Group Greedy Method for Sensor Placement

Chaoyang Jiang, *Member, IEEE*, Zhenghua Chen, Rong Su, *Senior Member, IEEE*, and Yeng Chai Soh, *Senior Member, IEEE*

***Abstract*—This paper discusses greedy methods for sensor placement in linear inverse problems. We comprehensively review the greedy methods in the sense of optimizing the mean squared error (MSE), the volume of the confidence ellipsoid, and the worst-case error variance. We show that the greedy method of optimizing an MSE related cost function can find a near-optimal solution. We then provide a new fast algorithm to optimize the MSE. In greedy methods, we select the sensing location one by one. In this way, the searching space is greatly reduced but many valid solutions are ignored. To further improve the current greedy methods, we propose a group-greedy strategy, which can be applied to optimize all the three criteria. In each step, we reserve a group of suboptimal sensor configurations which are used to generate the potential sensor configurations of the next step and the best one is used to check the terminal condition. Compared with the current greedy methods, the group-greedy strategy increases the searching space but greatly improve the solution performance. We find the necessary and sufficient conditions that the current greedy methods and the proposed group greedy method can obtain the optimal solution. The illustrative examples show that the group greedy method outperforms the corresponding greedy method. We also provide a practical way to find a proper group size with which the proposed group greedy method can find a solution that has almost the same performance as the optimal solution.**

***Index Terms*—Greedy method, group greedy method, submodularity, sensor placement, linear inverse problem.**

I. Introduction

SENSOR networks are often used to measure physical phenomena such as temperature, humidity, and concentration of the contaminant. Each sensor node can only observe the physical phenomenon at a particular location. With sensor networks, a physical field of interest can be estimated from sparse sensor observations by solving a linear inverse problem [1]–[10]. The number of sensor nodes and their sensing locations play a critical role in the physical field estimation.

Sensor placement is to find the least number of sensor nodes and their locations within all known potential sensing locations such that the estimation accuracy meets the requirement [1]–[8]. Sensor placement is mathematically similar to the sensor selection problem [11]–[15]. Sensor selection is to determine which nodes should be activated from all available sensor nodes such that the estimation accuracy and energy consumption meet the requirements. Sensor placement/selection for linear inverse problems can be formulated as a combinatorial optimization problem, which is NP-hard. Exhaustive search and branch-and-bound methods [16] can find the optimal solution but they both are computationally intractable, even for a moderate scale problem. Therefore, it is very interesting to find suboptimal solutions of the sensor placement problem via computationally more efficient methods.

A. Prior Arts

Three types of methods were commonly used to solve the sensor placement problem: heuristic methods, convex relaxation methods, and greedy methods.

Heuristics were used to reduce the computational cost of the exhaustive search. The simplest one is to place sensor nodes at the extrema of the principal component analysis modes of the physical field of interest [3] but this method is only useful for some special cases [4]. Other heuristics such as cross-entry optimization [15], genetic algorithms [17], particle swarm optimizer and tabu search [18] are computationally expensive and have no optimality guarantee.

The combinatorial optimization problem is nonconvex. Joshi and Boyd [13] approximated it to be a convex one by relaxing the Boolean constraints, which represent the sensing locations. Such a convex relaxation then has been applied in many other works [11], [14], [19]–[23]. The solutions of the convex optimization problems with a simple rounding procedure may lead to an ill-conditioned observation model due to the constraints approximation [2]. The local optimization technique [13] and the iterative rounding procedure [23] can enhance the results but with a very high computational cost. The storage requirement is also an issue when using the convex relaxation methods for large-scale problems.

Greedy methods determine the sensing locations one-by-one by optimizing some proxies of the estimation accuracy. The proxies include the mean squared error (MSE) which corresponds to A-optimality [6], the worst case error variance (WCEV) which corresponds to E-optimality [2], the

Manuscript received September 17, 2018; revised December 10, 2018; accepted February 19, 2019. Date of publication March 4, 2019; date of current version March 25, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xin Wang. This work was supported in part by the Building and Construction Authority (BCA) of Singapore through the NRF GBIC Program with the Project NRF2015ENC-GBICRD001-057, in part by the National Research Foundation Singapore through the NRF CRP Program under Grant NRF-CRP8-2011-03, in part by the A*STAR Industrial Internet of Things Research Program under the RIE2020 IAF-PP Grant A1788a0023, and in part by Beijing Institute of Technology Research Fund Program for Young Scholars. *(Corresponding authors: Chaoyang Jiang and Zhenghua Chen.)*

C. Jiang is with the School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China (e-mail: cjiang@bit.edu.cn).

Z. Chen is with the Institute for Infocomm Research, Agency for Science, Technology, and Research (A*STAR), Sinagpore 138632 (e-mail: chen0832@e.ntu.edu.sg).

R. Su and Y. C. Soh are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: rsu@ntu.edu.sg; eycsoh@ntu.edu.sg).

Digital Object Identifier 10.1109/TSP.2019.2903017

volume of the confidence ellipsoid (VCE) which corresponds to D-optimality [12], and the condition number [4], [5], [7] or the frame potential [1] of the observation matrix. The optimization of the condition number or the frame potential is only effective for the case where the observation vectors of all sensing locations have the same norm [1], [2]. One method is called near-optimal if it always provides a solution with guaranteed quality. If a proxy, i.e. the cost function, is monotone submodular function in terms of the set of sensing locations, the greedy method results in an $(1 - 1/e)$ optimal solution [12]. Simulation results showed that optimizing MSE [6], VCE [12] and WCEV [2] via greedy methods can obtain suboptimal solutions. However, only the VCE index was shown to be a monotone submodular function. Minimizing VCE via the greedy method results in a near-optimal solution [12].

