Learning-based Sparse Sensing with Performance Guarantees

Reza Vafaee, and Milad Siami

***Abstract*—In this study, we address the challenge of sensor scheduling in discrete-time linear dynamical networks. We propose a novel learning-based rounding method aimed at converting a provided weighted sensor schedule into a sparse, unweighted schedule while preserving a comparable level of observability performance to the original weighted schedule. We introduce the notion of L-systemic performance measures, which enjoy characteristics such as homogeneity, monotonicity, convexity, and Lipschitz continuity, covering a range of well-known measures. We integrate the initialization of the weighted sensor schedule, achieved via a convex relaxation of a combinatorial optimization problem based on an L-systemic measure, into our rounding approach. We show that this produces an unweighted sensor schedule that achieves a $(1+\epsilon)$ near-optimal approximation solution while ensuring system observability. Our polynomial-time deterministic framework provides a performance guarantee compared to the optimal solution for all types of L-systemic performance measures, including a class of non-submodular metrics. The effectiveness of the theoretical findings is evaluated for a benchmark numerical example in distributed frequency control.**

I. Introduction

In closed-loop control systems, each subsequent control decision relies on estimations obtained for various system variables. Typically, these estimations are derived from monitoring different aspects of the system using a set of strategically placed sensors. While attributes like measurement quality, temporal resolution, and latency are dictated by the sensor types, the spatial characteristics pertinent to the system are primarily defined by the sensor locations. In practical scenarios, limitations such as cost-intensive individual measurements or computational restrictions of the hardware often compel us to consider only a subset of the available sensors. However, finding an optimal sparse set of sensors is an *unsolved* challenge [1].

The challenge of optimal sensor selection is often cast as a combinatorial optimization problem:

$$\begin{aligned} \text{minimize} \quad & \rho\left(\mathsf{A}\right) \\ \text{subject to} \quad & \mathsf{A} \subset \mathsf{U},\ \text{card}(\mathsf{A}) \leq \alpha, \end{aligned} \tag{1}$$

where $\rho(\cdot)$ represents the objective, or cost, function that typically measures the observability performance of the system, U denotes the set of available sensors, $\text{card}(\cdot)$ indicates the cardinality of a set, and α represents the allocated budget.

In a simple scenario where a well-defined cost function is in place, discovering the optimal solution for the optimization problem (1) translates into an infeasible brute-force exploration of the combinatorial options. Notably, there exist $\binom{\text{card}(\mathsf{U})}{\alpha}$ potential combinations of α sensor selections from the pool of $\text{card}(\mathsf{U})$ available sensors. This exponential growth in complexity renders the problem intractable and NP-hard. Consequently, a majority of proposed strategies have aimed to approximate the optimal solution using heuristics and intuitive techniques.

Initial attempts, exemplified by [2], applied *non-linear integer programming* for approximate solutions. However, these methods might face scalability issues in real-world scenarios, such as in smart power grids and robotics. Subsequent research has explored various greedy algorithms to approximate the optimal solution, often with confidence bounds. These algorithms often address one or two specific performance measures and have limitations when dealing with non-submodular metrics [3]–[5].

When the cost function in (1) is convex, a common approach involves solving its corresponding convex relaxation. This suggests to obtaining a combinatorial solution through a (polynomial-time) algorithm that rounds the relaxation solution. For submodular costs, established methods like pipage and randomized rounding applied to the fractional solution from the semidefinite programming (SDP) relaxation yield fast approximation algorithms [6]. However, these techniques fall short for non-submodular costs.

The Hadamard decomposition of the observability Gramian matrix is utilized in [7] to derive optimal sets for single-input linear time-invariant (LTI) networks, covering A- and D-optimality performance measures. The approach extends to multi-input systems with a focus on T-optimality. Additionally, related work by [8] addresses placing time-varying actuators, noting that when the system matrix B has no zero columns, the optimal integer solution coincides with the continuously relaxed solution for T-optimality. In contrast, our introduced learning framework achieves near-optimal sets for both single and multi-input systems, including various performance measures within a significant family that includes A-, D-, and T-optimality.

In [9] and [10], the authors address a different challenge: identifying sparse actuator/sensor sets for LTI dynamics that mimic the performance of a *full set* in terms of controllability/observability. For example, [9] introduces deterministic and randomized algorithms that strategically sample actuators in space and time, approximating the controllability Gramian matrix for fully actuated systems. Similarly, [10] proposes a framework for simultaneous actuator and sensor sampling, approximating Hankel singular values of LTI dynamics. In contrast to our introduced framework, both of these approaches provide approximation guarantees

This material is based upon work supported in part by the U.S. Office of Naval Research under Grant Award N00014-21-1-2431; in part by the U.S. National Science Foundation under Grant Award 2121121 and Grant Award 2208182; and in part by the U.S. Department of Homeland Security under Grant Award 22STESE00001-03-02. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

Department of Electrical & Computer Engineering, Northeastern University, Boston, MA, 02115, USA (e-mails: {vafaee.r, m.siami}@northeastern.edu)

exclusively in comparison to dynamics featuring complete actuator/sensor sets.

In [11], a probabilistic approach establishes a high probability bound on the observability Gramian for a resulting sparse system, relative to full sensing. However, this analysis addresses only E- and A-optimality performance measures and applies to a scaled version of the original LTI system. Another closely related concern, distinct from the primary focus of this paper, involves identifying the *minimal* sensor set required to achieve observability in a system [12].

Our Contributions: Our work innovates by exploiting connections between sensor sparsification and positive semidefinite matrix-based regret minimization. Our paper's key contributions are:

- We establish a polynomial-time framework via a two-player online regret minimization game. This framework transforms weighted sensor schedules into unweighted, sparse configurations with limited active sensors, while preserving a comparable level of observability performance to the original weighted schedule.
- We introduce L-systemic observability metrics for linear systems, characterized by properties like homogeneity, monotonicity, convexity, and Lipschitz continuity w.r.t. the observability Gramian. We show that this family includes various well-known measures.
- We initialize weighted sensor schedules via convex relaxation with an L-systemic objective. We demonstrate that integrating this into our learning-based rounding yields unweighted schedules approximating $(1+\epsilon)$ optimality.
- We verify our contributions with a benchmark numerical example focused on distributed frequency control.

This paper extends the preliminary results presented in [13]. The manuscript contains several new results (Theorems 1 & 2, Propositions 1, 3 & 4, and Lemmas 1, 2 & 3), a numerical example section (Section V), algorithms (Algorithms 1 & 2), a new definition (Definition 1), and new figures (Figs. 1 & 2). To ensure a concise and focused narrative, specific aspects of the theoretical background, such as proofs and lemmas, have been relocated to the appendix.

II. Mathematical Terms and Definitions

A. Mathematical Notations

Sets. We denote the sets of integers and real numbers as $\mathbb{Z}$ and $\mathbb{R}$, respectively. The set of integers (real numbers) greater than or equal to $a \in \mathbb{R}$ is denoted as $\mathbb{Z}_{\geq a}$ ($\mathbb{R}_{\geq a}$). We use uppercase sans-serif letters to represent finite sets (e.g., A). For any integer $n \geq 1$, we define $[n] = \{1, \ldots, n\}$. Notation card(A) returns the cardinality or number of elements in set A.

Matrices. We use uppercase letters to represent real-valued matrices (e.g., A). I and $\mathbf{0}$ denote the identity and zero matrices, respectively, with dimensions inferred from context. For a square matrix X, we denote its determinant and trace as $\det X$ and $\operatorname{Trace} X$, respectively. The transpose of matrix A is represented by $A^\top$, and its Moore-Penrose pseudoinverse is denoted as $A^\dagger$, with $A^{-1/2} = (A^\dagger)^{1/2}$.

Vectors. Lowercase bold letters represent vectors (e.g., $\boldsymbol{b}$). The i-th basis vector is denoted as $\boldsymbol{e}_i$ with dimensions inferred from context, where $\boldsymbol{e}_i(j) = 0$ for $j \neq i$ and $\boldsymbol{e}_i(i) = 1$. The notation $\mathbb{1}$ is used to denote a vector of all ones.

Positive Semidefinite Ordering. For symmetric matrices $A, B \in \mathbb{R}^{n\times n}$, we use $A \preceq B$ to indicate that $\boldsymbol{x}^\top A \boldsymbol{x} \leq \boldsymbol{x}^\top B \boldsymbol{x}$ holds for all $\boldsymbol{x}$. Similarly, we define $\succeq$, $\prec$, and $\succ$. We refer to a symmetric matrix $A \in \mathbb{R}^{n\times n}$ as positive (semi)definite, denoted as PD (PSD), when $A \succ \mathbf{0}$ ($A \succeq \mathbf{0}$). The sets $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$ represent the PSD and PD cones of n-by-n matrices, respectively.

Operator Norms. Vector norms $\|\cdot\|_0$, $\|\cdot\|_1$, and $\|\cdot\|$ count the number of nonzero elements, sum the absolute values, and compute the Euclidean norm of the input vector, respectively. For matrix norms, $\|\cdot\|_1$, $\|\cdot\|$, and $\|\cdot\|_F$ find the maximum absolute column sum, maximum singular value, and Frobenius norm of the input matrix, respectively.

Diagonals. For $\boldsymbol{x} \in \mathbb{R}^n$, we use $\operatorname{diag}(\boldsymbol{x}) \in \mathbb{R}^{n\times n}$ to represent the diagonal matrix with $\operatorname{diag}(\boldsymbol{x})(i,i) = \boldsymbol{x}(i)$. For $A \in \mathbb{R}^{n\times n}$, we denote $\operatorname{diag}(A) \in \mathbb{R}^n$ as the vector corresponding to the diagonal of A, where $\operatorname{diag}(A)(i) = A(i,i)$.

Misc. Lowercase non-bold letters are used for scalars, indices (e.g., j), and functions (e.g., $f(\cdot)$), except for T, which denotes the total number of iterations in the regret game. The inner product of matrices A and B is defined as $\langle A, B\rangle \coloneqq \operatorname{Trace} A^\top B$. For a matrix $Z \in \mathbb{R}^{n\times m}$, we define $\operatorname{vec}(Z)$ as the vectorized form, where

$$\operatorname{vec}(Z) = [z_{1,1}, \ldots, z_{n,1}, z_{1,2}, \ldots, z_{n,2}, \ldots, z_{1,m}, \ldots, z_{n,m}]^\top,$$

and vec^{-1} performs the inverse operation. The function $\max\{a, b\}$ returns the larger value between the scalars a and b. Big O notation, denoted by O, is used to express the asymptotic behavior of functions. It is employed to denote that a function $f(n)$ is $O(g(n))$ when there exist positive constants c and n_0 such that for all $n \geq n_0$, $f(n) \leq c \cdot g(n)$. In this context, $f(n)$ represents a function of n whose behavior is being analyzed in terms of its growth rate. Additionally, tilde big O notation, represented by $\widetilde{O}$, is utilized to describe functions while hiding constant factors and logarithmic factors, focusing solely on the dominant growth rate. Big Omega notation is represented by Ω. It signifies that $f(n)$ is $\Omega(g(n))$ when there exist positive constants c and n_0 such that for all $n \geq n_0$, $f(n) \geq c \cdot g(n)$. Big Theta notation is shown as Θ. It implies that $f(n)$ is $\Theta(g(n))$ if both $f(n)$ is $O(g(n))$ and $f(n)$ is $\Omega(g(n))$.

B. Linear System, Estimation, and Observability

A canonical discrete-time, LTI dynamic system is described by

$$\boldsymbol{x}(k+1) = A\,\boldsymbol{x}(k) + B\,\boldsymbol{u}(k) \quad \text{and} \quad \boldsymbol{y}(k) = C\,\boldsymbol{x}(k), \tag{2}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, and $k \in \mathbb{Z}_{\geq 0}$. The state matrix A characterizes the system's internal structure and agent interactions, the input matrix B specifies the controlled nodes by external controllers, and the output matrix C depicts the relationship between the output vector $\boldsymbol{y}$ and the state vector.

With $t > 0$ as the estimation horizon (referred to as the time-to-estimate) and given sequences $\boldsymbol{u}(k)$ and $\boldsymbol{y}(k)$ spanning this horizon, (2) can be expanded as

$$\underset{\sim}{\boldsymbol{y}}(t) = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{t-1} \end{bmatrix} \boldsymbol{x}_0 + \underset{\sim}{T}(t) \begin{bmatrix} \boldsymbol{u}(0) \\ \boldsymbol{u}(1) \\ \vdots \\ \boldsymbol{u}(t-1) \end{bmatrix}, \tag{3}$$

where $\underset{\sim}{\boldsymbol{y}}(t)$ represents the collected vector of measurements over the estimation horizon, $\boldsymbol{x}_0 = \boldsymbol{x}(0)$ is the initial state, and $\underset{\sim}{T}(t)$ is a known block matrix with elements determined by combinations of system matrices A, B, and C. The second term on the right-hand side of (3) is readily identified; thus, we can subtract it from the vector of measurements on the left-hand side, yielding

$$\underset{\sim}{\boldsymbol{y}}(t) = \mathcal{O}(t)\,\boldsymbol{x}_0, \tag{4}$$

where we define $\underset{\sim}{\boldsymbol{y}}(t)$ as the resultant vector, and introduce the *t-step observability matrix* $\mathcal{O}(t) \coloneqq \big[C^\top, (CA)^\top, \ldots, (CA^{t-1})^\top\big]^\top$.

The LTI system (2) is called "observable" over t steps if it allows us to uniquely determine the initial state $\boldsymbol{x}_0$ based on the system of equations (4). We know that this is the case if and only if the t-step observability matrix is *full column rank*.

Assumption 1: We assume the observability of system (2) throughout this paper.

Let F be any left inverse of $\mathcal{O}(t)$, i.e., $F\,\mathcal{O}(t) = I$, then we have the observer $\boldsymbol{x}_0 = F \cdot \underset{\sim}{\boldsymbol{y}}(t)$, which determines $\boldsymbol{x}_0$ (exactly) from inputs and outputs over estimation horizon.

A more realistic scenario is to assume that the measurements are corrupted by noise, represented as $\boldsymbol{y}(k) = C\,\boldsymbol{x}(k) + \boldsymbol{v}(k)$, where $\boldsymbol{v}$ is the *sensor noise* or error. Equation (4) with sensor noises is given by

$$\underset{\sim}{\boldsymbol{y}}(t) = \mathcal{O}(t)\,\boldsymbol{x}_0 + \underset{\sim}{\boldsymbol{v}}(t), \tag{5}$$

where $\underset{\sim}{\boldsymbol{v}}(t) \coloneqq [\,\boldsymbol{v}^\top(0), \boldsymbol{v}^\top(1), \ldots, \boldsymbol{v}^\top(t-1)\,]^\top$. The *Least-Squares Observer* can be utilized to obtain an estimate for $\boldsymbol{x}_0$ in the presence of noise. This observer uses the pseudo-inverse $\mathcal{O}^\dagger(t)$ as the left inverse matrix F, i.e., $F = (\mathcal{O}^\top(t)\mathcal{O}(t))^{-1}\mathcal{O}^\top(t)$. When applied to (5), the observer yields $\hat{\boldsymbol{x}}_0 = \boldsymbol{x}_0 + \mathcal{O}^\dagger(t)\underset{\sim}{\boldsymbol{v}}(t)$, where $\hat{\boldsymbol{x}}_0$ is the least-squares estimate of the initial state.

