Greedy Sensor Selection
Leveraging Submodularity Based on Volume Ratio of Information Ellipsoid

Liu, Lingya; Hua, Cunqing; Xu, Jing; Leus, Geert; Wang, Yiyin

DOI
10.1109/TSP.2023.3283047

Publication date
2023

Document Version
Final published version

Published in
IEEE Transactions on Signal Processing

Greedy Sensor Selection: Leveraging Submodularity Based on Volume Ratio of Information Ellipsoid

Lingya Liu, *Member, IEEE*, Cunqing Hua, *Member, IEEE*, Jing Xu, Geert Leus, *Fellow, IEEE*, and Yiyin Wang, *Member, IEEE*

***Abstract*—This article focuses on greedy approaches to select the most informative k sensors from N candidates to maximize the Fisher information, i.e., the determinant of the Fisher information matrix (FIM), which indicates the volume of the information ellipsoid (VIE) constructed by the FIM. However, it is a critical issue for conventional greedy approaches to quantify the Fisher information properly when the FIM of the selected subset is rank-deficient in the first $(n-1)$ steps, where n is the problem dimension. In this work, we propose a new metric, i.e., the Fisher information intensity (FII), to quantify the Fisher information contained in the subset $\mathcal{S}$ with respect to that in the ground set $\mathcal{N}$ specifically in the subspace spanned by the vectors associated with $\mathcal{S}$. Based on the FII, we propose to optimize the ratio between VIEs corresponding to $\mathcal{S}$ and $\mathcal{N}$. This volume ratio is composed of a nonzero (i.e., the FII) and a zero part. Moreover, the volume ratio can be easily calculated based on a change of basis. A cost function is developed based on the volume ratio and proven monotone submodular. A greedy algorithm and its fast version are proposed accordingly to guarantee a near-optimal solution with a complexity of $\mathcal{O}(Nkn^3)$ and $\mathcal{O}(Nkn^2)$, respectively. Numerical results demonstrate the superiority of the proposed algorithms under various measurement settings.**

***Index Terms*—Greedy sensor selection, Fisher information intensity, change of basis, volume ratio, submodularity.**

I. Introduction

WIRELESS sensor networks (WSNs) have attracted great attention in applications and services related to surveillance, environmental and climate monitoring, etc. Sensors are deployed for distributed sensing at particular locations to extract relevant information [1]. Specifically, for linear observation models, the physical field of interest can be estimated from sensor observations/measurements by solving a linear inverse problem [2], [3]. Thus, sensor selection, i.e., the problem of selecting k sensors that acquire the most informative measurements from an available pool of N sensors to minimize the estimation error, is a fundamental design task in WSNs. It is essentially a combinatorial optimization problem involving $\binom{N}{k}$ searches and intractable even for small-scale networks [4]. This motivates numerous heuristics and approximate algorithms for the sensor selection problem.

A. Related Prior Works

The sensor selection problem is formulated as an optimization problem in [2], where the objective is to maximize the logarithm of the determinant ($\log\det$) of the inverse error covariance matrix. The problem is approximated to a convex problem by relaxing the Boolean constraints, which indicate whether the sensor is selected or not. Such kind of convex relaxation has been applied in many other sensor selection works [4], [5], [6], [7]. For example, it is used in [4] for a general non-linear observation model to optimize different performance criteria, e.g., the D-optimality, A-optimality, and E-optimality criterion. In this context, solutions to the related convex optimization problem with a simple rounding procedure may lead to an ill-conditioned observation model due to the approximation of constraints [8]. The local optimization technique [2] and the iterative rounding procedure [7] can enhance the results but with a high computational cost. As a result, convex relaxation methods are very effective for small-scale problems [2], [9], but less compelling for large-scale ones [3], [8].

On the other hand, greedy methods enjoy much less computational cost compared with convex relaxation methods. Greedy methods determine the sensors one by one via optimizing some proxies of the estimation accuracy, such as the volume of the confidence ellipsoid (VCE) [3], the mean squared error (MSE) [10], and the worst-case error variance (WCEV) [8]. These proxies correspond to the D-, A-, and E-optimality criterion, respectively. The theoretical foundation of some greedy approaches is drawn from submodular function optimization [11]. For instance, the VCE index is a monotone submodular function and optimized in the pioneering work [3] by a greedy algorithm, which guarantees a $(1-1/e)$ optimal solution with a run time of $\mathcal{O}(Nkn^2)$, where e is Euler's number and n is the problem dimension. Numerical results verify that optimizing the MSE [10], [12] and WCEV [8] via greedy methods can also obtain similar

Manuscript received 17 September 2022; revised 24 March 2023; accepted 17 May 2023. Date of publication 21 June 2023; date of current version 6 July 2023. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yuxin Chen. This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1807504, in part by the Natural Science Foundation of China under Grants 61773264, 61801295, 62171278, and 62101327, and in part by the Oceanic Interdisciplinary Program of Shanghai Jiao Tong University under Grant SL2020MS011. *(Corresponding author: Yiyin Wang.)*

Lingya Liu and Jing Xu are with the School of Communication and Electronic Engineering, East China Normal University, Shanghai 200241, China (e-mail: lyliu@cee.ecnu.edu.cn; jxu@ce.ecnu.edu.cn).

Cunqing Hua and Yiyin Wang are with the School of Electronic, Information, and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: cqhua@sjtu.edu.cn; yiyinwang@sjtu.edu.cn).

Geert Leus is with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 Delft, The Netherlands (e-mail: g.j.t.leus@tudelft.nl).

Digital Object Identifier 10.1109/TSP.2023.3283047

suboptimal solutions. For example, a frame potential-based cost function called FrameSense is proposed in [12] to approximate the MSE but it has the submodularity property, and thus guarantees a near-optimal solution. However, FrameSense is unable to exploit the information contained in the norms of the measurement vectors, as it can only be applied to measurement vectors with uniform norms. The maximal projection on minimum eigenspace (MPME) algorithm is proposed for optimizing the WCEV in [8]. In each step, it selects the sensor whose observation has the maximum projection onto the minimum eigenspace spanned by the currently selected sensor set. In the examples of [8], the MPME algorithm outperforms the convex relaxation method [2], FrameSense [12], and SparSenSe [13] in terms of the WCEV and MSE, where SparSenSe determines the minimum number of required sensors for a prescribed MSE by utilizing the sparsity of the selection vector. The MPME algorithm is also computationally efficient with a complexity of $\mathcal{O}(Nkn^2)$. More recently, a fast MSE-based sampling algorithm is proposed for linear-model signals varying from vectors to tensors in [14], where the complexity is reduced to $\mathcal{O}(Nk^2)$. It achieves the same performance as [3] for vector signals.

In the aforementioned greedy methods, the optimization metric is closely related to the currently selected sensors. This raises two fundamental issues that directly affect the performance of greedy algorithms. Firstly, the optimization metric in the first $(n-1)$ steps is ill-conditioned due to the rank-deficiency of the FIM. It is remedied by adding a full-rank matrix with very small eigenvalues [3]. However, the influence of this remedy is yet to be investigated as it changes the rank of the observation matrix. Moreover, the dependency of the metric on the current set leads to a cumulative effect. An improper selection in previous steps, where some better solutions are missed, may deteriorate the overall solution. In some recent works, it has been alleviated either by iteratively reserving a group of top L (i.e., the group size) suboptimal candidates but at the cost of an increased complexity of $\mathcal{O}(NkLn^2)$ [15], or by introducing randomization into the selection to reduce the algorithm complexity for problems with large N [16].

B. Contributions

In this article, we are particularly interested in greedy sensor selection approaches for linear measurement models [2], [3]. We define the determinant of the Fisher information matrix (FIM) as the volume of its corresponding information ellipsoid (Info-E). The volume of the Info-E (VIE) is a valid counterpart of the VCE. That is, minimizing the VCE is equivalent to maximizing $\log\det(\mathrm{FIM})$, which can be geometrically interpreted as to maximize the corresponding VIE. With all measurement vectors of the network available, the FIM associated with the ground sensor set, hereinafter called the full FIM, is necessarily nonsingular. The Info-E determined by the full FIM confines all the Info-E's related to the subsets of the ground sensor set. In order to make the best use of the profile of the largest Info-E corresponding to the full FIM, we propose to optimize the *volume ratio* of the VIE for the selected sensors to the maximum VIE. We summarize the contributions of this paper as follows.

1) We propose a novel metric, i.e., the *Fisher information intensity* (FII), which has a clear geometric interpretation. It represents the ratio between the VIEs associated with the selected subset ($\mathcal{S}$) and the ground set ($\mathcal{N}$) in the subspace spanned by the vectors related to $\mathcal{S}$. The FII is more elaborate than the Fisher information especially when the FIM of $\mathcal{S}$ is singular. We further propose to optimize the ratio between VIEs corresponding to $\mathcal{S}$ and $\mathcal{N}$ in the n-dimensional (n-D) space, where the nonzero part of the volume ratio is the FII.
2) The volume ratio can be achieved using the transformed measurement vectors. The linear transform is accomplished using the basis that results from the eigenvalue decomposition of the full FIM. The change of basis can also be adopted by existing greedy algorithms for performance improvement.
3) A novel cost function is constructed based on the volume ratio. It consists of two components corresponding to the nonzero and zero parts of the volume ratio, respectively. The nonzero component is exactly the FII, while the zero component is replaced by a very small constant when the number of selected sensors is smaller than n. The proposed cost function is proven monotone submodular.
4) A greedy algorithm is proposed to optimize the proposed cost function in two domains, which correspond to the first n steps and the remaining $(k-n)$ steps, respectively. As the cost function is monotone submodular, a $(1-1/e)$ optimal solution is guaranteed. A fast greedy algorithm is designed to further reduce the complexity to $\mathcal{O}(Nkn^2)$.

The rest of the paper is organized as follows. The sensor selection problem based on linear measurements is briefly reviewed in Section II, where its geometric interpretations are presented. The problem is then transformed into the volume ratio formulation based on the FII. Section III presents the greedy algorithm and its fast counterpart. Section IV provides a theoretical analysis of the proposed algorithm comparing it with a conventional submodular algorithm. Numerical results are illustrated in Section V, whereas Section VI concludes the paper.

Notations: Upper case calligraphic and bold face upper (lower) case letters, e.g., $\mathcal{A}$ and $\mathbf{A}$ ($\mathbf{a}$), denote sets and matrices (column vectors), respectively. We use $[\mathbf{a}]_i$ to denote the ith element of the vector $\mathbf{a}$. An identity and a zero matrix with proper dimensions are represented by $\mathbf{I}$ and $\mathbf{0}$, respectively. The operators $(\cdot)^{\mathrm{T}}$, $\|\cdot\|$, $\det(\cdot)$ and $\mathrm{rank}(\cdot)$ correspond to the transpose, norm, determinant and rank operations. The expression $\mathbf{A}\succ\mathbf{0}$ ($\mathbf{A}\succeq\mathbf{0}$) denotes $\mathbf{A}$ is a positive (semi)definite matrix. The canonical basis is given by $\{\mathbf{e}_i\}_{i=1}^n$, where $\mathbf{e}_i$ is a zero vector of length n except for the ith entry which is 1.

II. Problem Analysis and Statement

In this section, we will comprehensively review the sensor selection problem based on linear measurement models and show new insights from a geometric perspective. We further design a novel performance metric, i.e., the volume ratio of the VIEs, to recast the sensor selection problem as the maximization

of the volume ratio between the VIE of the selected sensors and the maximum VIE associated with the ground set of sensors.

A. Sensor Selection Based on Linear Models

Consider a physical field, where $\boldsymbol{\alpha} \in \mathbb{R}^n$ collects the parameters of interest, measured by a full network of N ($N \gg n$) sensors with linear measurements given by

$$y_i = \boldsymbol{\phi}_i^{\mathrm{T}} \boldsymbol{\alpha} + z_i, \quad i = 1, \ldots, N, \tag{1}$$

where $\boldsymbol{\phi}_i \in \mathbb{R}^n$ is the measurement vector of the ith sensor. The noise z_i of the ith sensor is independent identically distributed (i.i.d.) additive white Gaussian noise (AWGN) with zero mean and unit variance[1], i.e., $z_i \sim \mathcal{CN}(0, 1)$. The full measurement matrix is then given by $\boldsymbol{\Phi}_{\mathcal{N}} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_N]^{\mathrm{T}} \in \mathbb{R}^{N \times n}$, where $\mathcal{N}$ is the collection of N sensors. The known measurement matrix $\boldsymbol{\Phi}_{\mathcal{N}}$ is assumed to be a full column rank matrix.

Assuming a subset $\mathcal{S} \subseteq \mathcal{N}$, the measurement matrix $\boldsymbol{\Phi}_{\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}| \times n}$ ($|\mathcal{S}|$ is the cardinality of $\mathcal{S}$) is the submatrix of $\boldsymbol{\Phi}_{\mathcal{N}}$ that selects the rows related to $\mathcal{S}$. Denoting the maximum-likelihood estimate of $\boldsymbol{\alpha}$ measured by the sensors in $\mathcal{S}$ as $\hat{\boldsymbol{\alpha}}$, the estimation error $\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}$ has zero mean and covariance

$$\boldsymbol{\Sigma}_{\mathcal{S}} = \left(\boldsymbol{\Phi}_{\mathcal{S}}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}}\right)^{-1}. \tag{2}$$

Note that $\boldsymbol{\Phi}_{\mathcal{S}}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}}$ is required to be full-rank. Otherwise, $\boldsymbol{\alpha}$ cannot be estimated due to the rank-deficiency of $\boldsymbol{\Phi}_{\mathcal{S}}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}}$. According to the data model (1), the error covariance $\boldsymbol{\Sigma}_{\mathcal{S}}$ attains the Cramér-Rao bound (CRB) defined by the Fisher information matrix (FIM), i.e., $\boldsymbol{\Sigma}_{\mathcal{S}} = \mathbf{F}^{-1}(\mathcal{S})$, where $\mathbf{F}(\mathcal{S})$ is the FIM associated with $\mathcal{S}$ and given by

$$\mathbf{F}(\mathcal{S}) = \boldsymbol{\Phi}_{\mathcal{S}}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}} = \sum_{i \in \mathcal{S}} \boldsymbol{\phi}_i \boldsymbol{\phi}_i^{\mathrm{T}}. \tag{3}$$

We now introduce the *information ellipsoid*[2] based on $\mathbf{F}(\mathcal{S})$ as follows.

Definition 1 (Information Ellipsoid): The information ellipsoid (Info-E) of a FIM $\mathbf{F}(\mathcal{S})$ is defined as the set of points fulfilling the following condition

$$\left\{ \boldsymbol{w} \in \mathbb{R}^n \mid \boldsymbol{w}^{\mathrm{T}} \mathbf{F}^{-1}(\mathcal{S})\, \boldsymbol{w} \leq 1 \right\}, \tag{4}$$

where the center of the Info-E is at the origin, and $\mathbf{F}(\mathcal{S})$ determines how far the ellipsoid extends in every direction from the origin.

Obviously, the directions and the lengths of the semi-axes of the Info-E are decided by the eigenvectors and the square roots of the eigenvalues of $\mathbf{F}(\mathcal{S})$, respectively. Fig. 1(a) depicts an example of the Info-E with $n = 2$, where the Info-E determined by $\mathbf{F}(\mathcal{N})$ with $\mathcal{N} = \{1, 2\}$ is illustrated.

Note that the Info-E is a concept similar to the *η-confidence ellipsoid* [2], which represents the ellipsoid that contains $\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}$ with probability η. It is well-known that the volume of the η-confidence ellipsoid (VCE) is a scalar measure of the estimation quality. Based on the definition of the Info-E, minimizing the VCE is now equivalent to maximizing the volume of the Info-E (VIE), i.e., $\det(\mathbf{F}(\mathcal{S}))$. Thus, the VCE and VIE are dual counterparts. The sensor selection problem based on a linear model is to find a subset $\mathcal{S}$ of k ($k \geq n$) sensors out of N available candidates, to maximize the VIE associated with $\mathcal{S}$. It is formulated as [2]

$$\mathbf{P}_0: \quad \max_{\mathcal{S} \subseteq \mathcal{N}} \log\det\left(\mathbf{F}(\mathcal{S})\right) \tag{5}$$

$$\text{s.t.} \quad |\mathcal{S}| = k. \tag{5a}$$

In particular, $\mathbf{P}_0$ selects $\mathcal{S}$ to maximize $\log\det(\mathbf{F}(\mathcal{S}))$, which indeed gives a quantification of how informative the collection of the measurements $\boldsymbol{\phi}_i$ ($i \in \mathcal{S}$) is. In a nutshell, $\mathbf{P}_0$ aims to select the most informative subset in terms of the Fisher information.