The aforementioned studies mainly considered the centralized context. In some applications, the central processing units, which can gather all information of the sensor network, are unavailable, and hence, finding the solutions of sensor placement/selection becomes more challenging. For these cases, some effective distributed algorithms have been well-developed in [24]–[26]. Most of the current works assumed that the measurement noises are uncorrelated, and the solutions for the cases of correlated measurement noises can be found in [26], [27]. Besides the sensor placement for linear inverse problems, many other excellent sensor placement studies have focused on the continuous system [19], nonlinear model [20], energy saving [21], state estimation for dynamic system [21], [22], [28]–[31], and Gaussian process interpolation [32]–[35].

Sensor placement is of great interest in the applications of physical field reconstruction [3]–[8], [10], [19], while sensor selection is of more interest in the areas of state estimation [21], [30], target tracking [22], [24], [33], etc. The two problems can be formulated to be the same optimization problem. In general, however, the problem size of sensor placement is much larger than that of the sensor selection because the number of potential sensing locations for physical field reconstruction is usually much larger than the number of sensor nodes available for state estimation [10], [19]. The convex relaxation methods have been shown very effective for small scale problems [13], [14], [26]. If we increase the problem size, the greedy methods will outperform the convex relaxation methods [1], [2], [12]. Therefore, compared to convex relaxation methods, greedy approaches are more appealing for sensor placement in a centralized context, especially for large-scale problems.

B. Our Contributions

In this paper, we focus on greedy methods for sensor placement. Greedy methods can find suboptimal solutions with near optimality guarantee using less computational cost compared with convex relaxation methods [1], [12]. However, the greedy methods still have further room to improve their performance. In a greedy algorithm, we iteratively find a new sensing location which greatest improve the metric of interest until the performance meets the requirement. In each step, we determine one new sensing location, and such a strategy may miss some better solutions.

Inspired by the beam search algorithm in the area of natural language processing [36], to improve the solutions of the greedy methods, we propose a group greedy method in which we iteratively reserve a group of (top L in the sense of certain metric) suboptimal sensor configurations. Each sensor configuration consists of a sensor configuration reserved in the last step and a selected new sensing location. Once one of the group members (i.e. sensor configurations) meet the performance requirement, we stop the iteration and the member is our solution. In practice, adding a new sensor to a suboptimal sensor configuration of the last step may outperform adding a new sensor to the best sensor configuration of the last step. The proposed group greedy strategy is able to find these better sensor configurations.

In this work, we comprehensively review the greedy methods to minimize MSE, VCE, and WCEV for sensor placement problems. To improve the greedy methods, we then propose a group greedy method. The main contributions of this work are summarized as follows:

- We show that minimizing an MSE related cost function via the greedy method can obtain a near-optimal solution.
- We propose a new greedy algorithm which can find the same near-optimal solution in the sense of minimizing MSE but is computationally more efficient.
- We find the necessary and sufficient condition that the greedy methods can obtain the optimal solution.
- Considering the necessary and sufficient condition, we propose a group greedy strategy, which can be applied to all current greedy methods and improve their solutions.
- Compared with the solution of a greedy method, to reach the same performance, the solution of the group greedy method requires fewer sensor nodes.
- We also find the necessary and sufficient condition that the proposed group greedy method can obtain the optimal solution.
- We show that the proposed group greedy method also outperforms the greedy method for the case of correlated measurement noises,
- We provide a practical way to find a moderate group size with which the solution of the proposed group greedy method almost has the same performance with the optimal solution. Larger group size leads to better performance at the expense of additional computational load. With such a moderate group size, we can obtain a solution with high performance at a moderate computational cost.

C. Outline and Notations

The rest of this paper is organized as follows. In Section II, we introduce the linear inverse problem and the corresponding sensor placement problem. In Section III, we comprehensively review the current greedy method. In this section, we show that minimizing an MSE related cost function via the greedy method can find a near-optimal solution and we then provide a computationally more efficient algorithm for the MSE minimization. In Section IV, we develop the group greedy method. We show the effectiveness of the proposed group greedy method and analyze the influence of the group size via

illustrated examples in Section V. Finally, conclusions are given in Section VI.

This paper uses the following notations: Upper (lower) bold letters, e.g. $\mathbf{A}$ ($\mathbf{a}$) or $\mathbf{\Phi}$ ($\boldsymbol{\phi}$), indicate matrices (column vectors). $\mathbf{I}$ represents an identity matrix with proper dimension whose i-th column vector is denoted by $\mathbf{e}_i$. We use $(\cdot)^{\mathrm{T}}$, $\mathrm{tr}(\cdot)$, $\|\cdot\|$, and $\det(\cdot)$ to represent the transposition, trace, norm, and determinant operators, respectively.

II. Problem Statement

A. Linear Inverse Problem

We consider a physical field $\boldsymbol{\xi} \in \mathbb{R}^N$ described as $\boldsymbol{\xi} = \tilde{\mathbf{\Phi}}\boldsymbol{\alpha}$. $\tilde{\mathbf{\Phi}} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_N]^{\mathrm{T}} \in \mathbb{R}^{N\times n}$ is a known full column-rank matrix, $\boldsymbol{\alpha} \in \mathbb{R}^n$ is unknown, and $n \ll N$. We are to estimate the parameter vector $\boldsymbol{\alpha}$ from the measurements

$$\mathbf{y} = \mathbf{H}(\boldsymbol{\xi} + \boldsymbol{\nu}) = \mathbf{H}\tilde{\mathbf{\Phi}}\boldsymbol{\alpha} + \mathbf{H}\boldsymbol{\nu} = \mathbf{\Phi}\boldsymbol{\alpha} + \mathbf{H}\boldsymbol{\nu}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^M$, M is the number of sensor observations. We use M sensor observations to estimate the n-dimensional vector $\boldsymbol{\alpha}$, and $M \geq n$. In (1), $\mathbf{H} \in \mathbb{R}^{M\times N}$ whose i-th row is the unit vector $\mathbf{e}_{s_i}^{\mathrm{T}}$, $\mathbf{\Phi} = \mathbf{H}\tilde{\mathbf{\Phi}} = [\boldsymbol{\phi}_{s_1}, \boldsymbol{\phi}_{s_2}, ..., \boldsymbol{\phi}_{s_M}]^{\mathrm{T}} \in \mathbb{R}^{M\times n}$ is the *observation matrix* whose rows are chosen from the rows of $\tilde{\mathbf{\Phi}}$, $s_i \in \mathcal{N} = \{1, 2, ..., N\}$ corresponds to the i-th sensing location, $\boldsymbol{\phi}_{s_i}^{\mathrm{T}}$ represents the observation model of the i-th sensor node, which we call the *observation vector*. We consider the measurement noise $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$. Like the prior studies [1], [2], [12], [13], [23], let $\mathbf{C} = \sigma^2\mathbf{I}$. Hence, the noise level is independent of the sensing location, and without loss generality we can assume that $\sigma^2 = 1$.