Assuming sensor noises follow independent and identically distributed (i.i.d.) normal distributions $\mathcal{N}(\mathbf{0}, \sigma I)$, with σ representing the variance of each measurement, we have the covariance of the estimated initial state $\hat{\boldsymbol{x}}_0$ as

$$\mathrm{Cov}(\,\hat{\boldsymbol{x}}_0\,) = \sigma \cdot \mathcal{O}^\dagger(t){\mathcal{O}^\dagger}^\top(t) = \sigma \cdot (\mathcal{O}^\top(t)\mathcal{O}(t))^{-1},$$

where $\mathrm{Cov}(\cdot)$ returns the covariance of a vector. This implies that the covariance matrix is proportional to the inverse of the matrix $\mathcal{X}(t) \coloneqq \mathcal{O}^\top(t)\mathcal{O}(t)$. In the study of observability analysis, the matrix $\mathcal{X}(t)$ plays a pivotal role and is referred to as the *t-step observability Gramian* matrix for the dynamics described by (2). For the purpose of analysis in this paper, we may express the t-step observability Gramian matrix as

$$\mathcal{X}(t) = \sum_{k+1\in[t]} \sum_{j\in[p]} (\boldsymbol{c}_j^\top A^k)^\top (\boldsymbol{c}_j^\top A^k) = \sum_{i\in[tp]} \boldsymbol{o}_i \boldsymbol{o}_i^\top, \tag{6}$$

where $\boldsymbol{c}_j^\top$ represents the rows of the matrix $C \in \mathbb{R}^{p\times n}$, $\boldsymbol{o}_i^\top$ denotes the i-th row of the t-step observability matrix (4), with $i = kp + j$, and p denotes the number of sensors.

The least-squares estimator is known to be efficient, i.e., the Cramér-Rao lower bound is achieved [14]. This connects the Fisher information matrix (FIM) to the inverse of the covariance matrix. As for the case of i.i.d. measurements, we saw that the t-step observability Gramian is inversely proportional to the estimation covariance, so it is directly proportional to the FIM. Hence, manipulation or specification of the eigenvalues of the t-step observability Gramian directly influences the corresponding FIM.

The spectral spectrum of $\mathcal{X}(t)$ also determines the relative observability of individual system modes. For instance, the *minimum* eigenvalue, $\lambda_{\min}(\mathcal{X}(t))$, signifies the *maximum* estimation uncertainty and the Euclidean distance to the set of singular matrices, indicating the distance to the *unobservable* subspace [15]. Networks with observable dynamics featuring small Gramian eigenvalues often result in uncertain estimations. To mitigate this uncertainty, our paper assumes not only the observability of dynamics (2), i.e., a non-singular t-step observability Gramian matrix $\mathcal{X}(t)$, but also $\mathcal{X}(t) \succeq \delta I$ for some small δ.

C. L-systemic Observability Measure

We define L-systemic observability measures as real-valued operators applied to the space of linear dynamical systems (2). These measures assess the degree of uncertainty in estimation. Specifically, we characterize them as operators acting on the observability Gramian matrices of all n-dimensional observable systems with a minimum eigenvalue bounded by a small constant $\delta > 0$, denoted as $\mathcal{X}(t) \succeq \delta I \succ \mathbf{0}$.

Definition 1 (L-systemic Performance Measure): A Gramian-based metric $\rho : \mathsf{D} \to \mathbb{R}_{>0}$, where

$$\mathsf{D} \coloneqq \{P \in \mathbb{S}^n_{++} : P \succeq \delta I\}, \tag{7}$$

for some small $\delta > 0$, is L-systemic if and only if for all $\mathcal{A}, \mathcal{B} \in \mathsf{D}$ it satisfies:

- *Homogeneity*: $\rho(\beta\mathcal{A}) = \beta^{-1}\rho(\mathcal{A})$, $\forall\ \beta > 0$[1];
- *Monotonicity*: if $\mathcal{B} \preceq \mathcal{A}$, then $\rho(\mathcal{B}) \geq \rho(\mathcal{A})$;
- *Convexity*:

$$\rho\big(\alpha\,\mathcal{A} + (1-\alpha)\,\mathcal{B}\big) \leq \alpha\,\rho(\mathcal{A}) + (1-\alpha)\,\rho(\mathcal{B}),\ \forall \alpha \in [0,1];$$

- *Lipschitz Continuity*: $|\rho(\mathcal{A}) - \rho(\mathcal{B})| \leq L \cdot \|\mathcal{A} - \mathcal{B}\|$, with respect to ℓ_1 norm where L is the so-called *Lipschitz* constant.

In the next proposition, we establish that the properties listed in Definition 1 hold for several well-known choices of ρ, such as A(verage), D(eterminant), T(race), and E(igen)-optimality (refer to Table I).

Proposition 1: For the dynamics given by (2), with the t-step observability Gramian matrix $\mathcal{X}(t) \succeq \delta \cdot I \succ \mathbf{0}$, the performance measures listed in Table I exhibit L-systemic properties.

In the upcoming section, we provide the formulation of the central problem addressed in this paper.

III. Scheduling Problem

We tackle the challenge of *binary conversion of weighted sensor scheduling*. Specifically, for a given time-to-estimate $t \geq n$ and a weighted sensor schedule $\Pi = [\pi_{j,k+1}]$[2], where

[1] Function $\rho(\cdot)$ is said to be homogeneous if it satisfies the condition $\rho(\beta\mathcal{A}) = \beta^{-\gamma} \cdot \rho(\mathcal{A})$, where γ represents the degree of homogeneity. However, for the purposes of this paper, when we refer to a metric being homogeneous, we will consider it to be homogeneous of degree 1.

[2] This weighted schedule can be obtained from solving a convex relaxation of a combinatorial optimization problem or from the operational schedule of an existing physical system.

TABLE I: Important examples of L-systemic measures

Measure	Significance	Optimality-criteria
$\operatorname{Trace}\mathcal{X}^{-1}(t)$	Average estimation uncertainty	A-optimality
$\lambda_{\min}^{-1}(\mathcal{X}(t))$	Maximum estimation uncertainty	E-optimality
$1/\operatorname{Trace}\mathcal{X}(t)$	Lower bound on the average estimation uncertainty	T-optimality
$\det^{-1}\mathcal{X}(t)$	Volume of the estimation uncertainty ellipsoid	D-optimality

the bounded scalars $\pi_{j,k+1} \in [0,1]$ denote the strength of the j-th sensor ($j \in [p]$) at time k ($k+1 \in [t]$), and $\|\text{vec}(\Pi)\|_1 \leq q$, for the dynamics in (2), our objective is to devise an unweighted sensor schedule that approximates the provided weighted sensor schedule in a specific observability context.

In precise terms, for a given weighted sensor schedule Π, our goal is to determine a sparse and rounded sensor schedule $S = [s_{j,k+1}]$ consisting of binary scalars $s_{j,k+1} \in \{0,1\}$ and possessing a sparsity of $\|\text{vec}(S)\|_0 = q \ll tp$. For this schedule

$$\frac{\rho(\mathcal{X}_s(t)) - \rho(\mathcal{X}_\pi(t))}{\rho(\mathcal{X}_s(t))} \leq \epsilon, \tag{8}$$

where

$$\mathcal{X}_s(t) = \sum_{i\in[tp]} \text{vec}(S)(i)\boldsymbol{\mathfrak{o}}_i\boldsymbol{\mathfrak{o}}_i^\top, \text{ and } \mathcal{X}_\pi(t) = \sum_{i\in[tp]} \text{vec}(\Pi)(i)\boldsymbol{\mathfrak{o}}_i\boldsymbol{\mathfrak{o}}_i^\top, \tag{9}$$

the operator $\rho(\cdot)$ corresponds to an L-systemic observability performance measure as defined in Definition 1, while $\epsilon \in (0,1)$ represents the approximation factor. Indeed, in this problem, we seek to find an unweighted sensor schedule that preserves a certain level of observability compared to the given weighted schedule, subject to a sparsity constraint.

Assumption 2: We assume that the weighted t-step observability Gramian matrix (9), associated with the provided weighted sensor schedule, is of *full rank*.

This assumption is not overly restrictive. This is because when the weighted sensor schedule Π is derived through the solution of the convex relaxation of a combinatorial optimization problem based on an L-systemic observability measure listed in Table I, excluding T-optimality, the outcome will yield a weighted t-step observability Gramian matrix that is of full rank. This is due to the fact that all A-, E-, and D-optimality performance measures require the invertibility of their input arguments. Additionally, weighted sensor schedules not obtained via convex relaxation, such as those reflecting the operational schedule of a physical system, should inherently avoid generating a Gramian matrix with deficient rank, as such a scenario would lead to a loss of observability.

Based on Assumption 2, the existence of $\mathcal{X}_\pi^{-1/2}(t) = U\Lambda^{-1/2}U^\top$ is guaranteed, where $U\Lambda U^\top$ represents the eigen decomposition of $\mathcal{X}_\pi(t)$. In the upcoming lemma, we demonstrate that the explicit knowledge of the L-systemic performance measure $\rho(\cdot)$ is not a prerequisite for finding the desired rounded sparse sensor schedule S.

Claim 1: To obtain an unweighted sparse sensor schedule S satisfying (8), it is *necessary and sufficient* to establish a schedule S that fulfills the condition:

$$\mathcal{X}_s(t) \succeq (1-\epsilon)\cdot\mathcal{X}_\pi(t). \tag{10}$$

Proof: Without loss of generality, we assume that the positive constant δ in the definition of the L-systemic performance measure (Definition 1) is consistently less than or equal to $(1-\epsilon)\cdot\lambda_{\min}(\mathcal{X}_\pi(t))$. Then, the proof of the backward direction can be derived by applying the monotonicity and homogeneity properties of the L-systemic performance measure to (10), leading to

$$\rho(\mathcal{X}_s(t)) \overset{\text{Monotonicity}}{\leq} \rho\Big((1-\epsilon)\cdot\mathcal{X}_\pi(t)\Big) \overset{\text{Homogeneity}}{=} (1-\epsilon)^{-1}\cdot\rho(\mathcal{X}_\pi(t)). \tag{11}$$

It is not difficult to show that (11) and (10) are equivalent. Furthermore, reversing the steps yields the proof for the other direction. ■

The following results suggest further simplifications to the problem.

Proposition 2: Let

$$\hat{\mathcal{X}}_s(t) \coloneqq \sum_{k+1\in[t]}\sum_{j\in[p]} s_{j,k+1}(\hat{\boldsymbol{c}}_j^\top\hat{A}^k)^\top(\hat{\boldsymbol{c}}_j^\top\hat{A}^k), \tag{12}$$

where $\hat{A}$ and $\hat{\boldsymbol{c}}_j^\top$, representing rows of the matrix $\hat{C}$, are defined by the transformed realization (30) with the whitening transformation $\mathcal{T} = \mathcal{X}_\pi^{-1/2}(t)$, and $s_{j,k+1}$ denote the entries of schedule S. Then, $\lambda_{\min}(\hat{\mathcal{X}}_s(t)) \geq 1-\epsilon$, if and only if (10).

Combining the reductions proposed by Claim 1 and Proposition 2, we provide a precise definition of our main rounding problem:

Problem 1 (From Weighted to Unweighted): Given the approximation factor $\epsilon \in (0,1)$, a time horizon for estimation $t \geq n$, and a weighted sensor schedule $\Pi = [\pi_{j,k+1}]$ characterized by bounded scalars $\pi_{j,k+1} \in [0,1]$ and $\|\text{vec}(\Pi)\|_1 \leq q$ for dynamics (2), our goal is to determine an unweighted sparse sensor schedule $S = [s_{j,k+1}]$ consisting of binary scalars $s_{j,k+1} \in \{0,1\}$ and with sparsity $\|\text{vec}(S)\|_0 \leq q$, such that

$$\lambda_{\min}(\hat{\mathcal{X}}_s(t)) \geq 1-\epsilon, \tag{13}$$

where $\hat{\mathcal{X}}_s(t)$ is defined in (12).

Remark 1: Since the weighted t-step observability Gramian matrix in (9) is full rank, in accordance with Assumption 2, it follows that the unweighted t-step Gramian matrix in (9), where the sensor schedule S is the solution to Problem 1, is also full rank (as implied by (10)). As a result, the solution maintains observability.

In the subsequent sections of this paper, we illustrate how a two-player *regret minimization problem* centered around density matrices can be utilized to effectively address Problem 1.

IV. Proposed Solutions

In the problem of sequential decision making within an *online environment*, the decision maker (the *player*) must select actions, receiving costs (rewards) based on both their actions and those of the environment. The evolution law for the environment's actions is assumed unknown a priori and depends on the player's actions and an unobservable environmental state. This scenario resembles a two-player repeated game, with the environment as the *adversary*. The player aims to minimize cost accumulation, but the lack of assumptions about the adversary complicates objectives like minimizing expected cost. One prevalent criterion in online learning is *regret minimization*, where regret signifies the difference between the potential cost by always choosing a best fixed action against the adversary's choices and the actual obtained cost.

In this paper, we employ a framework of regret minimization using density matrices. We provide a concise overview of this method, originally introduced in [16] for spectral graph sparsification, in Section A2 of the appendix. In the subsequent sections, we adapt the game components to our scheduling scenario in order to tackle Problem 1. We recommend that readers who are not familiar with the concepts of regret minimization over density matrices first consult the general definition provided in the appendix before proceeding to our proposed game setup.

A. Toward a Solution

We define the elements of the regret game as follows to tackle the sensor scheduling problem. We introduce the action space

$$\mathsf{Y}_{n\times n} = \{\Sigma \in \mathbb{S}_+^n : \operatorname{Trace}\Sigma = 1\}, \tag{14}$$

as the set of covariance matrices of the zero-mean Gaussian initial state $\boldsymbol{x}_0 \in \mathbb{R}^n$ with $\mathbb{E}\,\|\boldsymbol{x}_0\|^2 = 1$. We start the game with schedule $S^{(0)} = [s_{j,k+1}^{(0)}]$, where all binary scalars are set to zero, i.e., $s_{j,k+1}^{(0)} = 0$ for all j and k. Then, we proceed with the following iterative process. At each iteration $\ell = 1, \ldots, T$,

- The player picks an initial state $\boldsymbol{x}_0{}^{(\ell)} \sim \mathcal{N}(0, \Sigma^{(\ell)})$, where the covariance matrix is chosen from the set of action space (14), for the transformed realization (30). The player selects $\boldsymbol{x}_0{}^{(\ell)}$ in order to *minimize* a cost.
- The adversary independently adjusts the sensor schedule $S^{(\ell-1)}$ by adding a sensor to build the schedule $S^{(\ell)}$ in an attempt to supposedly *maximize* the cost.
- The cost is evaluated by monitoring the change in the *transient energy* (see Section A3) of the dynamics (2), defined as

$$\mathbb{E}_{\boldsymbol{x}_0}\left[\tfrac{\partial}{\partial S^{(\ell)}}\left\{L_o(\boldsymbol{x}_0{}^{(\ell)}, S^{(\ell)}, t)\right\}\right] = \langle \hat{\mathcal{X}}_s^{(\ell)}(t) - \hat{\mathcal{X}}_s^{(\ell-1)}(t), \Sigma^{(\ell)}\rangle, \tag{15}$$

where $L_o(\boldsymbol{x}_0, S, t) \coloneqq \widetilde{\boldsymbol{y}}^\top(t)\,\operatorname{diag}(\operatorname{vec}(S))\,\widetilde{\boldsymbol{y}}(t)$ for the measurement vector $\widetilde{\boldsymbol{y}}(t)$ defined in (3) with the system being initialized by $\boldsymbol{x}_0$ and $\boldsymbol{u}(k) = \mathbf{0}$ for $k = 0, \ldots, t-1$. Additionally, $\hat{\mathcal{X}}_s^{(\beta)}(t)$ denotes the t-step transformed observability Gramian matrix, given by $\hat{\mathcal{X}}_s^{(\beta)}(t) = \sum_{i\in[tp]} \operatorname{vec}(S^{(\beta)})(i)\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top$, with $\hat{\mathbf{o}}_i^\top$ indicating the i-th row of the transformed t-step observability matrix $\hat{\mathcal{O}}(t) \coloneqq \mathcal{O}(t)\mathcal{X}_\pi^{-1/2}(t)$. Lastly, $\frac{\partial}{\partial x^{(k)}}\{f\} = f(x^{(k)}) - f(x^{(k-1)})$ represents the *partial backward difference* for the function $f(\cdot)$.