Problem $\mathbf{P}_0$ indeed is a combinatorial optimization, which is hard to solve. In [3], a greedy algorithm is proposed to address the approximated $\mathbf{P}_0$, which is denoted as $\mathbf{P}_{\text{approx}}$ and given by

$$\mathbf{P}_{\text{approx}}: \quad \max_{\mathcal{S} \subseteq \mathcal{N}} \log\det\left(\mathbf{F}(\mathcal{S}) + \varepsilon \mathbf{I}\right) \tag{6}$$

$$\text{s.t.} \quad |\mathcal{S}| = k, \tag{6a}$$

with $\varepsilon > 0$ being a very small constant. Note that $\mathbf{F}(\mathcal{S})$ is replaced by $\mathbf{F}(\mathcal{S}) + \varepsilon \mathbf{I}$ in (6). This modification ensures the applicability of the greedy approach to $\mathbf{P}_{\text{approx}}$, since $\log\det(\mathbf{F}(\mathcal{S}) + \varepsilon \mathbf{I})$ is a monotone submodular function w.r.t. $\mathcal{S}$ [3, Lemma 1]. Let us briefly review the greedy algorithm [3, Algorithm 1] for $\mathbf{P}_{\text{approx}}$. Adapting [3, eq. (13)] to the notations of this article, and assuming $l < k$ sensors have already been

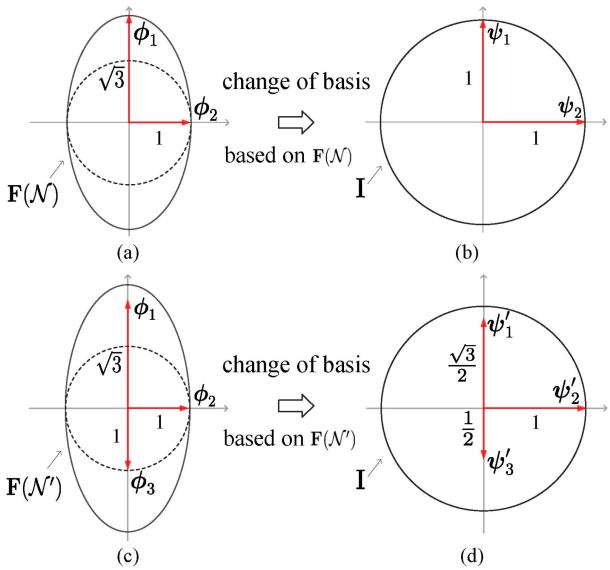


Fig. 1. Geometrical illustration of the Info-E: two 2-D measurement examples. (a) and (c) show the Info-E's associated with $\mathbf{F}(\mathcal{N})$ and $\mathbf{F}(\mathcal{N}')$ (the solid line) with $\mathcal{N} = \{1, 2\}$ and $\mathcal{N}' = \{1, 2, 3\}$, respectively, where $\boldsymbol{\phi}_1 = [0, \sqrt{3}]$, $\boldsymbol{\phi}_2 = [1, 0]$, and $\boldsymbol{\phi}_3 = [0, -1]$. (b) and (d) illustrate the Info-E's normalized as unit balls with transformed vectors $\boldsymbol{\psi}_i$'s and $\boldsymbol{\psi}'_i$'s, respectively. The vectors $\boldsymbol{\phi}_i$'s (or $\boldsymbol{\psi}_i$'s and $\boldsymbol{\psi}'_i$'s) are marked by red arrows, while the numbers attached are the $\|\boldsymbol{\phi}_i\|$'s (or $\|\boldsymbol{\psi}_i\|$'s and $\|\boldsymbol{\psi}'_i\|$'s). .

[1]With the knowledge of the noise variances $\{\sigma_i^2\}_{i=1}^n$, we can always prewhiten the noise (i.e., set the variance to 1) by $\boldsymbol{\phi}_i := \bar{\boldsymbol{\phi}}_i / \sigma_i$, where $\bar{\boldsymbol{\phi}}_i$ is the original ith measurement vector.

[2]The concept of *information ellipsoid* can be found in [17], [18], where it is specifically defined in the 2-dimensional (2-D) space for 2-D network positioning.

found, i.e., $|\mathcal{S}| = l$, the following problem needs to be solved in the $(l+1)$th iteration:

$$\begin{aligned} &\max\nolimits_{i\in\mathcal{N}\setminus\mathcal{S}} \operatorname{logdet}\left(\mathbf{F}(\mathcal{S}) + \varepsilon\mathbf{I} + \boldsymbol{\phi}_i\boldsymbol{\phi}_i^{\mathrm{T}}\right) \\ &\doteq \max\nolimits_{i\in\mathcal{N}\setminus\mathcal{S}} \boldsymbol{\phi}_i^{\mathrm{T}}(\mathbf{F}(\mathcal{S}) + \varepsilon\mathbf{I})^{-1}\boldsymbol{\phi}_i, \end{aligned} \tag{7}$$

where "$\doteq$" indicates the equivalence of two optimization problems. Here, the addition of $\varepsilon\mathbf{I}$ to $\mathbf{F}(\mathcal{S})$ helps to resolve the rank-deficiency of $\mathbf{F}(\mathcal{S})$ when $|\mathcal{S}| = l < n$. Thus, the impact of $\varepsilon\mathbf{I}$ is essential when $\mathbf{F}(\mathcal{S})$ is rank-deficient. As mentioned by [3], the auxiliary term $\varepsilon\mathbf{I}$ plays a role of prior knowledge about the covariance matrix of $\boldsymbol{\alpha}$. However, this kind of assumption cannot be validated. Furthermore, when the greedy algorithm starts from the null set $\mathcal{S} = \varnothing$ (i.e., $|\mathcal{S}| = l = 0$), the first sensor is selected by solving $\max_{i\in\mathcal{N}} \boldsymbol{\phi}_i^{\mathrm{T}}(\varepsilon\mathbf{I})^{-1}\boldsymbol{\phi}_i \doteq \max_{i\in\mathcal{N}} \|\boldsymbol{\phi}_i\|^2$. Such a selection is unfair since it only compares the norms of the measurement vectors, and neglects their direction correlations. Due to the cumulative effect, the improper selection in the first step may deteriorate the overall solution.

In order to solve $\mathbf{P}_0$ directly instead of solving $\mathbf{P}_{\text{approx}}$ and avoid the aforementioned issues, we recall the fact that $\mathcal{S}$ should be selected from the ground set $\mathcal{N}$, which provides a priori information. The profile of the Info-E with respect to (w.r.t.) $\mathbf{F}(\mathcal{N})$ would thus affect the optimal solution to $\mathbf{P}_0$. Hence, we raise the key question of how to make use of $\mathbf{F}(\mathcal{N})$ to properly quantify the measurement vectors for a fair comparison in the sensor selection. We answer this question in the following subsection.

B. Problem Reformulation and Analysis

1) Leveraging the Full FIM $\mathbf{F}(\mathcal{N})$: The FIM $\mathbf{F}(\mathcal{N})$ (i.e., $\boldsymbol{\Phi}_{\mathcal{N}}^{\mathrm{T}}\boldsymbol{\Phi}_{\mathcal{N}}$) creates an n-D Info-E with the maximum volume according to Definition 1. The Info-E w.r.t. $\mathbf{F}(\mathcal{S})$ is bounded by the one w.r.t. $\mathbf{F}(\mathcal{N})$, and they are exactly the same when $\mathcal{S} = \mathcal{N}$. Accordingly, $0 \le \det(\mathbf{F}(\mathcal{S}))/\det(\mathbf{F}(\mathcal{N})) \le 1$ for any $\mathcal{S} \subseteq \mathcal{N}$. It is thus intuitive to consider the profile of the Info-E determined by $\mathbf{F}(\mathcal{N})$ and investigate $\det(\mathbf{F}(\mathcal{S}))$ accordingly. This can be accomplished by *a change of basis*.

Note that $\mathbf{F}(\mathcal{N})$ is a real symmetric matrix. Thus, the eigenvalue decomposition of $\mathbf{F}(\mathcal{N})$ is denoted by

$$\mathbf{F}(\mathcal{N}) = \boldsymbol{\Phi}_{\mathcal{N}}^{\mathrm{T}}\boldsymbol{\Phi}_{\mathcal{N}} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathrm{T}}, \tag{8}$$

where $\boldsymbol{\Lambda} = \operatorname{diag}([\lambda_1, \ldots, \lambda_n])$ and $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ with λ_i's and $\mathbf{v}_i$'s $(i = 1, \ldots, n)$ being the eigenvalues and corresponding eigenvectors, respectively. Throughout this article, the eigenvalues of a matrix are by default arranged in nonincreasing order, i.e., $\lambda_i \ge \lambda_{i+1}$. Let us then decompose $\boldsymbol{\phi}_i$ in the orthogonal basis $\{\sqrt{\lambda_i}\mathbf{v}_i\}_{i=1}^n$ as

$$\begin{aligned} \boldsymbol{\phi}_i &= \mathbf{V}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{V}^{\mathrm{T}}\boldsymbol{\phi}_i \\ &= \underbrace{\left[\sqrt{\lambda_1}\mathbf{v}_1, \ldots, \sqrt{\lambda_n}\mathbf{v}_n\right]}_{\text{basis w.r.t. } \mathbf{F}(\mathcal{N})} \boldsymbol{\psi}_i \end{aligned} \tag{9}$$

with

$$\boldsymbol{\psi}_i := \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{V}^{\mathrm{T}}\boldsymbol{\phi}_i \tag{10}$$

being the transform of $\boldsymbol{\phi}_i$ w.r.t. the orthogonal basis $\{\sqrt{\lambda_i}\mathbf{v}_i\}_{i=1}^n$. Note that the norms of the orthogonal basis vectors are not necessarily 1. The linear transform (10) from $\boldsymbol{\phi}_i$ to $\boldsymbol{\psi}_i$ is a one-to-one mapping, which results in a matrix $\boldsymbol{\Psi}_{\mathcal{N}} \in \mathbb{R}^{N\times n}$ being the transform of $\boldsymbol{\Phi}_{\mathcal{N}}$ w.r.t. the basis $\{\sqrt{\lambda_i}\mathbf{v}_i\}_{i=1}^n$. It is given by

$$\boldsymbol{\Psi}_{\mathcal{N}} := [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_N]^{\mathrm{T}} = \boldsymbol{\Phi}_{\mathcal{N}}\mathbf{V}\boldsymbol{\Lambda}^{-\frac{1}{2}}, \tag{11}$$

with its ith row being $\boldsymbol{\psi}_i^{\mathrm{T}}$. Furthermore, we can also extend (9) to a matrix form as $\boldsymbol{\Phi}_{\mathcal{S}} = \boldsymbol{\Psi}_{\mathcal{S}}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{V}^{\mathrm{T}}$, where $\boldsymbol{\Psi}_{\mathcal{S}}$ is defined as $\boldsymbol{\Psi}_{\mathcal{N}}$ with $\mathcal{S}$ replacing $\mathcal{N}$ in (11). Thus, it can readily be derived that

$$\det(\mathbf{F}(\mathcal{S})) = \det\left(\mathbf{V}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Psi}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Psi}_{\mathcal{S}}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{V}^{\mathrm{T}}\right) \tag{12}$$

$$= \underbrace{\det\left(\boldsymbol{\Psi}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Psi}_{\mathcal{S}}\right)}_{\text{ratio factor}} \cdot \underbrace{\det\left(\boldsymbol{\Lambda}\right)}_{\det(\mathbf{F}(\mathcal{N}))}, \tag{13}$$

where $\det(\mathbf{F}(\mathcal{S}))$ is decomposed into two independent parts, i.e., $\det(\boldsymbol{\Psi}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Psi}_{\mathcal{S}})$ and $\det(\boldsymbol{\Lambda})$. The FIM $\mathbf{F}(\mathcal{N})$ is thereby elegantly taken into account in (13). Meanwhile, $\det(\boldsymbol{\Psi}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\Psi}_{\mathcal{S}})$ can be interpreted as a ratio factor. It takes a value from $[0, 1]$ and its maximum value is reached when $\mathcal{S} = \mathcal{N}$ as

$$\det(\boldsymbol{\Psi}_{\mathcal{N}}^{\mathrm{T}}\boldsymbol{\Psi}_{\mathcal{N}}) = \det(\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{V}^{\mathrm{T}}\boldsymbol{\Phi}_{\mathcal{N}}^{\mathrm{T}}\boldsymbol{\Phi}_{\mathcal{N}}\mathbf{V}\boldsymbol{\Lambda}^{-\frac{1}{2}}) = \det(\mathbf{I}) = 1. \tag{14}$$

We can better understand the impact of the change of basis with the help of two 2-D selection examples. Fig. 1(a) shows the first example, where $\boldsymbol{\phi}_1$ is orthogonal to $\boldsymbol{\phi}_2$ and $\mathcal{N} = \{1, 2\}$. Because $\|\boldsymbol{\phi}_1\| > \|\boldsymbol{\phi}_2\|$, the greedy algorithm [3] first chooses $\boldsymbol{\phi}_1$. However, we consider the first and the second sensors to be equally important despite the fact that $\|\boldsymbol{\phi}_1\| > \|\boldsymbol{\phi}_2\|$, as they contribute the same (i.e., 100%) in their specific direction. It can be proven that when the info-E of $\mathbf{F}(\mathcal{N})$ is normalized to a unit ball via the change of basis as shown in Fig. 1(b), then the norms of the transformed vectors are the same, i.e., $\|\boldsymbol{\psi}_1\| = \|\boldsymbol{\psi}_2\|$. Fig. 1(c) shows the second example, where a new vector $\boldsymbol{\phi}_3$ in parallel to $\boldsymbol{\phi}_1$ is added and $\mathcal{N}' = \{1, 2, 3\}$. The greedy algorithm [3] keeps $\boldsymbol{\phi}_1$ as the first selection, as it has the maximum norm. However, only the second sensor associated with $\boldsymbol{\phi}_2$ contributes 100% in its direction. On the other hand, both $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_3$ contribute to the FIM in the direction of $\boldsymbol{\phi}_1$, where their importance is proportional to their norms. Therefore, we consider the second sensor to be the most important and choose it at the first step. This can also be justified by the fact that when the info-E of $\mathbf{F}(\mathcal{N}')$ is normalized to a unit ball as shown in Fig. 1(d), then the transformed vectors $\|\boldsymbol{\psi}_1'\| = \sqrt{3}/2$, $\|\boldsymbol{\psi}_2'\| = 1$, and $\|\boldsymbol{\psi}_3'\| = 1/2$. According to Fig. 1(d), the second sensor has the largest norm. The following remark is now in order.

Remark 1 (Why is the change of basis (10) effective?): Based on the change of basis, the eigenvalues of $\boldsymbol{\Psi}_{\mathcal{N}}^{\mathrm{T}}\boldsymbol{\Psi}_{\mathcal{N}}$ are normalized to ones, and all directions are made equally important. The change of basis preserves the relative relationship of the projections of $\boldsymbol{\phi}_i$ on each direction. It thus enables fair comparisons of the norms/determinants of vectors/matrices in different directions/subspaces of the same dimension. It is particularly effective when $\mathbf{F}(\mathcal{N})$ has divergent eigenvalues. If $\mathbf{F}(\mathcal{N})$ has identical λ_i's $(\lambda_i = \lambda, \forall i = 1, \ldots, n)$, the change of basis only scales the

measurement vectors by the same amount. i.e., $\boldsymbol{\psi}_i = \frac{1}{\lambda}\boldsymbol{\phi}_i$. It would not have any impact on the greedy sensor selection.

Given a collection of $\boldsymbol{\phi}_i$'s ($i \in \mathcal{S}$), let $\bar{\mathbf{V}}_{r_s} = [\bar{\mathbf{v}}_1, \ldots, \bar{\mathbf{v}}_{r_s}]$ collect the eigenvectors corresponding to the r_s nonzero eigenvalues of $\mathbf{F}(\mathcal{S})$, where $r_s := \text{rank}(\mathbf{F}(\mathcal{S}))$ and $0 < r_s \leq n$. As a result, $\{\boldsymbol{\phi}_i \mid i \in \mathcal{S}\}$ span an r_s-D subspace denoted by $\mathbb{S} = \text{span}\{\bar{\mathbf{v}}_1, \ldots, \bar{\mathbf{v}}_{r_s}\}$. Note that $\det((\mathbf{F}(\mathcal{S})) = 0$, when $r_s < n$. To address the rank-deficiency of $\mathbf{F}(\mathcal{S})$, we propose to measure the Fisher information contained in $\mathcal{S}$ w.r.t. $\mathcal{N}$ focusing on the r_s-D subspace instead of the full n-D space. Note that $\bar{\mathbf{V}}_{r_s}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}}^{\mathrm{T}}$ and $\bar{\mathbf{V}}_{r_s}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{N}}^{\mathrm{T}}$ are the projections of $\boldsymbol{\Phi}_{\mathcal{S}}^{\mathrm{T}}$ and $\boldsymbol{\Phi}_{\mathcal{N}}^{\mathrm{T}}$ onto the subspace $\mathbb{S}$, respectively. The following definition is established.

Definition 2 (Fisher Information Intensity): We define the ratio between the squared volumes of the r_s-D parallelepipeds spanned by $\bar{\mathbf{V}}_{r_s}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}}^{\mathrm{T}}$ and $\bar{\mathbf{V}}_{r_s}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{N}}^{\mathrm{T}}$ as the *Fisher information intensity* (FII) of $\mathcal{S}$ w.r.t. $\mathcal{N}$. It is given by

$$\kappa(\mathcal{S}|\mathcal{N}) := \frac{\det\left(\bar{\mathbf{V}}_{r_s}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}} \bar{\mathbf{V}}_{r_s}\right)}{\det\left(\bar{\mathbf{V}}_{r_s}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{N}}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{N}} \bar{\mathbf{V}}_{r_s}\right)} = \frac{\det\left(\bar{\mathbf{V}}_{r_s}^{\mathrm{T}} \mathbf{F}(\mathcal{S}) \bar{\mathbf{V}}_{r_s}\right)}{\det\left(\bar{\mathbf{V}}_{r_s}^{\mathrm{T}} \mathbf{F}(\mathcal{N}) \bar{\mathbf{V}}_{r_s}\right)}. \tag{15}$$

For any $\mathcal{S} \subseteq \mathcal{N}$, the FII $\kappa(\mathcal{S} \mid \mathcal{N})$ takes a value from [0,1]. It, in essence, provides the ratio of the Fisher information (in terms of determinant) contributed by $\mathcal{S}$ to $\mathcal{N}$ in the directions $\bar{\mathbf{v}}_1, \ldots, \bar{\mathbf{v}}_{r_s}$. Note that $\det(\bar{\mathbf{V}}_{r_s}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}} \bar{\mathbf{V}}_{r_s})$ counts for the nonzero eigenvalues of $\mathbf{F}(\mathcal{S})$. Hence, the FII $\kappa(\mathcal{S} \mid \mathcal{N})$ is closely related to the VIE determined by $\{\boldsymbol{\psi}_i \mid i \in \mathcal{S}\}$. Then, the following definition is introduced.