The *minimum variance unbiased estimate* of $\boldsymbol{\alpha}$ is

$$\hat{\boldsymbol{\alpha}} = (\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{y}, \tag{2}$$

and the variance of $\hat{\boldsymbol{\alpha}}$ is $\sigma^2(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}$. With $\hat{\boldsymbol{\alpha}}$, we can easily reconstruct the physical field $\boldsymbol{\xi}$. Three criteria are commonly used to quantify the recovery performance [13], [20], [23]:

1) mean squared error (MSE)

$$\mathrm{MSE}(\hat{\boldsymbol{\alpha}}) = \sigma^2\mathrm{tr}\left((\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\right) = \sum_{i=1}^{n}\frac{\sigma^2}{\lambda_i}, \tag{3}$$

2) log volume of the confidence ellipsoid (VCE)

$$\mathrm{VCE}(\hat{\boldsymbol{\alpha}}) = \beta - \frac{1}{2}\log\det(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}) = \beta - \frac{1}{2}\log\left(\prod_{i=1}^{n}\lambda_i\right), \tag{4}$$

3) worst case error variance (WCEV)

$$\mathrm{WCEV}(\hat{\boldsymbol{\alpha}}) = \sigma^2\lambda_{\max}\left((\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\right) = \frac{\sigma^2}{\lambda_n}, \tag{5}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the eigenvalues of $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$. The parameter β in (4) is a constant which depends on σ and is independent of the sensing locations. Details about β can be found in [13].

The three performance indicators are closely related [2], [23]. In Fig. 1, we explain the three criteria from a geometrical perspective via an illustrative example. For this example,

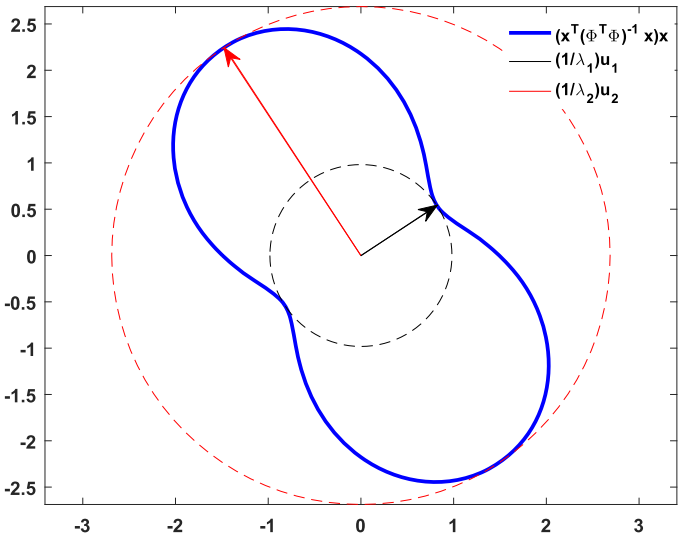


Fig. 1. A geometrical explanation of MSE, VCE, and WCEV. $\mathbf{\Phi}^{\mathrm{T}} = [\boldsymbol{\phi}_1\ \boldsymbol{\phi}_2]$, $\boldsymbol{\phi}_1 = [0.8546\ \ 0.0771]^{\mathrm{T}}$ and $\boldsymbol{\phi}_2 = [0.3077\ \ 0.7481]^{\mathrm{T}}$, which are randomly generated from the uniform distribution $\mathcal{U} \sim [0, 1]$. The MSE equals to the summation of the lengths of the red vector and the black vector. The VCE corresponds to the area in the blue curve. The WCEV equals to the length of the red vector. $\forall\mathbf{x} \in \mathbb{R}^2$ and $\|\mathbf{x}\|_2 = 1$, $(\mathbf{x}^{\mathrm{T}}(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{x})\mathbf{x}$ constitute the blue curve. $(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}$ and $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$ share the same normalized eigenvectors, i.e. $\mathbf{u}_1$ and $\mathbf{u}_2$.

considering the Courant-Fischer Minimax Theorem given in Appendix A, we can obtain that $\forall\|\mathbf{x}\|_2 = 1$,

$$\lambda_1((\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}) = \frac{1}{\lambda_2} = \max\ \mathbf{x}^{\mathrm{T}}(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{x}$$

is the maximum eigenvalue of $(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}$ and

$$\lambda_2((\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}) = \frac{1}{\lambda_1} = \min\ \mathbf{x}^{\mathrm{T}}(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{x}$$

is the minimum eigenvalue of $(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}$, which equal the magnitude of the red vector and the black vector, respectively. The direction of the vector represents the corresponding eigenvector. From this example we find that:

1) MSE corresponds to the summation of the magnitude of all (two) axes,
2) VCE corresponds to the area in the blue curve,
3) WCEV corresponds to the magnitude of the principal axis (red vector) which implies the area in the red circle.

Both equations (3)–(5) and Fig. 1 indicate that given the variance σ^2, the recovery performance depends on the eigenvalues of $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$. It is clear that given the matrix $\tilde{\mathbf{\Phi}}$, these eigenvalues depends on the sensing locations.

