Complexity of constrained sensor placement problems for optimal observability ⋆

Priyanka Dey [a], Niranjan Balachandran [b], Debasish Chatterjee [a]

[a] *Systems & Control Engineering, Indian Institute of Technology Bombay, Powai, Mumbai - 400076, India*

[b] *Department of Mathematics, Indian Institute of Technology Bombay, Powai, Mumbai - 400076, India*

Abstract

This article studies two problems related to observability and efficient constrained sensor placement in linear time-invariant discrete-time systems with partial state observations: (i) We impose the condition that both the set of outputs and the state that each output can measure are pre-specified. We establish that for any fixed $k > 2$, the problem of placing the minimum number of sensors/outputs required to ensure that the structural observability index is at most k, is NP-complete. Conversely, we identify a subclass of systems whose structures are directed trees with self-loops at every state vertex, for which the problem can be solved in linear time. (ii) Assuming that the set of states that each given output can measure is given, we prove that the problem of selecting a pre-assigned number of outputs in order to maximize the number of states of the system that are structurally observable (i.e., to maximize the size of the observable subgraph) is also NP-hard. As an application, we identify suitable conditions on the system structure under which there exists an efficient greedy strategy, which we provide, to obtain a $(1 - \frac{1}{e})$-approximate solution. An illustration of the techniques developed for this problem is given on the benchmark IEEE 118-bus power network containing roughly 400 states in its linearized model.

Key words: structural observability, graph theory, matching, submodular functions.

1 Introduction

The ever-increasing demand for low-cost control and quick reconstruction of past states from the observations for large-scale systems has brought to the foreground the problem of identifying a subset of the states with fewest elements that are required to "efficiently" control or observe these systems assuming exact knowledge of the system parameters. This problem may look deceptively simple, but it is a computationally difficult one. Indeed, [23] proves that finding the smallest number of actuators (resp. sensors) to make a linear system controllable (resp. observable) is NP-hard.

Due to the sheer size and the ubiquitous modeling uncertainties of these systems, it is difficult to accurately survey the system parameters that govern their dynamics. Moreover, the parameters are prone to drift over time due to ageing, structural alterations, etc. Therefore, it is important to control large-scale systems with the knowledge of only the interconnections among the various states of the dynamical system. This is still possible by using tools from structural systems theory that relies on the zero-nonzero pattern of the system matrices of a linear system, providing a fundamental bedrock on which conventional control theory may be enabled. A survey on various optimization problems studied via structural system theory can be found in [17]. In view of structural observabilty, a collection of interesting problems have been recently addressed in [5], [8], [9] and the references therein. These works mainly focus on optimal placement of sensors to ensure structural observability via efficient polynomial time algorithms and classification of sensors according to the influence of their failure on observability.

Since a significant fraction of real world systems admit only partial state observations, one of the central problems in systems theory is the efficient recovery of the actual system states from the observations. Moreover, for many networks it is of practical importance to recover the states quickly and efficiently within a small time window. Thus, it is crucial to understand *how quickly* the states can be recovered from the observations of a discrete-time linear system. The *observability index* characterises this speed of recovery by determining the minimum number of iterations required to fully reconstruct the states of a discrete-time linear system. Due to inevitable system uncertainties, we focus on the structural counterpart of the observability index, namely, the structural observability index. We address the following problem:

> *Minimal Sensor Placement Problem:* Determining the minimal number and placement of sensors/outputs required to guarantee a desired bound on the structural observability index when the given output matrix has a rectangular diagonal structure, i.e., the state to which each output is directly connected is pre-specified. [1]

Let d be the number of states in the linear system. Corre-

⋆ Corresponding author Priyanka Dey.

Email addresses: dey_priyanka@sc.iitb.ac.in (Priyanka Dey), niranj@math.iitb.ac.in (Niranjan Balachandran), dchatter@iitb.ac.in (Debasish Chatterjee).

[1] Note that the given output matrix can also be a diagonal structured matrix.

sponding to this problem, we have the following results:

- We provide an algorithm with run time complexity $O(d^3 \log d)$ to determine a solution of the minimal sensor placement problem when the desired bound on the structural observability index is equal to 2.
- We prove that, in the general set up, the minimal sensor placement problem is NP-hard whenever the desired bound on the structural observability index is at least 3, thereby illustrating a sharp transition in the hardness of the problem as the bound changes from 2 to 3.

Since the general problem turn out to be difficult, we identify conditions under which the minimal sensor placement problem is polynomially solvable. We consider a practically relevant special class of systems whose structure is a directed tree with a self-loop at every state. [2] We establish the following result for this subclass:

- The minimal sensor placement problem is solvable in polynomial time when the underlying graph structure of the system is a directed tree with a self-loop at every state and we give an $O(d)$ algorithm to solve it.

We move to the second problem addressed in this article:

Cardinality Constrained Sensor Placement Problem: Identifying a pre-assigned number of outputs from the given set of outputs so that maximum number of states are *structurally observable by them* [3] in the system when the set of states that each output can measure is pre-specified.

It becomes extremely relevant when the permissible number of sensors/outputs may not be adequate to observe the entire system/network. Therefore, a design strategy to select the outputs in such a way that as much of the network as possible is observable is needed. Corresponding to this problem, we have the following result:

- We establish that the cardinality constrained sensor placement problem is NP-hard. We observe that the problem remains NP-hard even if we impose a mild condition on the system where the digraph associated with the state matrix A is such that each state vertex has a self-loop. [4]

We confine ourselves to this special class where each state vertex has a self-loop since a wide class of systems exhibits this self-damped dynamics [2], for example, epidemic spread in networks [22], ecosystems [18], power grids [12], and even social networks [24]. For this class of systems, we give the following result:

- We provide a greedy algorithm to obtain an $(1 - 1/e)$-approximate solution for the cardinality constrained sensor placement problem. This is the *best possible result* that can be obtained via greedy algorithms and at the level of generality considered here.

The rest of this article unfolds as follows. §2 discusses a few existing results and related work in this area. §3 reviews certain useful concepts and results. The precise problem statements of the minimal sensor placement problem and the cardinality constrained sensor placement problem, and our corresponding main results are given in §4 and §5 respectively. We conclude a summary of our results along with possible future directions in §6.

2 Related work

The definition of structural observability index was introduced in [19], and a few methods required for its computation were proposed in [31]. By employing graph-theoretic techniques bounds on the (controllability and) observability index for structured linear systems were provided in [33]. [27] considered the problem of identifying the minimum number of states to be connected to distinct inputs (resp. outputs) to ensure a given bound on the structural controllability (resp. observability) index, and established that the problem is NP-hard. In addition, the trade-off between the structural controllability index and the minimum number of states that need to be actuated was explored on a variety of artificial and synthetic networks by using a heuristic algorithm and it was observed that the number of actuated states obtained is close to optimal. The problem we address is a generalization of the problem considered in [27]: the selection of states to be measured by distinct outputs is *constrained* to a specific preassigned family of states. This restriction makes the problem more realistic and increases the level of its difficulty; see Remark 2 for further technical details. In addition, we identify a practically relevant subclass of systems for which the minimal sensor placement problem is *optimally* solvable, viz., systems whose structures are directed trees with a self-loop at each vertex.

In control theory, several problems involving selection of a pre-specified/minimum number of states (or inputs/outputs) to optimize a certain objective function have been studied via submodularity tools and can be found in [3]. In particular, the problem of identifying a pre-assigned number of states to be actuated in order to optimize some of the energy metrics was investigated in [32] via the notion of submodularity. The problem of selecting the minimum number of sensors to optimize the Kalman filter with respect to the estimation error was studied employing submodularity in [34]. In the context of our problem, a relaxed version of the cardinality constrained placement problem (in the controllability framework) where each input (resp. output) is directly connected to only one state was treated in [4] via submodularity when the digraph associated with the state matrix A is strongly connected. [5] However, they do not comment on the complexity of this problem. In contrast to [4], we address this problem in a different subclass where it is difficult

[2] See §3 for a formal definition of the directed tree. This particular structure plays an important role for a large class of systems including leader-follower networks [26], [14], biological networks [1], transportation systems [36], [35], etc. Furthermore, in the multi-agent community, several well-known strong results [29], [30] rely on the existence of spanning trees in the network, thereby conforming to this category of structural hypotheses.

[3] See §5 for a formal definition.

[4] The systems satisfying this mild condition that every state vertex has a self-loop are often referred as *self-damped systems* in the literature.