Definition 3 (Volume Ratio): With the basis $\{\sqrt{\lambda_i}\mathbf{v}_i\}_{i=1}^n$, the Info-E associated with $\mathbf{F}(\mathcal{S})$ is reformed into $\tilde{\mathbf{F}}(\mathcal{S})$, which is given by

$$\tilde{\mathbf{F}}(\mathcal{S}) := \boldsymbol{\Psi}_{\mathcal{S}}^{\mathrm{T}} \boldsymbol{\Psi}_{\mathcal{S}} = \sum_{i \in \mathcal{S}} \boldsymbol{\psi}_i \boldsymbol{\psi}_i^{\mathrm{T}}, \tag{16}$$

and the corresponding VIE is derived as

$$\det\left(\tilde{\mathbf{F}}(\mathcal{S})\right) : = \det\left(\sum_{i \in \mathcal{S}} \boldsymbol{\psi}_i \boldsymbol{\psi}_i^{\mathrm{T}}\right) \tag{17}$$

$$= \det\left(\boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{V}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}} \mathbf{V} \boldsymbol{\Lambda}^{-\frac{1}{2}}\right) \tag{18}$$

$$= \det\left(\mathbf{F}(\mathcal{S}) \mathbf{F}^{-1}(\mathcal{N})\right). \tag{19}$$

It exactly represents the *volume ratio* of the VIE associated with $\mathbf{F}(\mathcal{S})$ to the VIE of $\mathbf{F}(\mathcal{N})$, i.e., $\det(\mathbf{F}(\mathcal{S}))/\det(\mathbf{F}(\mathcal{N}))$.

Let $\tilde{\mathbf{F}}(\mathcal{S}) = \tilde{\mathbf{V}} \tilde{\boldsymbol{\Lambda}} \tilde{\mathbf{V}}^{\mathrm{T}}$, where $\tilde{\boldsymbol{\Lambda}} = \text{diag}([\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n])$ and $\tilde{\mathbf{V}} = [\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_n]$ with the $\tilde{\lambda}_i$'s and $\tilde{\mathbf{v}}_i$'s being the eigenvalues and corresponding eigenvectors, respectively. Note that $\tilde{\mathbf{F}}(\mathcal{S})$ has the same rank as $\mathbf{F}(\mathcal{S})$, i.e., $\text{rank}(\tilde{\mathbf{F}}(\mathcal{S})) = r_s$. Let us define $\tilde{\boldsymbol{\Lambda}}_{r_s} = \text{diag}([\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{r_s}])$, and $\tilde{\mathbf{V}}_{r_s}$ and $\tilde{\mathbf{V}}_{n-r_s}$ collect the eigenvectors corresponding to the r_s nonzero and the $(n - r_s)$ zero eigenvalues of $\tilde{\mathbf{F}}(\mathcal{S})$, respectively. Separating the nonzero and zero eigenvalues of $\tilde{\mathbf{F}}(\mathcal{S})$, we obtain

$$\begin{aligned} \det(\tilde{\mathbf{F}}(\mathcal{S})) &= \det(\tilde{\mathbf{V}}_{r_s}^{\mathrm{T}} \tilde{\mathbf{F}}(\mathcal{S}) \tilde{\mathbf{V}}_{r_s}) \det(\tilde{\mathbf{V}}_{n-r_s}^{\mathrm{T}} \tilde{\mathbf{F}}(\mathcal{S}) \tilde{\mathbf{V}}_{n-r_s}) \\ &= \det(\tilde{\boldsymbol{\Lambda}}_{r_s})\ 0^{n-r_s}. \end{aligned} \tag{20}$$

On the other hand, according to (19), the VIE $\det(\tilde{\mathbf{F}}(\mathcal{S}))$ can also be decomposed as

$$\begin{aligned} &\det(\tilde{\mathbf{F}}(\mathcal{S})) \\ &= \det(\bar{\mathbf{V}}_{r_s}^{\mathrm{T}} \mathbf{F}(\mathcal{S}) \mathbf{F}^{-1}(\mathcal{N}) \bar{\mathbf{V}}_{r_s}) \det(\bar{\mathbf{V}}_{n-r_s}^{\mathrm{T}} \mathbf{F}(\mathcal{S}) \mathbf{F}^{-1}(\mathcal{N}) \bar{\mathbf{V}}_{n-r_s}) \\ &= \kappa(\mathcal{S} \mid \mathcal{N})\ 0^{n-r_s}, \end{aligned} \tag{21}$$

where $\kappa(\mathcal{S} \mid \mathcal{N})$ is defined in (15).

Note that (20) and (21) provide two ways of calculating $\det(\tilde{\mathbf{F}}(\mathcal{S}))$, which is decomposed into the nonzero and zero parts. The VIE $\det(\tilde{\mathbf{F}}(\mathcal{S}))$ is the product of the determinants of the $\tilde{\mathbf{F}}(\mathcal{S})$'s projections onto the r_s-D subspace and its corresponding $(n - r_s)$-D nullspace. As $\det(\tilde{\boldsymbol{\Lambda}}_{r_s})$ in (20) and $\kappa(\mathcal{S} \mid \mathcal{N})$ in (21) represent the determinant of the identical projection on the same r_s-D subspace, i.e., the nonzero part of $\det(\tilde{\mathbf{F}}(\mathcal{S}))$, we arrive at

$$\kappa(\mathcal{S} \mid \mathcal{N}) = \det(\tilde{\boldsymbol{\Lambda}}_{r_s}), \tag{22}$$

which reveals that the FII $\kappa(\mathcal{S} \mid \mathcal{N})$ is specified by the r_s nonzero eigenvalues of $\tilde{\mathbf{F}}(\mathcal{S})$. Therefore, the linear transform (10) from $\boldsymbol{\phi}_i$ to $\boldsymbol{\psi}_i$ naturally leads to the transform from the Fisher information to the FII, which provides clear geometric interpretations for any $\mathcal{S}$ as analyzed below.

- When $\mathcal{S} = \{i\}$ and $r_s = 1$, (15) and (22) reduce to

$$\kappa(\mathcal{S} \mid \mathcal{N}) = \frac{\|\boldsymbol{\phi}_i\|^2}{\sum_{j \in \mathcal{N}} \left\|\text{Proj}_{\boldsymbol{\phi}_i} \boldsymbol{\phi}_j\right\|^2} = \|\boldsymbol{\psi}_i\|^2, \tag{23}$$

  where $\text{Proj}_{\boldsymbol{\phi}_i} \boldsymbol{\phi}_j$ is the projection of $\boldsymbol{\phi}_j$ onto $\boldsymbol{\phi}_i$. It implies that the squared norm of $\boldsymbol{\psi}_i$ w.r.t. the basis $\{\sqrt{\lambda_i}\mathbf{v}_i\}_{i=1}^n$ provides the FII of the ith measurement $\boldsymbol{\phi}_i$ in the direction of $\boldsymbol{\phi}_i$. Apparently, $0 \leq \|\boldsymbol{\psi}_i\|^2 \leq 1$ for any $i \in \mathcal{N}$. The maximum value 1 of $\|\boldsymbol{\psi}_i\|^2$ is reached when $\boldsymbol{\phi}_i \neq \mathbf{0}$ and all the other measurements are orthogonal to $\boldsymbol{\phi}_i$, i.e., $\forall j \neq i, \boldsymbol{\phi}_i^{\mathrm{T}} \boldsymbol{\phi}_j = 0$. It implies that the Fisher information in the direction of $\boldsymbol{\phi}_i$ is fully contributed by the ith sensor (i.e., $\boldsymbol{\phi}_i$) itself. Note that $\boldsymbol{\phi}_2$ in the toy examples of Fig. 1 is exactly the show case here.
- When $r_s < n$, the FII $\kappa(\mathcal{S} \mid \mathcal{N})$ can be calculated either by (15) or (22). It is the ratio between the VIEs associated with the projections of $\mathbf{F}(\mathcal{S})$ and $\mathbf{F}(\mathcal{N})$ on the r_s-D subspace spanned by $\{\boldsymbol{\psi}_i \mid i \in \mathcal{S}\}$.
- When $r_s = n$, the FII $\kappa(\mathcal{S} \mid \mathcal{N})$ is extended to the full n-D space. It essentially turns into the ratio of the VIEs associated with $\mathbf{F}(\mathcal{S})$ and $\mathbf{F}(\mathcal{N})$, i.e.,

$$\kappa(\mathcal{S} \mid \mathcal{N}) = \det\left(\mathbf{F}(\mathcal{S}) \mathbf{F}^{-1}(\mathcal{N})\right) = \det\left(\tilde{\mathbf{F}}(\mathcal{S})\right). \tag{24}$$

  We remark here that the volume ratio is discussed in the n-D space, while the FII focuses on the r_s-D subspace that is spanned by $\{\boldsymbol{\psi}_i \mid i \in \mathcal{S}\}$. They are exactly the same for $r_s = n$.

2) Problem Reformulation Based on the Volume Ratio: Inspired by (20) and (21), we investigate the volume ratio $\det(\tilde{\mathbf{F}}(\mathcal{S}))$ as the nonzero part (i.e., the FII) and the zero part separately. In order to prevent the zero eigenvalues from nullifying the nonzero ones, we replace zero with a very small positive constant ϵ to distinguish it from ε in (6). The novel cost function

w.r.t. $\mathcal{S}$ is proposed as follows

$$f(\mathcal{S}) := \log\det\left(\tilde{\mathbf{V}}\begin{bmatrix}\tilde{\mathbf{\Lambda}}_{r_s} & \mathbf{0}\\ \mathbf{0} & \epsilon\mathbf{I}_{n-r_s}\end{bmatrix}\tilde{\mathbf{V}}^{\mathrm{T}}\right) \tag{25}$$

$$= \log(\kappa\left(\mathcal{S}\mid\mathcal{N}\right)\ \epsilon^{n-r_s}) \tag{26}$$

$$= \log\kappa\left(\mathcal{S}\mid\mathcal{N}\right) + (n-r_s)\log\epsilon, \tag{27}$$

where $\kappa(\mathcal{S}\mid\mathcal{N}) = \prod_{i=1}^{r_s}\tilde{\lambda}_i$ is the FII of $\mathcal{S}$. When $r_s < n$, $f(\mathcal{S})$ includes two terms, which are related to the r_s nonzero eigenvalues $\tilde{\lambda}_1,\ldots,\tilde{\lambda}_{r_s}$ and $(n-r_s)$ zero eigenvalues, respectively. When $r_s = n$, it follows from (27) that the auxiliary term $(n-r_s)\log\epsilon$ disappears. Thus, $f(\mathcal{S})$ is exactly equal to $\log\det(\tilde{\mathbf{F}}(\mathcal{S}))$ for the nonsingular $\tilde{\mathbf{F}}(\mathcal{S})$. Note that $\epsilon > 0$ is chosen to be a very small constant and used to replace $\log 0$ by $\log\epsilon$ (as $\log 0$ is an invalid operation) in $f(\mathcal{S})$. It is merely for theoretical analysis purposes, not to change the rank of $\mathbf{F}(\mathcal{S})$. In this article, we choose a sufficiently small ϵ, which satisfies

$$\epsilon < \min\left\{\left(\frac{\kappa\left(\mathcal{T}\mid\mathcal{N}\right)}{\kappa\left(\mathcal{S}\mid\mathcal{N}\right)}\right)^{\frac{1}{r_t-r_s}}, \min_{j\in\mathcal{N}}\|\boldsymbol{\psi}_j\|\right\}, \tag{28}$$

for any $\mathcal{S},\mathcal{T}\subset\mathcal{N}\setminus\{\varnothing\}$ and $r_s = \mathrm{rank}(\tilde{\mathbf{F}}(\mathcal{S})) < \mathrm{rank}(\tilde{\mathbf{F}}(\mathcal{T})) = r_t$, where $\kappa(\mathcal{T}\mid\mathcal{N})$ is defined in the same way as $\kappa(\mathcal{S}\mid\mathcal{N})$ with $\mathcal{T}$ replacing $\mathcal{S}$ in (15).

Relying on (28), we can readily verify the following lemma.

Lemma 1: If (28) holds true for ϵ, then $f(\mathcal{S}) < f(\mathcal{T})$ holds for any $\mathcal{S},\mathcal{T}\subset\mathcal{N}$ and $\mathrm{rank}(\tilde{\mathbf{F}}(\mathcal{S})) < \mathrm{rank}(\tilde{\mathbf{F}}(\mathcal{T}))$.

Proof: See Appendix A. ■

Consequently, the cost function $f(\mathcal{S})$ is critical for the design of greedy algorithms to solve the sensor selection problem. According to Lemma 1, the Fisher information contained in any arbitrary FIM can now be properly quantified and compared using $f(\mathcal{S})$, even when the FIM is singular. It can be derived from (27) that $f(\varnothing) \le f(\mathcal{S}) \le f(\mathcal{N})$, where $f(\varnothing) = n\log\epsilon$ and $f(\mathcal{N}) = 0$. Therefore, in this paper, we propose to reformulate the sensor selection problem as follows

$$\mathbf{P}_{\mathrm{ratio}}: \quad \max_{\mathcal{S}\subseteq\mathcal{N}} f(\mathcal{S}) \tag{29}$$

$$|\mathcal{S}| = k, \tag{29b}$$

which is well-defined for any $\mathcal{S}$.

Lemma 1 is particularly valuable for the greedy sensor selection of the first n steps. The following remark is important.

Remark 2 (How to select the first n sensors for $\mathbf{P}_{\mathrm{ratio}}$?): Singular FIMs related to candidate subsets should be first compared by means of the rank. The one with the larger rank results in a larger cost function according to Lemma 1, and thus is the better one. FIMs of the same rank need to be further compared by the FII of $\mathcal{S}$, i.e., $\kappa(\mathcal{S}\mid\mathcal{N})$, which collects the nonzero eigenvalues of $\tilde{\mathbf{F}}(\mathcal{S})$.

In addition to Lemma 1, we have another important property of $f(\mathcal{S})$.

Lemma 2: If (28) holds true for ϵ, then the cost function $f(\mathcal{S})$ (27) is a monotone submodular function w.r.t. $\mathcal{S}$.

Proof: See Appendix B. ■

Lemma 2 shows that $f(\mathcal{S})$ (conditioned on (28)) is a monotone submodular function. It thus ensures the applicability of the submodular approach to $\mathbf{P}_{\mathrm{ratio}}$ and guarantees a $(1-1/e)$ optimal solution, as will be illustrated in Section IV.

III. Algorithm Design: Optimizing Volume Ratio

In this section, we develop greedy algorithms to solve $\mathbf{P}_{\mathrm{ratio}}$ (29) iteratively. A greedy algorithm and its fast extension are proposed accordingly, where the fast one has lower complexity than its counterpart by further exploring the specific structure of the cost function.

A. A Greedy Algorithm for $\mathbf{P}_{ratio}$

The greedy algorithm is proposed over the two ranges of $|\mathcal{S}|$, i.e., $0 \le |\mathcal{S}| < n$ and $n \le |\mathcal{S}| < k$, respectively. Denote the subset determined at the lth $(1 \le l \le k)$ step by $\mathcal{S}^{(l)} = \{s_1,\ldots,s_l\}$ and let $\mathcal{S}^{(0)} = \varnothing$.

1) $0 \le |\mathcal{S}| < n$: Given $\mathcal{S}^{(0)} = \varnothing$, the 1st sensor s_1 is determined by

$$s_1 = \operatorname*{argmax}_{i\in\mathcal{N}} f(\{i\}) = \operatorname*{argmax}_{i\in\mathcal{N}} \log\left(\kappa\left(\{i\}\mid\mathcal{N}\right)\epsilon^{n-1}\right) = \operatorname*{argmax}_{i\in\mathcal{N}}\|\boldsymbol{\psi}_i\|^2, \tag{30}$$

which depends not only on $\boldsymbol{\phi}_i$ but also on the full measurements $\mathbf{\Phi}_{\mathcal{N}}$ as can be seen from (23). Thus, both the norms and directional correlations of the measurement vectors are taken into account. According to (23), the selection of s_1 based on (30) is also reasonable due to the geometric interpretation of $\|\boldsymbol{\psi}_i\|^2$. This makes our proposed algorithm different from the one in [3] from the beginning.

Subsequently, given $\mathcal{S}^{(l)}$ at the $(l+1)$th $(1 \le l < n)$ step, the sensor s_{l+1} can be determined by optimizing

$$\max_{i\in\mathcal{N}\setminus\mathcal{S}^{(l)}} f(\mathcal{S}^{(l)}\cup\{i\})$$

$$\doteq \max_{i\in\mathcal{N}\setminus\mathcal{S}^{(l)}} \log\left(\kappa\left(\mathcal{S}^{(l)}\cup\{i\}\mid\mathcal{N}\right)\epsilon^{n-(l+1)}\right) \tag{31}$$

$$\doteq \max_{i\in\mathcal{N}\setminus\mathcal{S}^{(l)}} \log\det\left(\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}^{\mathrm{T}}\right), \tag{32}$$

where (32) follows from (31) as $\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}^{\mathrm{T}}$ has the same nonzero eigenvalues (counting multiplicity) as $\tilde{\mathbf{F}}(\mathcal{S}^{(l)}\cup\{i\}) = \mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}^{\mathrm{T}}\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}$. Particularly, when $l < n$, note that (31) and (32) force $\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}^{\mathrm{T}}$ to have only nonzero eigenvalues so that $\det(\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}^{\mathrm{T}}) \neq 0$. Thus $\log\det(\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}^{\mathrm{T}}) = \log\prod_{i=1}^{r_s}\tilde{\lambda}_i = \log\kappa(\mathcal{S}^{(l)}\cup\{i\}\mid\mathcal{N})$. In addition, since $\epsilon^{n-(l+1)}$ is not dependent on i, it does not appear in (32). Note that to maximize $f(\mathcal{S}^{(l)}\cup\{i\})$, the participation of a new sensor i is desired to increase the rank of $\tilde{\mathbf{F}}(\mathcal{S}^{(l)}\cup\{i\})$ by 1 in the case of $1 \le l < n$. Otherwise, it will violate Lemma 1. This result is consistent with the insights gained from Remark 2.

2) $n \le |\mathcal{S}| < k$: Following the above greedy steps, we are led to the $(l+1)$th step $(l \ge n)$ given $\mathcal{S}^{(l)}$ and $r_s = n$. As $\tilde{\mathbf{F}}(\mathcal{S}^{(l)})$ is nonsingular, $f(\mathcal{S}^{(l)})$ is exactly equal to

Algorithm 1: Ratio Greedy Algorithm.