B. Sensor Placement Problem

We denote the set of selected sensing locations by $\mathcal{S} = \{s_1, s_2, ..., s_M\}$, and the set of potential sensing locations by $\mathcal{N} = \{1, 2, ..., N\}$, which correspond to the row indices of $\mathbf{\Phi}$ and $\tilde{\mathbf{\Phi}}$, respectively. Then, we formulate the following sensor placement problem.

Problem 1: Given the matrix $\tilde{\mathbf{\Phi}} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_N]^{\mathrm{T}} \in \mathbb{R}^{N\times n}$, select M rows of $\tilde{\mathbf{\Phi}}$ indexed by $s_1, s_2, ..., s_M$ to

construct the observation matrix $\mathbf{\Phi} = [\boldsymbol{\phi}_{s_1}, \boldsymbol{\phi}_{s_2}, ..., \boldsymbol{\phi}_{s_M}]^{\mathrm{T}} \in \mathbb{R}^{M\times n}$, such that the error of the estimated parameters $\hat{\boldsymbol{\alpha}}$ in (2) is less than the pre-defined threshold and the number of rows of $\mathbf{\Phi}$, i.e. M, is minimized.

We aim to find the minimum number of sensing locations with which the error of $\hat{\boldsymbol{\alpha}}$ is less than a predefined threshold. We use the MSE, VCE or the WCEV as the error indicator. The sensor placement problem can be formulated as the following combinatorial optimization problem

$$\begin{aligned} \text{minimize} \quad & M = |\mathcal{S}| \\ \text{subject to} \quad & \mathcal{S} \subseteq \mathcal{N} \\ & f(\mathcal{S}) \geq \gamma \end{aligned} \tag{6}$$

where $|\cdot|$ returns the cardinality of a set, and the constraint is one of the following equations:

$$f_{\mathrm{MSE}}(\mathcal{S}) = \frac{n}{\epsilon} - \mathrm{tr}(\mathbf{\Psi}^{-1}) \geq \frac{n}{\epsilon} - \gamma_{\mathrm{A}} \tag{7}$$

$$f_{\mathrm{VCE}}(\mathcal{S}) = \log\det \mathbf{\Psi} \geq \gamma_{\mathrm{D}} \tag{8}$$

$$f_{\mathrm{WCEV}}(\mathcal{S}) = \lambda_n \geq \frac{1}{\gamma_{\mathrm{E}}} \tag{9}$$

where $\gamma_{\mathrm{A}}, \gamma_{\mathrm{D}}$, and γ_{E} correspond to the predefined threshold[1] for the three metric in (3)–(5), respectively. For simplicity, in (7)–(8), we introduce the following *dual observation matrix*

$$\mathbf{\Psi} = \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi} + \epsilon\mathbf{I}, \tag{10}$$

where, $\epsilon \geq 0$ is a small number. When the number of sensor nodes is less than n, $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$ is a singular matrix. To adapt this situation, we introduce the small number ϵ.

III. Greedy Methods for Sensor Placement

The optimal solution of the combinatorial optimization problem (6) can be found by the exhaustive search in which we need to search all 2^N potential configurations. The computational cost is unaffordable for moderate scale problems.

A. General Greedy Method for Sensor Placement

The greedy method determines the sensing locations one-by-one by optimizing certain criterion. Although such a strategy ignores some valid sensor configurations, which may include suboptimal and even optimal solutions of the sensor placement problem, the greedy method has the following advantages:

- The number of searched sensor configurations can be reduced to $\sum_{i=0}^{M-1}(N-i)$ [2], which greatly reduce the searching space and hence reduce the computational cost.
- The minimum number of required sensor nodes, M, can be easily found by judging whether the constraint in (6) is satisfied after each sensing location is determined.
- If the cost function is a monotonic submodular set function, the greedy method can guarantee a near-optimal solution, which will be detailed later.

[1]Assumed that $\sigma = 1$, γ_{A} and γ_{E} are the accepted maximum MSE and WCEV, respectively. $\gamma_{\mathrm{D}} = 2(\beta - \bar{\gamma_{\mathrm{D}}})$ where $\bar{\gamma_{D}}$ is the accepted maximum log volume of the *confidence ellipsoid* of $\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}$.

Algorithm 1: The General Greedy Method.

Input: $\tilde{\mathbf{\Phi}} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_N]^{\mathrm{T}} \in \mathbb{R}^{N\times n}$, $\mathcal{N} = \{1, ..., N\}$.
Output: M and $\mathcal{S}$.
1 **Initialization**: $\mathcal{S} = \emptyset, M = 0$;
2 **while** *constraint is not satisfied* **do**
3 $\quad s^* = \arg\max_{j\in\mathcal{N}\setminus\mathcal{S}} f(\mathcal{S}\cup\{j\})$;
4 $\quad \mathcal{S} = \mathcal{S}\cup\{s^*\}, M = M+1$;
5 **end**

The general greedy method is shown in Algorithm 1. The constraint in line 2 is the equation (6), i.e. one of the constraints in (7)–(9). The cost function f in line 3, corresponding to the criterion constraint. When the constraint is satisfied, we stop and the current sensor configuration is one suboptimal solution of the sensor placement problem.

To show the optimality guarantee of the greedy method, we introduce the following dual optimization problem of (6):

$$\begin{aligned} \text{maximize} \quad & f(\mathcal{S}) \\ \text{subject to} \quad & \mathcal{S} \subseteq \mathcal{N} \\ & |\mathcal{S}| = M^* \end{aligned} \tag{11}$$

where M^* is the least number of required sensor nodes, which is unknown a prior and can be obtained from Algorithm 1.