[5] A digraph $G = (V, E)$ is *strongly connected* if for each ordered pair of vertices (x_i, x_j), G has a directed path from x_i to x_j.

to solve and there is no restriction on the given outputs to measure only one state.

3 Preliminaries

The set of real numbers, non-negative integers, and positive integers are denoted here by $\mathbb{R}$, $\mathbb{N}$, and $\mathbb{N}^\star$ respectively. Let $[r] \coloneqq \{1, 2, \ldots, r\}$ for each $r \in \mathbb{N}^\star$. The cardinality of set X is denoted by $|X|$. For a matrix A of appropriate dimension, A_{ij} or $[A]_{ij}$ represents the (i, j)-entry of this matrix.

Consider a linear time-invariant system

$$x(t+1) = \bar{A}x(t),\ \ y(t) = \bar{C}x(t), \quad t \in \mathbb{N}, \tag{1}$$

where $x(t) \in \mathbb{R}^d$ and $y(t) \in \mathbb{R}^p$ are the state and output vectors at time t. The state and output matrices are given by $\bar{A} \in \mathbb{R}^{d\times d}$ and $\bar{C} \in \mathbb{R}^{p\times d}$ respectively. In this article, the system (1) is sometimes described by the pair $(\bar{A}, \bar{C})$. In our analysis only the information about the locations of the fixed zeros in $\bar{A}$ and $\bar{C}$ is crucial and the precise numerical values of the non-entries of $\bar{A}$ and $\bar{C}$ are not relevant. For any matrix H, its *sparsity matrix* is defined as a matrix of the same dimension as H with either a zero or an independent free parameter (denoted by $*$) at each entry depending on whether the corresponding entry in H is zero or not. A *numerical realisation* of a sparsity matrix is obtained by giving numerical values to its $*$ entries. Let the sparsity matrices of the state and the output matrices in (1) are represented by $A \in \{0, *\}^{d\times d}$ and $C \in \{0, *\}^{p\times d}$, and let $[A]$ and $[C]$ be the collection of all numerical matrices of the same dimension and structure/sparsity as $A \in \{0, *\}^{d\times d}$ and $C \in \{0, *\}^{p\times d}$ respectively. We say that a pair (A, C) is *structurally observable* if there exists at least one observable numerical realization of (A, C). [6]

A linear time-invariant system (1) is associated with a digraph $G(A, C)$ by using the following natural way: Let $\mathcal{A} = \{x_1, x_2, \ldots, x_d\}$ and $\mathcal{C} = \{1, 2, \ldots, p\}$ be the state vertices and the output vertices corresponding to the states $x(t) \in \mathbb{R}^d$ and the outputs $y(t) \in \mathbb{R}^p$ of the system (1). Let $E_A = \{(x_j, x_i) \,|\, A_{ij} \neq 0\}$ and $E_C = \{(x_j, i) \,|\, C_{ij} \neq 0\}$. The digraph $G(A, C) = (\mathcal{A} \sqcup \mathcal{C}, E_A \sqcup E_C)$, and $\sqcup$ denotes the disjoint union. The sets E_A and E_C denote the edges between the state vertices, and the edges from the state vertices to the output vertices in the digraph $G(A, C)$ respectively. Similarly, we can define digraph $G(A) = (\mathcal{A}, E_A)$ with vertex set $\mathcal{A}$ and edge set E_A. Given $G(A, C)$, the *induced subgraph* by $U \subset \mathcal{A} \sqcup \mathcal{C}$ is a digraph consisting of vertex set U and all those edges of the digraph $G(A, C)$ with both end points in U. In particular, $G(A)$ is the induced subgraph of $G(A, C)$ by $\mathcal{A}$.

A sequence of edges $\{(x_1, x_2), (x_2, x_3), \ldots, (x_{k-1}, x_k)\}$, where each $x_i \in \mathcal{A}$ is distinct and $(x_i, x_{i+1}) \in E_A$ for $i = 1, 2, \ldots, k-1$, is called a *directed path* from x_1 to x_k in $G(A)$. A *cycle* is a directed path where the initial vertex x_1 coincides with the end vertex x_k. A digraph is acyclic if it contains no cycles. A digraph is a *directed tree towards* x if it is an acyclic graph where every vertex has a directed path towards x and every vertex except x has out-degree exactly equal to 1. Sometimes we refer to this digraph as just directed tree. The vertices with no incoming edges are termed as the *leaves* of the tree. The digraph $G(A, C)$ is said to have a *spanning forest topped* at output vertices C if it has a disjoint union of set of directed trees, where each tree is directed towards a vertex in C and this union contains all the state vertices.

Definition 1 *For the digraph $G(A, C) = (\mathcal{A} \sqcup C, E_A \sqcup E_C)$ associated with system* (1) *and a subset $S \subset \mathcal{A}$, the* out-neighbourhood *of S is the set $N^+(S) = \{v \,\big|\, (x_i, v) \in E_A \sqcup E_C, x_i \in S, v \in \mathcal{A} \sqcup C\}$. The digraph $G(A, C)$ is said to have a* contraction *if there exists a set $S \subset \mathcal{A}$ with $|N^+(S)| < |S|$.*

The following subgraphs associated with digraph $G(A)$ and $G(A, C)$ are defined in [25].

- State stem is a directed path, consists of only state vertices. An isolated state vertex is also considered as state stem. [7]
- Output Stem is a directed path obtained by connecting a directed edge from the tip of a state stem to an output vertex.
- An Output Cactus, defined recursively as follows: An output stem with at least one state vertex is an output cactus. An output cactus connected by a directed edge from a state vertex of a disjoint cycle (comprising of only state vertices) to either any state vertex or the output vertex of the cactus is also an output cactus.

The relation between certain properties of $G(A, C)$ and the structural observability of the pair (A, C) is given by:

Theorem 1 *[16][8] The following are equivalent:*

(a) *The pair (A, C) is structurally observable.*
(b) *In the digraph $G(A, C)$ derived from* (1)*, every state vertex x_i has a directed path from it to at least one of the output vertices, and $G(A, C)$ is free of contractions.*
(c) *The digraph $G(A, C)$ is spanned by a disjoint union of output cacti.*

3.1 Structural observability index

The *observability index* $\mu(\bar{A}, \bar{C})$ of (1) is [8]

$$\mu(\bar{A}, \bar{C}) \coloneqq \inf\left\{k \in [d] \,\middle|\, \operatorname{rank}(\bar{C}^\top\ (\bar{C}\bar{A})^\top\ \cdots\ (\bar{C}\bar{A}^{k-1})^\top)^\top = d\right\}.$$

In other words, $\mu(\bar{A}, \bar{C})$ is the minimum number of iterations required to recover/determine uniquely the initial state x_0 from $y_0, y_1, \ldots$. In other words, x_0 may be obtained by left-inversion in the linear equation

$$\begin{pmatrix} \bar{C} \\ \bar{C}\bar{A} \\ \vdots \\ \bar{C}\bar{A}^{\mu(\bar{A},\bar{C})-1} \end{pmatrix} x_0 = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{\mu(\bar{A},\bar{C})-1} \end{pmatrix}.$$

The k-step observability matrix associated with the pair $(\bar{A}, \bar{C})$ is given by $O_k(\bar{A}, \bar{C}) \coloneqq (\bar{C}^\top\ (\bar{C}\bar{A})^\top\ \cdots\ (\bar{C}\bar{A}^{k-1})^\top)^\top$ The structural counterpart of the observability index, namely, *structural observability index*, is defined as

$$\mu(A, C) \coloneqq \inf\left\{k \in [d] \,\middle|\, \sup_{\substack{A_1\in[A]\\ C_1\in[C]}} \operatorname{rank}(O_k(A_1, C_1)) = d\right\}. \tag{2}$$

[6] It is known that if one realization of (A, C) is observable, then *almost all* numerical realizations of (A, C) are observable; see [16].

[7] The tip of a state stem is a state vertex that does not have any outgoing edges from it to any other state vertex in that stem.

[8] The convention that $\inf \emptyset = +\infty$ is assumed to be in place.

The value of the infimum is $+\infty$ when none of the pairs in $([A],[C])$ is observable, i.e., the pair (A,C) is structurally unobservable. If the pair (A,C) has structural observability index $\mu(A,C)=\ell$, where $\ell \leq d$ is some positive integer, then almost all numerical realisations of pair (A,C) have observability index ℓ off a manifold with zero Lebesgue measure [28, p. 44]. The following result provides a graph theoretic interpretation of the structural observability index of a pair (A,C).