The adversary adds a new sensor in time and place to the sensor schedule at each iteration. This action is equivalent to defining $F^{(\ell)} = \hat{\mathbf{o}}_{i^{(\ell)}}\hat{\mathbf{o}}_{i^{(\ell)}}{}^\top$ as the rank-1 feedback matrix reflected by the adversary, where $i^{(\ell)}$ represents the index for the added sensor at iteration ℓ. The feedback matrix $F^{(\ell)}$ meets the condition outlined in Lemma 1 due to the fact that $\alpha\cdot\langle A^{(\ell)}{}^{1/2}, F^{(\ell)}\rangle = \alpha\cdot\hat{\mathbf{o}}_{i^{(\ell)}}{}^\top\Sigma^{(\ell)}{}^{1/2}\hat{\mathbf{o}}_{i^{(\ell)}} \geq 0$, as $\Sigma^{(\ell)}$ is PSD. Hence, the regret bound can be acquired using (35) within the framework of the regret game setup elucidated in this section.

One can leverage this regret bound along with (33) to obtain the following for our game

$$\lambda_{\min}\Big(\sum_{\ell=1}^{T} F^{(\ell)}\Big) \geq \sum_{\ell=1}^{T} \frac{\langle \Sigma^{(\ell)}, F^{(\ell)}\rangle}{1 + \alpha\cdot\langle \Sigma^{(\ell)}{}^{1/2}, F^{(\ell)}\rangle} - \frac{2\sqrt{n}}{\alpha}. \tag{16}$$

The fraction under the summation on the right-hand side of (16) is paid at each iteration, and we call it the *width term*. The second term on the right-hand side, however, is a fixed start-up cost that we refer to as the *diameter term*. One can show that $\sum_{r=1}^{\ell} F^{(r)} = \hat{\mathcal{X}}_s^{(\ell)}(t)$ because $S^{(0)}$ is set to be a zero matrix; therefore, (16) gives the lower bound for the minimum eigenvalue of $\hat{\mathcal{X}}_s^{(T)}(t)$.

Our objective is to optimize the selection of sensor $i^{(\ell)}$ at each iteration ℓ to ensure a minimum eigenvalue of $\hat{\mathcal{X}}_s^{(T)}(t)$ greater than or equal to $1-\epsilon$ for some $\epsilon \in (0,1)$, a requirement crucial for solving Problem 1. The following result demonstrates that maximizing the width term among all available sensors at each iteration ℓ provides the optimal approach for determining the sensor to be added.

Theorem 1: Assume the approximation factor $\epsilon \in (0,1)$, and a weighted sensor schedule $\Pi = [\pi_{j,k+1}]$ with scalars $\pi_{j,k+1} \in [0,1]$ and $\|\operatorname{vec}(\Pi)\|_1 \leq q$ for the dynamics (2). When $q \in [8n/\epsilon^2, +\infty)$, $T = q$ iterations of the regret minimization game, as described below, result in a sensor schedule S with sparsity $\|\operatorname{vec}(S)\|_0 \leq q$. This schedule satisfies, for $\hat{\mathcal{X}}_s(t)$ as defined in (12),

$$\lambda_{\min}(\hat{\mathcal{X}}_s(t)) \geq 1 - \epsilon. \tag{17}$$

Two-player regret game: Set $S^{(0)} = \mathbf{0}$. At each iteration $\ell \in [T]$,

- The player picks action $\Sigma^{(\ell)}$ based on the *Follow-The-Regularized-Leader* (FTRL) strategy defined in (34) for $\alpha = 4\sqrt{n}/\epsilon$ to minimize the cost (15).
- The adversary independently finds the index

$$i^{(\ell)} = \arg\max_{i\in[tp]} \frac{\langle \Sigma^{(\ell)}, \hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top\rangle}{1 + \alpha\cdot\langle \Sigma^{(\ell)}{}^{1/2}, \hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top\rangle}, \tag{18}$$

and then reflects the rank-1 feedback matrix $F^{(\ell)} = \hat{\mathbf{o}}_{i^{(\ell)}}\hat{\mathbf{o}}_{i^{(\ell)}}{}^\top$ to maximize the cost (15).

Although the desired bound is achieved in Theorem 1, we cannot immediately conclude that the resulting sparse sensor schedule S is the solution to Problem 1. This is because S is a multi-set, meaning $\operatorname{vec}(S) \in \mathbb{Z}_{\geq 0}^{tp}$. The

multi-set issue arises due to accepting replacements in (18). A simple solution is to avoid replacements during sensor selection. This concept is explored by the authors in [17], where they demonstrate that the multi-set issue can be resolved by selecting sensors without replacement. However, their approach introduces a new challenge—an oversampling issue. In other words, to simultaneously satisfy the desired bound (17) and resolve the multi-set issue, we need to sample $O(q)$ active sensors.

A commonly employed heuristic for tackling combinatorial problems is the local search, often referred to as the *Fedorov's exchange* heuristic [18]. This approach starts with a schedule of q sensors selected from the available set of tp sensors and aims to enhance the objective by iteratively exchanging one sensor in the schedule. In the subsequent section, we leverage the concept of Fedorov's exchange heuristic to carefully address both the oversampling and multi-set issues.

B. Regret Solution

We commence this section by exploring certain modifications to the iterative procedure outlined in Section IV-A. Subsequently, we discuss the analysis of the adjusted process. The proposed modifications are as follows:

- We initiate the regret game with an arbitrary unweighted sensor schedule $S^{(0)} = [s^{(0)}_{j,k+1}]$ consisting of binary scalars $s^{(0)}_{j,k+1} \in \{0,1\}$ and sparsity $\|\mathrm{vec}(S^{(0)})\|_0 \leq q$, in contrast to a zero schedule.
- Drawing inspiration from the Fedorov's exchange heuristic, at each iteration ℓ, the adversary autonomously updates the sensor schedule by swapping an active sensor with an inactive one. This swap is aimed at purportedly maximizing the cost (15).

Let $\mathsf{M}^{(\ell)} \coloneqq \{i : \mathrm{vec}(S^{(\ell)})(i) = 1\} \subseteq [tp]$ denote the set of selected sensors at the conclusion of the ℓ-th iteration of the game. Based on the introduced modifications, we initiate with an arbitrary initial set $\mathsf{M}^{(0)}$ with a cardinality of q, where the minimum eigenvalue of $\hat{\mathcal{X}}_s^{(0)}(t) = \sum_{i\in\mathsf{M}^{(0)}} \hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top$ may or may not meet the condition in bound (13). Assuming that the player employs actions derived from the $\ell_{1/2}$-strategy (34) to minimize the cost (15), our flexibility as algorithm designers lies in determining the adversary's actions.

During each iteration $\ell \in [T]$, we require the adversary to choose a pair of indices: $i^{(\ell)} \in \mathsf{M}^{(\ell-1)}$ and $j^{(\ell)} \in [tp]\backslash\mathsf{M}^{(\ell-1)}$, and execute the exchange $\mathsf{M}^{(\ell)} = \mathsf{M}^{(\ell-1)}\cup\{j^{(\ell)}\}\backslash\{i^{(\ell)}\}$. Correspondingly, this implies that at iteration ℓ, the adversary reflects the rank-2 feedback matrix as $F^{(\ell)} = \hat{\mathbf{o}}_{j^{(\ell)}}\hat{\mathbf{o}}_{j^{(\ell)}}{}^\top - \hat{\mathbf{o}}_{i^{(\ell)}}\hat{\mathbf{o}}_{i^{(\ell)}}{}^\top$.

To comprehend the impact of sensor exchanges on the minimum eigenvalue, we establish the initial state as $H^0 = \hat{\mathcal{X}}_s^{(0)}(t)$ and employ the regret bound specified in (36) to elucidate the relationship

$$\lambda_{\min}(\hat{\mathcal{X}}_s^{(T)}(t)) \geq \sum_{\ell=1}^{T}\Big(\frac{\langle\Sigma^{(\ell)}, \hat{\mathbf{o}}_{j^{(\ell)}}\hat{\mathbf{o}}_{j^{(\ell)}}{}^\top\rangle}{1+2\alpha\,\langle{\Sigma^{(\ell)}}^{1/2}, \hat{\mathbf{o}}_{j^{(\ell)}}\hat{\mathbf{o}}_{j^{(\ell)}}{}^\top\rangle} - \frac{\langle\Sigma^{(\ell)}, \hat{\mathbf{o}}_{i^{(\ell)}}\hat{\mathbf{o}}_{i^{(\ell)}}{}^\top\rangle}{1-2\alpha\,\langle{\Sigma^{(\ell)}}^{1/2}, \hat{\mathbf{o}}_{i^{(\ell)}}\hat{\mathbf{o}}_{i^{(\ell)}}{}^\top\rangle}\Big) - \frac{2\sqrt{n}}{\alpha}, \quad (19)$$

where $\{\Sigma^{(\ell)}\}_{\ell=1}^T$ represent FTRL solutions

$$\Sigma^{(\ell)} = \Big(c^{(\ell)}I + \alpha\cdot\hat{\mathcal{X}}_s^{(\ell-1)}(t)\Big)^{-2},$$

and the action $\mathcal{U}$ is taken to be the optimal fixed action $\mathcal{U}^*$ (see Section A2 for more details). This time, the width term in (19) consists of two fractions. In the subsequent outcome, we illustrate that it is feasible to select the pair of indices for swapping as the indices associated with the extreme points of the fractions within the width term during each iteration. This selection procedure allows us to acquire a solution to Problem 1 by the end of the game.

Theorem 2 (Regret Solution to Problem 1): Assume the approximation factor $\epsilon \in (0,1)$ and a weighted sensor schedule $\Pi = [\pi_{j,k+1}]$ with scalars $\pi_{j,k+1} \in [0,1]$ and $\|\mathrm{vec}(\Pi)\|_1 \leq q$ for the dynamics (2). When $q \in [45n/\epsilon^2, +\infty)$, executing $T = 3q/\epsilon$ iterations of the two-player regret game, as described below, deterministically yields an unweighted sensor schedule S with sparsity $\|\mathrm{vec}(S)\|_0 \leq q$. This schedule ensures, for $\hat{\mathcal{X}}_s(t)$ as defined in (12),

$$\lambda_{\min}(\hat{\mathcal{X}}_s(t)) \geq 1-\epsilon. \quad (20)$$

Two-player regret game: Initialize $S^{(0)}$ with an arbitrary unweighted sensor schedule of sparsity q. During each iteration $\ell \in [T]$,

- The player selects action $\Sigma^{(\ell)}$ using the $\ell_{1/2}$-strategy (34) and learning rate $\alpha = 3\sqrt{n}/\epsilon$ to minimize the cost (15).
- The adversary independently identifies indices

$$i^{(\ell)} = \arg\min_{i\in\mathsf{N}^{(\ell-1)}} \frac{\langle\Sigma^{(\ell)}, \hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle}{1-2\alpha\,\langle{\Sigma^{(\ell)}}^{1/2}, \hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle}, \quad (21)$$

where

$$\mathsf{N}^{(\ell-1)} \coloneqq \Big\{i \in \mathsf{M}^{(\ell-1)} : \alpha\cdot\langle{\Sigma^{(\ell)}}^{1/2}, \hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle < 1/2\Big\}, \quad (22)$$

and

$$j^{(\ell)} = \arg\max_{i\in[tp]\backslash\mathsf{M}^{(\ell-1)}} \frac{\langle\Sigma^{(\ell)}, \hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle}{1+2\alpha\,\langle{\Sigma^{(\ell)}}^{1/2}, \hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle}, \quad (23)$$

and then reflects the rank-2 feedback matrix $F^{(\ell)} = \hat{\mathbf{o}}_{j^{(\ell)}}\hat{\mathbf{o}}_{j^{(\ell)}}{}^\top - \hat{\mathbf{o}}_{i^{(\ell)}}\hat{\mathbf{o}}_{i^{(\ell)}}{}^\top$ to maximize the cost (15).

The proof is long and deferred to the appendix. To keep the proofs simple, we have not attempted to optimize the constants in the above statement.

Remark 2: The matrix $F^{(\ell)}$ has a rank of at most 2. This can be understood by noting that when subtracting one rank-1 matrix from another, the result may have a rank of 0 if they are equal; 1 if one is a scalar multiple of the other; or 2 if they are linearly independent. Thus, when referring to the feedback matrix as a rank-2 matrix, it is implied that its rank does not exceed 2.

Remark 3: Theorem 2 necessitates $q \geq \Omega(n/\epsilon^2)$ to yield a solution for Problem 1. If the aim is to incorporate, on average, d sensors at each time step $k+1 \in [t]$ within the unweighted sparse sensor schedule, then $t \geq \Omega(n/d\epsilon^2)$. This

introduces a trade-off between the approximation factor ϵ and the time horizon t, signifying that any reduction in the level of approximation factor comes at the cost of extending the time horizon t to maintain the same average count of active sensors d. Furthermore, the precision of the approximation enhances with an increase in d. Evidently, augmenting t necessitates a broader estimation time window, while elevating d requires the inclusion of a greater number of active sensors in the schedule.

Remark 4: Note that the total count of active sensors, q, within the resulting unweighted sensor schedule must always be greater than or equal to n. This requirement arises from the fact that if $q < n$, the unweighted t-step Gramian matrix (9) associated with any unweighted sensor schedule S having a sparsity of $\|\mathrm{vec}(S)\|_0 \leq q$ becomes rank-deficient. However, the invertibility of the input Gramian matrix is essential for the utilization of certain L-systemic performance metrics, such as A-, D-, and E-optimality.

Finding Constant c: In the player action $\Sigma^{(\ell)} = (c^{(\ell)} I + \alpha \cdot \hat{\mathcal{X}}_s^{(\ell-1)}(t))^{-2}$, the constant $c^{(\ell)}$ is a unique real number that ensures $c^{(\ell)} I + \alpha \cdot \hat{\mathcal{X}}_s^{(\ell-1)}(t) \succ 0$ and $\mathrm{Trace}\, \Sigma^{(\ell)} = 1$.

It can be demonstrated that for any symmetric matrix $X \in \mathbb{R}^{n\times n}$, the function defined as $g(c) := \mathrm{Trace}\,(cI + \alpha X)^{-2}$ is a continuous, strictly monotone decreasing function of c on the open interval $(-\alpha \cdot \lambda_{\min}(X), +\infty)$. Based on the *intermediate value theorem*, there must exist a unique constant c within this interval such that $\mathrm{Trace}\,(cI + \alpha X)^{-2} = 1$, since $\lim_{c\to(-\alpha\,\lambda_{\min}(X))^+} g(c) = +\infty$, and $\lim_{c\to+\infty} g(c) = 0$. The range of feasible c values also implies that $cI + \alpha X \succ 0$.

Furthermore, when $c = \sqrt{n}$, it can be shown that $\mathrm{Trace}(\sqrt{n} I + \alpha X)^{-2} \leq \mathrm{Trace}(\sqrt{n} I)^{-2} = 1$, which implies that finding c only necessitates searching the interval $(-\alpha \cdot \lambda_{\min}(X), \sqrt{n}]$. To accomplish this, we design a binary search routine in Algorithm 2 to identify a unique constant $c^{(\ell)}$ within the interval $(-\alpha \cdot \lambda_{\min}(\hat{\mathcal{X}}_s^{(\ell-1)}(t)), \sqrt{n}]$ at each iteration such that $\mathrm{Trace}\, \Sigma^{(\ell)}$ is closely approximated to 1.