Definition 1 (Submodular Function): Let $\mathcal{N}$ be a finite set, a set function $f: 2^{\mathcal{N}} \to \mathbb{R}$ is said to be submodular iff $\forall \mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{N}$ and $j \in \mathcal{N}\setminus\mathcal{B}$,

$$f(\mathcal{A}\cup\{j\}) - f(\mathcal{A}) \geq f(\mathcal{B}\cup\{j\}) - f(\mathcal{B}).$$

Theorem 1 ([37]): Let $f: 2^{\mathcal{N}} \to \mathbb{R}$ be a nondecreasing submodular function, and $f(\emptyset) = 0$. $\mathcal{N}$ is a finite set. Let $\mathcal{S}^*$ be the set chosen by the greedy method, $|\mathcal{S}^*| = M^*$, and let

$$\mathrm{OPT} = \underset{\mathcal{S}\subseteq\mathcal{N}, |\mathcal{S}|=M^*}{\arg\max} f(\mathcal{S})$$

be the optimal solution of the combinatorial optimization problem (11). Then,

$$f(\mathcal{S}^*) \geq \left(1 - \frac{1}{e}\right) f(\mathrm{OPT})$$

where e is the Euler's number.

Proof: See [37] or Theorem 3.5 in [38]. ■

A method is called near-optimal if it always provides a solution with guaranteed quality. Theorem 1 indicates that if $f(\mathcal{S})$ is a nondecreasing submodular function, the greedy method is near optimal. In many applications, even the cost function is not submodular, the greedy method can still provide very good solution [2], [4]–[8]. Next, we show that minimizing the MSE related cost function in (7) via the greedy method can find a near-optimal solution.

Algorithm 2: MSE Pursuit

Input: $\tilde{\mathbf{\Phi}} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_N]^{\mathrm{T}} \in \mathbb{R}^{N\times n}$, $\mathcal{N} = \{1, ..., N\}$.
Output: M and $\mathcal{S}$.
Initialization: $\mathcal{S} = \emptyset, M = 0$;
while $\mathrm{tr}(\mathbf{\Psi}_{\mathcal{S}}^{-1}) > \gamma_{\mathrm{A}}$ **do**
  $s^* = \arg\max_{j\in\mathcal{N}\setminus\mathcal{S}} f_{\mathrm{MSE}}(\mathcal{S}\cup\{j\})$;
  $\mathbf{\Psi}_{\mathcal{S}} = \mathbf{\Psi}_{\mathcal{S}} + \boldsymbol{\phi}_{s^*}\boldsymbol{\phi}_{s^*}^{\mathrm{T}}$, $\mathcal{S} = \mathcal{S}\cup\{s^*\}, M = M+1$;
end

Algorithm 3: Fast MSE Pursuit

Input: $\tilde{\mathbf{\Phi}} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_N]^{\mathrm{T}} \in \mathbb{R}^{N\times n}$, $\mathcal{N} = \{1, ..., N\}$.
Output: M and $\mathcal{S}$.
Initialization: $\mathcal{S} = \emptyset, M = 0$, and $\mathbf{\Psi}_{\mathcal{S}} = \epsilon\mathbf{I}$;
while $\mathrm{tr}(\mathbf{\Psi}_{\mathcal{S}}^{-1}) > \gamma_{\mathrm{A}}$ **do**
  $s^* = \arg\max_{j\in\mathcal{N}\setminus\mathcal{S}} g_{\mathrm{MSE}}(\mathcal{S}, j)$;
  $\mathbf{\Psi}_{\mathcal{S}} = \mathbf{\Psi}_{\mathcal{S}} + \boldsymbol{\phi}_{s^*}\boldsymbol{\phi}_{s^*}^{\mathrm{T}}$, $\mathcal{S} = \mathcal{S}\cup\{s^*\}, M = M+1$;
end

B. MSE Pursuit for Sensor Placement

If we minimize the MSE of $\hat{\boldsymbol{\alpha}}$ in (2), the constraint in (6) will be specified to be (7). We can use Algorithm 2, a special case of Algorithm 1, to solve the optimization problem.

Theorem 2: Let $\mathcal{S}^*$ be the solution of Algorithm 2,

$$f_{\mathrm{MSE}}(\mathcal{S}^*) \geq \left(1 - \frac{1}{e}\right) f_{\mathrm{MSE}}(\mathrm{OPT}) \tag{12}$$

where

$$\mathrm{OPT} = \underset{\mathcal{S}\subseteq\mathcal{N}, |\mathcal{S}|=M^*}{\arg\max} f_{\mathrm{MSE}}(\mathcal{S}).$$

Proof: See Appendix B. ■

Theorem 2 indicates that Algorithm 2 can provide a near-optimal solution for the sensor placement problem.

Although the greedy method has greatly reduced the computational cost compared with the exhaustive search and branch-and-bound methods [16], the computational cost of Algorithm 2 can be further reduced. Applying (38), (32) and (35) into (7) yields

$$\begin{aligned} f_{\mathrm{MSE}}(\mathcal{S}\cup\{j\}) &= \frac{n}{\epsilon} - \mathrm{tr}\left(\mathbf{\Psi}_{\mathcal{S}\cup\{j\}}^{-1}\right) \\ &\overset{(38)}{=} \frac{n}{\epsilon} - \mathrm{tr}\left((\mathbf{\Psi}_{\mathcal{S}} + \boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}})^{-1}\right) \\ &\overset{(32)}{=} \frac{n}{\epsilon} - \mathrm{tr}\left(\mathbf{\Psi}_{\mathcal{S}}^{-1} - \frac{\mathbf{\Psi}_{\mathcal{S}}^{-1}\boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}}\mathbf{\Psi}_{\mathcal{S}}^{-1}}{1+\boldsymbol{\phi}_j^{\mathrm{T}}\mathbf{\Psi}_{\mathcal{S}}^{-1}\boldsymbol{\phi}_j}\right) \\ &\overset{(35)}{=} \frac{n}{\epsilon} - \mathrm{tr}\left(\mathbf{\Psi}_{\mathcal{S}}^{-1}\right) + \frac{\boldsymbol{\phi}_j^{\mathrm{T}}\mathbf{\Psi}_{\mathcal{S}}^{-2}\boldsymbol{\phi}_j}{1+\boldsymbol{\phi}_j^{\mathrm{T}}\mathbf{\Psi}_{\mathcal{S}}^{-1}\boldsymbol{\phi}_j} \end{aligned} \tag{13}$$