Theorem 2 *[27, Theorem 2] A pair (A,C) is structurally observable with index ℓ if and only if the digraph $G(A,C)$ is spanned by a disjoint union of output cacti, where every output cactus contains at most ℓ state vertices.*

The graph-theoretic notion of structural observability index is demonstrated via a digraph $G(A,C)$ shown in Fig. 1. Given $A \in \{0,*\}^{d\times d}$ and $C \in \{0,*\}^{p\times d}$, there exists several possible disjoint unions of output cacti spanning the digraph $G(A,C)$.

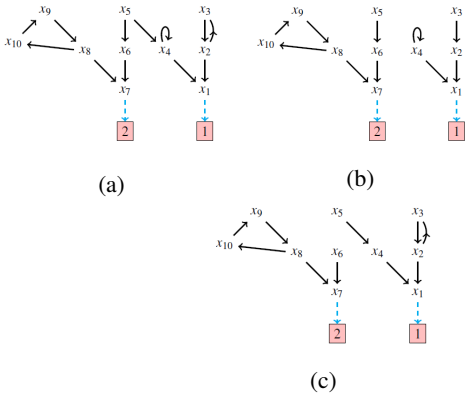

Fig. 1. Illustration of a digraph $G(A,C)$ in (a) and two possible output cacti in (b) and (c). There are four state vertices (b) in one output cactus and six in the other, whereas (c) contains one output cactus with five state vertices and the other with five. Moreover, these two are the only spanning cacti of $G(A,C)$, so the structural observability index is five.

Lemma 1 *[27, Corollary 1] Let $A \in \{0,*\}^{d\times d}$ and $C \in \{0,*\}^{p\times d}$, with $A_{ii}=*$ for all $i=1,2,\ldots,d$. Let $\{\mathcal{H}_i\}_{i\in\mathcal{K}}$ be the collection of all spanning forests of $G(A,C)$ containing only directed trees towards the output vertices, where $\mathcal{K}$ contains the indices of such spanning forests. Let $z_i \in \mathbb{N}^\star$ be the number of directed trees in $\mathcal{H}_i$, i.e., $\mathcal{H}_i=\{\mathcal{T}_j^i\}_{j=1}^{z_i}$. Then, the structural observability index of the pair (A,C) is*

$$\mu(A,C) := \min_{i\in\mathcal{K}} \max_{\mathcal{T}\in\mathcal{H}_i} |\mathcal{T}|_s\,,$$

where $|\mathcal{T}|_s$ is the number of state vertices in a tree $\mathcal{T}$ towards an output vertex.

3.2 Submodular functions

Let V be a non-empty finite set, and let 2^V denote the collection of all subsets of V.

Definition 2 *A function $f: 2^V \to \mathbb{R}$ is* submodular *if for all $S \subset T \subset V$ and $v \in V \setminus T$ we have $f(S\cup\{v\})-f(S) \geq f(T\cup\{v\})-f(T)$.*

Definition 3 *A function $f: 2^V \to \mathbb{R}$ is* monotone non-decreasing *if for every $S \subset T \subset V$ we have $f(S) \leq f(T)$.*

When a function f is submodular, one can use greedy algorithms that yield, in reasonable time [21], approximate solutions that are often very close to an optimal solution.

4 Minimal sensor placement problem

Before stating the precise problem statement, we define the identity structured output matrix $C=I_d \in \{0,*\}^{d\times d}$ as follows: $[I_d]_{ij}=*$ if $i=j$ and 0 otherwise, where $i,j\in[d]$. Recall that $\mathcal{A}=\{x_1,x_2,\ldots,x_d\}$. An output matrix $C=I_F \in \{0,*\}^{|F|\times d}$ is obtained from I_d, as defined above, by retaining the rows corresponding to the state vertices in $F\subset\mathcal{A}$.

Given $A\in\{0,*\}^{d\times d}$, $C=I_F\in\{0,*\}^{|F|\times d}$, and $\ell\in[d]$ such that $\mu(A,I_F)\leq\ell$, find J^* that solves

$$\begin{cases} \underset{J\subset F}{\text{minimize}} & |J| \\ \text{subject to} & \mu(A,I_J)\leq\ell, \end{cases} \qquad (\mathcal{P}_1)$$

where $F\subset\mathcal{A}$, I_J is the submatrix of the output matrix I_F obtained by retaining the rows corresponding to state vertices in J. $(\mathcal{P}_1)$ determines the minimal subset of states required to admit outputs from the set of available states F to ensure a desired bound on structural observability index.

A special instance of $(\mathcal{P}_1)$ corresponds to the case where the given output matrix $C=I_d\in\{0,*\}^{d\times d}$, i.e., $F=\mathcal{A}$ and we refer to this problem as $(\mathcal{P}_1')$ in this sequel.

Next we show that $(\mathcal{P}_1)$ (and hence $(\mathcal{P}_1')$) can be solved optimally when the bound $\ell\in\{1,2\}$. Clearly, solving $(\mathcal{P}_1)$ for $\ell=1$ is trivial since it requires each state vertex to be directly connected to a distinct output, i.e., the solution is the output matrix C of identity structure I_d (as defined above).

The concept of a matching is needed to solve $(\mathcal{P}_1)$ for $\ell=2$. Recall that for an undirected graph $G=(V,E)$, a *matching* is a set of edges with no shared endpoints. The vertices incident to the edges of a matching M are said to be *saturated* by M; the other are *unsaturated*. A *maximum matching* is one that has the largest cardinality among all possible matchings in a graph. We define a weight function $w: E\to\mathbb{R}$ that assigns weights to the edges of the graph G. Subsequently, we introduce the *maximum weight matching* problem, concerned with finding a matching of G that has the maximum weight-sum of its edges; in other words, determining a matching M^* such that $\sum_{e\in M^*} w(e) \geq \sum_{e\in\bar{M}} w(e)$ for any matching $\bar{M}$ in G. We solve $(\mathcal{P}_1)$ for $\ell=2$ by employing Algorithm 1.
In Step 1, we construct an undirected graph G_u from $G(A)$ with vertex set $\mathcal{A}$ and edge set $E_u=E_F\sqcup E_{\mathcal{A}\setminus F}$. An edge exists between a pair of vertices x_i, x_j in G_u if any one of the conditions is satisfied: (i) both the vertices lie in F and either $(x_i,x_j)\in E_A$ or $(x_j,x_i)\in E_A$; (ii) one vertex say x_i lies in $\mathcal{A}\setminus F$ and the other one x_j is in F and a directed edge exists from x_i to x_j in $G(A)$. We assign weights to the edges of E_u and compute a maximum weight matching M in G_u in Steps 2 and 3 respectively. In Step 5, we form a set J^* by collecting the vertices unsaturated by M and by selecting one vertex from each undirected edge in M in the following way: for $e=(x_i,x_j)\in M$, if $(x_i,x_j)\in E_A$ then $x_j\in J^*$; if $(x_j,x_i)\in E_A$ then $x_i\in J^*$; if both (x_i,x_j) and (x_j,x_i) are in E_A then any one among x_i or x_j is collected in J^*. The construction of the undirected graph G_u from $G(A)$ requires $O(d^2)$ computations. A maximum weight matching

Algorithm 1: Algorithm to solve $(\mathcal{P}_1)$ for $\ell = 2$.

Input: $G(A) = (\mathcal{A}, E_A)$, $F \subset \mathcal{A}$ such that $\mu(A, I_F) \leq 2$.
Output: A solution of $(\mathcal{P}_1)$, $J^* \subset F$.

1 Construct an undirected graph $G_u = (V, E_u)$ from $G(A)$ with $V = \mathcal{A}$ and $E_u = E_F \sqcup E_{\mathcal{A}\setminus F}$. $(x_i, x_j) \in E_F$, if $x_i, x_j \in F$ and either $(x_i, x_j) \in E_A$ or $(x_j, x_i) \in E_A$. $(x_k, x_m) \in E_{\mathcal{A}\setminus F}$, if $x_k \in \mathcal{A} \setminus F$, $x_m \in F$ and $(x_k, x_m) \in E_A$. Neglect self-loops, if present.