In the `Weighted_To_Unweighted` algorithm, we present a pseudocode that outlines the steps of Theorem 2 and incorporates the binary search routine from Algorithm 2 to derive an unweighted sensor schedule that solves Problem 1. Theoretical analysis demonstrates that if the constant $c^{(\ell)}$ is approximated up to an additive error of $\gamma = \Theta(1/\mathrm{poly}(\epsilon^{-1}, q))$ (where poly stands for polynomial), both sides of the inequality (19) are altered by, at most, an additive approximation factor of ϵ. For a stability analysis of the $\ell_{1/2}$-strategy, refer to the Appendix of [16].

The `BinarySearch` algorithm for computing $c^{(\ell)}$ terminates in $O(\log((f_u - f_l)/\gamma)) = \widetilde{O}(1)$ iterations, where the notation $\widetilde{O}(\cdot)$ hides logarithmic factors. Additionally, each iteration of the algorithm requires a maximum of $O(n^3)$ operations to perform matrix inversion for a matrix of size n. Consequently, the total computational complexity of each run of the algorithm is $\widetilde{O}(n^3)$ operations.

In Algorithm 1, the initialization step on line 3 takes $O(n^3)$ time due to the matrix square root computation. The process of finding the swap indices $i^{(\ell)}$ and $j^{(\ell)}$, carried out during each iteration of the "while loop", takes $O(tpn^2)$ time, when $\Sigma^{(\ell)}$ and ${\Sigma^{(\ell)}}^{1/2}$ are computed. Taking all these factors into account, it can be demonstrated that Algorithm 1 requires $\widetilde{O}(tpqn^2/\epsilon)$ time for each execution.

Algorithm 1 `Weighted_To_Unweighted`(Π,ϵ,q)

Input: The weighted sensor schedule Π, approximation factor $\epsilon \in (0,1)$, and desired number of active sensors in the unweighted sensor schedule $q \in [45n/\epsilon^2, +\infty)$.

Output: Unweighted sparse sensor schedule $S = [s_{j,k+1}]$ with scalars $s_{j,k+1} \in \{0,1\}$ and sparsity $\|\mathrm{vec}(S)\|_0 = q$ such that for $\hat{\mathcal{X}}_s(t)$ as defined in (12) inequality (20) holds.

Initialization:

1: $\alpha = 3\sqrt{n}/\epsilon$, $T = 3q/\epsilon$, $\ell = 1$;
2: $[\hat{\mathbf{o}}_1 \ \cdots \ \hat{\mathbf{o}}_{tp}]^\top = \mathcal{O}(t)\mathcal{X}_\pi^{-1/2}(t)$;
3: Set S to an arbitrary unweighted sensor schedule with q active sensors;
4: $\hat{\mathcal{X}}_s(t) = \sum_{i\in[tp]} \mathrm{vec}(S)(i)\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top$;
5: **while** $\ell \leq T$ and $\lambda_{\min}(\hat{\mathcal{X}}_s(t)) \leq 1-\epsilon$ **do**
6: $\quad c \leftarrow$ `BinarySearch`$(\hat{\mathcal{X}}_s(t), \alpha)$;
7: $\quad \Sigma \leftarrow (c\,I + \alpha\,\hat{\mathcal{X}}_s(t))^{-2}$;
8: $\quad$ Update (i,j) according to (21) and (23);
9: $\quad \mathrm{vec}(S) \leftarrow \mathrm{vec}(S) + \boldsymbol{e}_j - \boldsymbol{e}_i$;
10: $\quad \hat{\mathcal{X}}_s(t) \leftarrow \hat{\mathcal{X}}_s(t) + \hat{\mathbf{o}}_j\hat{\mathbf{o}}_j^\top - \hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top$;
11: $\quad \ell \leftarrow \ell + 1$;
12: **end while**
13: **return** S.

In the following section, we demonstrate the efficient solution of the convex relaxation for a combinatorial optimization problem that features an L-systemic performance measure as its cost function. Subsequently, we leverage our rounding algorithm to acquire what is known as a $(1+\epsilon)$ near-optimal approximation solution for the given combinatorial optimization problem.

Algorithm 2 `BinarySearch`(X, α)

Initialization:

1: $f_l = -\alpha \cdot \lambda_{\min}(X)$, $f_u = \sqrt{n}$;
2: $\gamma = 10^{-12}$; ▷ γ is the desired error/accuracy.
3: **while** $|f_l - f_u| > \gamma$ **do**
4: $\quad \bar{f} \leftarrow (f_l + f_u)/2$;
5: $\quad$ **if** $\mathrm{Trace}\,(\bar{f} I + \alpha X)^{-2} > 1$ **then**
6: $\quad\quad f_l \leftarrow \bar{f}$;
7: $\quad$ **else**
8: $\quad\quad f_u \leftarrow \bar{f}$;
9: $\quad$ **end if**
10: **end while**
11: **return** $c = (f_l + f_u)/2$.

C. $(1+\epsilon)$ *Approximation*

Given the LTI dynamics (2), a time horizon for estimation $t \geq n$, the L-systemic observability performance measure $\rho(\cdot)$, and a desired sparsity level for active sensors $q \ll tp$, the unweighted sparse sensor schedule S with sparsity q can be found by solving the following combinatorial optimiza-

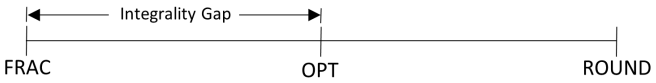

Fig. 1: This plot displays the optimality gaps for a minimization problem. The $(1-\epsilon)^{-1}$ in the figure represents the gap generated by the rounding process using Algorithm 1.

tion problem [19]:

$$\begin{aligned} \underset{S}{\text{minimize}} \quad & \rho\Big(\sum_{i\in[tp]} \text{vec}(S)(i)\boldsymbol{\mathfrak{o}}_i\boldsymbol{\mathfrak{o}}_i^\top\Big) \\ \text{subject to} \quad & \text{vec}(S)\in\{0,1\}^{tp} \wedge \|\text{vec}(S)\|_0 \le q, \end{aligned} \tag{24}$$

where $\boldsymbol{\mathfrak{o}}_i^\top$ is the i-th row of the observability matrix (4).

The combinatorial optimization problem (24) is known to be intractable and NP-hard. Convex relaxations are among the most powerful techniques for devising polynomial-time approximation algorithms for NP-hard optimization problems. In these approaches, the integer program is relaxed to a convex program as follows:

$$\begin{aligned} \underset{\Pi}{\text{minimize}} \quad & \rho\Big(\sum_{i\in[tp]} \text{vec}(\Pi)(i)\boldsymbol{\mathfrak{o}}_i\boldsymbol{\mathfrak{o}}_i^\top\Big) \\ \text{subject to} \quad & \text{vec}(\Pi)\in[0,1]^{tp} \wedge \|\text{vec}(\Pi)\|_1 \le q. \end{aligned} \tag{25}$$

This convex program can be solved in polynomial-time, such as through a linear program (LP) or semidefinite program (SDP). A solution to the combinatorial problem can then be obtained by designing a polynomial-time algorithm to round the solution of this convex relaxation to an integer solution.

While combining convex relaxations and classical rounding methods has yielded approximation solutions for various submodular measures [6], these approaches have faced limited success when dealing with non-submodular performance measures [17]. However, in the following, we present a polynomial-time approximation algorithm capable of obtaining the so-called $(1+\epsilon)$ near-optimal integer solution for all L-systemic performance measures, including a class of non-submodular measures.

Let OPT represent the true combinatorial optimum of the minimization problem (24). The convex relaxation has an optimum value OPT_FRAC, which is at most as large as OPT (as the integer solution is also a feasible solution to the convex program). The rounding algorithm utilizes the solution to the convex relaxation (25) with an objective value of FRAC to produce an integer solution valued at ROUND. In approximation approaches, the ratio ROUND/OPT is important. However, most approximation algorithms require providing an upper bound on ROUND/OPT_FRAC, which in turn bounds ROUND/OPT since OPT remains unknown for the majority of instances of the combinatorial problem. Previously, we discussed a rounding procedure (Algorithm 1) that leverages a regret minimization framework over density matrices. This procedure provides a rounded sparse sensor schedule that achieves a $(1-\epsilon)^{-1}$ approximation of a given weighted sensor schedule (see (11)). In other words, for respective objective values ROUND and FRAC for unweighted and weighted sensor schedules, the ratio ROUND/FRAC is upper-bounded by $(1-\epsilon)^{-1}$ in our rounding result. Fig. 1 displays the optimality gaps for a minimization combinatorial problem.

Let us we turn our attention to the gap between FRAC and OPT_FRAC. We will demonstrate that if the objective $\rho(\cdot)$ in (25) is L-systemic, a fractional solution denoted as Π' with an objective value of FRAC can be obtained such that FRAC/OPT_FRAC is upper-bounded by $(1+\mu)$ for any arbitrary $\mu\in(0,1)$. We succinctly refer to this fractional solution as the $(1+\mu)$ approximation.

Given any arbitrary $\mu\in(0,1)$, the authors of [20] investigate a polynomial-time algorithm aimed at computing a $(1+\mu)$ approximation solution to the convex relaxation (25) under the condition that the objective function is convex and adheres to specific properties [3]:

(a) For any $\varrho>0$, there exists a parameter L_ϱ such that the *smoothed* objective

$$\rho_\varrho:\Pi\mapsto\rho\Big(\sum_{i\in[tp]}(\text{vec}(\Pi)(i)+\frac{\varrho}{tp})\boldsymbol{\mathfrak{o}}_i\boldsymbol{\mathfrak{o}}_i^\top\Big),$$

is L_ϱ-Lipschitz continuous with respect to the ℓ_1 norm. That is, for all $\Pi,\hat{\Pi}$ in

$$\mathsf{C}\coloneqq\{Z=[z_{i,j}]\in\mathbb{R}^{p\times t}: z_{i,j}\in[0,1] \ \wedge \ \|\text{vec}(Z)\|_1\le q\},$$

we have $|\rho_\varrho(\Pi)-\rho_\varrho(\hat{\Pi})|\le \mathrm{L}_\varrho\|\text{vec}(\Pi)-\text{vec}(\hat{\Pi})\|_1$.

(b) There exists $\zeta_0>0$ such that $\inf_{\Pi\in\mathsf{C}}\rho(\Pi)\ge\zeta_0$.

The following results show that if the objective is L-systemic, we can employ the relaxed approximation algorithm from [20] to compute the $(1+\mu)$ approximation solution to (25).

Proposition 3: Any L-systemic observability performance measure satisfies both properties (a) and (b).

Proof: To demonstrate that any L-systemic measure satisfies property (a), let $\mathcal{A}=\sum_{i\in[tp]}(\text{vec}(\Pi)(i)+\frac{\varrho}{tp})\boldsymbol{\mathfrak{o}}_i\boldsymbol{\mathfrak{o}}_i^\top$ and $\mathcal{B}=\sum_{i\in[tp]}(\text{vec}(\hat{\Pi})(i)+\frac{\varrho}{tp})\boldsymbol{\mathfrak{o}}_i\boldsymbol{\mathfrak{o}}_i^\top$ represent the symmetric matrices for any two schedules $\Pi,\hat{\Pi}\in\mathsf{C}$. Note that $\mathcal{A}$ and $\mathcal{B}$ belong to the set D defined in (7). Therefore, since the L-systemic measure ρ is Lipschitz continuous with respect to the ℓ_1 norm, there exists a Lipschitz constant L such that

$$|\rho(\mathcal{A})-\rho(\mathcal{B})|\le L\cdot\|\mathcal{A}-\mathcal{B}\|_1. \tag{26}$$

Substituting the expressions for $\mathcal{A}$ and $\mathcal{B}$, we can rewrite (26) as follows:

$$\begin{aligned} |\rho_\varrho(\Pi)-\rho_\varrho(\hat{\Pi})| &\le L\cdot\Big\|\sum_{i\in[tp]}(\text{vec}(\Pi)(i)-\text{vec}(\hat{\Pi})(i))\boldsymbol{\mathfrak{o}}_i\boldsymbol{\mathfrak{o}}_i^\top\Big\|_1 \\ &\overset{(c)}{\le} \sqrt{n}\cdot L\cdot\Big\|\sum_{i\in[tp]}(\text{vec}(\Pi)(i)-\text{vec}(\hat{\Pi})(i))\boldsymbol{\mathfrak{o}}_i\boldsymbol{\mathfrak{o}}_i^\top\Big\|_F \\ &\overset{(d)}{\le} \sqrt{n}\cdot L\cdot\sum_{i\in[tp]}\Big\|(\text{vec}(\Pi)(i)-\text{vec}(\hat{\Pi})(i))\boldsymbol{\mathfrak{o}}_i\boldsymbol{\mathfrak{o}}_i^\top\Big\|_F \\ &\overset{(e)}{\le} \sqrt{n}\cdot L\cdot\|\boldsymbol{\mathfrak{o}}_{\max}\|^2\cdot\|\text{vec}(\Pi)-\text{vec}(\hat{\Pi})\|_1, \end{aligned} \tag{27}$$

which completes the proof of property (a). In (27), inequality (c) holds due to (40) and the fact that for any square matrix $X\in\mathbb{R}^{n\times n}$ with rank $r\le n$, $\|X\|=\sigma_{\max}\le\sqrt{\sigma_1^2+\cdots+\sigma_r^2}=\|X\|_F$, where $\sigma_1,\ldots,\sigma_r$ are the singular values and $\sigma_{\max}$ is the maximum singular value

[3]While the algorithm is founded on *entropic mirror descent*, it is noted that other approaches, including projected gradient descent and conic programming, can also be employed [20, Section 3].

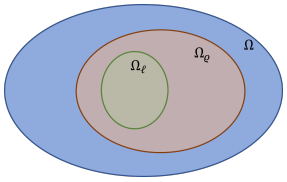

Fig. 2: Venn diagram that illustrates the relationship among the set of general observability performance measures Ω, the family of L-systemic performance measures Ω_ℓ, and the set of observability performance measures that satisfy properties (a) and (b) of Proposition 3, Ω_ϱ. We note that Ω_ℓ is a subset of Ω_ϱ.

of X. Inequality (d) holds because the Frobenius norm is sub-additive, and (e) holds because $\|\boldsymbol{o}_i\boldsymbol{o}_i^\top\|_F = \|\boldsymbol{o}_i\|^2$, and $\|\boldsymbol{o}_{\max}\|^2$ is defined as the maximum value of $\|\boldsymbol{o}_i\|^2$ over the set $\{\boldsymbol{o}_i\}_{i=1}^{tp}$.

Since the scalars $\pi_{j,k+1}$ are positive real numbers less than or equal to 1, we can show that

$$\sum_{i\in[tp]} \mathrm{vec}(\Pi)(i)\boldsymbol{o}_i\boldsymbol{o}_i^\top \preceq \mathcal{X}(t) = \sum_{i\in[tp]} \boldsymbol{o}_i\boldsymbol{o}_i^\top \preceq \lambda_{\max}(\mathcal{X}(t))\cdot I,$$

for any weighted schedule $\Pi \in \mathsf{C}$. The *Gershgorin Circle Theorem* establishes that the largest eigenvalue of a real square matrix is constrained by the largest absolute row sum. Consequently, given that the entries of the t-step observability Gramian matrix are finite, it follows that all of its eigenvalues are also finite, i.e., $\lambda_{\max}(\mathcal{X}(t)) < \infty$. Therefore, one can show that property (b) holds for A-, E-, T-, and D-optimality for ζ_0 equal to $n/\lambda_{\max}(\mathcal{X}(t))$, $1/\lambda_{\max}(\mathcal{X}(t))$, $1/(n\cdot\lambda_{\max}(\mathcal{X}(t)))$, and $1/\lambda^n_{\max}(\mathcal{X}(t))$, respectively. ■

General observability performance measures can be categorized based on their functional properties, as defined in Definition 1 and Proposition 3. We define Ω_ϱ as the set of performance metrics satisfying properties (a) and (b) in Proposition 3, and Ω_ℓ as the L-systemic observability performance measures from Definition 1. In Proposition 3, we show that L-systemic measures adhere to both Definition 1 and Proposition 3's properties (a) and (b). Fig. 2 illustrates the relationship among Ω_ϱ, Ω_ℓ, and general observability performance measures.

