ye for helpful discussions pertaining to the Linear Degeneracy problem.

References

Amin, Saurabh, Litrico, Xavier, Sastry, S Shankar, & Bayen, Alexandre M (2010).Stealthy deception attacks on water SCADA systems. In *Proc. of the 13th ACM int. conference on hybrid systems: Computation and control* (pp. 161–170).

Candes, Emmanuel J., & Tao, Terence (2005). Decoding by linear programming. *IEEE Transactions on Information Theory*, *51*(12), 4203–4215.

Cárdenas, Alvaro A., Amin, Saurabh, & Sastry, Shankar (2008). Research challenges for the security of control systems. In *HotSec*.

Chen, Chi-Tsong (1998). *Linear system theory and design*. Inc.: Oxford University Press.

Chen, Yuan, Kar, Soummya, & Moura, Jose M. F. (2018a). Resilient distributed estimation through adversary detection. *IEEE Transactions on Signal Processing*, *66*(9), 2455–2469.

Chen, Yuan, Kar, Soummya, & Moura, José M. F. (2018b). Topology free resilient distributed estimation. arXiv:1812.08902.

Chong, Michelle S., Wakaiki, Masashi, & Hespanha, Joao P. (2015). Observability of linear systems under adversarial attacks. In *Proc. of the American control conference* (pp. 2439–2444). IEEE.

Cormen, Thomas H, Leiserson, Charles E, Rivest, Ronald L, & Stein, Clifford (2009). *Introduction to algorithms*. MIT Press.

Deghat, Mohammad, Ugrinovskii, Valery, Shames, Iman, & Langbort, Cedric (2019). Detection and mitigation of biasing attacks on distributed estimation networks. *Automatica*, *99*, 369–381.

Fawzi, Hamza, Tabuada, Paulo, & Diggavi, Suhas (2011). Secure state-estimation for dynamical systems under active adversaries. In *Proc. of the 49th annual allerton conference on communication, control, and computing* (pp. 337–344).

Fawzi, Hamza, Tabuada, Paulo, & Diggavi, Suhas (2014). Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, *59*(6), 1454–1467.

Giraldo, Jairo, Urbina, David, Cardenas, Alvaro, Valente, Junia, Faisal, Mustafa, Ruths, Justin, et al. (2018). A survey of physics-based attack detection in cyber-physical systems. *ACM Computing Surveys*, *51*(4), 1–36.

Gupta, Abhishek, Langbort, Cédric, & Başar, Tamer (2010). Optimal control in the presence of an intelligent jammer with limited actions. In *Proc. of the 49th IEEE conference on decision and control* (pp. 1096–1101).

Khachiyan, Leonid (1995). On the complexity of approximating extremal determinants in matrices. *Journal of Complexity*, *11*(1), 138–153.

Langner, Ralph (2011). Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security & Privacy*, *9*(3), 49–51.

Laub, Alan J. (2004). *Matrix analysis for scientists and engineers*. SIAM.

LeBlanc, Heath J, Zhang, Haotian, Koutsoukos, Xenofon, & Sundaram, Shreyas (2013). Resilient asymptotic consensus in robust networks. *IEEE Journal on Selected Areas in Communications*, *31*(4), 766–781.

LeBlanc, Heath J, Zhang, Haotian, Sundaram, Shreyas, & Koutsoukos, Xenofon (2013). Resilient continuous-time consensus in fractional robust networks. In *Proc. of the American control conference* (pp. 1237–1242). IEEE.

Lu, An-Yang, & Yang, Guang-Hong (2017). Secure state estimation for cyber-physical systems under sparse sensor attacks via a switched Luenberger observer. *Information Sciences*, *417*, 454–464.

Mao, Yanwen, Mitra, Aritra, Sundaram, Shreyas, & Tabuada, Paulo (2019). When is the secure state-reconstruction problem hard? In *Proc. of the 58th IEEE conference on decision and control* (pp. 5368–5373). IEEE.

Mitra, Aritra, & Sundaram, Shreyas (2016). Secure distributed observers for a class of linear time invariant systems in the presence of Byzantine adversaries. In *Proc. of the 55th IEEE conference on decision and control* (pp. 2709–2714).

Mitra, Aritra, & Sundaram, Shreyas (2018). Distributed observers for LTI systems. *IEEE Transactions on Automatic Control*, *63*(11), 3689–3704.

Mitra, Aritra, & Sundaram, Shreyas (2019). Byzantine-resilient distributed observers for LTI systems. *Automatica*, *108*, Article 108487.

Natarajan, Balas K. (1995). Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, *24*(2), 227–234.

Pasqualetti, Fabio, Dörfler, Florian, & Bullo, Francesco (2012). Cyber-physical security via geometric control: Distributed monitoring and malicious attacks. In *Proc. of the 51st IEEE conference on decision and control* (pp. 3418–3425).

Pasqualetti, Fabio, Dörfler, Florian, & Bullo, Francesco (2013). Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, *58*(11), 2715–2729.

Sandberg, Henrik, Teixeira, André, & Johansson, Karl H. (2010). On security indices for state estimators in power networks. In *First workshop on secure control systems*.

Shoukry, Yasser, Nuzzo, Pierluigi, Sangiovanni-Vincentelli, Alberto L, Seshia, Sanjit A, Pappas, George J, & Tabuada, Paulo (2018). Smc: Satisfiability modulo convex programming. *Proceedings of the IEEE*, *106*(9), 1655–1679.