2 Define the weight w for the edges of G_u:

$$w(e) \leftarrow \begin{cases} 1 & \text{for } e \in E_F, \\ |E_u| + 1 & \text{for } e \in E_{\mathcal{A}\setminus F}. \end{cases} \tag{3}$$

3 Determine a maximum weight matching M in G_u under weight w.

4 For each edge $e = (x_i, x_j) \in M$, the direction of the edge e is chosen in accordance to the direction of e in $G(A)$.

5 If $e = (x_i, x_j) \in M$ with $(x_i, x_j) \in E_A$ then $x_j \in J^*$ and if x_k is unsaturated by M then $x_k \in J^*$.

is computed in $O(d^3 \log d)$ computations [13, Chapter 11] and the rest of the steps have linear complexity. Therefore, the overall complexity of Algorithm 1 is $O(d^3 \log d)$.

Lemma 2 *Let $A \in \{0, *\}^{d\times d}$, $C{=}I_F \in \{0, *\}^{|F|\times d}$, and $G(A) = (\mathcal{A}, E_A)$ be the state matrix, the output matrix, and the digraph associated with A and $d \geq 2$. Suppose $\mathcal{A} \setminus F \neq \emptyset$ and G_u be the undirected graph associated with $G(A)$ obtained via Algorithm 1. Then J^* obtained from Algorithm 1 solves $(\mathcal{P}_1)$ with $\ell = 2$.*

PROOF. Assume, as hypothesized that $\mu(A, I_F) \leq 2$. By Theorem 2, the digraph $G(A, I_F)$ is spanned by a disjoint union of output cacti such that each cactus has at most 2 state vertices. In fact, each cactus is an output stem with at most 2 state vertices. Thus, the condition $\mu(A, I_F) \leq 2$ and the construction of G_u in Algorithm 1 confirms that each $x_i \in \mathcal{A} \setminus F$ can be associated with a distinct $x_j \in F$ such that $(x_i, x_j) \in E_{\mathcal{A}\setminus F}$. Let M be a maximum weight matching in G_u under the weight function w defined in (3). The weight structure (Step 2) of Algorithm 1 guarantees that every vertex $x_i \in \mathcal{A} \setminus F$ is saturated by an edge $e = (x_i, x_j) \in M$ for some $x_j \in J^*$. The set of vertices unsaturated by M lies in F and belongs to J^* (Step 5). The output J^* of Algorithm 1 ensures that we obtain a disjoint union of output stems covering all state vertices and each stem has at most 2 state vertices. Thus, $\mu(A, I_{J^*}) \leq 2$.

Let $|\mathcal{A} \setminus F| = r$. Suppose that there exists $J' \subset F$ such that $|J'| < |J^*|$ and $\mu(A, I_{J'}) \leq 2$. Since each output is directly connected to a distinct state in J' in $G(A, I_{J'})$, we can obtain a disjoint union of cacti or output stems P spanning $G(A, I_{J'})$ such that each output cactus has at most 2 state vertices with each output being directly connected to the tip of a state stem in P. Note that each output cactus in P can be one of the following three types: 1) An output stem P_i consisting of two state vertices connected by an directed edge of the form (x_a, x_b), where $x_a \in \mathcal{A} \setminus F$ and $x_b \in J'$; 2) An output stem P_k comprising of two state vertices connected by a directed edge of the form (x_c, x_d), where $x_c \in F$ and $x_d \in J'$; 3) An output stem P_j consisting of only one state vertex $x_e \in J'$. Accordingly, P is the union of $\{P_i\}_{i=1}^{r} \cup \{P_k\}_{k=r+1}^{n} \cup \{P_j\}_{j=n+1}^{|J'|}$ and $n \leq |J'|$. Here each P_i is associated with an edge $e_i \in E_{\mathcal{A}\setminus F}$ of G_u, where $1 \leq i \leq r$. Similarly, P_k is associated with an edge $e_k \in E_F$ of G_u. Observe that the collection of edges $\{e_i\}_{i=1}^{r} \cup \{e_k\}_{k=r+1}^{n}$ refers to a matching M' in the undirected graph G_u. Clearly, both M and M' saturates all the vertices in $\mathcal{A} \setminus F$. The condition $|J'| < |J^*|$ implies that the edges e_k of the form (x_c, x_d), where $x_c, x_d \in F$, must be greater in number in M' than in M. Since the edges in $\{e_k\}_{k=r+1}^{n}$ have unit weights each $w(M') > w(M)$. This gives a contradiction and completes the proof. □

Remark 1 *Since $(\mathcal{P}'_1)$ is a special case of $(\mathcal{P}_1)$, a solution of $(\mathcal{P}'_1)$ can be obtained by using Algorithm 1 by setting $F = \mathcal{A}$. Indeed, it follows that the edge set $E_u = E_F$ since $E_{\mathcal{A}\setminus F}$ is empty, and since unit weight is assigned to each edge in E_F, determining a maximum weight matching in Step 3 reduces to finding a maximum matching M in G_u. Therefore, the set of state vertices J^* (obtained in Step 5 of Algorithm 1) such that $\mu(A, I_{J^*}) \leq 2$, satisfies $|J^*| = |M| + (d - 2|M|) = d - |M|$, where M is a maximum matching in G_u and $(d - 2|M|)$ is the number of state vertices unsaturated by M in G_u. The proof of optimality of the obtained solution J^* is omitted due to similarity with the proof of Lemma 2.*

We shall now discuss the complexity of $(\mathcal{P}'_1)$ and show that it is hard whenever the bound on the structural observability index $\ell \geq 3$. The decision version of $(\mathcal{P}'_1)$ is given by:

Instance: $A \in \{0, *\}^{d\times d}, C = I_d \in \{0, *\}^{d\times d}$, and two positive integers $\ell \in [d]$ and $K \leq d$.
Question: Does there exists a set $J \subset \mathcal{A}$ with $|J| \leq K$ such that $\mu(A, I_J) \leq \ell$.

We use the following NP-complete problem to show the hardness of $(\mathcal{P}'_1)$ whose decision version is given by:

Instance: An undirected graph $G = (V, E)$, weight $w(v) \in \mathbb{N}$ for each $v \in V$, positive integers ℓ and $K \leq |V|$.
Question: Can the vertices in V be partitioned into $k \leq K$ disjoint sets $V_1, V_2, \ldots, V_k$ such that, for $1 \leq i \leq k$, the induced subgraph of G by V_i is connected and the sum of the weights of the vertices in V_i does not exceed ℓ?

The preceding problem is the *Bounded Component Spanning Forest (BCSF)* problem, and it remains NP-complete even if the weights of all the vertices in G is equal to 1 and ℓ is any fixed integer larger than 2 [10].

Theorem 3 *$(\mathcal{P}'_1)$ is NP-complete if the desired bound on the structural observability index is any fixed integer greater than or equal to 3.*

PROOF. Given A and a positive number K, for any $J \subset \mathcal{A}$, it can be verified whether $|J| \leq K$ and $\mu(A, I_J) \leq \ell$ in polynomial time [31]. Therefore, the decision version of $(\mathcal{P}'_1)$ is in NP.

Let $G = (V, E)$ be an undirected graph with $|V| = r$. We define a weight function $w : V \to \mathbb{N}$ such that $w(v) = 1$ for all $v \in V$. For each connected component induced by V_i in G, $\sum_{v_j \in V_i} w(v_j) = |V_i|$ for $1 \leq i \leq k$, i.e., the sum of the weights of the vertices in V_i is equal to the number of vertices in the component induced by V_i. We construct a state matrix

$A \in \{0, *\}^{r \times r}$ as follows: for every vertex $v_i \in V$, we associate a state x_i and with each undirected edge $e_{ij} = (v_j, v_i)$ we associate two non-zero entries $A_{ij} = *$ and $A_{ji} = *$. We assume that $A_{ii} = *$ for all $i \in [r]$. Therefore, $G(A)$ has only bidirectional edges, and a self-loop at every vertex. Let C be a matrix with a diagonal structure and dimension r, i.e., $C = I_r \in \{0, *\}^{r \times r}$. Let ℓ be a fixed integer larger than 2.