Input: k, $\mathbf{\Phi}_{\mathcal{N}} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \cdots, \boldsymbol{\phi}_N]^{\mathrm{T}} \in \mathbb{R}^{N\times n}$;
Output: $\mathcal{S}^{(k)}$;
1 Obtain the full FIM $\mathbf{F}(\mathcal{N}) = \mathbf{\Phi}_{\mathcal{N}}^{\mathrm{T}}\mathbf{\Phi}_{\mathcal{N}}$ and its eigen decomposition $\mathbf{F}(\mathcal{N}) = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}$;
2 Obtain $\mathbf{\Psi}_{\mathcal{N}} = \mathbf{\Phi}_{\mathcal{N}}\mathbf{V}\mathbf{\Lambda}^{-\frac{1}{2}}$;
3 Initialize: $\mathcal{S}^{(0)} = \varnothing$, $l = 0$;
4 **for** $l < n$ **do**
5 Determine the next sensor by (32), i.e., $s_{l+1} = \arg\max_{i\in\mathcal{N}\setminus\mathcal{S}^{(l)}} \log\det\left(\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}^{\mathrm{T}}\right)$;
6 $\mathcal{S}^{(l+1)} = \mathcal{S}^{(l)} \cup \{s_{l+1}\}$, $l = l+1$;
7 **end**
8 Matrix inversion $\tilde{\mathbf{F}}^{-1}(\mathcal{S}^{(n)})$;
9 **for** $n \le l < k$ **do**
10 Determine the next sensor by (35), i.e., $s_{l+1} = \arg\max_{i\in\mathcal{N}\setminus\mathcal{S}^{(l)}} \boldsymbol{\psi}_i^{\mathrm{T}}\tilde{\mathbf{F}}^{-1}(\mathcal{S}^{(l)})\boldsymbol{\psi}_i$;
11 $\mathcal{S}^{(l+1)} = \mathcal{S}^{(l)} \cup \{s_{l+1}\}$, $l = l+1$.
12 **end**

$\log\det(\tilde{\mathbf{F}}(\mathcal{S}^{(l)}))$. Thus, $f(\mathcal{S}^{(l)}\cup\{i\})$ can be derived as

$$\begin{aligned} f\left(\mathcal{S}^{(l)}\cup\{i\}\right) &= \log\det\left(\tilde{\mathbf{F}}(\mathcal{S}^{(l)}\cup\{i\})\right) \\ &= \log\left(\det\left(\left(\mathbf{I}+\boldsymbol{\psi}_i\boldsymbol{\psi}_i^{\mathrm{T}}\tilde{\mathbf{F}}^{-1}(\mathcal{S}^{(l)})\right)\tilde{\mathbf{F}}(\mathcal{S}^{(l)})\right)\right) \quad (33) \\ &= \log\left(\left(1+\boldsymbol{\psi}_i^{\mathrm{T}}\tilde{\mathbf{F}}^{-1}(\mathcal{S}^{(l)})\boldsymbol{\psi}_i\right)\det\left(\tilde{\mathbf{F}}(\mathcal{S}^{(l)})\right)\right) \\ &= \log\left(1+\boldsymbol{\psi}_i^{\mathrm{T}}\tilde{\mathbf{F}}^{-1}(\mathcal{S}^{(l)})\boldsymbol{\psi}_i\right) + f(\mathcal{S}^{(l)}), \quad (34) \end{aligned}$$

where (33) is obtained by plugging in $\tilde{\mathbf{F}}(\mathcal{S}^{(l)}\cup\{i\}) = \tilde{\mathbf{F}}(\mathcal{S}) + \boldsymbol{\psi}_i\boldsymbol{\psi}_i^{\mathrm{T}}$ and (34) is due to the fact that $\det(\mathbf{I}+\mathbf{a}\mathbf{b}^{\mathrm{T}}) = 1+\mathbf{b}^{\mathrm{T}}\mathbf{a}$ for any $\mathbf{a},\mathbf{b}\in\mathbb{R}^n$. Consequently, given $\mathcal{S}^{(l)}$, the $(l+1)$th sensor can be determined by optimizing

$$\max_{i\in\mathcal{N}\setminus\mathcal{S}^{(l)}} f(\mathcal{S}^{(l)}\cup\{i\}) \doteq \max_{i\in\mathcal{N}\setminus\mathcal{S}^{(l)}} \boldsymbol{\psi}_i^{\mathrm{T}}\tilde{\mathbf{F}}^{-1}(\mathcal{S}^{(l)})\boldsymbol{\psi}_i. \quad (35)$$

It is notably in the same form[3] as (7), except that there is no $\epsilon\mathbf{I}$ here for the nonsingular $\tilde{\mathbf{F}}(\mathcal{S}^{(l)})$.

In this way, $\mathbf{P}_{\mathrm{ratio}}$ can be addressed in a greedy fashion. The overall greedy algorithm is sketched in Algorithm 1, where lines $4\sim7$ are for $0\le|\mathcal{S}|<n$ to determine the first n sensors by maximizing (32) iteratively. Meanwhile, lines $9\sim12$ of Algorithm 1 determine the remaining sensors according to (35). Note that $f(\mathcal{S})$ is greedily maximized and increases along the iteration of Algorithm 1.

It is worth mentioning that ϵ does not appear in Algorithm 1. This is because we leverage Lemma 1 to bypass the rank-deficiency of $\tilde{\mathbf{F}}(\mathcal{S})$ by expanding the r_s-D subspace dimension-by-dimension (Algorithm 1 lines $4\sim7$), and to ensure that $f(\mathcal{S})$ greedily increases. In this way, we do not need the inverse of the FIM to derive the next sensor when $r_s<n$, and ϵ is not required when implementing the ratio greedy algorithm. We find out that the first n iterations of Algorithm 1 coincide with those of the maximal projection on minimum eigenspace (MPME) algorithm in [8]. At each of the first n iterations, the MPME chooses the sensor, whose observation has the maximum projection onto the minimum eigenspace of the current FIM (which is termed the dual observation matrix in [8]). Note that the minimum eigenspace [8, Definition 1] of the FIM $\tilde{\mathbf{F}}(\mathcal{S})$ is exactly the nullspace of $\mathbf{\Psi}_{\mathcal{S}}^{\mathrm{T}}$ in our work. Thus, the selection principle of the first n sensors by the MPME algorithm is coincidentally similar to the proposed ratio greedy algorithm. The main difference lies in whether the linear transform (10) of measurement vectors is carried out or not. The MPME algorithm uses $\boldsymbol{\phi}_i$, meanwhile we employ $\boldsymbol{\psi}_i$, which is a linear transform of $\boldsymbol{\phi}_i$. This will particularly affect the first n steps of the sensor selection, and thus results in different solutions. Moreover, the MPME algorithm optimizes the WCEV criterion in the remaining $(k-n)$ steps, and no theoretical performance analysis is provided in [8].

Algorithm 1 first obtains the full FIM $\mathbf{F}(\mathcal{N})$ and its eigenvalue decomposition, which costs $\mathcal{O}(Nn^2+n^3)$ flops. Obtaining $\mathbf{\Psi}_{\mathcal{N}} = \mathbf{\Phi}_{\mathcal{N}}\mathbf{V}\mathbf{\Lambda}^{-\frac{1}{2}}$ requires $\mathcal{O}(Nn^2)$ flops. The greedy iteration (lines $4\sim12$) costs $\mathcal{O}(Nkn^3)$ flops. As $N>n$ in general, the computational cost of Algorithm 1 is $\mathcal{O}(Nkn^3)$. The most high-cost calculations of Algorithm 1 are the determinant operation of $\det(\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}^{\mathrm{T}})$ (line 5) and the propagation of $\tilde{\mathbf{F}}^{-1}(\mathcal{S}^{(l)})$ (line 10), which can be simplified to develop a fast greedy algorithm.

B. A Fast Greedy Algorithm for $\mathbf{P}_{ratio}$

In this subsection, we develop a fast implementation of Algorithm 1 by simplifying the calculation of $\det(\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}^{\mathrm{T}})$ and the propagation of $\tilde{\mathbf{F}}^{-1}(\mathcal{S}^{(l)})$ therein. Likewise, we will investigate the two ranges of $0\le|\mathcal{S}|<n$ and $n\le|\mathcal{S}|<k$, respectively.

1) $0\le|\mathcal{S}|<n$*:* We will first simplify the calculation of $\det(\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}^{\mathrm{T}})$ by leveraging its specific structure. Specifically, adopting the Cholesky decomposition of $\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}^{\mathrm{T}}$ results in

$$\begin{aligned} &\det\left(\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}^{\mathrm{T}}\right) \\ &= \begin{vmatrix} \mathbf{\Psi}_{\mathcal{S}^{(l)}}\mathbf{\Psi}_{\mathcal{S}^{(l)}}^{\mathrm{T}} & \mathbf{\Psi}_{\mathcal{S}^{(l)}}\boldsymbol{\psi}_i \\ \boldsymbol{\psi}_i^{\mathrm{T}}\mathbf{\Psi}_{\mathcal{S}^{(l)}}^{\mathrm{T}} & \boldsymbol{\psi}_i^{\mathrm{T}}\boldsymbol{\psi}_i \end{vmatrix} = \begin{vmatrix} \mathbf{L}^{(l)} & \mathbf{0} \\ (\mathbf{c}_i^{(l)})^{\mathrm{T}} & d_i^{(l)} \end{vmatrix} \begin{vmatrix} \left(\mathbf{L}^{(l)}\right)^{\mathrm{T}} & \mathbf{c}_i^{(l)} \\ \mathbf{0}^{\mathrm{T}} & d_i^{(l)} \end{vmatrix} \\ &= \left(d_i^{(l)}\right)^2 \det\left(\mathbf{L}^{(l)}\left(\mathbf{L}^{(l)}\right)^{\mathrm{T}}\right) = \left(d_i^{(l)}\right)^2 \det\left(\mathbf{\Psi}_{\mathcal{S}^{(l)}}\mathbf{\Psi}_{\mathcal{S}^{(l)}}^{\mathrm{T}}\right), \end{aligned} \quad (36)$$

where $\mathbf{L}^{(l)}(\mathbf{L}^{(l)})^{\mathrm{T}}$ is the Cholesky decomposition of $\mathbf{\Psi}_{\mathcal{S}^{(l)}}\mathbf{\Psi}_{\mathcal{S}^{(l)}}^{\mathrm{T}}$, and $\mathbf{L}^{(l)}$ is an invertible lower triangular matrix. It can be derived from (36) that the vector $\mathbf{c}_i^{(l)}\in\mathbb{R}^l$ and the scalar $d_i^{(l)}>0$ satisfy

$$\mathbf{L}^{(l)}\mathbf{c}_i^{(l)} = \mathbf{\Psi}_{\mathcal{S}^{(l)}}\boldsymbol{\psi}_i, \quad (37)$$

$$\left(d_i^{(l)}\right)^2 = \left\|\boldsymbol{\psi}_i\right\|^2 - \left\|\mathbf{c}_i^{(l)}\right\|^2. \quad (38)$$

Therefore, given $\mathcal{S}^{(l)}$, optimizing $\log\det(\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}\mathbf{\Psi}_{\mathcal{S}^{(l)}\cup\{i\}}^{\mathrm{T}})$ is equivalent to optimizing $\log d_i^{(l)}$ based on (36). With $|\mathcal{S}^{(l)}|=l$, the index of the next selected sensor is thus given by

$$s_{l+1} = \arg\max_{i\in\mathcal{N}\setminus\mathcal{S}^{(l)}} \log d_i^{(l)}. \quad (39)$$

[3]It is derived in [3] that $\det(\mathbf{F}(\mathcal{S}\cup\{i\})+\varepsilon\mathbf{I}) = (1+\boldsymbol{\phi}_i^{\mathrm{T}}(\mathbf{F}(\mathcal{S})+\varepsilon\mathbf{I})^{-1}\boldsymbol{\phi}_i)\det(\mathbf{F}(\mathcal{S})+\varepsilon\mathbf{I})$, which leads to (7).

Algorithm 2: Fast Ratio Greedy Algorithm.

Input: k, $\mathbf{\Phi}_{\mathcal{N}} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \cdots, \boldsymbol{\phi}_N]^{\mathrm{T}} \in \mathbb{R}^{N\times n}$;
Output: $\mathcal{S}^{(k)}$;
1 Obtain the full FIM $\mathbf{F}(\mathcal{N}) = \mathbf{\Phi}_{\mathcal{N}}^{\mathrm{T}}\mathbf{\Phi}_{\mathcal{N}}$ and its eigen decomposition $\mathbf{F}(\mathcal{N}) = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}$;
2 Obtain $\mathbf{\Psi}_{\mathcal{N}} = \mathbf{\Phi}_{\mathcal{N}}\mathbf{V}\mathbf{\Lambda}^{-\frac{1}{2}}$;
3 Initialize: $s_1 = \arg\max_{i\in\mathcal{N}} \log d_i^{(0)}$, $\mathcal{S}^{(1)} = \{s_1\}$, $\mathbf{c}_i^{(0)} = [\,]$, $\left(d_i^{(0)}\right)^2 = \|\boldsymbol{\psi}_i\|^2$, $q_i^{(0)} = 0$, $l = 1$;
4 **for** $l < n$ **do**
5 **for** $i \in \mathcal{N}\setminus\mathcal{S}^{(l)}$ **do**
6 $q_i^{(l)} = (\boldsymbol{\psi}_{s_l}^{\mathrm{T}}\boldsymbol{\psi}_i - (\mathbf{c}_{s_l}^{(l-1)})^{\mathrm{T}}\mathbf{c}_i^{(l-1)})/d_{s_l}^{(l-1)}$;
7 Update $\mathbf{c}_i^{(l)} = \left[(\mathbf{c}_i^{(l-1)})^{\mathrm{T}},\ q_i^{(l)}\right]^{\mathrm{T}}$, $\left(d_i^{(l)}\right)^2 = \left(d_i^{(l-1)}\right)^2 - \left(q_i^{(l)}\right)^2$ by (41) and (42);
8 **end**
9 Determine the next sensor by (39), i.e., $s_{l+1} = \arg\max_{i\in\mathcal{N}\setminus\mathcal{S}^{(l)}} \log d_i^{(l)}$;
10 $\mathcal{S}^{(l+1)} = \mathcal{S}^{(l)} \cup \{s_{l+1}\}$, $l = l+1$;
11 **end**
12 Matrix inversion $\tilde{\mathbf{F}}^{-1}(\mathcal{S}^{(n)})$;
13 **for** $n \le l < k$ **do**
14 Determine the next sensor by (35), i.e., $s_{l+1} = \arg\max_{i\in\mathcal{N}\setminus\mathcal{S}^{(l)}} \boldsymbol{\psi}_i^{\mathrm{T}}\tilde{\mathbf{F}}^{-1}(\mathcal{S}^{(l)})\boldsymbol{\psi}_i$;
15 $\mathcal{S}^{(l+1)} = \mathcal{S}^{(l)} \cup \{s_{l+1}\}$;
16 Update $\tilde{\mathbf{F}}^{-1}(\mathcal{S}^{(l)}) = \tilde{\mathbf{F}}^{-1}(\mathcal{S}^{(l-1)}) - \frac{\tilde{\mathbf{F}}^{-1}(\mathcal{S}^{(l-1)})\boldsymbol{\psi}_{s_l}\boldsymbol{\psi}_{s_l}^{\mathrm{T}}\tilde{\mathbf{F}}^{-1}(\mathcal{S}^{(l-1)})}{1+\boldsymbol{\psi}_{s_l}^{\mathrm{T}}\tilde{\mathbf{F}}^{-1}(\mathcal{S}^{(l-1)})\boldsymbol{\psi}_{s_l}}$ by (43);
17 $l = l+1$.
18 **end**

For each candidate i, $\mathbf{c}_i^{(l)}$ and $d_i^{(l)}$ can be updated incrementally using the information from iteration $l-1$. Recall that $\mathcal{S}^{(l)} = \mathcal{S}^{(l-1)} \cup \{s_l\}$. Based on (37), we arrive at

$$\mathbf{L}^{(l)}\mathbf{c}_i^{(l)} = \begin{bmatrix} \mathbf{L}^{(l-1)} & \mathbf{0} \\ (\mathbf{c}_{s_l}^{(l-1)})^{\mathrm{T}} & d_{s_l}^{(l-1)} \end{bmatrix}\mathbf{c}_i^{(l)} = \mathbf{\Psi}_{\mathcal{S}^{(l)}}\boldsymbol{\psi}_i = \begin{bmatrix} \mathbf{\Psi}_{\mathcal{S}^{(l-1)}}\boldsymbol{\psi}_i \\ \boldsymbol{\psi}_{s_l}^{\mathrm{T}}\boldsymbol{\psi}_i \end{bmatrix}. \tag{40}$$

Observing (40), we can rewrite $\mathbf{c}_i^{(l)}$ in an iterative way, i.e., $\mathbf{c}_i^{(l)} = [(\mathbf{c}_i^{(l-1)})^{\mathrm{T}},\ q_i^{(l)}]^{\mathrm{T}}$ with

$$q_i^{(l)} = \left(\boldsymbol{\psi}_{s_l}^{\mathrm{T}}\boldsymbol{\psi}_i - \left(\mathbf{c}_{s_l}^{(l-1)}\right)^{\mathrm{T}}\mathbf{c}_i^{(l-1)}\right)/d_{s_l}^{(l-1)}. \tag{41}$$

Plugging (41) into (38), we arrive at

$$\begin{aligned}\left(d_i^{(l)}\right)^2 &= \left\|\boldsymbol{\psi}_i\right\|^2 - \left\|\mathbf{c}_i^{(l)}\right\|^2 = \left\|\boldsymbol{\psi}_i\right\|^2 - \left\|\mathbf{c}_i^{(l-1)}\right\|^2 - (q_i^{(l)})^2 \\ &= \left(d_i^{(l-1)}\right)^2 - \left(q_i^{(l)}\right)^2.\end{aligned} \tag{42}$$

The propagation of (31) is simplified in this way, and summarized in lines $4 \sim 11$ of Algorithm 2.