In the following corollary, we combine results from Section IV-B and this section to achieve an unweighted sparse sensor schedule for optimization problem (24). This ensures that the ratio $\mathsf{ROUND/OPT\_FRAC}$ is upper-bounded by $(1+\epsilon)$, providing a $(1+\epsilon)$ near-optimal solution to (24).

Corollary 1: Given $\epsilon \in (0, 1/2]$, a time horizon $t \geq n$, an L-systemic observability performance measure $\rho(\cdot)$, and a desired sparsity level $q \ll tp$ for LTI dynamics (2), if $q \in [45n/\epsilon^2, +\infty)$, there exists a polynomial-time approximation algorithm that computes an unweighted sparse sensor schedule $\hat{S}$ containing at most q active sensors, satisfying $\rho(\hat{S}) \leq (1+3\epsilon)\cdot \mathsf{OPT}$, where OPT represents the true combinatorial optimum of problem (24).

Proof: By Proposition 3, for an L-systemic objective in combinatorial problem (24), for any arbitrary $\mu \in (0,1)$, a polynomial-time approximation algorithm computes an approximate weighted sparse sensor schedule $\Pi' = [\pi'_{j,k+1}]$ with objective value FRAC and $\|\mathrm{vec}(\Pi')\|_1 \leq q$ that satisfies $\mathsf{FRAC/FRAC\_OPT} \leq (1+\mu)$, where $\mathsf{FRAC\_OPT}$ is the optimum value of the convex relaxation (25).

Furthermore, according to Theorem 2, for any $\epsilon \in (0,1)$ and $q \geq 45n/\epsilon^2$, Algorithm 1 can be applied to the fractional solution Π'. It yields an unweighted sparse sensor schedule $\hat{S}$ with sparsity at most q such that $\mathsf{FRAC/ROUND}$ is lower bounded by $(1-\epsilon)$. This implies $\mathsf{ROUND/FRAC} \leq 1/(1-\epsilon) \leq 1+2\epsilon$, if we narrow the interval for ϵ to $\epsilon \in \left(0, \frac{1}{2}\right]$.

Therefore, for the ratio $\mathsf{ROUND/FRAC\_OPT}$, we have

$$\frac{\mathsf{ROUND}}{\mathsf{FRAC\_OPT}} \leq (1+2\epsilon)\cdot(1+\mu) \leq (1+3\epsilon), \tag{28}$$

where the last inequality follows by setting $\mu = \frac{\epsilon}{2}$ and noting that we narrow down the interval for ϵ to $(0, \frac{1}{2}]$. Since $\mathsf{FRAC\_OPT} \leq \mathsf{OPT}$ for any minimization problem, (28) concludes the proof. ■

Remark 5: Corollary 1 establishes the requirement that a minimum of $\Omega(n/\epsilon^2)$ active sensors is necessary to achieve a $(1+\epsilon)$ near optimal approximation. Without any condition on q, it has been proven that obtaining any $(1+\epsilon)$ approximation is NP-hard, especially in the context of D- and E-optimality [21], [22]. Additionally, as pointed out in Remark 4, the number of active sensors q must be at least equal to the dimension of the system n. This is because the nonsingularity condition of the t-step Gramian observability matrix is required for some optimality criteria within the L-systemic family.

Remark 6: All L-systemic observability performance measures listed in Table I necessitate non-singularity in their input arguments, except for T-optimality. Hence, any polynomial-time relaxed approximation algorithm produces a weighted sparse sensor schedule for (25), ensuring that its corresponding weighted t-step observability Gramian matrix, denoted as $\mathcal{X}_{\pi'}(t)$, maintains full rank. Moreover, the polynomial-time rounding procedure of Algorithm 1 effectively maintains observability. This is evident as, for the unweighted t-step observability Gramian matrix $\mathcal{X}_{\hat{s}}(t)$ associated with the resultant rounded sensor schedule $\hat{S}$, as specified in (10), we can establish that $\lambda_{\min}(\mathcal{X}_{\hat{s}}(t)) \geq (1-\epsilon)\cdot\lambda_{\min}(\mathcal{X}_{\pi'}(t))$. Consequently, Corollary 1 establishes an unweighted sparse sensor schedule for A-, E-, and D-optimality that ensures system observability.

1) Discussion on the Tightness of the Results in Corollary 1: In the following, we utilize results from graph theory to demonstrate that the derived lower bound on the number of active sensors q in Corollary 1 for the general case of an L-systemic objective is within a constant factor of the optimal. In this discussion, we assume $t = n+1$, meaning the time-to-estimate is $n+1$ steps. We provide a class of linear systems with state matrices A and C where no unweighted sparse sensor schedule with sparsity q can achieve a $(1+\epsilon)$ approximation unless $q = \Omega(n/\epsilon^2)$. This shows that our bound on the number of active sensors cannot be improved.

There is a close connection between sparse sensor scheduling and graph sparsification. The contribution of each link in the Laplacian matrix is a rank-1 matrix, just as the contribution of each sensor at each timestep is in the Gramian matrix. Therefore, to sparsify a graph (i.e., reducing the number of links), we need to sparsify the summation of rank-1 matrices. In this paper, we utilize a similar strategy to sparsify sensors in both time and space.

To define a class of linear systems with state matrices A and C where no unweighted sparse sensor schedule with sparsity q can achieve a $(1+\epsilon)$ approximation unless q is at least $\Omega(n/\epsilon^2)$, we start by considering the dynamics in (2)

with system matrices $\widetilde{A} = [\boldsymbol{e}_2 \quad \boldsymbol{e}_3 \quad \dots \quad \boldsymbol{e}_{n+1} \quad \boldsymbol{e}_1]$, and

$$\widetilde{C} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

where $\boldsymbol{e}_i$'s are the standard basis for $\mathbb{R}^{n+1}$. For this system, the t-step observability matrix $\widetilde{\mathcal{O}}(t)$, as defined in (4), has a unique structure: the rows of this matrix, $\widetilde{\boldsymbol{o}}_i^\top$'s, are the columns of the vertex-edge incidence matrix of a complete digraph of size $n+1$. The t-step observability Gramian matrix $\widetilde{\mathcal{X}}(t) = 2(n+1)I_{n+1} - 2J_{n+1}$ is twice the (unnormalized) Laplacian of the undirected complete graph K_{n+1}. Here, I_n denotes the $n \times n$ identity matrix, and J_n denotes the $n \times n$ all-ones matrix.

All the $\widetilde{\boldsymbol{o}}_i$ are orthogonal to the all-ones vector $\mathbb{1}$, which makes the observability Gramian rank deficient and thus unobservable. We define the observable instance of the system by expressing it in an orthogonal basis of this subspace: select any orthonormal basis $\boldsymbol{z}_1, \dots, \boldsymbol{z}_n$ of the subspace of $\mathbb{R}^{n+1}$ orthogonal to $\mathbb{1}$, and we now define our system matrices of the observable system as $A \coloneqq Z^\top \widetilde{A} Z$ and $C \coloneqq \widetilde{C} Z$ for $Z = [\boldsymbol{z}_1, \dots, \boldsymbol{z}_n] \in \mathbb{R}^{(n+1)\times n}$. Remember that $t = n+1$, so we can determine the t-step observability Gramian matrix $\mathcal{X}(t)$ for the projected system using the equation $\mathcal{O}^\top(t)\mathcal{O}(t) = \sum_i \boldsymbol{o}_i \boldsymbol{o}_i^\top = Z^\top \widetilde{\mathcal{X}}(t) Z = 2(n+1) \cdot I_n$, where $\boldsymbol{o}_i^\top = \widetilde{\boldsymbol{o}}_i^\top Z$ denotes the i-th row of the observability matrix for the projected system.

For the observable system, we consider the weighted sensor schedule $\mathrm{vec}(\Pi)(i) = \frac{q}{P(n+1,2)}$, where notation $P(n,r)$ signifies the permutation of r elements from a set of n elements. Then the weighted t-step observability Gramian matrix is given by $\sum_i \mathrm{vec}(\Pi)(i) \boldsymbol{o}_i \boldsymbol{o}_i^\top = \frac{2q}{n} I_n$. It is evident that $\|\mathrm{vec}(\Pi)\|_1 = q$.

Consider now any unweighted sensor scheduling solution $S \in P([n+1], 2)$ [4] of the optimization problem (24) with L-systemic design. We can consider the set S as the edges of an unweighted graph G with a vertex set of $[n+1]$. The Laplacian matrix L_G for this graph is defined as $L_G = \sum_{\{i,j\} \in S} \widetilde{\boldsymbol{o}}_{ij} \widetilde{\boldsymbol{o}}_{ij}^\top$, where each $\widetilde{\boldsymbol{o}}_{ij}$ represents a vector associated with the edge $\{i,j\} \in S$. This vector takes the value 1 in the i-th coordinate, -1 in the j-th coordinate, and 0 in all other coordinates. If the objective value of S is within an ϵ constant of the objective value of Π (as required by (8)), then according to Claim 1, the smallest eigenvalue of the unweighted sparse t-step observability Gramian matrix, i.e., $\sum_{\{i,j\} \in S} \boldsymbol{o}_{ij} \boldsymbol{o}_{ij}^\top$, where $\boldsymbol{o}_{ij}^\top = \widetilde{\boldsymbol{o}}_{ij}^\top Z$, is at least $(1-\epsilon)\frac{2q}{n}$. Since $\sum_{\{i,j\} \in S} \boldsymbol{o}_{ij} \boldsymbol{o}_{ij}^\top = Z^\top L_G Z$, this means that the second smallest eigenvalue of L_G is at least $(1-\epsilon)\frac{2q}{n}$. The average degree Δ of G is $\frac{2q}{n+1}$. Therefore, we have a graph G on $n+1$ vertices with average degree Δ for which the second smallest eigenvalue of its Laplacian is at least $(1-\epsilon)\left(\frac{n+1}{n}\right)\Delta \geq (1-\epsilon)\Delta$.

The authors of [23] have recently extended the Alon-Boppana bound to not necessarily regular graphs, but with worse constants (refer to Theorem C.2). Note that the proof of Theorem C.2 does not require the graph G to be simple, i.e., parallel edges are allowed. According to their findings, the second smallest eigenvalue of the Laplacian of graph G is bounded by $\Delta - c\sqrt{\Delta}$, where c is an absolute positive constant, for sufficiently large n relative to Δ. Recall that the existence of rounded solutions that perform closely to our weighted instance sensor schedule Π requires that, for all sufficiently large n, graph G with an average degree $\Delta = \frac{2q}{n+1}$ satisfies $\lambda_2(L_G) \geq (1-\epsilon)\Delta$, where $\lambda_2(\cdot)$ returns the second smallest eigenvalue. Combining this with the results of [23], we have $\Delta \geq \frac{c^2}{\epsilon^2}$ or $q \geq \frac{c^2}{2}\frac{n}{\epsilon^2}$. These results show that for this class of observable linear systems, the number of active sensors has to be at least $\Omega(n/\epsilon^2)$ to guarantee a $(1+\epsilon)$ approximation. Therefore, our results in Corollary 1 are within a constant factor of optimal and cannot be improved.

[4] $P([n+1], 2)$ is the set containing all ordered pairs that result from selecting 2 distinct elements from a set of $\{1, \dots, n+1\}$.

V. Numerical Example

In this section, we assess the practical implications of our theoretical findings through a benchmark example in distributed frequency control. We then proceed to compare the obtained results with those achieved by several competing algorithms.

A. Power Network

For this example, we use the IEEE 39-bus test system (also known as the 10-machine New England power system) [9], [24]. Our goal is to select sensors for wide-area damping control in the power system, aimed at suppressing generator fluctuations and ensuring synchronization.

The continuous-time swing dynamics for the power system are given by $\mathfrak{m}_i \ddot{\theta}_i + \mathfrak{d}_i \dot{\theta}_i = -\sum_{j \sim i} k_{ij}(\theta_i - \theta_j) + u_i$, where θ_i represents the voltage (or rotor) angle of the generator at bus i (in rad), $\mathfrak{m}_i$ is the inertia coefficient of the generator at bus i (in pu-sec^2/rad), $\mathfrak{d}_i$ is the damping coefficient of the generator at bus i (in pu-sec/rad), and the notation $j \sim i$ indicates that bus j and i are adjacent (connected).

We introduce the notation $\omega_i \coloneqq \dot{\theta}_i$, representing the frequency of the generator at bus i. We assume that the power grid comprises $\hat{n} = 10$ generators. The state-space representation of the swing equation used for frequency control in power networks is described by:

$$\begin{bmatrix} \dot{\boldsymbol{\theta}}(t) \\ \dot{\boldsymbol{\omega}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{0} & I \\ -M^{-1}L & -M^{-1}D \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}(t) \\ \boldsymbol{\omega}(t) \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ M^{-1} \end{bmatrix} \boldsymbol{u}(t)$$
$$\boldsymbol{y}(t) = C_c \begin{bmatrix} \boldsymbol{\theta}(t) \\ \boldsymbol{\omega}(t) \end{bmatrix}. \tag{29}$$

In this representation, $M = \mathrm{diag}(\mathfrak{m}_1, \dots, \mathfrak{m}_{\hat{n}})$, $D = \mathrm{diag}(\mathfrak{d}_1, \dots, \mathfrak{d}_{\hat{n}})$, and L is the Laplacian matrix (see [25] for detailed calculations). The vectors $\boldsymbol{\theta}(t) = [\theta_1, \dots, \theta_{\hat{n}}]^\top(t)$ and $\boldsymbol{\omega}(t) = [\omega_1, \dots, \omega_{\hat{n}}]^\top(t)$ represent the rotor angle and frequency vectors, respectively. The vector $\boldsymbol{u}(t)$ represents the mechanical power input, and the output matrix is denoted by C_c.

In numerical implementations, we assume that only the rotor frequency is available for measurement at each generator, i.e., $C_c = [\ \mathbf{0} \quad I\ ]$. The system is then discretized into a discrete-time LTI system with system matrices A, B, and C and a sampling time of 0.2 seconds. These matrices have been obtained from [25].

In the next, we apply our sparsification framework to the IEEE 39-bus test system. We solve the convex relaxation

(25) using CVX, a package designed for specifying and solving convex programs [26]. To obtain the relaxed solution for A- and E-optimality, we utilize the SDP solver included in the CVX package. However, for D-optimality, we employ an experimental successive approximation algorithm due to the involvement of the log function in the model. We formulate each instance of the convex relaxation problem as an SDP and then apply the solver. For this purpose, we leverage the SDP formulations discussed in [27, Section 7.5].

B. Results and Discussions

Since it is only assumed that frequencies of the generators are available to measure, the number of available sensors p is 10. Recall that the performance loss ϵ is some number in the interval $(0, 1/2]$ according to Corollary 1. If we assume ϵ and q take their maximum allowed values, i.e., $\epsilon = 1/2$ and $q = 10t$, then, one can show that based on the theoretical limit $q \geq 45n/\epsilon^2$ of this corollary, t must be greater than or equal to $18n$. However, to obtain sparse schedule, q should be less than $10t$, and we prefer to choose as small as possible number for ϵ. If we deviate from the maximum values for ϵ and q, the estimation time horizon t must take a considerably larger value than $18n$. To illustrate this, let us consider a scenario where $q = 5t$ and $\epsilon = 1/100$. In this case, t needs to be on the order of magnitude of $10^5 \times n$, which can be challenging to achieve in practical applications.