We will prove the following statement: G has a partition of at most K connected components such that the sum of the weights of the vertices in each component is at most ℓ if and only if there exists a $J \subset \mathcal{A}$ such that $|J| \leq K$ and $\mu(A, I_J) \leq \ell$. Let us suppose that G has a partition of at most K connected components. Let $V_1, V_2, \ldots, V_k$ be the vertex sets of the components of that partition, where $k \leq K$. Since $w(v) = 1$ for all $v \in V$, $|V_i| \leq \ell$ for all $i \in [k]$. Since each component is connected, each has an undirected spanning tree. Select one vertex say v_i from each component V_i of G. Let $J = \bigcup_{i=1}^{k} \{x_i\}$, where x_i is the state vertex associated with v_i selected from the component induced by V_i. Observe that $|J| = k \leq K$. These vertices in J are directly connected to distinct output vertices. Thus, $G(A, I_J)$ has a spanning forest topped at output vertices such that each directed tree has at most ℓ state vertices. Hence, $\mu(A, I_J) \leq \ell$. Conversely, suppose that there exists a $J \subset \mathcal{A}$ such that $|J| \leq K$ and $\mu(A, I_J) \leq \ell$. Then by Lemma 1, $G(A, I_J)$ has a spanning forest topped at output vertices such that the number of state vertices in each directed tree being at most ℓ and each state vertex in J is directly connected to a distinct output. Since edges between distinct vertices in $G(A)$ are bidirectional, each directed tree without the output vertex corresponds to a connected component of G with at most ℓ state vertices. Therefore, we obtain a partition, of cardinality at most K and weight of each component at most ℓ, of the graph G.

The bounded component spanning forest problem remains NP-complete when the weights $w(v) = 1$ for all $v \in V$ and the bound ℓ is fixed in $\{3, 4, \ldots\}$. Under these conditions on the weights and the bound, in the above analysis, we provided a polynomial time reduction showing that the bounded component spanning forest problem has a solution of size at most K with the sum of the weights of the vertices in each component at most ℓ if and only if $(\mathcal{P}'_1)$ has a solution for the constructed instance of cardinality at most K with the bound on the structural observability index being ℓ. Thus, it follows from the above two points that $(\mathcal{P}'_1)$ is NP-complete whenever the bound ℓ on the structural observability index is fixed to a value ≥ 3, thereby completing the proof. □

Specifically, for parameters A and $C = I_d$, a solution to $(\mathcal{P}_1)$ provides, in particular, a solution to $(\mathcal{P}'_1)$. Therefore, $(\mathcal{P}_1)$ is at least as difficult as $(\mathcal{P}'_1)$. Consequently, $(\mathcal{P}_1)$ is NP-complete even if the desired bound on the structural observability index is any fixed integer larger than two.

Remark 2 *As discussed in §2, NP-hardness of $(\mathcal{P}'_1)$ is proved in [27] by reducing from the Graph-Partitioning problem [10]. The proof given in [27] demonstrates that an optimal solution of $(\mathcal{P}'_1)$ leads to a feasible solution to the Graph-Partitioning problem, but optimality of the obtained solution was not addressed there. In contrast, we use the decision versions of $(\mathcal{P}'_1)$ and the BCSF problem to demonstrate the equivalence between the solution of the given instance of the BCSF problem and the solution of $(\mathcal{P}'_1)$ for the constructed instance.*

While subclasses for which one can obtain an optimal solution of $(\mathcal{P}'_1)$ are not considered in [27], it follows from our proof of Theorem 3 that $(\mathcal{P}'_1)$ is NP-hard even when $\ell \geqslant 3$, the digraph $G(A)$ has only bidirectional edges, and every state vertex has a self-loop. This is because the constructed instance of $(\mathcal{P}'_1)$ lies in this subclass. In fact, we can address this subclass via the BCSF problem because an instance of $(\mathcal{P}'_1)$ under this assumption can be reduced to an instance of the BCSF problem in polynomial time. By neglecting the self-loops, we view the digraph $G(A)$ as a weighted undirected graph G_u by setting $w(x_i) = 1$ for all $i \in [d]$. Given an integer ℓ, the output of the BCSF problem on the instance (G_u, ℓ) finds a decomposition of G_u with the minimum number of connected partitions such that each partition has at most ℓ vertices. Then it is easy to see from our proof of Theorem 3 that:

(i) *a minimal partition given by the BCSF problem provides an optimal solution of $(\mathcal{P}'_1)$, and*
(ii) *for any $\alpha \geq 1$, an α-optimal solution of the BCSF problem on the instance (G_u, ℓ) gives an α-optimal solution of $(\mathcal{P}'_1)$ for the given instance.* [9]

Thus, under the above assumption on the digraph $G(A)$, any approximation for the BCSF problem gives an approximation for $(\mathcal{P}'_1)$ with the same approximation ratio. The proof technique employed here, therefore, sheds more light into the problem than the one in [27].

Next we impose the following assumption on the digraph $G(A)$ associated with the state matrix A.

Assumption 1 *We stipulate that the digraph $G(A) = (\mathcal{A}, E_A)$ is a directed tree towards a state vertex $x \in \mathcal{A}$ with a self-loop at every state vertex.*

We have the following Lemma that plays an important role to solve $(\mathcal{P}_1)$ when Assumption 1 holds.

Lemma 3 *Let $A \in \{0, *\}^{d \times d}$ and $C = I_F \in \{0, *\}^{|F| \times d}$. Suppose Assumption 1 holds. Let $\ell \in [d]$ be the desired bound on the structural observability index with $\mu(A, I_F) \leq \ell$. Let $P = \{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_{|J^*|}\}$ be a partition of $G(A) = (\mathcal{A}, E_A)$ into the minimum number of subtrees, where each $\mathcal{T}_i$ represents a subtree such that its tip is a vertex present in $F \subset \mathcal{A}$ and the number of state vertices in no subtree exceeds ℓ. The collection $J^* \subset F$ of tips of each subtree solves $(\mathcal{P}_1)$.* [10]

PROOF. Since $\mu(A, I_F) \leq \ell$, by Lemma 1, there exists a partition or a spanning forest topped at output vertices for $G(A, I_F)$ such that each directed tree is towards an output vertex and has at most ℓ state vertices. Consider the partition P of $G(A)$ such that each subtree $\mathcal{T}_i = (X_i, E_{X_i})$ has at most

[9] Recall that an α-optimal solution is a feasible solution whose value is at most α times the optimal value.

[10] A *tip* of a tree $\mathcal{T}$ is the vertex that does not have any outgoing edges to any other vertex in $\mathcal{T}$ and has a directed path from every vertex to it in $\mathcal{T}$. A *subtree* is a subgraph of the directed tree which satisfies all the properties of tree.

ℓ state vertices, where $X_i \subset \mathcal{A}$, E_{X_i} is the edges in the subtree induced by X_i for $1 \leq i \leq |J^*|$ (excluding self-loops). Let $x_i \in F$ be the tip of the subtree $\mathcal{T}_i$. It follows that $G_i = (X_i \sqcup \{i\}, E_{X_i} \sqcup (x_i, i))$, where $x_i \in J^*$ and $i \in C$ (output set), is a directed tree towards $i \in C$. J^* contains the tip of each subtree. Therefore, the digraph $G(A, I_{J^*})$ has a spanning forest with a collection of directed trees $\{G_i\}_{i=1}^{|J^*|}$ with at most ℓ state vertices in each G_i. Since every state vertex has a self-loop, by Lemma 1 $\mu(A, I_{J^*}) \leq \ell$. The minimality assertion follows by the fact that if $J' \subset F$ has size less than J^*, then there exists another partition of the tree into subtrees of smaller cardinality than P where the tip of each subtree is a vertex lying in F. This leads to a contradiction and confirms that J^* solves $(\mathcal{P}_1)$. □

In the further analysis to solve $(\mathcal{P}_1)$, when Assumption 1 holds, we neglect the self-loops present at every state vertex of $G(A)$. We provide the merging procedure that will be used later in Algorithm 3. Given $F \subset \mathcal{A}$, we use this procedure in Algorithm 2 to transform the given directed tree $G(A)$ into another directed tree $G'(A)$ whose vertex set is F. Given the directed tree $G(A) = (\mathcal{A}, E_A)$ and $F \subset \mathcal{A}$, define a variable h associated with every $x_i \in \mathcal{A} \setminus F$ such that $h(x_i) = x_j$ if $(x_i, x_j) \in E_A$ and $x_i \neq x_j$ (i.e., excluding self-loops).