2) $n \le |\mathcal{S}| < k$: Simplifying the propagation of $\tilde{\mathbf{F}}^{-1}(\mathcal{S})$ in line 10 of Algorithm 1 seems more straightforward. Specifically, given $\tilde{\mathbf{F}}^{-1}(\mathcal{S})$ and a new sensor i, the FIM $\tilde{\mathbf{F}}^{-1}(\mathcal{S}\cup\{i\})$ can be obtained by recursion [3] based on the *Sherman-Morrison Formula* as follows

$$\tilde{\mathbf{F}}^{-1}\left(\mathcal{S}\cup\{i\}\right) = \tilde{\mathbf{F}}^{-1}(\mathcal{S}) - \frac{\tilde{\mathbf{F}}^{-1}(\mathcal{S})\boldsymbol{\psi}_i\boldsymbol{\psi}_i^{\mathrm{T}}\tilde{\mathbf{F}}^{-1}(\mathcal{S})}{1+\boldsymbol{\psi}_i^{\mathrm{T}}\tilde{\mathbf{F}}^{-1}(\mathcal{S})\boldsymbol{\psi}_i}, \tag{43}$$

thus reducing the computational cost.

The fast greedy algorithm is summarized in Algorithm 2, which leverages (36) and (43) to reduce the complexity of the iteration down to $\mathcal{O}(Nkn^2)$. Thus, Algorithm 2 has an overall complexity of $\mathcal{O}(Nkn^2)$.

IV. Theoretical Performance Analysis

In this section, we provide a performance analysis of the proposed greedy algorithm for $\mathbf{P}_{\mathrm{ratio}}$. As a comparison, we will also analyze the near-optimal result of the classical submodular algorithm [3] that solves $\mathbf{P}_{\mathrm{approx}}$. Note that the theoretical analysis is presented here with $k \ge n$, as it is required for the measurement problem. In the rest of this article, we name the proposed algorithm (outlined in Algorithm 1 and its fast counterpart in Algorithm 2) the `Ratio` algorithm. The greedy algorithm proposed in [3] is referred to as the `Approx` algorithm.

A. Effectiveness of the Basis Change: A Toy Example

Let us first show the effectiveness of the proposed change of basis via a toy example. The `Ratio` algorithm leverages the knowledge of $\mathbf{F}(\mathcal{N})$ for sensor selection, and takes both the norms and direction correlations of the measurement vectors into account using the FII. This greatly affects the step-by-step selection mechanism of greedy methods as the sequential sensors are selected based on the former ones. It can be visualized geometrically via the following 2-D measurement example.

Fig. 2 illustrates a toy example of a 2-D measurement set-up with $N = 5$ and $n = 2$, where the evolution of the Info-E obtained by the `Approx` and `Ratio` algorithms is shown in Fig. 2(a)$\sim$(d) and (e)$\sim$(h), respectively. In general, the VIE related to the set obtained by both algorithms increases along with the participation of more sensors and gradually approaches the full Info-E (marked by purple ellipsoids in Fig. 2), corresponding to $\det(\mathbf{F}(\mathcal{S})\mathbf{F}^{-1}(\mathcal{N})) = 1$ with $\mathcal{S} = \mathcal{N}$.

The `Approx` algorithm selects sensors in the order of $\{5, 3, 2, 1\}$, where the 1st sensor $s_1 = 5$ is the one with the maximum measurement norm and the subsequent s_{l+1}'s are sequentially selected using the inverse of the updated $\mathbf{F}(\mathcal{S}^{(l+1)}) + \varepsilon\mathbf{I}$. On the other hand, the `Ratio` algorithm selects sensors in the order of $\{2, 1, 5, 4\}$, which happens to be different from the `Approx` algorithm at each step. As shown in Fig. 2, the `Ratio` algorithm achieves a larger volume ratio $\det(\mathbf{F}(\mathcal{S})\mathbf{F}^{-1}(\mathcal{N}))$ over the entire range, i.e., $1 \le l < N$. It implies that the modification of the first n steps is essential, as it could lead to a significant change to the overall solution.

B. Near-Optimal Results for $\mathbf{P}_{\mathrm{ratio}}$

Since the cost function $f(\mathcal{S})$ is monotone submodular (see Lemma 2), the `Ratio` algorithm guarantees a $(1-1/e)$ approximation of $\mathbf{P}_{\mathrm{ratio}}$'s optimal result. Let $\mathcal{S}_{\mathrm{opt}}^{(k)}$ denote the optimal solution to the original problem $\mathbf{P}_0$. For $\mathbf{P}_{\mathrm{ratio}}$ (29) with the cost function $f(\mathcal{S})$, we have the following result.

Lemma 3: The optimal solution $\mathcal{S}_{\mathrm{opt}}^{(k)}$ to problem $\mathbf{P}_0$ is also optimal to problem $\mathbf{P}_{\mathrm{ratio}}$.

Proof: Given any $\mathcal{S}^{(k)} \subset \mathcal{N}$ with $|\mathcal{S}^{(k)}| = k$ and $\mathcal{S}^{(k)} \neq \mathcal{S}_{\mathrm{opt}}^{(k)}$, $\det(\mathbf{F}(\mathcal{S}_{\mathrm{opt}}^{(k)})) > \det(\mathbf{F}(\mathcal{S}^{(k)}))$, implying that $\mathcal{S}_{\mathrm{opt}}^{(k)}$ is the

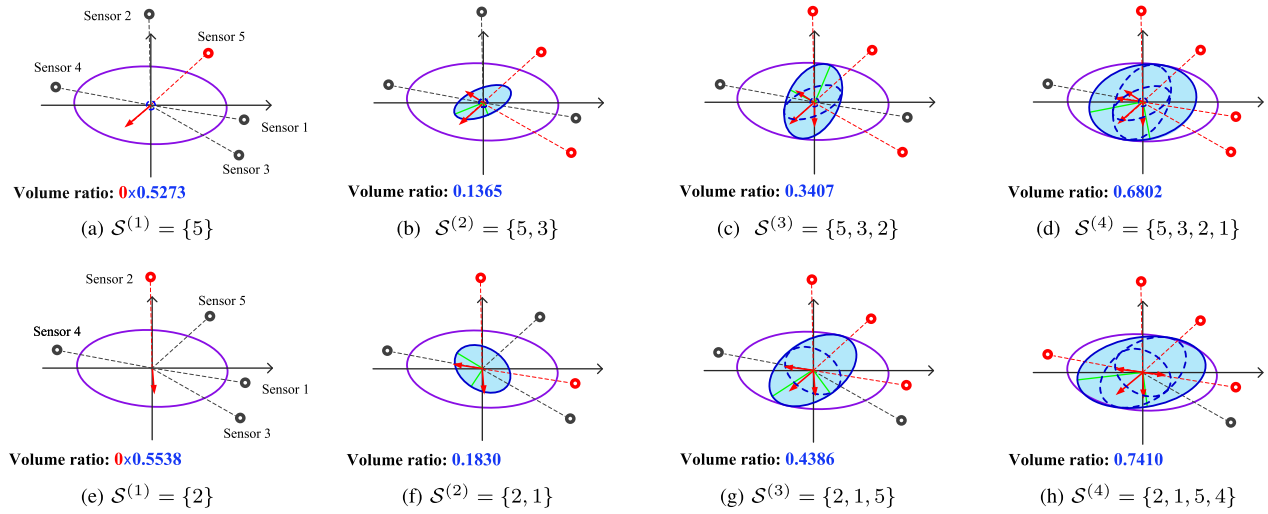


Fig. 2. Evolution of the Info-E along with the greedy sensor selection: a toy example of a 2-D measurement set-up, where $N=5$ and $\mathbf{\Phi}_{\mathcal{N}}=[-0.932,0.121;0.015,-0.859;-0.625,0.393;0.890,-0.166;-0.752,-0.631]$. (a)∼(d) The `Approx` [3] solution. (e)∼(h) The `Ratio` (proposed) solution. Candidate and selected sensors are marked by grey and red circles, respectively, while selected $\boldsymbol{\phi}_i$'s are marked by red arrows. The purple and blue ellipsoids denote the Info-E associated with $\mathcal{N}$ and $\mathcal{S}^{(l)}$, respectively. The number displayed in each subgraph is the value of $\det(\tilde{\mathbf{F}}(\mathcal{S}^{(l)}))$.

optimal solution to $\mathbf{P}_0$. To prove $\mathcal{S}_{\rm opt}^{(k)}$ is also optimal to $\mathbf{P}_{\rm ratio}$, we need to prove $f(\mathcal{S}_{\rm opt}^{(k)})>f(\mathcal{S}^{(k)})$. When $\mathrm{rank}(\tilde{\mathbf{F}}(\mathcal{S}^{(k)}))<n$, according to Lemma 1, $f(\mathcal{S}_{\rm opt}^{(k)})>f(\mathcal{S}^{(k)})$ readily holds noting that $\mathrm{rank}(\tilde{\mathbf{F}}(\mathcal{S}_{\rm opt}^{(k)}))=n$. When $\mathrm{rank}(\tilde{\mathbf{F}}(\mathcal{S}^{(k)}))=n$, $f(\mathcal{S}_{\rm opt}^{(k)})>f(\mathcal{S}^{(k)})$ is equivalently transformed into $\log\det(\mathbf{F}(\mathcal{S}_{\rm opt}^{(k)})\mathbf{F}^{-1}(\mathcal{N}))>\log\det(\mathbf{F}(\mathcal{S}^{(k)})\mathbf{F}^{-1}(\mathcal{N}))$, which is readily established. Thus, $\mathcal{S}_{\rm opt}^{(k)}$ is optimal to $\mathbf{P}_{\rm ratio}$ as well. This completes the proof of Lemma 3. ■

Theorem 1: The `Ratio` algorithm greedily maximizes $f(\mathcal{S})$ and guarantees

$$f\left(\mathcal{S}_{\rm ratio}^{(k)}\right)\geq\left(1-\frac{1}{e}\right)f\left(\mathcal{S}_{\rm opt}^{(k)}\right)+\left(1-\frac{1}{k}\right)^k n\log\epsilon, \quad (44)$$

where $\mathcal{S}_{\rm ratio}^{(k)}$ is the solution obtained by the `Ratio` algorithm and $\mathcal{S}_{\rm opt}^{(k)}$ is the theoretical optimal solution to $\mathbf{P}_0$. Let us define $\tilde{k}=(1-1/k)^k$. It follows from (44) that

$$\log\det\left(\mathbf{F}\left(\mathcal{S}_{\rm ratio}^{(k)}\right)\right)$$
$$\geq\left(1-\frac{1}{e}\right)\log\det\left(\mathbf{F}\left(\mathcal{S}_{\rm opt}^{(k)}\right)\right)+\frac{1}{e}\log\det\mathbf{F}(\mathcal{N})+\tilde{k}n\log\epsilon. \quad (45)$$

Proof: The result of (44) follows from the fact that $(f(\mathcal{S}_{\rm opt}^{(k)})-f(\mathcal{S}_{\rm ratio}^{(k)}))/(f(\mathcal{S}_{\rm opt}^{(k)})-f(\varnothing))\leq(1-1/k)^k$, which is straightforward as $f(\mathcal{S})$ is a monotone submodular function [11], [19]. Since $\mathbf{F}(\mathcal{S}_{\rm ratio}^{(k)})$ and $\mathbf{F}(\mathcal{S}_{\rm opt}^{(k)})$ are nonsingular FIMs, we can plug $f(\mathcal{S}_{\rm ratio}^{(k)})=\log\det(\mathbf{F}(\mathcal{S}_{\rm ratio}^{(k)})\mathbf{F}^{-1}(\mathcal{N}))$ and $f(\mathcal{S}_{\rm opt}^{(k)})=\log\det(\mathbf{F}(\mathcal{S}_{\rm opt}^{(k)})\mathbf{F}^{-1}(\mathcal{N}))$ into the earlier equation. It readily yields (44) and (45). This completes the proof of Theorem 1. ■

The last term in (44) and (45) is due to the fact that $f(\mathcal{S})$ is not normalized, i.e., $f(\varnothing)=n\log\epsilon\neq 0$. Moreover, the last two terms of (44) and (45) are all related to the ground set. Theorem 1 implies that the performance bound of the submodular algorithm is closely related to the ground set $\mathcal{N}$. This is reasonable.

C. Near-Optimal Results for $\mathbf{P}_{\rm approx}$

In this subsection, we attempt to study the impact of ε on the `Approx` Algorithm and its near-optimal solution, which to the best of our knowledge has not been discussed in existing works.

Unlike the `Ratio` algorithm where ϵ is not really used, the `Approx` algorithm needs to use $(\mathbf{F}(\mathcal{S})+\varepsilon\mathbf{I})^{-1}$ throughout the greedy iterations to optimize $\mathbf{P}_{\rm approx}$ according to (7). Although the value of ε does not affect the monotonicity and submodularity of the cost function $\log\det(\mathbf{F}(\mathcal{S})+\varepsilon\mathbf{I})$ to guarantee a $(1-1/e)$ approximation factor, it decides the approximation of $\mathbf{P}_{\rm approx}$ to $\mathbf{P}_0$. This is because the optimal solution to $\mathbf{P}_{\rm approx}$ (denoted by $\mathcal{S}_{\rm approx-opt}^{(k)}$) is sensitive to ε.

The `Approx` algorithm [3, Algorithm 1] that optimizes $\log\det(\mathbf{F}(\mathcal{S})+\varepsilon\mathbf{I})$ guarantees a $(1-1/e)$ optimal solution $\mathcal{S}_{\rm approx}^{(k)}$ as follows

$$\log\det\left(\mathbf{F}\left(\mathcal{S}_{\rm approx}^{(k)}\right)+\varepsilon\mathbf{I}\right)$$
$$\geq\left(1-\frac{1}{e}\right)\log\det\left(\mathbf{F}\left(\mathcal{S}_{\rm approx-opt}^{(k)}\right)+\varepsilon\mathbf{I}\right)+\tilde{k}n\log\varepsilon. \quad (46)$$

It can be clearly seen from (46) that $\mathcal{N}$ also has an impact on the performance of the `Approx` algorithm via ε. In addition, there is an auxiliary term $\varepsilon\mathbf{I}$ on both sides of (46), which is inherited from $\mathbf{P}_{\rm approx}$'s cost function $\log\det(\mathbf{F}(\mathcal{S})+\varepsilon\mathbf{I})$ and cannot be removed from (46). This can be easily verified by a counter example, i.e., $\log(x+\varepsilon)\geq(1-1/e)\log(y+\varepsilon)\nRightarrow\log(x)\geq(1-1/e)\log(y)$ with $x,y\in\mathbb{R}^+$. Therefore, (46) is

TABLE I
SUMMARY OF THE SENSOR SELECTION GREEDY ALGORITHMS IN THE SIMULATION

Algorithm	Cost function		Complexity
	VCE (Fig. 4)	MSE (Fig. 5 and Fig. 6)	
`Ratio` (proposed)	$\log\kappa(\mathcal{S}\mid\mathcal{N})+(n-r_s)\log\epsilon$	$\log\kappa(\mathcal{S}\mid\mathcal{N})+(n-r_s)\log\epsilon$ for the first n steps, $\mathrm{Tr}\left(\tilde{\mathbf{F}}(\mathcal{S})\right)$ for the remaining $(k-n)$ steps	$\mathcal{O}(Nkn^2)$
`Ratio-Approx` (proposed)	$\log\det\left(\tilde{\mathbf{F}}(\mathcal{S})+\varepsilon\mathbf{I}\right)$	$\mathrm{Tr}\left(\tilde{\mathbf{F}}(\mathcal{S})+\varepsilon\mathbf{I}\right)$	
`MPME` [8] (modified)	$\log\kappa(\mathcal{S})+(n-r_s)\log\epsilon$	$\log\kappa(\mathcal{S})+(n-r_s)\log\epsilon$ for the first n steps, $\mathrm{Tr}\left(\mathbf{F}(\mathcal{S})\right)$ for the remaining $(k-n)$ steps	
`Approx` [3]	$\log\det\left(\mathbf{F}(\mathcal{S})+\varepsilon\mathbf{I}\right)$	$\mathrm{Tr}\left(\mathbf{F}(\mathcal{S})+\varepsilon\mathbf{I}\right)$	

an irreducible result in the sense that it cannot decouple $\varepsilon\mathbf{I}$ from $\mathbf{F}(\mathcal{S}^{(k)}_{\text{approx}})$ or $\mathbf{F}(\mathcal{S}^{(k)}_{\text{approx-opt}})$.

Note that for any nonzero and small ε, the performance bound (46) holds. However, we are particularly interested in the case when $\mathbf{P}_{\text{approx}}$ has the same optimal solution as $\mathbf{P}_0$, i.e., $\mathcal{S}^{(k)}_{\text{approx-opt}}=\mathcal{S}^{(k)}_{\text{opt}}$. The following proposition is established to serve for the aforementioned purpose.

Proposition 1: Problem $\mathbf{P}_{\text{approx}}$ has the same optimal solution as $\mathbf{P}_0$, iff for any $\mathcal{S}^{(k)}\subset\mathcal{N}$ of cardinality k ($k\geq n$) and $\mathcal{S}^{(k)}\neq\mathcal{S}^{(k)}_{\text{opt}}$, the parameter ε satisfies

$$\sum_{i=1}^{n}\left(h^i\left(\boldsymbol{\lambda}'\right)-h^i(\boldsymbol{\lambda}^*)\right)\varepsilon^i<\det\left(\mathbf{F}\left(\mathcal{S}^{(k)}_{\text{opt}}\right)\right)-\det\left(\mathbf{F}\left(\mathcal{S}^{(k)}\right)\right), \tag{47}$$

where $\boldsymbol{\lambda}^*=[\lambda_1^*,\ldots,\lambda_n^*]^{\mathrm{T}}$ and $\boldsymbol{\lambda}'=[\lambda_1',\ldots,\lambda_n']^{\mathrm{T}}$ collect the eigenvalues of $\mathbf{F}(\mathcal{S}^{(k)}_{\text{opt}})$ and $\mathbf{F}(\mathcal{S}^{(k)})$, respectively, and $h^i(\boldsymbol{\zeta})$ is the polynomial coefficient of ε^i, where $\sum_{i=0}^{n}h^i(\boldsymbol{\zeta})\varepsilon^i=\prod_{j=1}^{n}(\zeta_j+\varepsilon)$.