Equation (13) implies that given $\mathcal{S}$, to maximize $f_{\mathrm{MSE}}(\mathcal{S}\cup\{j\})$, we can directly maximize the proxy

$$g_{\mathrm{MSE}}(\mathcal{S}, j) = \frac{\boldsymbol{\phi}_j^{\mathrm{T}}\mathbf{\Psi}_{\mathcal{S}}^{-2}\boldsymbol{\phi}_j}{1+\boldsymbol{\phi}_j^{\mathrm{T}}\mathbf{\Psi}_{\mathcal{S}}^{-1}\boldsymbol{\phi}_j}. \tag{14}$$

If we replace the cost function $f_{\mathrm{MSE}}(\mathcal{S}\cup\{j\})$ in Algorithm 2 by $g_{\mathrm{MSE}}(\mathcal{S}, j)$, as shown in Algorithm 3, the new algorithm can obtain the same solution at a much lower computational cost, In Algorithm 2, to maximize $f_{\mathrm{MSE}}(\mathcal{S}\cup\{j\})$, we solve the inverse of $N - |\mathcal{S}|$ matrices with the dimension $n\times n$, i.e. $\mathbf{\Psi}_{\mathcal{S}\cup\{i\}}^{-1} = (\mathbf{\Psi}_{\mathcal{S}} + \boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}})^{-1}$ for all $i \in \mathcal{N}\setminus\mathcal{S}$. In Algorithm 3, however, to maximize $g_{\mathrm{MSE}}(\mathcal{S}, j)$, we only solve the inverse of one $n\times n$ matrix, i.e. $\mathbf{\Psi}_{\mathcal{S}}$, which is computationally more efficient than Algorithm 2.

C. Minimization of VCE and WCEV

Similar to the minimization of MSE, we can minimize the VCE and the WCEV via the greedy method. The corresponding fast algorithms can be found in [12] and [2], respectively.

$f_{\mathrm{VCE}}(\mathcal{S})$ in (8) is a submodular function. Hence, minimizing VCE, i.e. maximizing $f_{\mathrm{VCE}}(\mathcal{S})$, via the greedy method can also obtain a near-optimal solution, which has been shown in [12]. Considering (38) and (36), we can obtain that

$$\begin{aligned} f_{\mathrm{VCE}}(\mathcal{S}\cup\{j\}) &= \log\det(\mathbf{\Psi}_{\mathcal{S}\cup\{j\}}) \\ &= \log\det(\mathbf{\Psi}_{\mathcal{S}} + \boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}}) \\ &= \log\det(\mathbf{\Psi}_{\mathcal{S}})(1 + \boldsymbol{\phi}_j^{\mathrm{T}}\mathbf{\Psi}_{\mathcal{S}}^{-1}\boldsymbol{\phi}_j). \end{aligned}$$

Therefore, given $\mathbf{\Psi}_{\mathcal{S}}$, to maximize $f_{\mathrm{VCE}}(\mathbf{\Psi}_{\mathcal{S}\cup\{j\}})$, we can directly maximize the proxy

$$g_{\mathrm{VCE}}(\mathcal{S}, j) = \boldsymbol{\phi}_j^{\mathrm{T}}\mathbf{\Psi}_{\mathcal{S}}^{-1}\boldsymbol{\phi}_j. \tag{15}$$

In this way, computational cost can be greatly reduced [12]. when choosing each sensing location, we only solve the inverse of one $n\times n$ matrix, i.e. $\mathbf{\Psi}_{\mathcal{S}}$, instead of computing $N - |\mathcal{S}|$ determinants.

WCEV minimization, i.e. maximizing $f_{\mathrm{WCEV}}(\mathcal{S})$, via the greedy method has been given in [2]. $f_{\mathrm{WCEV}}(\mathcal{S})$ is not a submodular function. However, the examples in [2] have shown the effectiveness of the WCEV minimization. If the number of selected sensing locations is less than n, the minimum eigenvalue of $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$ is always zero. Hence, to select the first n sensing locations, we maximize the minimum nonzero eigenvalue.

Considering the Courant-Fischer Minimax Theorem in Appendix A, we can obtain that

$$\lambda_n(\mathbf{\Psi}_{\mathcal{S}}) = \min_{\|\mathbf{x}\|_2=1} \mathbf{x}^{\mathrm{T}}\mathbf{\Psi}_{\mathcal{S}}\mathbf{x} = \mathbf{u}_n^{\mathrm{T}}(\mathbf{\Psi}_{\mathcal{S}})\mathbf{\Psi}_{\mathcal{S}}\mathbf{u}_n(\mathbf{\Psi}_{\mathcal{S}}), \tag{16}$$

and

$$\begin{aligned} \lambda_n(\mathbf{\Psi}_{\mathcal{S}\cup\{j\}}) &= \min_{\|\mathbf{x}\|_2=1} \mathbf{x}^{\mathrm{T}}\mathbf{\Psi}_{\mathcal{S}\cup\{j\}}\mathbf{x} \\ &= \min_{\|\mathbf{x}\|_2=1} \mathbf{x}^{\mathrm{T}}\mathbf{\Psi}_{\mathcal{S}}\mathbf{x} + (\boldsymbol{\phi}_j^{\mathrm{T}}\mathbf{x})^2. \end{aligned} \tag{17}$$

$\mathbf{x}^{\mathrm{T}}\mathbf{\Psi}_{\mathcal{S}}\mathbf{x}$ is minimum when $\mathbf{x} = \mathbf{u}_n(\mathbf{\Psi}_{\mathcal{S}})$, which is the eigenvector associated with the minimum eigenvalue of $\mathbf{\Psi}_{\mathcal{S}}$. Therefore,