Algorithm 2: Merging procedure

Input: directed tree $G(A) = (\mathcal{A}, E_A)$, $F \subset \mathcal{A}$, variable h, and weight $w(x_i) = 1$ for all $x_i \in \mathcal{A}$
Output: directed tree $G'(A)$ obtained from $G(A)$ with vertex set as F and weight $w : F \to \mathbb{N}^\star$

1 If $F = \mathcal{A}$
  return $G(A)$ and $w(x_i) = 1$ for all $x_i \in F$
2 else
3 for k = maximum level in $G(A)$ down to 1 **do**
4 begin process kth level
5 **while** there exists $x_i \in \mathcal{A} \setminus F$ in level k **do**
6 if $x_j = h(x_i)$ then
  set $w(x_j) = w(x_j) + w(x_i)$ and merge x_i to x_j
7 end
8 end **while**
9 end process level
10 end for
11 end

In Algorithm 2, we begin with the maximum level (i.e., from the leaves) and go down to the tip of the tree (level 1). Note that when a state vertex $x_i \in \mathcal{A} \setminus F$ is merged with x_j such that $h(x_i) = x_j$ in Step 6, then the edges of the given tree incident to x_i are now incident to x_j, and the digraph gets modified. Since the maximum number of levels is bounded above by d and the procedure comprises of only 'for' and 'while' loops, the overall complexity of Algorithm 2 is $O(d)$.

By Lemma 3 it is clear that to solve $(\mathcal{P}_1)$ we need to find a minimal partition of $G(A)$ into subtrees such that the tip of each subtree lies in F and the number of state vertices in each subtree does not exceed the bound ℓ imposed on the structural observability index. For this purpose, we use the problem of finding a minimal partition P of a tree with positive weight assigned to every vertex in the directed tree such that the sum of the weights of the vertices of no subtree exceed a prespecified value ℓ; This problem is solved in [15]. The key idea of the algorithm in [15] is discussed briefly next. Let $\mathcal{T}$ be the given tree and $\mathcal{T}_m$ be a subtree with tip at x_m, $S(m)$ be the set of children of x_m, i.e., having a directed edge to x_m, $w(m)$ be the weight of the vertex x_m, and $W(m)$ be the sum of the weights of the vertices in the subtree $\mathcal{T}_m$. Each vertex has weight at most $\ell \in \mathbb{N}^\star$. Given the prespecified bound ℓ, if x_m is a vertex such that $W(m) > \ell$ and $W(k) \leq \ell$ for all $x_k \in S(m)$, then the edge (x_{k_0}, x_m) is removed where $W(k_0) = \max_{x_k \in S(m)} W(k)$. This results in the removal of the subtree whose tip is x_{k_0}, $\mathcal{T}_{k_0}$, from the tree $\mathcal{T}$ and $\mathcal{T}_{k_0}$ becomes a component of the partition P. By proceeding along the tree level by level, starting from the leaves and ending at the tip, we locate the vertex x_m in this procedure. The resultant tree, obtained after deletion of the subtree whose tip is x_{k_0}, is analysed further by employing the same procedure as given above. At any stage of the algorithm, a single tree is modified by the deletion of a subtree. We employ this linear time algorithm provided in [15] to find a solution of $(\mathcal{P}_1)$ via Algorithm 3.

Algorithm 3: Solve Problem $(\mathcal{P}_1)$ under Assumption 1

Input: $A \in \{0, *\}^{d \times d}$ and $F \subset \mathcal{A}$ such that $\mu(A, I_F) \leq \ell$
Output: A solution of $(\mathcal{P}_1)$, $J^* \subset F$.

1 Assign weight $w(x_i) = 1$ for all $x_i \in \mathcal{A}$
2 Use merging procedure given in Algorithm 2 to obtain $G'(A)$ and weight function $w : F \to \mathbb{N}^\star$.
3 Use Algorithm of [15] on $G'(A)$ with weight function w and bound ℓ to find an optimal partition P
4 J^* is the collection of tip of each subtree of P

Since each step in Algorithm 3 has linear complexity, we obtain a solution to $(\mathcal{P}_1)$ in linear time. It is easy to see that we can obtain a minimal partition of $G(A)$ which satisfies the condition that each subtree has at most ℓ number of state vertices (with its tip lying in F) from the minimal partition of $G'(A)$ obtained by using [15] in Step 3 of Algorithm 3. The demonstration of the steps of Algorithm 3 to identify a solution of $(\mathcal{P}_1)$ via an example is given in [7] and is omitted here due to space constraints.

5 Cardinality constrained sensor placement problem

Suppose $A \in \{0, *\}^{d \times d}$ and $C \in \{0, *\}^{p \times d}$ are given. Recall that $\mathcal{A} = \{x_1, x_2, \ldots, x_d\}$ and $C = \{1, 2, \ldots, p\}$ are the sets of state and output vertices of the digraph $G(A, C)$ respectively. Without loss of generality, we assume that each output measures at least one state. For $S \subset C$, $C(S)$ is a submatrix obtained by retaining the rows corresponding to the output vertices in S. A set of state vertices $\mathcal{A}' \subset \mathcal{A}$ in $G(A)$ is said to be *structurally observable by* S if the induced subgraph $G(A', C'(S)) = (\mathcal{A}' \sqcup S, E_{A'} \sqcup E'_S)$ of $G(A, C)$ by $\mathcal{A}' \sqcup S$ satisfies both the conditions given in Theorem 1(b), where $E_{A'}$ contains the edges between the state vertices in $\mathcal{A}'$ and $E'_S \subset E_C$ contains only the edges between $\mathcal{A}'$ and S and $A' \in \{0, *\}^{|\mathcal{A}'| \times |\mathcal{A}'|}$, $C'(S) \in \{0, *\}^{|S| \times |\mathcal{A}'|}$. In other words, $(A', C'(S))$ is structurally observable.

For $S \subset C$, we define $\Xi : 2^C \to \mathbb{R}$ by

$$\Xi(S) := \max\Big\{\big|\mathcal{A}'\big| \;\Big|\; \mathcal{A}' \subset \mathcal{A} \text{ and } (A', C'(S)) \text{ is structurally observable}\Big\}. \quad (4)$$

In simple words, $\Xi(S)$ is the size of the largest subgraph of $G(A)$ that is structurally observable by S. The value of $\frac{\Xi(S)}{d}$, clearly, lies in the interval $[0, 1]$; the fraction $\frac{\Xi(S)}{d}$ takes the value 1 if the set of all state variables of $G(A)$ is structurally observable by S, and 0 if no state variable is observable by S. Based on (4), given $A \in \{0, *\}^{d\times d}$ and $C \in \{0, *\}^{p\times d}$, we have the problem of selecting up to r ($1 \le r \le p$) output vertices so as maximize the following function:

$$\begin{array}{ll} \underset{S\subset C}{\text{maximize}} & \dfrac{\Xi(S)}{d} \\ \text{subject to} & |S| \le r. \end{array} \quad (\mathcal{P}_2)$$

Next, we prove that $(\mathcal{P}_2)$ is NP-hard by reducing to $(\mathcal{P}_2)$ from the *maximum cover problem*, the latter being a well-known NP-hard problem [11].

Theorem 4 *$(\mathcal{P}_2)$ is NP-hard.*

PROOF. Consider the maximum cover problem: Given a positive integer r and a collection of sets in $K = \{Z_1, Z_2, \ldots, Z_p\}$ such that each Z_i contains some elements, find a subset $\hat{K} \subset K$ of sets such that $\big|\hat{K}\big| \le r$ and the number of covered elements $\big|\bigcup_{Z_i\in\hat{K}} Z_i\big|$ is maximized. Let $\bigcup_{Z_i\in K} Z_i = \{s_j\}_{j=1}^{d}$. To prove the NP-hardness, we build an instance of $(\mathcal{P}_2)$ starting from an instance of the maximum cover problem in polynomial time. Each element s_j is associated with a state vertex x_j and each set Z_i is associated to an output vertex i giving the output set $C = \{1, 2, \ldots, p\}$. The corresponding state and output matrices are: $A = I_d \in \{0, *\}^{d\times d}$ as defined earlier in §4 and $C \in \{0, *\}^{p\times d}$ with $C_{ij} = *$ if $s_j \in Z_i$, and 0 otherwise, for $i \in [p]$, $j \in [d]$. Each state vertex has a self-loop in $G(A, C)$.