Proof: See Appendix C. ∎

When ε fulfills (47), we obtain $\mathcal{S}^{(k)}_{\text{approx-opt}}=\mathcal{S}^{(k)}_{\text{opt}}$. We can replace $\mathcal{S}^{(k)}_{\text{approx-opt}}$ in (46) with $\mathcal{S}^{(k)}_{\text{opt}}$. However, the lower bounds in (45) and (46) are still not comparable due to the difference between ϵ and ε and the coupling between $\varepsilon\mathbf{I}$ and $\mathbf{F}(\mathcal{S}^{(k)}_{\text{approx}})$ (or $\mathbf{F}(\mathcal{S}^{(k)}_{\text{opt}})$).

To summarize, the following remark is presented.

Remark 3 (Impact of ϵ and ε on different algorithms): The way that ϵ and ε appear in the cost function influences the greedy algorithm design and further affects the well-known $(1-1/e)$ near-optimal result.

- The `Ratio` algorithm does not employ ϵ. On the other hand, the `Approx` algorithm needs ε in each iteration.
- Both ϵ and ε help to solve the issue that the $\log\det$ operation of a singular matrix in the cost functions is not well-defined.
- The monotonicity of $f(\mathcal{S})$ is guaranteed by making ϵ satisfy (28). Moreover, $\mathbf{P}_{\text{ratio}}$ and $\mathbf{P}_0$ have the same optimal solution. Meanwhile, $\mathbf{P}_{\text{approx}}$ can have the same optimal solution as $\mathbf{P}_0$, when ε fulfills (47).
- The two terms $(1-1/k)^k n\log\epsilon+1/e\log\det\mathbf{F}(\mathcal{N})$ and $(1-1/k)^k n\log\varepsilon$ in (45) and (46), respectively, indicate that the performance gap between the optimal solutions and the ones provided by the submodular algorithms is closely related to the ground set $\mathcal{N}$.

V. NUMERICAL RESULTS AND DISCUSSIONS

In this section, Monte Carlo (MC) simulation results are provided to assess the VCE and root-mean-square error (RMSE) performance of the proposed `Ratio` algorithm. All the experiments are implemented under two typical distributions of the measurement vectors, i.e., the uniform distribution $\mathcal{U}[-1,1]^n$ and $\mathcal{U}[0,1]^n$. The measurement vectors $\boldsymbol{\phi}_i$ are generated i.i.d. following either one of the two distributions. All the experiments are averaged over 10^3 MC runs.

For performance comparison, several greedy algorithms are implemented as baselines. Table I summarizes the greedy algorithms involved in the simulation, where the complexity is referred to for the corresponding fast counterparts of the algorithms, and the newly defined $\kappa(\mathcal{S})$ is the product of the nonzero eigenvalues of $\mathbf{F}(\mathcal{S})$.

- `Ratio` (proposed): the ratio greedy algorithm outlined in Algorithm 1, where line 10 will be modified to $\mathrm{argmax}_{i\in\mathcal{N}\setminus\mathcal{S}}\boldsymbol{\psi}_i^{\mathrm{T}}\tilde{\mathbf{F}}^{-2}(\mathcal{S})\boldsymbol{\psi}_i/(1+\boldsymbol{\psi}_i^{\mathrm{T}}\tilde{\mathbf{F}}^{-1}(\mathcal{S})\boldsymbol{\psi}_i)$ to make the algorithm adapt to the MSE metric.[4]
- `Approx`: the greedy algorithm [3, Algorithm 1] that optimizes the VCE, where (7) will be replaced by $\mathrm{argmax}_{i\in\mathcal{N}\setminus\mathcal{S}}\ \boldsymbol{\phi}_i^{\mathrm{T}}\ (\mathbf{F}(\mathcal{S})\ +\varepsilon\mathbf{I})^{-2}\ \boldsymbol{\phi}_i/(1+\boldsymbol{\phi}_i^{\mathrm{T}}\ (\mathbf{F}\ (\mathcal{S})\ +\varepsilon\mathbf{I})^{-1}\ \boldsymbol{\phi}_i)$ in each iteration to adapt to the MSE metric. A simplified greedy algorithm [3, Algorithm 2] has also been proposed by reducing the complexity of the matrix inverse to $\mathcal{O}(n^2)$ in each iteration. Therefore, `Approx` also enjoys a complexity of $\mathcal{O}(Nkn^2)$ by employing the simplified determinant and (or) matrix inverse operations.
- `MPME` (modified): keeps the first n steps of [8, Algorithm 2] and modifies the remaining steps for the VCE (or MSE) metric. It can be seen as a counterpart of the `Ratio` algorithm but using the $\boldsymbol{\phi}_i$'s. Thus, it is easy to verify that the `MPME` algorithm also takes a complexity of $\mathcal{O}(Nkn^2)$ leveraging the simplified determinant computation proposed for the fast greedy ratio algorithm (see Algorithm 2).
- `Ratio-Approx` (proposed): maximizes $\det(\tilde{\mathbf{F}}(\mathcal{S})+\varepsilon\mathbf{I})$ (or $\mathrm{Tr}(\tilde{\mathbf{F}}(\mathcal{S})+\varepsilon\mathbf{I})$) for the VCE (or MSE) metric. It can be

[4]It is easy to replace the VCE metric [3, eq. (13)], [15, eq. (15)] by the MSE metric [15, eq. (14)], [16, eq. (10)] in each iteration when the current FIM is nonsingular.

seen as a straightforward extension of `Approx` but using the ψ_i's instead of the ϕ_i's. Comparing with `Approx`, `Ratio-Approx` has an extra cost of $\mathcal{O}(2Nn^2+n^3)$ flops for the operation of basis change as we have analyzed for Algorithm 1. This is a one-off cost and $N > n$ in general, thus the complexity of the `Ratio-Approx` algorithm remains $\mathcal{O}(Nkn^2)$.

- `Group`: performs the `Approx` algorithm by reserving a group of top L suboptimal candidates[5] [15]. It is investigated as a high-performance competitor but takes a complexity of $\mathcal{O}(NkLn^2)$. The results of the `Group` algorithm with a group size of $L = 2, 5$, and 10 will be presented, noting that `Group` ($L = 1$) reduces to `Approx`.

Note that the proposed `Ratio` (`Ratio-Approx`) algorithm and the `MPME` (modified) (`Approx`) algorithm can be seen as counterparts w.r.t. the basis $\{\sqrt{\lambda_i}\mathbf{v}_i\}_{i=1}^n$ and the canonical basis $\{\mathbf{e}_i\}_{i=1}^n$, respectively. As such, the effectiveness of the linear transform (10) can be observed by comparing the `Ratio` (or `Ratio-Approx`) algorithm with the `MPME` (modified) (or `Approx`) algorithm. Meanwhile, for RMSE results, comparisons between the `Ratio` (or `MPME` (modified)) algorithm and the `Ratio-Approx` (or `Approx`) algorithm demonstrate the difference brought by the hybrid VCE and MSE metrics from the pure MSE metric. Throughout the simulation, we set $\varepsilon = 10^{-4}$ which is used by the `Approx`, `Ratio-Approx`, and `Group` algorithms.

In the following, we will first study the impact of different measurement distributions on $\mathbf{F}(\mathcal{N})$ in Section V-A. The VCE and RMSE results of different greedy algorithms are compared in Sections V-B and V-C, respectively, where the impact of measurement distributions on the greedy algorithms will also be discussed accordingly.

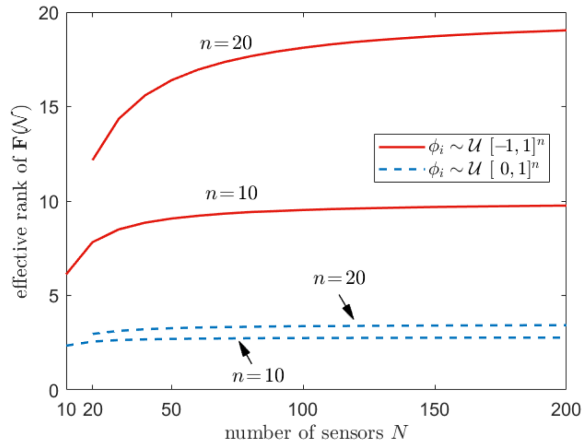
A. Impact of Different Measurement Distributions on $\mathbf{F}(\mathcal{N})$

The performance of sensor selection algorithms and their merits may heavily depend on the measurement settings, mainly in the form of $\mathbf{F}(\mathcal{N})$. Following Remark 1, we are particularly interested in the distribution of $\mathbf{F}(\mathcal{N})$'s eigenvalues, which can be measured by a convenient metric, i.e., the effective rank, introduced in [20] as follows

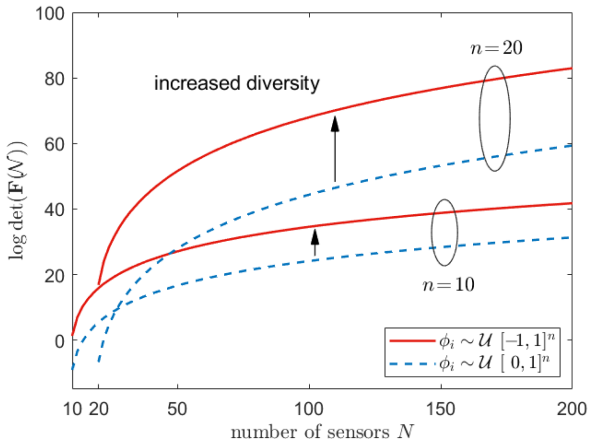
$$R_{\text{eff}}\left(\mathbf{F}(\mathcal{N})\right) = \exp\left(-\sum_{i=1}^{n} \check{\lambda}_i \log \check{\lambda}_i\right), \tag{48}$$

where $\check{\lambda}_i = \lambda_i/(\sum_i \lambda_i)$ are the normalized eigenvalues of $\mathbf{F}(\mathcal{N})$. The expression (48) suggests that $R_{\text{eff}}(\mathbf{F}(\mathcal{N}))$ contains the necessary information: the larger $R_{\text{eff}}(\mathbf{F}(\mathcal{N}))$ is, the flatter is the distribution of the λ_i's. A larger $R_{\text{eff}}(\mathbf{F}(\mathcal{N}))$ also indicates that $\mathbf{F}(\mathcal{N})$ is more diverse.

We present the value of $R_{\text{eff}}(\mathbf{F}(\mathcal{N}))$ versus N in Fig. 3(a), with $\phi_i \sim \mathcal{U}[-1,1]^n$ and $\phi_i \sim \mathcal{U}[0,1]^n$, respectively. It shows that $R_{\text{eff}}(\mathbf{F}(\mathcal{N}))$ with $\phi_i \sim \mathcal{U}[0,1]^n$ is much less than that with $\phi_i \sim \mathcal{U}[-1,1]^n$. In the later case, $R_{\text{eff}}(\mathbf{F}(\mathcal{N}))$ approaches n with increasing N. Note that $\mathbf{F}(\mathcal{N})$ can be seen as optimally diverse when $R_{\text{eff}}(\mathbf{F}(\mathcal{N})) = n$. Therefore, we expect more significant performance differences between the algorithms using the linear transformation (10) for $\phi_i \sim \mathcal{U}[0,1]^n$. Fig. 3(b) presents the value of $\log\det(\mathbf{F}(\mathcal{N}))$ in parallel with Fig. 3(a). It clearly shows that the average $\log\det(\mathbf{F}(\mathcal{N}))$ with $\phi_i \sim \mathcal{U}[-1,1]^n$ is larger than that with $\phi_i \sim \mathcal{U}[0,1]^n$ under the same n and N. This result is in accordance with Fig. 3(a), as in the two measurement settings, the norms of the ϕ_i's uniformly take a value from the same range [0,1]. We remark that there is potentially more directional diversity to be exploited from $\mathcal{N}$ with $\phi_i \sim \mathcal{U}[-1,1]^n$, corresponding to a larger $R_{\text{eff}}(\mathbf{F}(\mathcal{N}))$ and $\log\det(\mathbf{F}(\mathcal{N}))$. It in general leads to an improved VCE and RMSE performance regardless of the algorithm. We will discuss this in detail in the following subsections. In the rest of this article, we refer to the two distributions $\phi_i \sim \mathcal{U}[0,1]^n$ and $\phi_i \sim \mathcal{U}[-1,1]^n$ as the typical inhomogeneous and homogeneous measurement distributions, respectively.



(a) Average effective rank of $\mathbf{F}(\mathcal{N})$ vs. N.



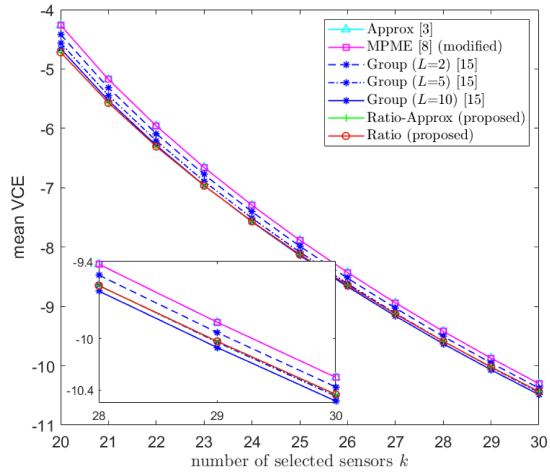
(b) Average $\log\det(\mathbf{F}(\mathcal{N}))$ vs. N.

Fig. 3. Some statistics of $\mathbf{F}(\mathcal{N})$ under different measurement distributions.

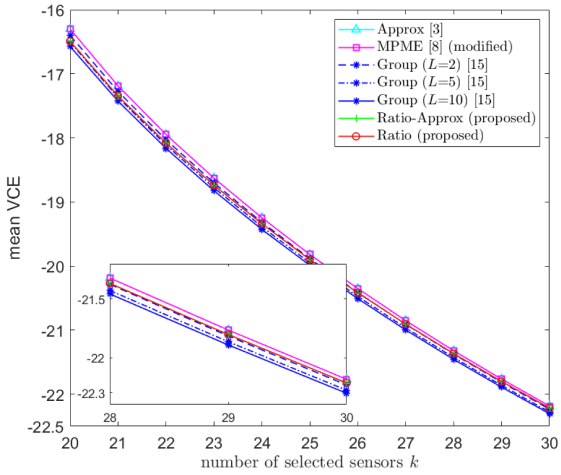
B. VCE Comparison to Baseline Methods

We present the VCE results, i.e., $-1/2\log\det(\mathbf{F}(\mathcal{S}))$, versus the number of sensors k (i.e., $|\mathcal{S}|$) for $n = 20$ and $N = 100$

[5]By simply reserving the top L candidates instead of the maximum one, the idea of *group greedy* can be applied to any of the greedy approaches as needed, including our proposed `Ratio` and `Ratio-Approx` algorithms.



(a) VCE vs. k, $\boldsymbol{\phi}_i \sim \mathcal{U}\,[0,1]^n$



(b) VCE vs. k, $\boldsymbol{\phi}_i \sim \mathcal{U}\,[-1,1]^n$

Fig. 4. Comparison of VCE under varying k, $n = 20$, $N = 100$.

in Fig. 4, where Fig. 4(a) and (b) are for $\boldsymbol{\phi}_i \sim \mathcal{U}[0,1]^n$ and $\boldsymbol{\phi}_i \sim \mathcal{U}[-1,1]^n$, respectively. With a small ε (i.e., $\varepsilon = 10^{-4}$, which is sufficiently small compared with the norms of $\boldsymbol{\phi}_i$'s), the `Approx` (`Ratio-Approx`) algorithm achieves almost the same VCE as the `MPME` (`Ratio`) algorithm. Note that the proposed `Ratio` (`Ratio-Approx`) algorithm can be seen as the counterpart of the `MPME` (`Approx`) algorithm, where the only difference is the basis of the measurement vectors. Meanwhile, the `Ratio` and `Ratio-Approx` algorithms notably outperform the `MPME` and `Approx` algorithms especially for $\boldsymbol{\phi}_i \sim \mathcal{U}[0,1]^n$ in Fig. 4(a). We attribute this improvement to the change of basis based on $\mathbf{F}(\mathcal{N})$ that leads us to the optimization of the volume ratio. The performance of the `Ratio` and `Ratio-Approx` algorithms are even better than the `Group` (with $L = 5$ and $L = 2$ in Fig. 4(a) and (b), respectively) algorithm, and close to the `Group` (with $L = 10$ and $L = 5$ in Fig. 4(a) and (b), respectively) algorithm when k is small; but the gap gradually reduces with increasing k. This further implies that the change of basis based on $\mathbf{F}(\mathcal{N})$ is valuable. Note that the `Ratio` algorithm ($\mathcal{O}(Nkn^2)$) has a much lower complexity than the `Group` ($\mathcal{O}(NkLn^2)$) algorithm.

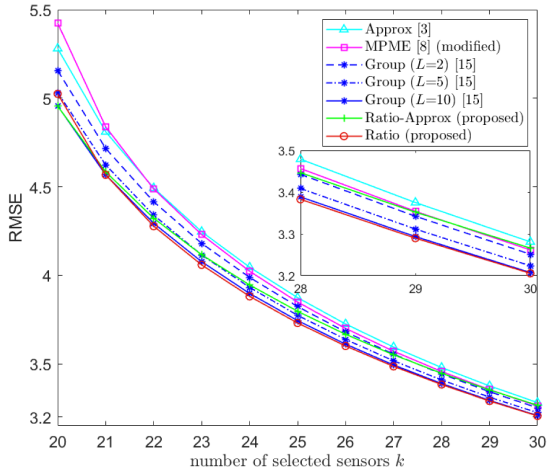
Another important observation can be obtained by comparing Fig. 4(a) with (b). That is, under the same parameter settings, the results for $\boldsymbol{\phi}_i \sim \mathcal{U}[-1,1]^n$ are in general better than that for $\boldsymbol{\phi}_i \sim \mathcal{U}[0,1]^n$. This is not surprising since the homogeneous measurements result in a larger $\log\det(\mathbf{F}(\mathcal{N}))$, as we have discussed in Section V-A. Moreover, as insighted in Remark 1, the `Ratio` (`Ratio-Approx`) algorithm that leverages the change of basis outperforms the `MPME` (`Approx`) algorithm much more in Fig. 4(a) than in Fig. 4(b), noting that $\boldsymbol{\phi}_i \sim \mathcal{U}[0,1]^n$ corresponds to the $\mathbf{F}(\mathcal{N})$ with nonuniform distributions of the eigenvalues.