Shoukry, Yasser, & Tabuada, Paulo (2015). Event-triggered state observers for sparse sensor noise/attacks. *IEEE Transactions on Automatic Control*, *61*(8), 2079–2091.

Smith, Roy S. (2011). A decoupled feedback structure for covertly appropriating networked control systems. *IFAC Proceedings Volumes*, *44*(1), 90–95.

Smith, Roy S. (2015). Covert misappropriation of networked control systems: Presenting a feedback structure. *IEEE Control Systems Magazine*, *35*(1), 82–92.

(2017). *Special issue on secure control of cyber physical systems, vol. 4.*

Su, Lili, & Shahrampour, Shahin (2019). Finite-time guarantees for Byzantine-resilient distributed state estimation with noisy measurements. *IEEE Transactions on Automatic Control*.

Sundaram, Shreyas, & Hadjicostis, Christoforos N. (2010). Distributed function calculation via linear iterative strategies in the presence of malicious agents. *IEEE Transactions on Automatic Control*, *56*(7), 1495–1508.

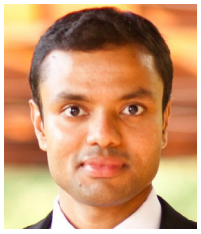
Tiwari, Ashish, Dutertre, Bruno, Jovanović, Dejan, de Candia, Thomas, Lincoln, Patrick D, Rushby, John, et al. (2014). Safety envelope for security. In *Proc. of the 3rd international conference on high confidence networked systems* (pp. 85–94). ACM.

Tseng, Lewis, & Vaidya, Nitin H. (2015). Fault-tolerant consensus in directed graphs. In *Proc. of the 2015 ACM symposium on principles of distributed computing* (pp. 451–460).

Yong, Sze Zheng, Foo, Ming Qing, & Frazzoli, Emilio (2016). Robust and resilient estimation for cyber-physical systems under adversarial attacks. In *Proc. of the American control conference* (pp. 308–315). IEEE.

Yanwen Mao is a Ph.D. candidate in the Department of Electrical and Computer engineering, University of California, Los Angeles. He received the M.S. degree from University of California, Los Angeles, in 2019, and the B.E. degree from Shanghai Jiao Tong University in 2017. His current research interests include decentralized optimization and robust networked control systems. He was a recipient of the Ultra High Voltage Scholarship at Shanghai Jiao Tong University.

Aritra Mitra is a Postdoctoral Researcher in the Department of Electrical and Systems Engineering, University of Pennsylvania. He received the Ph.D. degree from Purdue University, USA, the M.Tech. degree from the Indian Institute of Technology Kanpur, India, and the B.E. degree from Jadavpur University, Kolkata, India, in 2020, 2015, and 2013, respectively, all in Electrical Engineering. His current research interests include distributed learning and optimization, statistical signal processing, networked control systems, and secure control. He was a recipient of the University Gold Medal at Jadavpur University and the Academic Excellence Award at IIT Kanpur.



Shreyas Sundaram is the Marie Gordon Associate Professor in the Elmore Family School of Electrical and Computer Engineering at Purdue University. He received his MS and Ph.D. degrees in Electrical Engineering from the University of Illinois at Urbana-Champaign in 2005 and 2009, respectively. He was a Postdoctoral Researcher at the University of Pennsylvania from 2009 to 2010, and an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Waterloo from 2010 to 2014. He is a recipient of the NSF CAREER award, and an Air Force Research Lab Summer Faculty Fellowship. At Purdue, he received the Hesselberth Award for Teaching Excellence, the Ruth and Joel Spira Outstanding Teacher Award, the Outstanding Mentor of Engineering Graduate Students Award, and the HKN Outstanding Professor Award. At Waterloo, he received the Department of Electrical and Computer Engineering Research Award and the Faculty of Engineering Distinguished Performance Award. His research interests include resilient multi-agent systems, network science, analysis of large-scale dynamical systems, fault-tolerant and secure control, linear system and estimation theory, and game theory.

Paulo Tabuada was born in Lisbon, Portugal, one year after the Carnation Revolution. He received his "Licenciatura" degree in Aerospace Engineering from Instituto Superior Tecnico, Lisbon, Portugal in 1998 and his Ph.D. degree in Electrical and Computer Engineering in 2002 from the Institute for Systems and Robotics, a private research institute associated with Instituto Superior Tecnico. Between January 2002 and July 2003 he was a postdoctoral researcher at the University of Pennsylvania. After spending three years at the University of Notre Dame, as an Assistant Professor, he joined the Electrical and Computer Engineering Department at the University of California, Los Angeles, where he currently is the Vijay K. Dhir Professor of Engineering.

Paulo Tabuada's contributions to control and cyber–physical systems have been recognized by multiple awards including the NSF CAREER award in 2005, the Donald P. Eckman award in 2009, the George S. Axelby award in 2011, the Antonio Ruberti Prize in 2015, the grade of fellow awarded by IEEE in 2017 and by IFAC in 2019. He has been program chair and general chair for several conferences in the areas of control and of cyber–physical systems such as NecSys, HSCC, and ICCPS. He currently serves as the chair of HSCC's steering committee and served on the editorial board of the IEEE Embedded Systems Letters and the IEEE Transactions on Automatic Control.