For a given integer $m \le d$, we prove that there is a set $\hat{S} \subset C$ of size at most r such that $\Xi(\hat{S}) \ge m$ if and only if there exists a $\hat{K} \subset K$ of cardinality at most r such that $\big|\bigcup_{Z_i\in\hat{K}} Z_i\big| \ge m$. Let $\hat{S} \subset C$ with $\big|\hat{S}\big| \le r$ and $\Xi(\hat{S}) \ge m$. We will show that $\hat{K} = \{Z_i \mid i \in \hat{S}\}$ is a feasible solution of the maximum cover problem with $\big|\bigcup_{Z_i\in\hat{K}} Z_i\big| \ge m$. Clearly, $\big|\hat{K}\big| \le r$ as $\big|\hat{S}\big| \le r$. Since every state vertex has a self-loop in $G(A)$, $\Xi(\hat{S})$ is equal to the number of state vertices having a directed path (which in this case is an edge) to at least one of the output vertices in $\hat{S}$. Hence, by construction, it is easy to see that $\Xi(\hat{S}) = \big|\bigcup_{Z_i\in\hat{K}} Z_i\big|$. Thus, $\Xi(\hat{S}) \ge m$ implies that $\big|\bigcup_{Z_i\in\hat{K}} Z_i\big| \ge m$. Conversely, if $\hat{K}$ is such that $\big|\hat{K}\big| \le r$ and $\big|\bigcup_{Z_i\in\hat{K}} Z_i\big| \ge m$ then $\hat{S} = \{i \mid Z_i \in \hat{K}\}$ satisfies $\big|\hat{S}\big| \le r$ and $\Xi(\hat{S}) = \big|\bigcup_{Z_i\in\hat{K}} Z_i\big| \ge m$, and thereby completes the proof. □

In fact, Theorem 4 implies that $(\mathcal{P}_2)$ is NP-hard even when every state vertex has a self-loop in $G(A)$ since the constructed instance in the proof of Theorem 4 belongs to this subclass. This motivates us to impose the following condition on the digraph $G(A)$.

Assumption 2 *It is assumed that each state vertex has a self-loop in the graph $G(A)$. Consequently, the resultant $G(A, C)$ has no contraction.*

Given $S \subset C$, we say that a state vertex $x_i \in \mathcal{A}$ is *accessible* by S if there exists a directed path from it to at least one of the output vertices in S. Under Assumption 2, $\Xi(S)$ is equal to the number of state vertices accessible by S and leads to the following lemma that proves the submodularity of $\Xi(\cdot)$.

Lemma 4 *Consider the system* (1) *and the associated sets $\mathcal{A}$ and C as defined in §3. Suppose Assumption 2 holds. Then $\Xi : 2^C \to \mathbb{R}$ defined in* (4) *is a monotone non-decreasing submodular function.*

PROOF. By the definition of $\Xi(\cdot)$ it is clear that for $S \subset T$, $\Xi(S) \le \Xi(T)$ when Assumption 2 holds. Therefore, $\Xi(\cdot)$ is a monotone non-decreasing function.

Let $S \subset T \subset C$, and suppose that $v \notin T$. To show that $\Xi(\cdot)$ is submodular it is enough to establish that $\Xi(S \cup \{v\}) - \Xi(S) \ge \Xi(T \cup \{v\}) - \Xi(T)$.

Let m be the number of state vertices accessible by both S and v. Similarly, let n be the number of state vertices accessible by both T and v. We have

$$\begin{aligned} \Xi(S \cup \{v\}) &= \Xi(S) + \Xi(v) - m, \\ \Xi(T \cup \{v\}) &= \Xi(T) + \Xi(v) - n. \end{aligned} \quad (5)$$

Since $S \subset T$, we have $m \le n$. Consequently,

$$\Xi(v) - m \ge \Xi(v) - n. \quad (6)$$

Using (5) and (6) it follows that $\Xi(\cdot)$ is submodular. □

Although maximizing a submodular function $\Xi(\cdot)$ is an NP-hard problem, there exists efficient greedy algorithms for providing an approximate solution of $(\mathcal{P}_2)$ [21], and we employ one such mechanism in Algorithm 4 that follows:

Algorithm 4: Approximation algorithm for solving $(\mathcal{P}_2)$

Input: $G(A, C)$ and maximum number of outputs r
Output: A set $S^* \subset C$

Initialization: $S^* = \emptyset$, $i \leftarrow 0$
while $i < r$ **do**
  $s^* \leftarrow \arg\max_{s\in C\setminus S^*} \Xi(S^* \cup \{s\}) - \Xi(S^*)$
  $S^* \leftarrow S^* \cup \{s^*\}$
  $i \leftarrow i + 1$
  if $\Xi(S^*) = d$
    stop
  else
    go to step 2
  end
end while
return S^*; exit

The next theorem gives a qualitative estimate of how close a solution obtained by Algorithm 4, is from an optimal solution of $(\mathcal{P}_2)$:

Theorem 5 *Let $\hat{S}$ be an optimal solution of $(\mathcal{P}_2)$, and let S^* be a set returned by Algorithm 4. Then*

$$\tfrac{\Xi(S^*)}{d} \ge (1 - \tfrac{1}{e})\tfrac{\Xi(\hat{S})}{d}. \quad (7)$$

PROOF. [20, Chapter III, Section 3.9, Theorem 9.3] asserts that for any monotone non-decreasing submodular function $\Xi(\cdot)$, the greedy Algorithm 4 returns a set satisfying $\Xi(S^*) \ge$

$(1 - 1/e)\Xi(\hat{S})$, where $\Xi(\hat{S}) := \max\left\{\Xi(S) \,\middle|\, |S| \leq r\right\}$. The assertion follows. □

Algorithm 4 is a greedy method that progressively picks an output vertex from C and adds it to S^* so that the maximal increase of $\Xi(\cdot)$ is obtained at each iteration. Observe that at most r indices may be added to S^* (See Step 2). In Step 3 the algorithm loops over at most p indices and checks their possible contributions to increase $\Xi(\cdot)$ when an element of C is added to S^*. Therefore, any operation in the algorithm will be at most rp times. Given S^*, we also need to compute $\Xi(S^*)$ in Step 3. Under Assumption 2, this is equivalent to computing the total number of state vertices accessible by $S^* \subset C$. This set of state vertices is obtained by employing depth-first search which has $O(d^2 + dr)$ complexity [6]. Therefore, the overall complexity of Algorithm 4 is $O(rp(d^2 + dr))$. In practice, greedy-type Algorithm 4 for submodular maximization often outperform their worst-case theoretical approximation guarantees.

Example 1 *We demonstrate our result established in §5 on the benchmark electrical power grid, the IEEE 118-bus system. It consists of* 118 *buses,* 53 *power generators, and* 65 *power loads, connected to each other through network interconnections. A cyber-physical model of the generators and the loads proposed in [12] is adopted, where a Taylor linearization is performed at the nominal operating point to obtain a linear system. The obtained linear system $G(A)$ has total number of state vertices equal to* 407 *and the total number of edges between them is* 920*. The state variables of generators and loads of the IEEE* 118*-bus system are as follows: P_{T_G} is the mechanical power of turbine; P_G is the electrical power of generator; w_G is the generator's output frequency; a_G is the valve opening of generator; P_L is the electrical power delivered to load; w_L is the frequency measured at load; I_L is the real power consumed by load.*

Assume that a transmission line (i, j) exists between the generator i and load j, and is represented by a digraph shown in Fig. 2. The frequency component w_{L_j} of bus j influences the dynamics of the power component P_{G_i} of bus i and vice-versa. This shows that we have outgoing edges from the frequencies into the powers of the components in the neighbouring buses. According to the construction shown in Fig. 2,

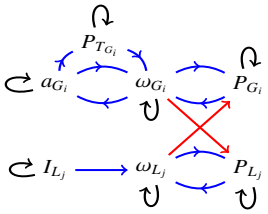

Fig. 2. Illustration of the digraph representation of generator i connected to load j through a transmission line (i, j).

each state vertex of the generators and the loads has a self-loop. Therefore, Assumption 2 is valid for $G(A)$. We consider the given output matrix $C = I_d$, where $d = 407$. We provide an approximate solution of the cardinality constrained sensor placement Problem ($\mathcal{P}_2$) by employing Algorithm 4. Fig. 3 depicts the variation $\Xi(\cdot)$ as the permissible number of outputs changes. If the number of permitted outputs is small, then the maximum size of the set of states structurally observable by the output set obtained from Algorithm 4 is small, which is natural. However, beyond a certain threshold of the permissible number of outputs, in this case 14*, the set of all the state vertices in $G(A)$ become structurally observable by the output set obtained from Algorithm 4.*

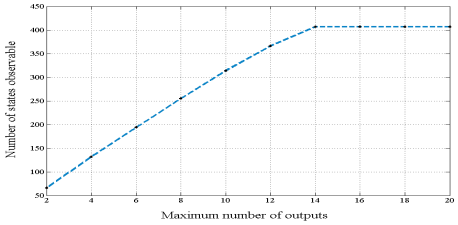

Fig. 3. The change in the cardinality of the set of states structurally observable by the output set (obtained from Algorithm 4) against the permissible number of outputs.