C. RMSE Comparison to Baseline Methods

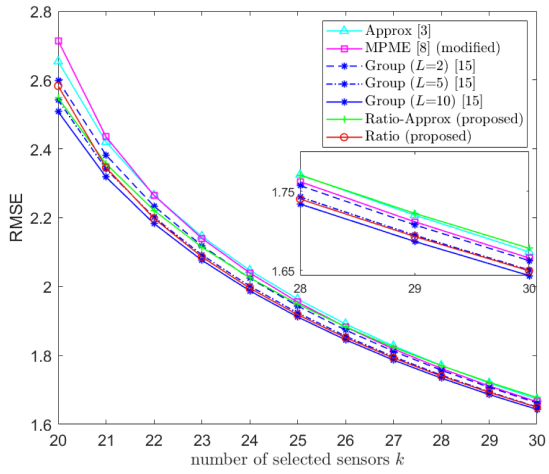
In practice, the MSE result is generally more concerned in measurement problems. Thus, we proceed to assess the RMSE performance of the greedy algorithms in this subsection. Instead of simply presenting the numerical RMSE in parallel with the VCE (Fig. 4) that corresponds to the identical solutions, all the algorithms are modified in this experiment to adapt to the MSE optimization. As shown in Table I, the `Approx` and `Ratio-Approx` algorithms adopt the MSE metric in each step as an invertable FIM is conducted for each iteration. The `Group` algorithm uses the same cost function as `Approx`. On the other hand, the `Ratio` and `MPME` algorithms keep their first n steps and only modifiy the remaining steps when the current FIM is nonsingular.

The RMSE results versus k with $\boldsymbol{\phi}_i \sim \mathcal{U}[0,1]^n$ and $\boldsymbol{\phi}_i \sim \mathcal{U}[-1,1]^n$ are presented in Figs. 5(a) and (b), respectively. The performance difference of the greedy algorithms is more obvious in terms of RMSE than VCE under the same parameter settings. The `MPME` algorithm is, in general, superior to the `Approx` algorithm in terms of RMSE, and so is the `Ratio` algorithm to the `Ratio-Approx` algorithm. This is due to the VCE metric used in the first n steps of the `Ratio` and `MPME` algorithms. As can be seen from Fig. 5, though the `Ratio` and `MPME` algorithms at $k = 20$ (i.e., $k = n$) are inferior to the `Ratio-Approx` and `Approx` algorithms, respectively, their sequential steps benefit from the first n steps for the minimization of MSE. This result is interesting, because it implies that the determinant of a singular matrix is likely to contain more information than the trace or minimum eigenvalue. Indeed, this well explains the superiority of the MPME algorithm [8] over the state-of-the-art algorithms at the time. Fig. 5 also shows that the `Ratio` and `Ratio-Approx` algorithms outperform the `MPME` and `Approx` algorithms, respectively, because they optimize the volume ratio rather than the absolute value of the Fisher information. Furthermore, in Fig. 5(a), the `Ratio-Approx` algorithm outperforms the `Group` ($L = 5$) algorithm for small k's and gets close to the `Group` ($L = 2$) algorithm for large k's, whereas the `Ratio` algorithm even outperforms the `Group` ($L = 10$) algorithm in the entire regime of k.

The RMSE performance versus N by fixing n and k is also investigated in Fig. 6, where $n = 20$ and a fixed number $k = 25$ of sensors are selected from a total number of N candidates.

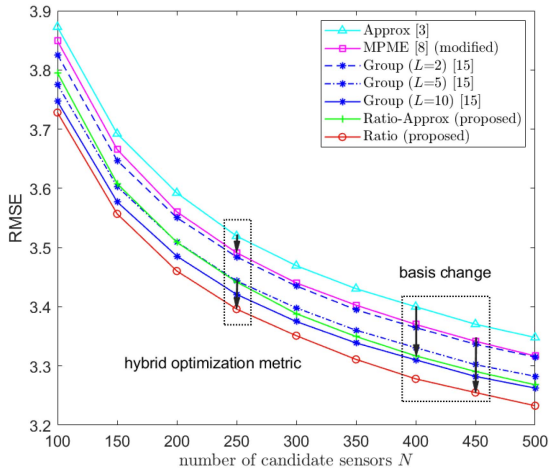


(a) RMSE vs. k, $\boldsymbol{\phi}_i \sim \mathcal{U}[0,1]^n$

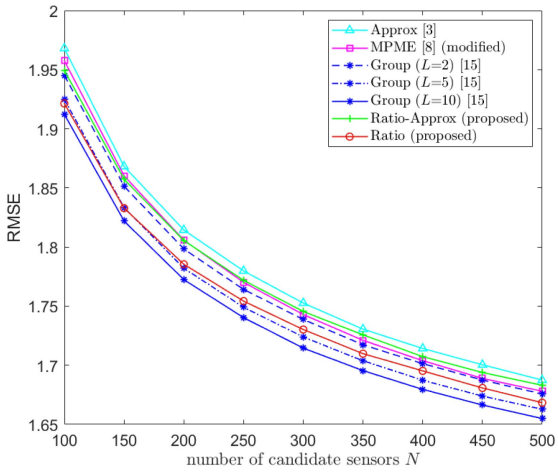


(b) RMSE vs. k, $\boldsymbol{\phi}_i \sim \mathcal{U}[-1,1]^n$

Fig. 5. Comparison of RMSE under varying k, $n = 20$, $N = 100$.



(a) RMSE vs. N, $\boldsymbol{\phi}_i \sim \mathcal{U}[0,1]^n$



(b) RMSE vs. N, $\boldsymbol{\phi}_i \sim \mathcal{U}[-1,1]^n$

Fig. 6. Comparison of RMSE under varying N, $n = 20$, $k = 25$.

In general, the RMSEs of all the involved algorithms decrease as N increases, as a larger set $\mathcal{N}$ provides more information to be exploited. However, note that the sensor selection task becomes more difficult for a large N given n and k, thus efficient sensor selection algorithms become more demanding. With $\boldsymbol{\phi}_i \sim \mathcal{U}[0,1]^n$, it shows in Fig. 6(a) that the `Ratio` algorithm notably outperforms the `Group` ($L = 10$) algorithm while the `Ratio-Approx` algorithm outperforms the `Group` ($L = 5$) algorithm, noting that the superiority gets larger as N increases. With $\boldsymbol{\phi}_i \sim \mathcal{U}[-1,1]^n$, the `Ratio` algorithm in Fig. 6(b) becomes slightly inferior to the `Group` ($L = 5$) algorithm, while the `Ratio-Approx` algorithm performs worse than `Group` ($L = 2$) but still better than `Approx`, i.e., `Group` ($L = 1$). These results well demonstrate the efficiency of the proposed `Ratio` and `Ratio-Approx` algorithms based on the FII.

The impact of different measurement distributions can also be studied by comparing the RMSEs in Figs. 5(a) and 6(a) with Figs. 5(b) and 6(b), respectively, where the differences among the algorithms are more obvious than in Fig. 4. For $\boldsymbol{\phi}_i \sim \mathcal{U}[-1,1]^n$, the RMSEs are improved for all algorithms, while the differences among them become smaller. This is easy to understand via an extreme case, where the measurement vectors are uniformly orthogonal. In this case, even a random sampling algorithm would achieve satisfactory performance. Because of this, it is worth mentioning that the sensor selection in the inhomogeneous measurement scenario with $\boldsymbol{\phi}_i \sim \mathcal{U}[0,1]^n$ is indeed more challenging due to reduced effective rank and directional diversity therein. We also find that the superiority of the `Ratio` (`Ratio-Approx`) algorithm over the `MPME` (`Approx`) algorithm is much larger for $\boldsymbol{\phi}_i \sim \mathcal{U}[0,1]^n$ than for $\boldsymbol{\phi}_i \sim \mathcal{U}[-1,1]^n$. This is because $\boldsymbol{\phi}_i \sim \mathcal{U}[-1,1]^n$ leads to an FIM with more uniform eigenvalues, making the change of basis (10) become less effective as we have illustrated in Remark 1. These results well demonstrate the advantage of the `Ratio` algorithm, as it exploits the full FIM $\mathbf{F}(\mathcal{N})$ to determine the first crucial n sensors, where the measurement intensity and directional diversity are jointly considered.

VI. Conclusion

This article investigates the sensor selection problem based on linear models to maximize the VIE. We propose a metric of FII based on the change of basis that leverages a priori information of the full FIM so that the profile of the full Info-E is taken into account. The Fisher information contained in an FIM is quantified more properly in terms of the FII, based on which the sensor selection problem is transformed into a volume ratio formulation, where the cost function is decoupled into a nonzero and a zero part of the volume ratio between the VIEs of the selected FIM and full FIM. The `Ratio` algorithm is proposed accordingly. It guarantees a $(1-1/e)$ optimal solution as the proposed cost function is shown monotone submodular. Its fast version with complexity $\mathcal{O}(Nkn^2)$ is further developed by exploiting the specific structure of the volume ratio. In addition, a comprehensive performance analysis of the proposed algorithm is provided and compared with the `Approx` algorithm. Numerical results are provided to investigate how different measurement distributions can affect the performance of greedy algorithms. Both VCE and RMSE results demonstrate that the proposed `Ratio` algorithm can perform close to the `Group` algorithm but with much lower complexity.

Appendix A
Proof of Lemma 1

Let r_s and r_t be the rank of $\tilde{\mathbf{F}}(\mathcal{S})$ and $\tilde{\mathbf{F}}(\mathcal{T})$, respectively, where $0 \le r_s < r_t \le n$. Following $f(\mathcal{S})$ (26), $f(\mathcal{T})$ can be calculated accordingly as

$$f(\mathcal{T}) = \log \kappa\left(\mathcal{T} \mid \mathcal{N}\right) + (n - r_t)\log\epsilon, \tag{49}$$

where $0 < \kappa(\mathcal{T} \mid \mathcal{N}) < 1$ represents of the FII of $\mathcal{T}$. Thus, the result $\log\kappa(\mathcal{T} \mid \mathcal{N})$ takes a finite negative value.

For the case $\mathcal{S} \neq \varnothing$, the FII $\kappa(\mathcal{S} \mid \mathcal{N}) \neq 0$. Let us subtract $f(\mathcal{S})$ (27) from $f(\mathcal{T})$ (49). It yields

$$\begin{aligned} f(\mathcal{T}) - f(\mathcal{S}) &= \log\kappa(\mathcal{T} \mid \mathcal{N}) - \log\kappa(\mathcal{S} \mid \mathcal{N}) - (r_t - r_s)\log\epsilon \\ &= \log\frac{\kappa\left(\mathcal{T} \mid \mathcal{N}\right)}{\kappa\left(\mathcal{S} \mid \mathcal{N}\right)} - (r_t - r_s)\log\epsilon > 0, \end{aligned} \tag{50}$$

where the inequality "$>$" holds due to (28).

For the case $\mathcal{S} = \varnothing$, we achieve $f(\varnothing) = n\log\epsilon$. As $r_s < r_t$, we obtain $\mathcal{T} \neq \varnothing$. Let us denote $\mathcal{T}_{\min}$, which satisfies $f(\mathcal{T}) \ge f(\mathcal{T}_{\min})$, $\forall \mathcal{T} \neq \varnothing$. Therefore, to prove $f(\mathcal{T}) > f(\mathcal{S})$ in this case, it suffices to prove $f(\mathcal{T}_{\min}) > f(\varnothing) = n\log\epsilon$. To find $\mathcal{T}_{\min}$, we first prove the monotonicity of $f(\mathcal{S})$, i.e.,

$$f\left(\mathcal{S}' \cup \{i\}\right) > f\left(\mathcal{S}'\right),\ \forall\, \mathcal{S}' \subset \mathcal{N},\ i \in \mathcal{N} \setminus \mathcal{S}'. \tag{51}$$

To verify (51), we need to consider two cases.

Case i): When $\operatorname{rank}(\tilde{\mathbf{F}}(\mathcal{S}' \cup \{i\})) = \operatorname{rank}(\tilde{\mathbf{F}}(\mathcal{S}')) + 1$, (51) is established by replacing $\mathcal{T}$ and $\mathcal{S}$ in (50) by $\mathcal{S}' \cup \{i\}$ and $\mathcal{S}'$, respectively.

Case ii): When $\operatorname{rank}(\tilde{\mathbf{F}}(\mathcal{S}' \cup \{i\})) = \operatorname{rank}(\tilde{\mathbf{F}}(\mathcal{S}'))$, both $\tilde{\mathbf{F}}(\mathcal{S}' \cup \{i\})$ and $\tilde{\mathbf{F}}(\mathcal{S}')$ are Hermitian matrices. Moreover, $\tilde{\mathbf{F}}(\mathcal{S}' \cup \{i\}) - \tilde{\mathbf{F}}(\mathcal{S}')$ is Hermitian and positive semi-definite (PSD), i.e.,

$$\tilde{\mathbf{F}}(\mathcal{S}' \cup \{i\}) - \tilde{\mathbf{F}}(\mathcal{S}') = \boldsymbol{\psi}_i\boldsymbol{\psi}_i^{\mathrm{T}} \succeq \mathbf{0}. \tag{52}$$

Let us define the function $\gamma_j(\mathbf{A})$ to obtain the jth $(1 \le j \le \operatorname{rank}(\mathbf{A}))$ nonincreasingly ordered eigenvalues of $\mathbf{A}$. As $\tilde{\mathbf{F}}(\mathcal{S}' \cup \{i\}) = \tilde{\mathbf{F}}(\mathcal{S}') + \boldsymbol{\psi}_i\boldsymbol{\psi}_i^{\mathrm{T}}$, according to the monotonicity theorem of Hermitian matrices [21, Corollary 4.3.12], for any $1 \le j \le n$ it holds that

$$\gamma_j(\tilde{\mathbf{F}}(\mathcal{S}' \cup \{i\})) \ge \gamma_j(\tilde{\mathbf{F}}(\mathcal{S}')). \tag{53}$$

Thus, we have $\kappa(\mathcal{S}' \cup \{i\} \mid \mathcal{N}) > \kappa(\mathcal{S}' \mid \mathcal{N})$. As a result,

$$\begin{aligned} f\left(\mathcal{S}' \cup \{i\}\right) &= \log\kappa\left(\mathcal{S}' \cup \{i\} \mid \mathcal{N}\right) + (n - r_{s'})\log\epsilon \\ &> \log\kappa\left(\mathcal{S}' \mid \mathcal{N}\right) + (n - r_{s'})\log\epsilon = f(\mathcal{S}'), \end{aligned}$$

where $r_{s'} = \operatorname{rank}(\tilde{\mathbf{F}}(\mathcal{S}'))$. At this point, (51) has been verified, i.e, $f(\mathcal{S})$ is monotone.

It follows from (51) that for any $\mathcal{T} \subseteq \mathcal{N} \setminus \{\varnothing\}$

$$f(\mathcal{T}) \ge f(\mathcal{T}_{\min}) = \log\min_{j\in\mathcal{N}}\left\{\|\boldsymbol{\psi}_j\|\right\} + (n-1)\log\epsilon. \tag{54}$$

According to (28), we arrive at $f(\mathcal{T}_{\min}) > f(\varnothing)$. This completes the proof of Lemma 1.

Appendix B
Proof of Lemma 2

We prove $f(\mathcal{S})$ (27) is a monotone submodular function.

Monotonicity: For monotonicity, it suffices to show that $f(\mathcal{S} \cup \{i\}) > f(\mathcal{S})$, for all $\mathcal{S} \subset \mathcal{N}$ and $i \in \mathcal{N} \setminus \mathcal{S}$. This has already been proved in (51) Appendix A.

Submodularity: Let $\mathcal{S} \subseteq \mathcal{Y} \subset \mathcal{N}$ and choose a generic element $i \in \mathcal{N} \setminus \mathcal{Y}$. To prove the submodularity of $f(\mathcal{S})$, we need to show that

$$f(\mathcal{S} \cup \{i\}) - f(\mathcal{S}) \ge f(\mathcal{Y} \cup \{i\}) - f(\mathcal{Y}). \tag{55}$$

Let r_y be the rank of $\tilde{\mathbf{F}}(\mathcal{Y})$. As $\tilde{\mathbf{F}}(\mathcal{Y}) = \tilde{\mathbf{F}}(\mathcal{S}) + \sum_{j\in\mathcal{Y}\setminus\mathcal{S}}\boldsymbol{\psi}_j\boldsymbol{\psi}_j^{\mathrm{T}}$, thus $r_s \le r_y \le n$. Moreover, let us define $r_{s+1} = \operatorname{rank}(\tilde{\mathbf{F}}(\mathcal{S} \cup \{i\}))$ and $r_{y+1} = \operatorname{rank}(\tilde{\mathbf{F}}(\mathcal{Y} \cup \{i\}))$, respectively. We then need to discuss $f(\mathcal{S} \cup \{i\}) - f(\mathcal{S})$ and $f(\mathcal{Y} \cup \{i\}) - f(\mathcal{Y})$ according to the values of r_{s+1} and r_{y+1}.