To address this issue, we have chosen to bypass the theoretical constraint on the parameter q and set t equal to n during simulation. However, during a specific iteration of Algorithm 1, it may become difficult to determine an index $i^{(\ell)}$ using (21) due to the modification we have made. This situation has motivated us to devise an alternative termination criterion for Algorithm 1. In practical applications, we terminate the algorithm when there are no values of $i^{(\ell)} \in \mathsf{M}^{(\ell-1)}$ that result in a positive denominator in (21). We have observed that this new stopping rule speeds up the algorithm by preventing it from running to completion. However, this acceleration comes at the expense of reduced algorithm performance.

In Fig. 3, the top row illustrates three weighted sparse sensor schedules computed for three distinct L-systemic measures (D-, A-, and E-optimality) applied to the IEEE 39-bus dynamics (29). Corresponding unweighted sparse sensor schedules, derived using Algorithm 1, are displayed in the second row of this figure. The performance of Algorithm 1 is also compared with four competing algorithms in Fig. 4. The plot displays the D-optimality performance measure of the t-step observability Gramian matrix of the sparse system against the average number of active sensors q/t. Smaller values of D-optimality, obtained by $(\det \mathcal{X}_{\hat{s}}(t))^{-1}$, indicate better performance.

Each data point on the uniform sampling curve is obtained by randomly sampling q sensors from the pool of tp available sensors uniformly. Conversely, the weighted sampling curve's data points are generated through sampling based on the weighted distribution $\mathrm{vec}(\Pi')/q$, where Π' represents the fractional solution of (25). For both of these curves, the sampling process is repeated 10 times at each data point, and the result with the lower D-optimality value is reported. In the case of the greedy static and greedy time-varying curves, we implement Algorithms 8 and 9 as outlined in [9]. In the greedy static approach, the same set of sensors is selected at each time step throughout the time horizon. In contrast, the greedy time-varying method optimizes both sensor selection and activation times to maximize the reduction in the L-systemic metric of the t-step observability Gramian matrix.

As expected, all the algorithms outperform the greedy static algorithm. It appears that our algorithm slightly underperforms compared to both the greedy time-varying and weighted sampling algorithms. However, drawing a fair conclusion from this observation is complicated for the following reasons. Firstly, the introduction of the new alternative stopping rule prevents our rounding algorithm from reaching its full potential, making the plot unable to showcase the best possible result for Algorithm 1. Secondly, in practical scenarios, we have observed that the choice of ϵ significantly impacts the algorithm's output performance. Nevertheless, to prevent premature termination of the algorithm due to the new stopping rule, we are constrained to run the algorithm with ϵ values greater than or equal to $1/1000$.

With an average of seven active sensors, the respective running times for the greedy time-varying, our algorithm, and the greedy static methods are 0.30, 0.17, and 0.02 seconds. Additionally, both sampling algorithms exhibit a running time of approximately 0.01 seconds[5]. For this comparison, we run MATLAB on a laptop with Intel Core i7-8750H CPU @ 2.20GHz 2.21GHz and 16.00 GB of RAM. The proposed rounding algorithm is computationally efficient as its time complexity is linear in the number of available sensors, tp (when both q and n are small). However, the time complexity of the greedy algorithm scales quadratically or even cubically with the number of available sensors and rapidly becomes intractable as the time horizon to estimate, t, increases [20].

We demonstrate that the resulting sparse sensor configuration obtained by the proposed algorithm renders an observable sparse system, but there is no guarantee that the other competitors (excluding the weighted sampling) provide an observable sensor configuration. Furthermore, Algorithm 1 does not require always initiating with a fractional solution obtained from a convex relaxation; however, it can be initialized by any weighted sensor schedule (e.g., the operating schedule of a real plant) and yields an unweighted sparse sensor schedule with a performance guarantee. Finally, Algorithm 1 provides a performance guarantee for all L-systemic performance measures, including a class of non-submodular metrics.

VI. Concluding Remarks

This paper addresses the problem of time-varying sensor selection for discrete-time LTI systems. We leverage recent advancements in theoretical computer science and online learning to establish a polynomial-time framework. This framework transforms a provided weighted sensor schedule into an unweighted sparse sensor schedule with a limited number of active sensors, approximating the performance of the weighted schedule in terms of observability.

Additionally, we illustrate that when we supply the proposed rounding algorithm with an approximate fractional solution obtained from a convex relaxation of a combinatorial optimization problem with an L-systemic objective, we

[5]The running times do not account for the time required for continuous optimization.

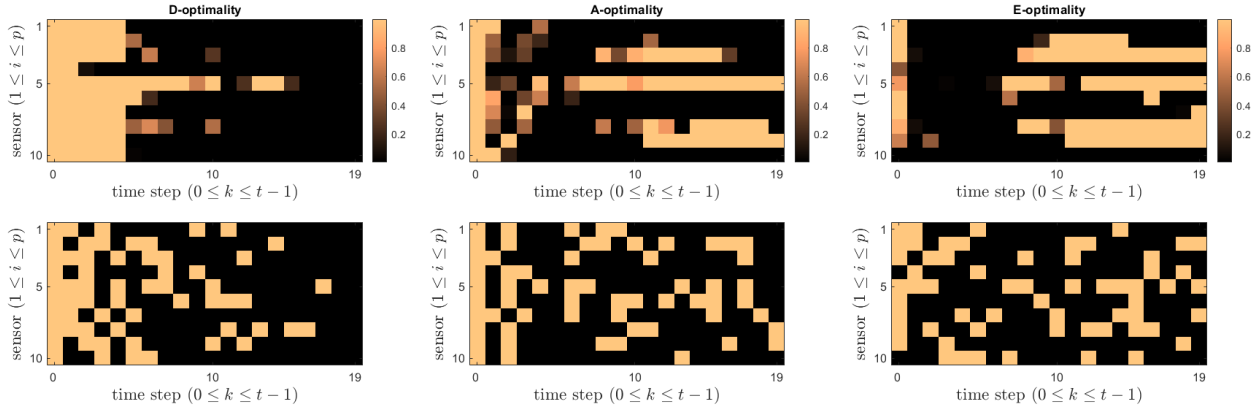

Fig. 3: These subplots, from left to right, illustrate three weighted sparse sensor schedules for D-, A- and E-optimality in the top row and their corresponding unweighted sparse sensor schedules in the bottom row for the 10-machine New England power system. Weighted sensor schedules are computed using the CVX package, while their unweighted counterparts are derived using Algorithm 1. Experimental settings include $q = 30$, $\epsilon = 1/6$, and a time horizon of $t = 20$. Color-coding in each subplot represents the corresponding scalar values of the schedule, with the brightest copper denoting a value of one. (Indices: $i \in [10]$, $k+1 \in [20]$).

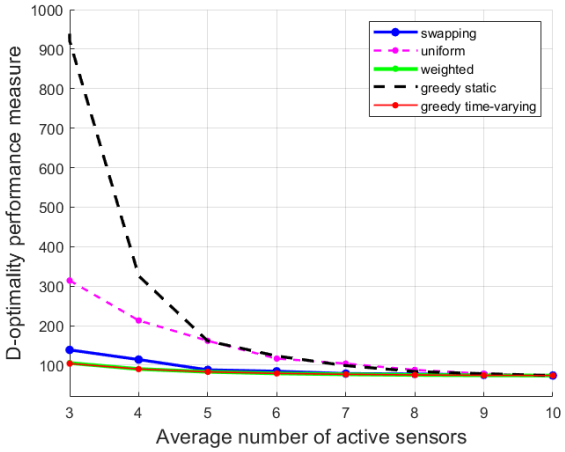

Fig. 4: This figure compares our swapping algorithm (Algorithm 1) with four competitor algorithms for creating an unweighted sparse sensor schedule in the 10-machine New England power system (29). It shows the D-optimality values of the sparse t-step observability Gramian matrix versus the average number of active sensors per time, q/t.

can attain a $(1+\epsilon)$ near-optimal approximation solution, guaranteeing the observability of the linear system. This algorithm exhibits computational efficiency, applicability to various non-submodular performance measures, and provides a performance bound relative to the optimal solution. Furthermore, it works for any weighted sensor schedule, not necessarily obtained from a convex relaxation. A potential future direction is to explore the extension of these results to nonlinear systems.

Appendix

A. Definition of Terms

This section aims to collect and elucidate the definitions of key concepts that are fundamental to comprehending the paper.

1) Similarity Transformation (Whitening): Given a non-singular coordinate transformation $\mathcal{T} \in \mathbb{R}^{n\times n}$, the new system realization of the state-space $\left[\begin{array}{c|c} A & B \\ \hline C & \mathbf{0} \end{array}\right]$ is given by

$$\left[\begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & \mathbf{0} \end{array}\right] = \left[\begin{array}{c|c} \mathcal{T}^{-1}A\mathcal{T} & \mathcal{T}^{-1}B \\ \hline C\mathcal{T} & \mathbf{0} \end{array}\right]. \tag{30}$$

Furthermore, the t-step observability Gramian matrix of the transformed realization (30) is given by $\hat{\mathcal{X}}(t) = \mathcal{T}^{\top}\mathcal{X}(t)\mathcal{T}$, where $\mathcal{X}(t)$ is the t-step observability Gramian matrix of the original dynamics.

Definition 2 (Whitening): We define the *whitening* transformation as $\mathcal{T} := \mathcal{X}^{-1/2}(t)$. This change of coordinates converts the t-step observability Gramian matrix of the transformed realization, $\hat{\mathcal{X}}(t)$, to the identity matrix, I.

2) Regret Minimization over Density Matrices: In this regret game, at each round $\ell \in [T]$, the player is tasked with selecting an action $\mathcal{A}^{(\ell)}$ from a set of density matrices $\mathsf{Y}_{n\times n}$ (also referred to as the *action space*), which consists of PSD matrices with a trace of one. Subsequently, the player incurs a loss $\langle \mathcal{A}^{(\ell)}, F^{(\ell)} \rangle$ which depends on their chosen action as well as the adversary's symmetric matrix action $F^{(\ell)} \in \mathbb{R}^{n\times n}$. The regret is defined as the difference between the total loss suffered by the player and the loss suffered by a posteriori best fixed action $\mathcal{U}^* := \arg\min_{\mathcal{U}\in\mathsf{Y}_{n\times n}} \sum_{\ell=1}^{T}\langle \mathcal{U}, F^{(\ell)}\rangle$, i.e.,

$$\text{Regret}_{\mathcal{U}^*}(T) := \sum_{\ell=1}^{T}\langle \mathcal{A}^{(\ell)}, F^{(\ell)}\rangle - \sum_{\ell=1}^{T}\langle \mathcal{U}^*, F^{(\ell)}\rangle. \tag{31}$$

For a PSD matrix $\mathcal{A}$ (the proof is omitted due to space limitations),

$$\lambda_{\min}(\mathcal{A}) = \min_{\mathcal{U}\in\mathsf{Y}_{n\times n}} \langle \mathcal{A}, \mathcal{U}\rangle. \tag{32}$$

Given (32), one can reformulate the regret as

$$\text{Regret}_{\mathcal{U}^*}(T) = \sum_{\ell=1}^{T} \langle \mathcal{A}^{(\ell)}, F^{(\ell)} \rangle - \lambda_{\min}\Big(\sum_{\ell=1}^{T} F^{(\ell)} \Big), \quad (33)$$

which implies that, in hindsight, the optimal choice of $\mathcal{U}$ can be interpreted as the rank-1 projection onto the eigenvector corresponding to the minimum eigenvalue of $\sum_{\ell=1}^{T} F^{(\ell)}$.

The objective of the player is to minimize the regret (33). To achieve this objective in this paper, we employ the Follow-The-Regularized-Leader strategy with the $\ell_{1/2}$-regularizer defined as $w(X) = -2 \operatorname{Trace}(X^{1/2})$. This strategy guides the player's actions according to the choice of action [16]:

$$\mathcal{A}^{(\ell)} = \Big(c^{(\ell)} I + \alpha\, H^0 + \alpha \cdot \sum_{i=1}^{\ell-1} F^{(i)} \Big)^{-2}, \quad (34)$$

where $\alpha > 0$ is the learning rate, H^0 is a PSD matrix such that for certain real constants $c^{(0)}$, $c^{(0)} I + \alpha\, H^0 \succ 0$, and $c^{(\ell)} \in \mathbb{R}$ is a unique constant that ensures $c^{(\ell)} I + \alpha\, H^0 + \alpha \cdot \sum_{i=0}^{\ell-1} F^{(i)} \succ 0$, and $\operatorname{Trace} \mathcal{A}^{(\ell)} = 1$.

The regret upper bound for $\ell_{1/2}$-strategy (34) according to different choices for feedback matrices $F^{(\ell)}$ are previously obtained in [16] and [20]. In the following lemma, we summarize these results.

Lemma 1 (Theorem 3.2 of [16] and Lemma 2.5 of [20]): Assume the regret game (31), where the player's actions $\{\mathcal{A}^{(\ell)}\}_{\ell \in [T]}$ are defined according to the $\ell_{1/2}$-strategy (34) for some positive learning rate α, and the feedback $F^{(\ell)} = \boldsymbol{u}^{(\ell)} \boldsymbol{u}^{(\ell)\top}$ for $\boldsymbol{u}^{(\ell)} \in \mathbb{R}^n$ are rank-1 matrices that satisfy $\alpha \cdot \langle A^{(\ell)^{1/2}}, F^{(\ell)} \rangle > -1$ for all ℓ. Then, for every $\mathcal{U} \in \mathsf{Y}_{n \times n}$:

$$\text{Regret}_{\mathcal{U}}(T) \le \alpha \sum_{\ell=1}^{T} \frac{\langle A^{(\ell)}, F^{(\ell)} \rangle \cdot \langle A^{(\ell)^{1/2}}, F^{(\ell)} \rangle}{1 + \alpha \cdot \langle A^{(\ell)^{1/2}}, F^{(\ell)} \rangle} + \frac{2\sqrt{n}}{\alpha}. \quad (35)$$

If instead $F^{(\ell)} = \boldsymbol{u}^{(\ell)} \boldsymbol{u}^{(\ell)\top} - \boldsymbol{v}^{(\ell)} \boldsymbol{v}^{(\ell)\top}$ for $\boldsymbol{u}^{(\ell)}, \boldsymbol{v}^{(\ell)} \in \mathbb{R}^n$ are rank-2 feedback matrices, as long as $\alpha \cdot \langle A^{(\ell)^{1/2}}, \boldsymbol{v}^{(\ell)} \boldsymbol{v}^{(\ell)\top} \rangle < 1/2$ for all ℓ, the following inequality holds for every $\mathcal{U} \in \mathsf{Y}_{n \times n}$:

$$\text{Regret}_{\mathcal{U}}(T) \le 2\alpha \sum_{\ell=1}^{T} \Bigg[-\frac{\langle A^{(\ell)}, \boldsymbol{u}^{(\ell)} \boldsymbol{u}^{(\ell)\top} \rangle \cdot \langle A^{(\ell)^{1/2}}, \boldsymbol{u}^{(\ell)} \boldsymbol{u}^{(\ell)\top} \rangle}{1 + 2\alpha \cdot \langle A^{(\ell)^{1/2}}, \boldsymbol{u}^{(\ell)} \boldsymbol{u}^{(\ell)\top} \rangle} + \frac{\langle A^{(\ell)}, \boldsymbol{v}^{(\ell)} \boldsymbol{v}^{(\ell)\top} \rangle \cdot \langle A^{(\ell)^{1/2}}, \boldsymbol{v}^{(\ell)} \boldsymbol{v}^{(\ell)\top} \rangle}{1 - 2\alpha \cdot \langle A^{(\ell)^{1/2}}, \boldsymbol{v}^{(\ell)} \boldsymbol{v}^{(\ell)\top} \rangle} \Bigg] + \frac{\psi_{\mathcal{A}^{(0)}}(\mathcal{U})}{\alpha}, \quad (36)$$

where $\psi_X(Y) \coloneqq \langle X^{-1/2}, Y \rangle + \operatorname{Trace} X^{1/2} - 2 \operatorname{Trace} Y^{1/2}$ is the so-called *Bregman divergence* for the $\ell_{1/2}$-regularizer, and $\psi_{\mathcal{A}^{(0)}}(\mathcal{U}) \le 2\sqrt{n} + \alpha \cdot \langle H^0, \mathcal{U} \rangle$.