6 Concluding remarks

In this article we have addressed two problems related to optimal sensor placement in linear systems: the minimal sensor placement problem and the cardinality constrained sensor placement problem.

- We produced an efficient polynomial time algorithm to solve the minimal sensor placement problem when the desired bound on the structural observability index is 2.
- We have demonstrated an interesting transition in the hardness of the minimal sensor placement problem as the desired bound changes from 2 to 3.
- The NP-hardness of the minimal sensor placement problem does not preclude the existence of classes of systems for which it is possible to determine solutions efficiently. In fact, we provided a linear time algorithm to solve this problem under a mild assumption that the system structure is a directed tree with self-loop at each state vertex.
- We proved that the cardinality constrained placement problem is a hard combinatorial optimization problem and remains computationally difficult for self-damped systems. We employed a simple greedy algorithm to find an $(1 - \frac{1}{e})$-approximate solution of this problem for self-damped systems.

By standard duality arguments, all our results have analogous counterparts and interpretations for controllability and actuator placement. Future work involves determining other interesting subclasses where the current problems can be solved efficiently, and identifying vulnerable connections between the states whose deletion leads to sudden jumps in the observability index of the system.

References

[1] T. Akutsu, M. Hayashida, W. Ching, and M. K. Ng. Control of boolean networks: Hardness results and algorithms for tree structured networks. *Journal of theoretical biology*, 244(4):670–679, 2007.

[2] A. Chapman and M. Mesbahi. On strong structural controllability of networked systems: A constrained matching approach. In *2013 American Control Conference*, pages 6126–6131, 2013.

[3] A. Clark, B. Alomair, L. Bushnell, and R. Poovendran. *Submodularity in dynamics and control of networked systems*. Communications and Control Engineering Series. Springer, Cham, 2016.

[4] A. Clark, L. Bushnell, and R. Poovendran. On leader selection for performance and controllability in multi-agent systems. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 86–93, 2012.

[5] C. Commault, J. Dion, and D. H. Trinh. Observability preservation under sensor failure. *IEEE Transactions on Automatic Control*, 53(6):1554–1559, 2008.

[6] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press, Cambridge, MA, third edition, 2009.

[7] P. Dey, N. Balachandran, and D. Chatterjee. Efficient constrained sensor placement for observability of linear systems. *arXiv e-prints, arXiv:1711.08264*, 2017.

[8] M. Doostmohammadian and U. A. Khan. On the genericity properties in distributed estimation: Topology design and sensor placement. *IEEE Journal of Selected Topics in Signal Processing*, 7(2):195–204, 2013.

[9] M. Doostmohammadian, H. R. Rabiee, H. Zarrabi, and U. A. Khan. Distributed estimation recovery under sensor failure. *IEEE Signal Processing Letters*, 24(10):1532–1536, 2017.

[10] M. R. Garey and D. S. Johnson. *Computers and Intractability*. W. H. Freeman and Co., San Francisco, Calif., 1979. A Guide to the Theory of NP-Completeness, A Series of Books in the Mathematical Sciences.

[11] D. S. Hochbaum. Approximating covering and packing problems: set cover, vertex cover, independent set, and related problems. *Approximation Algorithms for NP-Hard Problem*, pages 94–143, 1996.

[12] M. D. Ilic, L. Xie, U. A. Khan, and J. M. F. Moura. Modeling of future cyber-physical energy systems for distributed sensing and control. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 40(4):825–838, 2010.

[13] B. Korte and J. Vygen. *Combinatorial Optimization*, volume 21 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, third edition, 2006. Theory and algorithms.

[14] S. Kruzick, S. Pequito, S. Kar, J. M. F. Moura, and A. P. Aguiar. Structurally observable distributed networks of agents under cost and robustness constraints. *IEEE Transactions on Signal and Information Processing over Networks*, 4(2):236–247, 2018.

[15] S. Kundu and J. Misra. A linear tree partitioning algorithm. *SIAM Journal on Computing*, 6(1):151–154, 1977.

[16] C. T. Lin. Structural controllability. *IEEE Transactions on Automatic Control*, 19(3):201–208, 1974.

[17] Y. Y. Liu and A. L. Barabási. Control principles of complex systems. *Reviews of Modern Physics*, 88(3):035006, 2016.

[18] R. M. May. *Stability and complexity in model ecosystems*, volume 6. Princeton university press, 2001.

[19] H. Mortazavian. On k-controllability and k-observability of linear systems. In *Analysis and optimization of systems (Versailles, 1982)*, volume 44 of *Lecture Notes in Control and Information Sciences*, pages 600–612. Springer, Berlin, 1982.

[20] G. Nemhauser and L. Wolsey. *Integer and Combinatorial Optimization*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., New York, 1999. Reprint of the 1988 original, A Wiley-Interscience Publication.

[21] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions. I. *Math. Programming*, 14(3):265–294, 1978.

[22] C. Nowzari, V. M. Preciado, and G. J. Pappas. Analysis and control of epidemics: A survey of spreading processes on complex networks. *IEEE Control Systems Magazine*, 36(1):26–46, 2016.

[23] A. Olshevsky. Minimal controllability problems. *IEEE Transactions on Control of Network Systems*, 1(3):249–258, 2014.

[24] S. Pequito, S. Kar, and A. P. Aguiar. Minimum number of information gatherers to ensure full observability of a dynamic social network: A structural systems approach. In *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 750–753, 2014.

[25] S. Pequito, S. Kar, and A. P. Aguiar. A framework for structural input/output and control configuration selection in large-scale systems. *IEEE Transactions on Automatic Control*, 61(2):303–318, 2016.

[26] S. Pequito, S. Kruzick, S. Kar, J. M. F. Moura, and A. Pedro Aguiar. Optimal design of distributed sensor networks for field reconstruction. In *21st European Signal Processing Conference (EUSIPCO 2013)*, pages 1–5, 2013.

[27] S. Pequito, V. M. Preciado, A. L. Barabási, and G. J. Pappas. Trade-offs between driving nodes and time-to-control in complex networks. *Scientific Reports*, 7:39978, 2017.

[28] K. J. Reinschke. *Multivariable Control: a Graph-Theoretic Approach*, volume 108 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin, 1988.

[29] W. Ren and R. W. Beard. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Transactions on Automatic Control*, 50(5):655–661, 2005.

[30] W. Ren, R. W. Beard, and E. M. Atkins. Information consensus in multivehicle cooperative control. *IEEE Control Systems Magazine*, 27(2):71–82, 2007.

[31] C. Sueur and G. Dauphin-Tanguy. Controllability indices for structured systems. *Linear Algebra and its Applications*, 250:275–287, 1997.

[32] T. H. Summers, F. L. Cortesi, and J. Lygeros. On submodularity and controllability in complex dynamical networks. *IEEE Transactions on Control of Network Systems*, 3(1):91–101, 2016.

[33] S. Sundaram and C. N. Hadjicostis. Structural controllability and observability of linear systems over finite fields with applications to multi-agent systems. *IEEE Transactions on Automatic Control*, 58(1):60–73, 2013.

[34] V. Tzoumas, A. Jadbabaie, and G. J. Pappas. Sensor placement for optimal kalman filtering: Fundamental limits, submodularity, and algorithms. In *2016 American Control Conference (ACC)*, pages 191–196, 2016.

[35] L. Y. Wang, A. Syed, G. G. Yin, A. Pandya, and H. Zhang. Control of vehicle platoons for highway safety and efficient utility: consensus with communications and vehicle dynamics. *J. Syst. Sci. Complex.*, 27(4):605–631, 2014.

[36] Y. Zheng, S. E. Li, K. Li, F. Borrelli, and J. K. Hedrick. Distributed model predictive control for heterogeneous vehicle platoons under unidirectional topologies. *IEEE Transactions on Control Systems Technology*, 25(3):899–910, 2017.