Case i): When $r_{s+1} = r_s$, the additional vector $\boldsymbol{\psi}_i$ is in the range of $\tilde{\mathbf{V}}_{r_s}$ i.e., $\boldsymbol{\psi}_i = \tilde{\mathbf{V}}_{r_s}\mathbf{x}_{r_s}$, where $\mathbf{x}_{r_s}$ is the coefficient vector and $\mathbf{x}_{r_s} \in \mathbb{R}^{r_s}$. As a result, we have

$$\tilde{\mathbf{F}}(\mathcal{S}\cup\{i\}) = \begin{bmatrix}\tilde{\mathbf{V}}_{r_s} & \tilde{\mathbf{V}}_{n-r_s}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{\Lambda}}_{r_s} + \mathbf{x}_{r_s}\mathbf{x}_{r_s}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{V}}_{r_s}^{\mathrm{T}} \\ \tilde{\mathbf{V}}_{n-r_s}^{\mathrm{T}}\end{bmatrix}, \tag{56}$$

where note that $\tilde{\mathbf{V}}_{n-r_s}$ is perpendicular to $\tilde{\mathbf{V}}_{r_s}$, but not unique. Sequentially, we arrive at

$$\begin{aligned} &\log\kappa\left(\mathcal{S}\cup\{i\}|\mathcal{N}\right) = \log\det\left(\tilde{\boldsymbol{\Lambda}}_{r_s} + \mathbf{x}_{r_s}\mathbf{x}_{r_s}^{\mathrm{T}}\right) \\ &= \log\det\left(\left(\mathbf{I}_{r_s} + \mathbf{x}_{r_s}\mathbf{x}_{r_s}^{\mathrm{T}}\tilde{\boldsymbol{\Lambda}}_{r_s}^{-1}\right)\tilde{\boldsymbol{\Lambda}}_{r_s}\right) \\ &= \log\left(1 + \mathbf{x}_{r_s}^{\mathrm{T}}\tilde{\boldsymbol{\Lambda}}_{r_s}^{-1}\mathbf{x}_{r_s}\right) + \log\det\left(\tilde{\boldsymbol{\Lambda}}_{r_s}\right) \qquad (57) \\ &= \log\left(1 + \boldsymbol{\psi}_i^{\mathrm{T}}\tilde{\mathbf{V}}_{r_s}\tilde{\boldsymbol{\Lambda}}_{r_s}^{-1}\tilde{\mathbf{V}}_{r_s}^{\mathrm{T}}\boldsymbol{\psi}_i\right) + \log\kappa\left(\mathcal{S}|\mathcal{N}\right), \qquad (58) \end{aligned}$$

where (58) is obtained by plugging $\mathbf{x}_{r_s} = \tilde{\mathbf{V}}_{r_s}^{\mathrm{T}} \boldsymbol{\psi}_i$ into (57). Because $r_{s+1} = r_s$, it follows from (58) that

$$\begin{aligned} f(\mathcal{S} \cup \{i\}) - f(\mathcal{S}) &= \log \kappa \left(\mathcal{S} \cup \{i\} | \mathcal{N}\right) - \log \kappa \left(\mathcal{S} | \mathcal{N}\right) \\ &= \log \left(1 + \boldsymbol{\psi}_i^{\mathrm{T}} \tilde{\mathbf{V}}_{r_s} \tilde{\boldsymbol{\Lambda}}_{r_s}^{-1} \tilde{\mathbf{V}}_{r_s}^{\mathrm{T}} \boldsymbol{\psi}_i\right). \end{aligned} \tag{59}$$

Meanwhile, as $\mathcal{S} \subseteq \mathcal{Y}$, it readily holds that $r_{y+1} = r_y$. Following the derivation of (59), it is easy to derive that

$$f(\mathcal{Y} \cup \{i\}) - f(\mathcal{Y}) = \log \left(1 + \boldsymbol{\psi}_i^{\mathrm{T}} \breve{\mathbf{V}}_{r_y} \breve{\boldsymbol{\Lambda}}_{r_y}^{-1} \breve{\mathbf{V}}_{r_y}^{\mathrm{T}} \boldsymbol{\psi}_i\right), \tag{60}$$

where $\breve{\boldsymbol{\Lambda}}_{r_y} = \mathrm{diag}([\breve{\lambda}_1, \ldots, \breve{\lambda}_{r_y}])$ and $\breve{\mathbf{V}}_{r_y} = [\breve{\mathbf{v}}_1, \ldots, \breve{\mathbf{v}}_{r_y}]$ collecting the nonzero eigenvalues and corresponding eigenvectors of $\tilde{\mathbf{F}}(\mathcal{Y})$, respectively. Note that $\tilde{\mathbf{V}}_{r_s} \tilde{\boldsymbol{\Lambda}}_{r_s} \tilde{\mathbf{V}}_{r_s}^{\mathrm{T}} \preceq \breve{\mathbf{V}}_{r_y} \breve{\boldsymbol{\Lambda}}_{r_y} \breve{\mathbf{V}}_{r_y}^{\mathrm{T}}$ following from $\tilde{\mathbf{F}}(\mathcal{S}) \preceq \tilde{\mathbf{F}}(\mathcal{Y})$. Therefore, $\boldsymbol{\psi}_i^{\mathrm{T}} \tilde{\mathbf{V}}_{r_s} \tilde{\boldsymbol{\Lambda}}_{r_s}^{-1} \tilde{\mathbf{V}}_{r_s}^{\mathrm{T}} \boldsymbol{\psi}_i \geq \boldsymbol{\psi}_i^{\mathrm{T}} \breve{\mathbf{V}}_{r_y} \breve{\boldsymbol{\Lambda}}_{r_y}^{-1} \breve{\mathbf{V}}_{r_y}^{\mathrm{T}} \boldsymbol{\psi}_i$, resulting in (59) $\geq$ (60). Thus, the desired result (55) is obtained.

Case ii): When $r_{s+1} = r_s + 1$, i.e., $\boldsymbol{\psi}_i = \tilde{\mathbf{V}}_{r_s+1} \mathbf{x}_{r_s+1}$, where $\tilde{\mathbf{V}}_{r_s+1} = [\tilde{\mathbf{V}}_{r_s} \ \tilde{\mathbf{v}}_{r_s+1}]$ and $\mathbf{x}_{r_s+1} = [\mathbf{x}_{r_s}^{\mathrm{T}} \ x_{r_s+1}]^{\mathrm{T}} \in \mathbb{R}^{r_s+1}$ being the coefficient vector. Note that $\tilde{\mathbf{v}}_{r_s+1}$ is obtained from the Gram-Schmidt orthonormalization of $\tilde{\mathbf{V}}_{r_s}$ and $\boldsymbol{\psi}_i$, i.e., $\tilde{\mathbf{v}}_{r_s+1} = (\boldsymbol{\psi}_i - \tilde{\mathbf{V}}_{r_s} \mathbf{x}_{r_s}) / x_{r_s+1}$. This readily leads to a partition of $\tilde{\mathbf{V}} = [\tilde{\mathbf{V}}_{r_s} \ \tilde{\mathbf{v}}_{r_s+1} \ \tilde{\mathbf{V}}_{n-r_s-1}]$. Accordingly,

$$\begin{aligned} \tilde{\mathbf{F}}(\mathcal{S} \cup \{i\}) = &\left[\tilde{\mathbf{V}}_{r_s} \ \tilde{\mathbf{v}}_{r_s+1} \ \tilde{\mathbf{V}}_{n-r_{s+1}}\right] \\ &\times \begin{bmatrix} \tilde{\boldsymbol{\Lambda}}_{r_s} + \mathbf{x}_{r_s} \mathbf{x}_{r_s}^{\mathrm{T}} & x_{r_s+1} \mathbf{x}_{r_s} & \mathbf{0} \\ x_{r_s+1} \mathbf{x}_{r_s}^{\mathrm{T}} & x_{r_s+1}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}_{r_s}^{\mathrm{T}} \\ \tilde{\mathbf{v}}_{r_s+1}^{\mathrm{T}} \\ \tilde{\mathbf{V}}_{n-r_{s+1}}^{\mathrm{T}} \end{bmatrix}, \end{aligned}$$

which is easy to verify by plugging in the eigenvalue decomposition of $\tilde{\mathbf{F}}(\mathcal{S})$ and $\boldsymbol{\psi}_i = \tilde{\mathbf{V}}_{r_s+1} \mathbf{x}_{r_s+1}$. It follows that

$$\begin{aligned} &\log \kappa \left(\mathcal{S} \cup \{i\} | \mathcal{N}\right) \\ &= \log \det \left(\begin{bmatrix} \tilde{\boldsymbol{\Lambda}}_{r_s} + \mathbf{x}_{r_s} \mathbf{x}_{r_s}^{\mathrm{T}} & x_{r_s+1} \mathbf{x}_{r_s} \\ x_{r_s+1} \mathbf{x}_{r_s}^{\mathrm{T}} & x_{r_s+1}^2 \end{bmatrix} \right) && (61) \\ &= \log \left(x_{r_s+1}^2 \det \left(\tilde{\boldsymbol{\Lambda}}_{r_s} \right) \right) && (62) \\ &= \left(\|\boldsymbol{\psi}_i\|^2 - \left\| \mathrm{Proj}_{\{\boldsymbol{\psi}_j\}_{j \in \mathcal{S}}} \boldsymbol{\psi}_i \right\|^2 \right) + \log \kappa \left(\mathcal{S} | \mathcal{N}\right), && (63) \end{aligned}$$

where (62) is obtained using the formula of the determinant of block matrices [22, Chapter 8.8.2] on (61). Note that $x_{r_s+1} \tilde{\mathbf{v}}_{r_s+1} = \boldsymbol{\psi}_i - \tilde{\mathbf{V}}_{r_s} \mathbf{x}_{r_s}$, where $\tilde{\mathbf{V}}_{r_s} \mathbf{x}_{r_s}$ is the projection of $\boldsymbol{\psi}_i$ onto the vector space of $\{\boldsymbol{\psi}_j\}_{j \in \mathcal{S}}$ denoted by $\mathrm{Proj}_{\{\boldsymbol{\psi}_j\}_{j \in \mathcal{S}}} \boldsymbol{\psi}_i$. Thus, $x_{r_s+1}^2 = \|\boldsymbol{\psi}_i\|^2 - \|\mathrm{Proj}_{\{\boldsymbol{\psi}_j\}_{j \in \mathcal{S}}} \boldsymbol{\psi}_i\|^2$. Meanwhile, $\log \det(\tilde{\boldsymbol{\Lambda}}_{r_s}) = \log \kappa(\mathcal{S}|\mathcal{N})$, thus (63) readily follows from (62). As a result,

$$\begin{aligned} f(\mathcal{S} \cup \{i\}) - f(\mathcal{S}) = &\log \kappa \left(\mathcal{S} \cup \{i\} | \mathcal{N}\right) + (n - r_{s+1}) \log \epsilon \\ &- \log \kappa \left(\mathcal{S} | \mathcal{N}\right) - (n - r_s) \log \epsilon \\ = &\ \|\boldsymbol{\psi}_i\|^2 - \left\| \mathrm{Proj}_{\{\boldsymbol{\psi}_j\}_{j \in \mathcal{S}}} \boldsymbol{\psi}_i \right\|^2 - \log \epsilon. \end{aligned} \tag{64}$$

On the other hand, $f(\mathcal{Y} \cup \{i\}) - f(\mathcal{Y})$ depends on the value of r_{y+1}, i.e.,

$$\begin{aligned} &f(\mathcal{Y} \cup \{i\}) - f(\mathcal{Y}) \\ &= \begin{cases} \log \left(1 + \boldsymbol{\psi}_i^{\mathrm{T}} \breve{\mathbf{V}}_{r_y} \breve{\boldsymbol{\Lambda}}_{r_y}^{-1} \breve{\mathbf{V}}_{r_y}^{\mathrm{T}} \boldsymbol{\psi}_i\right), & \text{if } r_{y+1} = r_y, \\ \|\boldsymbol{\psi}_i\|^2 - \left\| \mathrm{Proj}_{\{\boldsymbol{\psi}_j\}_{j \in \mathcal{Y}}} \boldsymbol{\psi}_i \right\|^2 - \log \epsilon, & \text{if } r_{y+1} = r_y + 1, \end{cases} \end{aligned} \tag{65}$$

where (65) can be obtained following the derivation of (64) with the orthogonal vector $\tilde{\mathbf{v}}_{r_y+1}$ obtained from the Gram-Schmidt orthonormalization of $\tilde{\mathbf{V}}_{r_s}$ and $\boldsymbol{\psi}_i$. When $r_{y+1} = r_y$, it readily holds (64) $\geq$ (60), thus yielding the desired result of (55). When $r_{y+1} = r_y + 1$, we need to compare (64) with (65). Since $\mathcal{S} \subseteq \mathcal{Y}$, $\|\mathrm{Proj}_{\{\boldsymbol{\psi}_j\}_{j \in \mathcal{S}}} \boldsymbol{\psi}_i\|^2 \leq \|\mathrm{Proj}_{\{\boldsymbol{\psi}_j\}_{j \in \mathcal{Y}}} \boldsymbol{\psi}_i\|^2$. Thus, we have (64) $\geq$ (65), i.e., $f(\mathcal{S} \cup \{i\}) - f(\mathcal{S}) \geq f(\mathcal{Y} \cup \{i\}) - f(\mathcal{Y})$.

Combining the above results[6] of ***Case i)*** and ***Case ii)***, $f(\mathcal{S})$ has been proved a submodular function. This completes the proof of Lemma 2.

Appendix C
Proof of Proposition 1

With ε satisfying (47), we now prove $\mathcal{S}_{\mathrm{opt}}^{(k)}$ is also optimal to $\mathbf{P}_{\mathrm{approx}}$, i.e.,

$$\det \left(\mathbf{F} \left(\mathcal{S}_{\mathrm{opt}}^{(k)} \right) + \varepsilon \mathbf{I} \right) > \det \left(\mathbf{F} \left(\mathcal{S}^{(k)} \right) + \varepsilon \mathbf{I} \right) \tag{66}$$

holds for any $\mathcal{S}^{(k)}$. Let us define $h(\boldsymbol{\zeta}, \varepsilon) := \prod_{j=1}^{n} (\zeta_j + \varepsilon) = \sum_{i=0}^{n} h^i(\boldsymbol{\zeta}) \varepsilon^i$. It is easy to verify that $\det(\mathbf{F}(\mathcal{S}_{\mathrm{opt}}^{(k)}) + \varepsilon \mathbf{I}) = \prod_{j=1}^{n} (\lambda_j^* + \varepsilon)$ and $\det(\mathbf{F}(\mathcal{S}^{(k)}) + \varepsilon \mathbf{I}) = \prod_{j=1}^{n} (\lambda_j' + \varepsilon)$. As a result, the r.h.s and the l.h.s of (66) can be rewritten as

$$\det \left(\mathbf{F} \left(\mathcal{S}_{\mathrm{opt}}^{(k)} \right) + \varepsilon \mathbf{I} \right) = h \left(\boldsymbol{\lambda}^*, \varepsilon \right), \tag{67}$$

and

$$\det \left(\mathbf{F} \left(\mathcal{S}^{(k)} \right) + \varepsilon \mathbf{I} \right) = h \left(\boldsymbol{\lambda}', \varepsilon \right), \tag{68}$$

respectively. Let us subtract (68) from (67). Followed by algebra operations, it readily yields $h(\boldsymbol{\lambda}^*, \varepsilon) > h(\boldsymbol{\lambda}', \varepsilon)$ with ε fulfilling (47). As a result, (66) is verified, and $\mathcal{S}_{\mathrm{opt}}^{(k)}$ is optimal to $\mathbf{P}_{\mathrm{approx}}$. This completes the proof.

References

[1] V. Roy, A. Simonetto, and G. Leus, "Spatio-temporal sensor management for environmental field estimation," *IEEE Trans. Signal Process.*, vol. 128, no. 7, pp. 369–381, Nov. 2016.

[2] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.

[3] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: Leveraging submodularity," in *Proc. IEEE Conf. Decis. Control*, 2010, pp. 2572–2577.

[4] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 684–698, Feb. 2015.

[6] Note that $r_s = n$ (i.e., $\tilde{\mathbf{F}}(\mathcal{S})$ is nonsingular) is included in *Case i)*. The derivation of *Case ii)* and *Case i)* (in particular Eqs. (63) and (58)) is in accordance with the first n and remaining l ($l > n$) steps of the `Ratio` algorithm, respectively.

[5] Y. Mo, R. Ambrosino, and B. Sinopoli, "Sensor selection strategies for state estimation in energy constrained wireless sensor networks," *Automatica*, vol. 47, no. 7, pp. 1330–1338, 2011.
[6] X. Shen and P. K. Varshney, "Sensor selection based on generalized information gain for target tracking in large sensor networks," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 363–375, Jan. 2014.
[7] C. Rusu, J. Thompson, and N. M. Robertson, "Sensor scheduling with time, energy, and communication constraints," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 528–539, Jan. 2018.
[8] C. Jiang, Y. C. Soh, and H. Li, "Sensor placement by maximal projection on minimum eigenspace for linear inverse problems," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5595–5610, Nov. 2016.
[9] H. Jamali-Rad, A. Simonetto, X. Ma, and G. Leus, "Distributed sparsity-aware sensor selection," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 5951–5964, Nov. 2015.
[10] C. Jiang, Y. C. Soh, and H. Li, "Sensor and CFD data fusion for airflow field estimation," *Appl. Thermal Eng.*, vol. 92, no. 6, pp. 149–161, 2016.
[11] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions I," *Math. Prog..*, vol. 14, no. 1, pp. 265–294, 1978.
[12] J. Ranieri, A. Chebira, and M. Vetterli, "Near-optimal sensor placement for linear inverse problems," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1135–1146, Mar. 2014.
[13] H. Jamali-Rad, A. Simonetto, and G. Leus, "Sparsity-aware sensor selection: Centralized and distributed algorithms," *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 217–220, Feb. 2014.
[14] F. Wang, G. Cheung, T. Li, Y. Du, and Y.-P. Ruan, "Fast sampling and reconstruction for linear inverse problems: From vectors to tensors," *IEEE Trans. Signal Process.*, vol. 70, pp. 6376–6391, Aug. 2022.
[15] C. Jiang, Z. Chen, R. Su, and Y. C. Soh, "Group greedy method for sensor placement," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2249–2262, May 2019.
[16] A. Hashemi, M. Ghasemi, H. Vikalo, and U. Topcu, "Randomized greedy sensor selection: Leveraging weak submodularity," *IEEE Trans. Automat. Control*, vol. 66, no. 1, pp. 199–212, Jan. 2021.
[17] Y. Shen and M. Z. Win, "Fundamental limits of wideband localization—part I: A general framework," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4956–4980, Oct. 2010.
[18] M. Z. Win, Y. Shen, and W. Dai, "A theoretical foundation of network localization and navigation," *Proc. IEEE*, vol. 106, no. 7, pp. 1136–1165, Jul. 2018.
[19] G. L. Nemhauser and L. A. Wolsey, "Best algorithms for approximating the maximum of a submodular set function," *Math. Oper. Res.*, vol. 3, no. 3, pp. 177–188, 1978.
[20] O. Roy and M. Vetterli, "The effective rank: A measure of effective dimensionality," in *Proc. IEEE Eur. Signal Process. Conf.*, 2007, pp. 606–610.
[21] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2013.
[22] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," Nov. 2008. [Online]. Available: http://matrixcookbook.com