TABLE I
SUMMARY OF GREEDY METHODS FOR SENSOR PLACEMENT

metric	cost function	cost function for the fast algorithm	constraint
MSE	$f_{\mathrm{MSE}}(\mathcal{S}\cup j)$ in (7)	$g_{\mathrm{MSE}}(\mathcal{S},j)$ in (14)	(7)
VCE	$f_{\mathrm{VCE}}(\mathcal{S}\cup j)$ in (8)	$g_{\mathrm{VCE}}(\mathcal{S},j)$ in (15)	(8)
WCEV	$f_{\mathrm{WCEV}}(\mathcal{S}\cup j)$ in (9)	$g_{\mathrm{WCEV}}(\mathcal{S},j)$ in (18)	(9)

to maximize $\lambda_n(\mathbf{\Psi}_{\mathcal{S}\cup\{j\}})$, an approximated solution is to select the observation vector $\boldsymbol{\phi}_j$ that can maximize

$$g_{\mathrm{WCEV}}(\mathcal{S},j)=(\boldsymbol{\phi}_j^{\mathrm{T}}\mathbf{u}_n(\mathbf{\Psi}_{\mathcal{S}}))^2, \tag{18}$$

which corresponds to the projection of the ϕ_j onto the eigenspace associated with the minimum eigenvalue of $\mathbf{\Psi}_{\mathcal{S}}$. In this way, when selecting each sensing location we can rank all $g_{\mathrm{WCEV}}(\mathcal{S},j)$ for $j\in\mathcal{N}\backslash\mathcal{S}$ instead of solving $N-|\mathcal{S}|$ eigenvalue problems. Such a computationally efficient algorithm for WCEV minimization can be found in [2].

In summary, we can utilize the greedy method, i.e. Algorithm 1, to optimize the three metrics for sensor placement. For all the three metrics, except the selections of the constraint in line 2 and the cost function in line 3, the other steps are identical. The selections of the cost function and the constraint for the optimization of the three metrics are summarized in Table I.

IV. GROUP GREEDY METHODS FOR SENSOR PLACEMENT

Section III shows that the current greedy methods are effective to find the suboptimal solutions but still cannot find the optimal solution. In this section, we propose a group greedy strategy, which can be applied to all current greedy methods and further improve their performance.

The solution of the sensor placement problem (6) includes sensing locations set $\mathcal{S}$ and $M^*=|\mathcal{S}|$, i.e. the least number of required sensor nodes that corresponds to the predefined threshold of a performance metric. When using the greedy method, the value of M^* is not required. After we select each sensing location we check whether the estimation performance is satisfied. If the performance metric meets the requirement, we stop the algorithm. To easily analyze the greedy methods, we introduce the following dual optimization problem:

$$\begin{aligned}\text{maximize}\quad & f(\mathcal{S})\\ \text{subject to}\quad & \mathcal{S}\subseteq\mathcal{N}\\ & |\mathcal{S}|=M\end{aligned} \tag{19}$$

in which the number of sensor nodes M is assumed to be known a prior. We can use the same greedy algorithms to find the suboptimal solution of the optimization problems (6) and (19) but with different terminating conditions. The terminating condition of (6) is that the estimation performance is satisfied while the terminating condition of (19) is that the number of selected sensing locations equals to M. Next we analyze the solution of (19). If we can improve the solution of (19) obtained from greedy methods, we can do the same operations for (6).

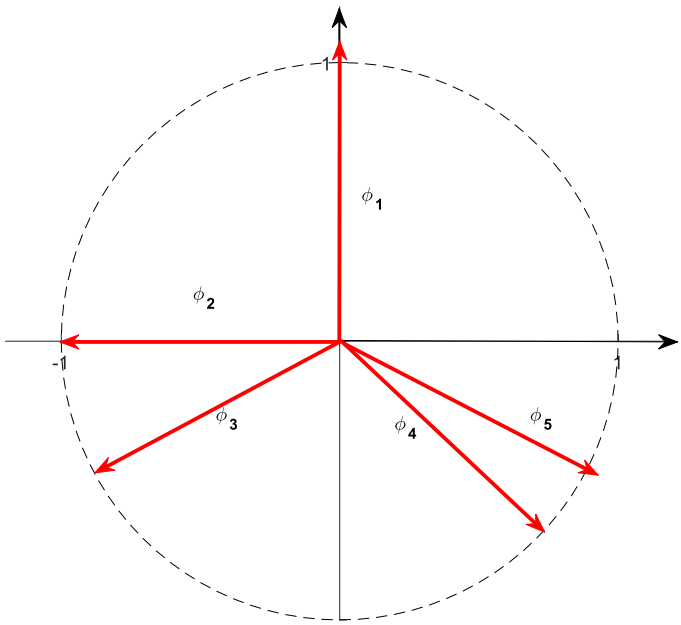


Fig. 2. An illustration example. $\|\boldsymbol{\phi}_1\|_2>\|\boldsymbol{\phi}_5\|_2>\|\boldsymbol{\phi}_2\|_2=\|\boldsymbol{\phi}_3\|_2=\|\boldsymbol{\phi}_4\|_2=1$.

A. Necessary and Sufficient Condition That Greedy Methods Can Find the Optimal Solution

Let $\mathcal{A}$ and $\mathcal{B}$ be both optimal solutions of (19) with $M_{\mathcal{A}}=|\mathcal{A}|<M_{\mathcal{B}}=|\mathcal{B}|$, i.e.

$$\mathcal{A}=\underset{|\mathcal{S}|=M_{\mathcal{A}},\mathcal{S}\subseteq\mathcal{N}}{\arg\max}\ f(\mathcal{S}), \tag{20}$$

$$\mathcal{B}=\underset{|\mathcal{S}|=M_{\mathcal{B}},\mathcal{S}\subseteq\mathcal{N}}{\arg\max}\ f(\mathcal{S}). \tag{21}$$

We denote the solution of (20) and (21) via the greedy method by $\mathcal{A}_{\mathrm{G}}$ and $\mathcal{B}_{\mathrm{G}}$, respectively. Since $M_{\mathcal{A}}<M_{\mathcal{B}}$, it is apparent that $\mathcal{A}_{\mathrm{G}}\subset\mathcal{B}_{\mathrm{G}}$. Assumed that we have found the optimal solution $\mathcal{A}$ via the greedy method. If we improve the estimation performance, i.e. decreasing γ_{A}, $\bar{\gamma}_{\mathrm{D}}$ or γ_{E}, which implies that more sensors are required, the optimal solution of (6) becomes $\mathcal{B}$. If $\mathcal{A}\nsubseteq\mathcal{B}$, we cannot find the optimal solution via the general greedy method. Next, we illustrate this situation via an example, which is shown in Fig. 2.