Note that Lemma 1 provides regret upper bounds for all actions, including the optimal fixed action $\mathcal{U}^*$.

3) Transient Observability Function: We define the *transient observability function* for dynamics (2) as $L_o(\boldsymbol{x}_0, t) = \widetilde{\boldsymbol{y}}^\top(t) \widetilde{\boldsymbol{y}}(t)$, where t is the time horizon for estimation, $\boldsymbol{x}_0$ denotes the initial state, and $\widetilde{\boldsymbol{y}}(t)$ is the vector of collected measurements defined in (3) when $\boldsymbol{u}(k) = \boldsymbol{0}$ for all $k+1 \in [t]$. If we assume that the initial state $\boldsymbol{x}_0$ is non-deterministic, following a normal distribution with zero mean and covariance matrix Σ, i.e., $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{0}, \Sigma)$, according to [28, Theorem 2], the expected value of the transient function is $\mathbb{E}_{\boldsymbol{x}_0} L_o(\boldsymbol{x}_0, t) = \langle \mathcal{X}(t), \Sigma \rangle$, where $\mathcal{X}(t)$ represents the t-step observability Gramian matrix (6), and $\Sigma = \mathbb{E}[\boldsymbol{x}_0 \boldsymbol{x}_0^\top]$.

B. Essential Results

In this section, we present key mathematical results necessary for comprehending the content of the paper. The following lemma repeats the classical result of [29, paragraph 48, page 104].

Lemma 2 (Hoffman and Wielandt, 1953): Let $\mathcal{C} = \mathcal{A} + \mathcal{B}$, where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are symmetric $n \times n$ matrices with eigenvalues α_i, β_i, and γ_i, respectively. Then, we have $\sum_{i=1}^{n} (\gamma_i - \alpha_i)^2 \le \|\mathcal{B}\|^2$.

Let $\lambda_k(\cdot)$ denote the k-th largest eigenvalue, and $\mathcal{A}$ and $\mathcal{B}$ be PD matrices. Utilizing Lemma 2, for all $k \in [n]$, we obtain

$$|\lambda_k(\mathcal{A}) - \lambda_k(\mathcal{B})|^2 \le \sum_{i=1}^{n} |\lambda_i(\mathcal{A}) - \lambda_i(\mathcal{B})|^2 \le \|\mathcal{A} - \mathcal{B}\|^2. \quad (37)$$

Proposition 4: For any PSD matrix $X \in \mathbb{R}^{n \times n}$,

$$\operatorname{Trace} X^{1/2} \le \sqrt{n \cdot \operatorname{Trace} X}. \quad (38)$$

Proof: Let $\sigma(\cdot)$ denote the function that outputs the vector of all eigenvalues of its input PSD matrix. Thus, $\operatorname{Trace} X^{1/2} = \big\|\sigma(X^{1/2})\big\|_1 \le \sqrt{n} \cdot \big\|\sigma(X^{1/2})\big\|_2 = \sqrt{n} \cdot \sqrt{\|\sigma(X)\|_1}$, where the inequality holds due to (40). ∎

Lemma 3 (Claim 2.14 of [20]): For any $X \in \{M \in \mathbb{S}_+^n : \lambda_{\min}(M) \le 1\}$, let $\mathcal{A} = (cI + \alpha X)^{-2}$ for some positive number α, where a unique number $c \in \mathbb{R}$ ensures the positive semidefiniteness of matrix $\mathcal{A}$ and $\operatorname{Trace} \mathcal{A} = 1$. Then, the following statements hold: $\langle \mathcal{A}^{1/2}, X \rangle \le \frac{n}{\alpha} + \sqrt{n}$; and $\langle \mathcal{A}, X \rangle \le \lambda_{\min}(X) + \sqrt{n}/\alpha$.

C. Missing Proofs

The purpose of this section is to present the proofs that complement the theoretical findings discussed in the main text of the paper.

Proof of Proposition 1. It can be shown that:

- Convexity is valid for all metrics except D-optimality. For this specific metric, it is a well-established practice to utilize the *negative log-determinant*, denoted as $-\log\det \mathcal{X}(t)$, which is convex [30].
- The properties of homogeneity and monotonicity defined in Definition 1 apply to all the measures (refer to [9, Section III]).

However, establishing the Lipschitz property for these measures necessitates careful consideration. We demonstrate this property for each measure in the table individually in the following. We begin by demonstrating the Lipschitz continuity of the A-optimal objective with respect to the ℓ_1 norm. Since trace is a linear operation, our focus shifts to proving the Lipschitz continuity of matrix inversion. More precisely, we aim to establish that for the mapping $\rho : X \mapsto X^{-1}$, there exists a constant $L > 0$ such that $\|\mathcal{A}^{-1} - \mathcal{B}^{-1}\|_1 \le L \cdot \|\mathcal{A} - \mathcal{B}\|_1$, for all matrices $\mathcal{A}, \mathcal{B} \succeq \delta \cdot I \succ \boldsymbol{0}$. To this end,

$$\|\mathcal{A}^{-1} - \mathcal{B}^{-1}\|_1 = \|\mathcal{A}^{-1}(\mathcal{B} - \mathcal{A})\mathcal{B}^{-1}\|_1 \le \|\mathcal{A}^{-1}\|_1 \|\mathcal{B}^{-1}\|_1 \|\mathcal{B} - \mathcal{A}\|_1,$$

where the inequality holds because all induced matrix norms are submultiplicative. Furthermore,

$$\|\mathcal{A}^{-1}\|_1\|\mathcal{B}^{-1}\|_1\|\mathcal{B}-\mathcal{A}\|_1 \le \frac{n}{\lambda_{\min}(\mathcal{A})\cdot\lambda_{\min}(\mathcal{B})}\cdot\|\mathcal{B}-\mathcal{A}\|_1 \le \frac{n}{\delta^2}\cdot\|\mathcal{B}-\mathcal{A}\|_1, \tag{39}$$

which establishes the Lipschitz property with $L = n/\delta^2$. In (39), the first inequality holds because for any matrix $M \in \mathbb{R}^{m\times n}$, it is known that

$$\frac{1}{\sqrt{m}}\|M\|_1 \le \|M\|_2 \le \sqrt{n}\|M\|_1. \tag{40}$$

To demonstrate the Lipschitz continuity of E-optimality, we can say

$$\Big|\frac{1}{\lambda_{\min}(\mathcal{A})}-\frac{1}{\lambda_{\min}(\mathcal{B})}\Big| = \frac{1}{|\lambda_{\min}(\mathcal{A})\cdot\lambda_{\min}(\mathcal{B})|}|\lambda_{\min}(\mathcal{A})-\lambda_{\min}(\mathcal{B})| \le \frac{1}{\delta^2}\cdot|\lambda_{\min}(\mathcal{A})-\lambda_{\min}(\mathcal{B})|, \tag{41}$$

but from (37) and (40), we know that

$$|\lambda_{\min}(\mathcal{A})-\lambda_{\min}(\mathcal{B})| \le \|\mathcal{A}-\mathcal{B}\| \le \sqrt{n}\cdot\|\mathcal{A}-\mathcal{B}\|_1, \tag{42}$$

establishing the Lipschitz property with $L = \sqrt{n}/\delta^2$.

T-optimality exhibits Lipschitz continuity because

$$\Big|\frac{1}{\operatorname{Trace}\mathcal{A}}-\frac{1}{\operatorname{Trace}\mathcal{B}}\Big| = \frac{1}{|\operatorname{Trace}\mathcal{A}\cdot\operatorname{Trace}\mathcal{B}|}|\operatorname{Trace}\mathcal{A}-\operatorname{Trace}\mathcal{B}| \le \frac{1}{(n\delta)^2}\cdot|\operatorname{Trace}\mathcal{A}-\operatorname{Trace}\mathcal{B}|,$$

and given that trace is a linear function. For D-optimality, we need to establish that the mapping $\rho : X \mapsto \log\det X$ is Lipschitz continuous. We know that the gradient of this mapping, i.e., $\partial(\log\det X)/\partial X = X^{-1}$, is upper-bounded in a positive semidefinite sense by $1/\delta\cdot I$. Therefore, the proof is concluded by $\|\partial(\log\det X)/\partial X\|_1 = \|X^{-1}\|_1 \le 1/\delta$.

Proof of Theorem 1. We present the proof in two steps:

- *First step:* We prove the following inequality hold:

$$\max_{i\in[tp]}\frac{\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top\rangle}{1+\alpha\cdot\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top\rangle} \ge \frac{1}{q+\alpha\sqrt{n}}. \tag{43}$$

- *Second step:* Using (43), we demonstrate the main result.

First step: We know $\hat{\mathcal{X}}_\pi(t) = \sum_{i\in[tp]}\operatorname{vec}(\Pi)(i)\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top = I$, and $\operatorname{Trace}\Sigma^{(\ell)} = 1$ since $\Sigma^{(\ell)} \in \mathsf{Y}_{n\times n}$. Therefore, it is not difficult to show that the weighted sum $\sum_{i\in[tp]}\operatorname{vec}(\Pi)(i)\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top\rangle$ is equal to one. Furthermore,

$$\textstyle\sum_{i\in[tp]}\operatorname{vec}(\Pi)(i)(1+\alpha\cdot\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top\rangle) \le q+\alpha\cdot\operatorname{Trace}{\Sigma^{(\ell)}}^{1/2},$$

but $\operatorname{Trace}{\Sigma^{(\ell)}}^{1/2} \le \sqrt{n\operatorname{Trace}\Sigma^{(\ell)}} = \sqrt{n}$ due to Proposition 4. This, according to averaging arguments, completes the proof for (43).

Second step: Given the result from (43), we can derive the regret bound according to (16) as follows

$$\lambda_{\min}(\hat{\mathcal{X}}_s^{(T)}(t)) \ge \frac{T}{q+\alpha\cdot\sqrt{n}} - \frac{2\sqrt{n}}{\alpha}. \tag{44}$$

To obtain the desired bound (17), we only need to set $\alpha = 4\sqrt{n}/\epsilon$ and $T = q \ge 8n/\epsilon^2$ in (44). Finally, let $\mathsf{M}^{(T)} = \{i^{(1)},\dots,i^{(T)}\}$ be the set of sensor indices that have been selected during the iterations of the game. Then, $S = \operatorname{vec}^{-1}(\sum_{k\in\mathsf{M}^{(T)}}\boldsymbol{e}_k)$ yields the sparse sensor schedule.

Proof of Theorem 2. Some parts of this proof are adopted from [20]. The proof of this theorem is quite extensive. To improve clarity and comprehension, we will break it down into the following main steps:

- **Step 1:** We demonstrate that as long as $\lambda_{\min}(\hat{\mathcal{X}}_s^{(\ell-1)}(t)) < 1-\epsilon$, it is possible to select indices $i^{(\ell)}$ and $j^{(\ell)}$ based on (21) and (23), respectively.
- **Step 2:** We establish that τ satisfies the inequality $\tau \le (3-\epsilon)/3q$, where τ is defined as:

$$\tau \coloneqq \min_{i\in\mathsf{N}^{(\ell-1)}}\frac{\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i{\hat{\mathbf{o}}_i}^\top\rangle}{1-2\alpha\,\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathbf{o}}_i{\hat{\mathbf{o}}_i}^\top\rangle}. \tag{45}$$

- **Step 3:** For $\varkappa \coloneqq \tau+\epsilon/3q$, we show the inequality:

$$\max_{i\in[tp]\setminus\mathsf{M}^{(\ell-1)}}\frac{\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i{\hat{\mathbf{o}}_i}^\top\rangle}{1+2\alpha\,\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathbf{o}}_i{\hat{\mathbf{o}}_i}^\top\rangle} \ge \varkappa \tag{46}$$

- **Step 4:** We prove the main result.

Step 1: Let $\mathsf{M}^{(\ell-1)}$ denote the set of selected sensors at or before iteration ℓ of the regret game explained in the theorem. Now, consider $\hat{\mathcal{X}}_s^{(\ell-1)}(t) = \sum_{i\in\mathsf{M}^{(\ell-1)}}\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top$. If $\lambda_{\min}(\hat{\mathcal{X}}_s^{(\ell-1)}(t)) \ge 1-\epsilon$, our goal is achieved, and we can set $S^{(\ell-1)} = \operatorname{vec}^{-1}(\sum_{k\in\mathsf{M}^{(\ell-1)}}\boldsymbol{e}_k)$, providing the solution to Problem 1. If the condition is not met, i.e.,

$$\lambda_{\min}(\hat{\mathcal{X}}_s^{(\ell-1)}(t)) < 1-\epsilon, \tag{47}$$

then, as the first step of the proof, we aim to establish that $\mathsf{N}^{(\ell-1)}$ is not empty. In other words, there must be at least one sensor i in $\mathsf{M}^{(\ell-1)}$ such that $2\alpha\cdot\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathbf{o}}_i{\hat{\mathbf{o}}_i}^\top\rangle < 1$.

By contradiction, we assume that there does not exist a sensor i satisfying the condition, i.e., $2\alpha\cdot\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top\rangle \ge 1$ for all i in $\mathsf{M}^{(\ell-1)}$. Therefore,

$$\sum_{i\in\mathsf{M}^{(\ell-1)}}2\alpha\cdot\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top\rangle = 2\alpha\cdot\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathcal{X}}_s^{(\ell-1)}(t)\rangle \ge q. \tag{48}$$

On the other hand, according to Lemma 3,

$$2\alpha\cdot\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathcal{X}}_s^{(\ell-1)}(t)\rangle \le 2n+2\alpha\sqrt{n}, \tag{49}$$

because $\hat{\mathcal{X}}_s^{(\ell-1)}(t) \succeq 0$, and $\lambda_{\min}(\hat{\mathcal{X}}_s^{(\ell-1)}(t)) < 1$ due to (47). However, (49), for the choices of $\alpha = 3\sqrt{n}/\epsilon$, $\epsilon\in(0,1)$, and $q \ge 45n/\epsilon^2 > 8n/\epsilon$, contradicts (48). Thus, $\mathsf{N}^{(\ell-1)}$ is not an empty set, and we also can conclude that

$$\sum_{i\in\mathsf{M}^{(\ell-1)}}1-2\alpha\,\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathbf{o}}_i{\hat{\mathbf{o}}_i}^\top\rangle > 0. \tag{50}$$

Step 2: Given (45), the following inequality holds for all $i\in\mathsf{M}^{(\ell-1)}$:

$$\Big(1-2\alpha\cdot\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top\rangle\Big)\cdot\tau \le \langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top\rangle, \tag{51}$$

because when $i\notin\mathsf{N}^{(\ell-1)}$, the left-hand side is non-positive, while the right-hand side $\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top\rangle$ is always non-negative since the matrix $\Sigma^{(\ell)}$ is, by definition, PSD. This implies

$(q-2\alpha\cdot\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathcal{X}}_s^{(\ell-1)}(t)\rangle)\cdot\tau\le\langle\Sigma^{(\ell)},\hat{\mathcal{X}}_s^{(\ell-1)}(t)\rangle$, and so

$$\tau\overset{(a)}{\le}\frac{\langle\Sigma^{(\ell)},\hat{\mathcal{X}}_s^{(\ell-1)}(t)\rangle}{q-2\alpha\cdot\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathcal{X}}_s^{(\ell-1)}(t)\rangle}\overset{(b)}{\le}\frac{\sqrt{n}/\alpha+\lambda_{\min}(\hat{\mathcal{X}}_s^{(\ell-1)}(t))}{q-2n-2\alpha\sqrt{n}}\overset{(c)}{\le}\frac{1-2\epsilon/3}{q(1-8\epsilon/45)}\overset{(d)}{\le}\frac{3-\epsilon}{3q},\quad(52)$$

which finishes the proof for Step 2. In (52), (a) holds because the denominator is strictly positive due to (50), (b) holds due to Lemma 3, inequality (c) holds by applying the choices $\alpha=3\sqrt{n}/\epsilon$, $q\ge 45n/\epsilon^2$, $\epsilon<1$, and (47), and finally, the last inequality holds because $\epsilon>0$.

Step 3: Based on *averaging arguments*, demonstrating inequality (46) is equivalent to proving

$$\sum_{i\in[tp]\setminus\mathsf{M}^{(\ell-1)}}\mathrm{vec}(\Pi)(i)\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle-\varkappa\cdot\sum_{i\in[tp]\setminus\mathsf{M}^{(\ell-1)}}\mathrm{vec}(\Pi)(i)\times\Big(1+2\alpha\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle\Big)\ge 0\quad(53)$$

where $\pi_{j,k+1}\ge 0$ for all $j\in[p]$ and $k+1\in[t]$. In (53),

$$\begin{aligned}&\sum_{i\in[tp]\setminus\mathsf{M}^{(\ell-1)}}\mathrm{vec}(\Pi)(i)\Big(1+2\alpha\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle\Big)\\&\le\Big(q-\sum_{i\in\mathsf{M}^{(\ell-1)}}\mathrm{vec}(\Pi)(i)\Big)+2\alpha\sum_{i\in[tp]}\mathrm{vec}(\Pi)(i)\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle\\&=q-\sum_{i\in\mathsf{M}^{(\ell-1)}}\mathrm{vec}(\Pi)(i)+2\alpha\,\mathrm{Trace}\,{\Sigma^{(\ell)}}^{1/2},\end{aligned}\quad(54)$$

and

$$\textstyle\sum_{i\in[tp]\setminus\mathsf{M}^{(\ell-1)}}\mathrm{vec}(\Pi)(i)\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle=\Big\langle\Sigma^{(\ell)},I-\sum_{i\in\mathsf{M}^{(\ell-1)}}\mathrm{vec}(\Pi)(i)\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\Big\rangle=1-\sum_{i\in\mathsf{M}^{(\ell-1)}}\mathrm{vec}(\Pi)(i)\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle.\quad(55)$$

In the above simplifications, we utilize the facts that $\|\mathrm{vec}(\Pi)\|_1\le q$, $\hat{\mathcal{X}}_\pi(t)=\sum_{i\in[tp]}\mathrm{vec}(\Pi)(i)\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i^\top=I$, and $\Sigma^{(\ell)}\in\mathsf{Y}_{n\times n}$. Substituting (55) and (54) in (53), one can show the left-hand side of (53) is greater than or equal to

$$1-\varkappa q-2\alpha\varkappa\sqrt{n}-\sum_{i\in\mathsf{M}^{(\ell-1)}}\mathrm{vec}(\Pi)(i)(\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle-\varkappa),\quad(56)$$

where we also use the result of Proposition 4. Since scalars $\pi_{j,k+1}\le 1$ for all $j\in[p]$ and $k+1\in[t]$, we can show that (56) is greater than or equal to

$$1-\varkappa q-2\alpha\varkappa\sqrt{n}-\sum_{i\in\mathsf{M}^{(\ell-1)}}\max\big\{(\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle-\varkappa),0\big\},\quad(57)$$

or equivalently

$$1-\varkappa q-2\alpha\varkappa\sqrt{n}-\sum_{i\in\mathsf{M}^{(\ell-1)}}(\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle-\varkappa)-\sum_{i\in\mathsf{M}^{(\ell-1)}}\max\Big\{(\varkappa-\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle),0\Big\}.\quad(58)$$

Furthermore, due to Lemma (3), (58) is greater than or equal to

$$1-2\alpha\varkappa\sqrt{n}-\sqrt{n}/\alpha-\lambda_{\min}(\hat{\mathcal{X}}_s^{\ell-1}(t))-\sum_{i\in\mathsf{M}^{(\ell-1)}}\max\Big\{(\varkappa-\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle),0\Big\}.\quad(59)$$

Now, let us take care of the last term in (59). To this end, we have

$$\sum_{i\in\mathsf{M}^{(\ell-1)}}\max\big\{(\varkappa-\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle),0\big\}=\sum_{i\in\overline{\mathsf{M}}^{(\ell-1)}}(\varkappa-\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle),$$

where $\overline{\mathsf{M}}^{(\ell-1)}\coloneqq\big\{i\in\mathsf{M}^{(\ell-1)}:\varkappa-\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle\ge 0\big\}$. Then,

$$\sum_{i\in\overline{\mathsf{M}}^{(\ell-1)}}(\varkappa-\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle)=(\varkappa-\tau)\,\mathrm{card}(\overline{\mathsf{M}}^{(\ell-1)})+\sum_{i\in\overline{\mathsf{M}}^{(\ell-1)}}(\tau-\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle),\quad(60)$$

$$\text{and}\quad\sum_{i\in\overline{\mathsf{M}}^{(\ell-1)}}(\tau-\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle)\overset{(e)}{\le}2\alpha\tau\sum_{i\in\overline{\mathsf{M}}^{(\ell-1)}}\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle\overset{(f)}{\le}2\tau(n+\alpha\sqrt{n}),\quad(61)$$

where inequality (e) holds due to (51), and for (f), we use the result of Lemma 3 and the fact that $\overline{\mathsf{M}}^{(\ell-1)}\subseteq\mathsf{M}^{(\ell-1)}$. One can put (61) back to (60) to obtain

$$\sum_{i\in\mathsf{M}^{(\ell-1)}}\max\big\{(\varkappa-\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_i\hat{\mathbf{o}}_i{}^\top\rangle),0\big\}\le(\varkappa-\tau)q+2\tau(n+\alpha\sqrt{n})\le\frac{\epsilon}{3}+\frac{8n}{\epsilon q},\quad(62)$$

where the inequalities hold because $\overline{\mathsf{M}}^{(\ell-1)}\subseteq\mathsf{M}^{(\ell-1)}$, by definition $\varkappa-\tau=\epsilon/3q$, $\alpha=3\sqrt{n}/\epsilon$, $\epsilon<1$, and $\tau\le(3-\epsilon)/3q<1/q$ according to (52). Finally, one can leverage $\alpha=3\sqrt{n}/\epsilon$, $\varkappa<1/q$ and (47) together with (62) to show that (59) is lower bounded by $\epsilon/3-14n/\epsilon q$. Because q is chosen to be greater than or equal to $45n/\epsilon^2$, this lower bound is always non-negative which finishes the proof for the third step.

Step 4: In the first step, we show that as long as $\lambda_{\min}(\hat{\mathcal{X}}_s^{(\ell-1)}(t))<1-\epsilon$, we have the option to choose indices $i^{(\ell)}$ and $j^{(\ell)}$ according to (21) and (23), respectively. The Second and the third steps of the proof imply that for the pair $(i^{(\ell)},j^{(\ell)})$

$$\frac{\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_{j^{(\ell)}}\hat{\mathbf{o}}_{j^{(\ell)}}{}^\top\rangle}{1+2\alpha\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathbf{o}}_{j^{(\ell)}}\hat{\mathbf{o}}_{j^{(\ell)}}{}^\top\rangle}-\frac{\langle\Sigma^{(\ell)},\hat{\mathbf{o}}_{i^{(\ell)}}\hat{\mathbf{o}}_{i^{(\ell)}}{}^\top\rangle}{1-2\alpha\langle{\Sigma^{(\ell)}}^{1/2},\hat{\mathbf{o}}_{i^{(\ell)}}\hat{\mathbf{o}}_{i^{(\ell)}}{}^\top\rangle}\ge\frac{\epsilon}{3q}.\quad(63)$$

All in all, at the end of some iteration $\ell\le T$, if $\lambda_{\min}(\hat{\mathcal{X}}_s^{(\ell)}(t))\ge 1-\epsilon$, then we are done. Otherwise, for all $\ell\in[T]$, there exist indices $i^{(\ell)}$ and $j^{(\ell)}$ such that equation (63) holds. Referring to (19) and (63) for the regret bound of the game, we can write $\lambda_{\min}(\hat{\mathcal{X}}_s^{(T)}(t))\ge\Big(\sum_{i=1}^T\frac{\epsilon}{3q}\Big)-\frac{2\sqrt{n}}{\alpha}=\frac{T\epsilon}{3q}-\frac{2}{3}\epsilon$, which gives the result because T is chosen to be $3q/\epsilon$.

References

[1] B. W. Brunton, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Sparse sensor placement optimization for classification," *SIAM Journal on Applied Mathematics*, vol. 76, no. 5, pp. 2099–2122, 2016.

[2] P. Müller and H. Weber, "Analysis and optimization of certain qualities of controllability and observability for linear dynamical systems," *Automatica*, vol. 8, no. 3, pp. 237–246, 1972.

[3] T. H. Summers, F. L. Cortesi, and J. Lygeros, "On submodularity and controllability in complex dynamical networks," *IEEE Transactions on Control of Network Systems*, vol. 3, no. 1, pp. 91–101, 2015.

[4] A. Olshevsky, "On (non)supermodularity of average control energy," *IEEE Transactions on Control of Network Systems*, 2018, to appear.

[5] V. Tzoumas, K. Gatsis, A. Jadbabaie, and G. J. Pappas, "Resilient monotone submodular function maximization," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 2017, pp. 1362–1367.

[6] S. N. Ravi, V. Ithapu, S. Johnson, and V. Singh, "Experimental design on a budget for sparse linear models and applications," in *International Conference on Machine Learning*, 2016, pp. 583–592.

[7] A. S. A. Dilip, "The controllability gramian, the hadamard product, and the optimal actuator/leader and sensor selection problem," *IEEE Control Systems Letters*, vol. 3, no. 4, pp. 883–888, 2019.

[8] A. Olshevsky, "On a relaxation of time-varying actuator placement," *IEEE Control Systems Letters*, vol. 4, no. 3, pp. 656–661, 2020.

[9] M. Siami, A. Olshevsky, and A. Jadbabaie, "Deterministic and randomized actuator scheduling with guaranteed performance bounds," *IEEE Transactions on Automatic Control*, vol. 66, no. 4, pp. 1686–1701, 2020.

[10] M. Siami and A. Jadbabaie, "A separation theorem for joint sensor and actuator scheduling with guaranteed performance bounds," *Automatica*, vol. 119, p. 109054, 2020.

[11] S. D. Bopardikar, "A randomized approach to sensor placement with observability assurance," *Automatica*, vol. 123, p. 109340, 2021.

[12] A. Olshevsky, "Minimal controllability problems," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 3, pp. 249–258, 2014.

[13] R. Vafaee and M. Siami, "Learning-based sensor selection with guaranteed performance bounds," in 2022 American Control Conference, accepted.

[14] J. L. Crassidis and J. L. Junkins, *Optimal estimation of dynamic systems*. Chapman and Hall/CRC, 2004.

[15] B. T. Hinson, "Observability-based guidance and sensor placement," Ph.D. dissertation, university of washington, 2014.

[16] Z. Allen-Zhu, Z. Liao, and L. Orecchia, "Spectral sparsification and regret minimization beyond matrix multiplicative updates," in *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, 2015, pp. 237–245.

[17] Z. Allen-Zhu, Y. Li, A. Singh, and Y. Wang, "Near-optimal design of experiments via regret minimization," in *International Conference on Machine Learning*. PMLR, 2017, pp. 126–135.

[18] V. V. Fedorov, *Theory of optimal experiments*. Elsevier, 2013.

[19] A. Jadbabaie, A. Olshevsky, and M. Siami, "Limitations and tradeoffs in minimum input selection problems," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 185–190.

[20] Z. Allen-Zhu, Y. Li, A. Singh, and Y. Wang, "Near-optimal discrete optimization for experimental design: A regret minimization approach," *Mathematical Programming*, vol. 186, no. 1, pp. 439–478, 2021.

[21] A. Civril and M. Magdon-Ismail, "On selecting a maximum volume sub-matrix of a matrix and related problems," *Theoretical Computer Science*, vol. 410, no. 47-49, pp. 4801–4811, 2009.

[22] M. D. Summa, F. Eisenbrand, Y. Faenza, and C. Moldenhauer, "On largest volume simplices and sub-determinants," in *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 2014, pp. 315–323.

[23] A. Nikolov, M. Singh, and U. Tantipongpipat, "Proportional volume sampling and approximation algorithms for a-optimal design," *Mathematics of Operations Research*, vol. 47, no. 2, pp. 847–877, 2022.

[24] I. E. Atawi, "An advance distributed control design for wide-area power system stability," Ph.D. dissertation, University of Pittsburgh, 2013.

[25] G. Fazelnia, R. Madani, A. Kalbat, and J. Lavaei, "Convex relaxation for optimal distributed control problems," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 206–221, 2016.

[26] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[27] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[28] D. Georges, "The use of observability and controllability gramians or functions for optimal sensor and actuator location in finite-dimensional systems," in *Proceedings of 1995 34th IEEE Conference on Decision and Control*, vol. 4. IEEE, 1995, pp. 3319–3324.

[29] J. Wilkinson, "The algebraic eigenvalue problem," in *Handbook for Automatic Computation, Volume II, Linear Algebra*. Springer-Verlag New York, 1971.

[30] F. Bach, "Submodular functions: from discrete to continuous domains," *Mathematical Programming*, vol. 175, no. 1, pp. 419–459, 2019.

Reza Vafaee is a PhD candidate in electrical engineering program at Northeastern University (NEU). He is conducting research focusing on the mathematical and algorithmic foundations of network optimization and control with applications in robotics and power systems. His research builds on spectral graph theory, optimization, estimation theory, and distributed computing. Prior to joining NEU, he received his second master with highest distinction in pure and applied mathematics from Montclair State University, US, in 2020.

Milad Siami (S'12-M'18-SM'23) received his dual B.Sc. degrees in electrical engineering and pure mathematics from Sharif University of Technology in 2009, and his M.Sc. degree in electrical engineering from Sharif University of Technology in 2011. He received his M.Sc. and Ph.D. degrees in mechanical engineering from Lehigh University in 2014 and 2017, respectively. From 2017 to 2019, he was a postdoctoral associate at MIT Institute for Data, Systems, and Society. Since 2019, he has been an Assistant Professor in the Department of Electrical & Computer Engineering at Northeastern University, Boston, MA. Dr. Siami's research interests include distributed control systems, distributed optimization, and sparse sensing, which is applied in robotics and cyber–physical systems. He has been recognized with several awards and fellowships, including a Gold Medal at the National Mathematics Olympiad in Iran, the Best Student Paper Award at the 5th IFAC Workshop on Distributed Estimation and Control in Networked Systems, and the Rossin College Doctoral Fellowship at Lehigh University.