In this example, we have five potential sensing locations associated with the five observation vectors: $\{\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\boldsymbol{\phi}_3,\boldsymbol{\phi}_4,\boldsymbol{\phi}_5\}$. If we select two sensing locations, i.e. $M=2$ in (19), $\mathcal{A}=\{1,2\}$ will be the optimal solution. If we select three sensing locations, i.e. $M=3$ in (19), $\mathcal{B}=\{1,3,5\}$ will be the optimal solution. When we use the greedy method to minimize any criterion, i.e. MSE, VCE or WCEV, the solutions will be $\mathcal{A}_{\mathrm{G}}=\{1,2\}$ and $\mathcal{B}_{\mathrm{G}}=\{1,2,4\}$ for the two cases, respectively.

In the greedy method, when selecting more sensor nodes, the previous solution must be a subset of the current solution, i.e. $\mathcal{A}_{\mathrm{G}}\subset\mathcal{B}_{\mathrm{G}}$. This rule makes that the greedy method easily misses the optimal solution, such as the example in Fig. 2: $\mathcal{A}=\mathcal{A}_G\subset\mathcal{B}_G\neq\mathcal{B}$ because $\mathcal{A}\nsubseteq\mathcal{B}$. On the contrary, we have the following theorem.

Theorem 3: Sensor index sets $\mathcal{A}$ and $\mathcal{B}$ are given in (20)–(21), respectively. The greedy method, i.e. Algorithm 1, can obtain the optimal solution of (19) iif $\forall M_{\mathcal{A}} < M_{\mathcal{B}}$, $\mathcal{A} \subset \mathcal{B}$.

Proof: See Appendix C. ■

Theorem 3 indicates that the necessary and sufficient condition that the general greedy method can obtain the optimal solution of (19) associated with any number of sensor nodes is $\mathcal{A} \subset \mathcal{B}$. In other words, the optimal solution of (19) associated with a small number of sensor nodes is always a subset of the optimal solution associated with a large number of sensor nodes. Unfortunately, in practice, this condition is not true. To reduce the influence of $\mathcal{A} \nsubseteq \mathcal{B}$, next, we provide a group strategy for the greedy method to optimize all the three criteria.

B. Group Greedy Method

In the greedy method, when selecting each sensing location, to reduce the influence of $\mathcal{A} \nsubseteq \mathcal{B}$, except the current optimal sensor configuration we can reserve some other suboptimal configurations, which may contain a subset of the final optimal solution. In this way, we can find a better solution.

Our main idea to improve the current greedy method is that when determining each sensing location we select a group of suboptimal sensor configurations instead of the current optimal one. The current optimal sensor configuration is used to check the terminating condition, and the group members are used to generate the potential sensor configurations with one more sensor nodes. In practice, $\mathcal{A}$ is not always a subset of $\mathcal{B}$ but $\mathcal{A}_{\rm G}$ is always a subset of $\mathcal{B}_{\rm G}$. In each step of the greedy method, we reserve a group of suboptimal solution, i.e. $\{\mathcal{A}_{\rm G1}, \mathcal{A}_{\rm G2}, ..., \mathcal{A}_{\rm GL}\}$. $\mathcal{A}_{{\rm G}i}$ and one more sensor node constitute a potential sensor configuration of the next step. In this way, we guess that if there exists $i \in [1, L]$, $\mathcal{A}_{{\rm G}i} \in \mathcal{B}$, the greedy method can find the optimal solution. Next, we detail the idea and confirm the guess.

We use L to denote the group size. In each step, we rank all potential sensor configurations in terms of one criterion in a descending order. Different with Algorithm 1 in which only the current optimal sensor configuration is reserved, we reserve the best L sensor configurations which are used to generate the potential sensor configurations with one more sensor node. From the L configurations, we can obtain the best choice of the current step, which is used to check the terminating condition.

The group greedy method is shown in Algorithm 4. In Algorithm 4, the constraint in line 2 for the optimization problem (6) is one of the criterion constraints in (7)–(9) while for the optimization problem (19) the constraint is $|\mathcal{S}| = M$. In line 3, '$\rm rankde_{\it i}$' represents an operator which ranks all elements of a set in a descending order and returns the i-th element. The cost function $f(\cdot)$ can be any criterion of interest, such as the cost functions in (7)–(9).

Next, we use the example in Fig. 2 to illustrate the group greedy method. We set $L = 2$ for this example. The procedure of sensor placement for the example is given in Table II.

In the first step ($k = 1$), all sensing locations are potential choices. The best $L(=2)$ choices are $\{1\}$ and $\{5\}$, which are used to generate the potential choices of the next step, i.e. the 7 sensor configurations in the second step, by combining with one more sensing location. In the second step ($k = 2$), the best sensor configuration is $\{1, 2\}$, which is used to check the terminating condition. If the condition is not satisfied, we use the best $L(=2)$ sensor configurations, i.e. $\{1, 2\}$ and $\{1, 5\}$, to generate the potential choice of the next step, i.e. the 5 potential choices shown in the third step ($k = 3$). From the 5 potential choices, we can obtain the best one $\{1, 3, 5\}$. In this way, we can obtain the optimal solution of this example for both $M = 2$ and $M = 3$.

In this example, let $M_{\mathcal{A}} = 2$ and $M_{\mathcal{B}} = 3$. The solutions of the general greedy method are $\mathcal{A}_{\rm G} = \{1, 2\}$ and $\mathcal{B}_{\rm G} = \{1, 2, 4\}$ respectively, while the optimal solutions are $\mathcal{A} = \{1, 2\}$ and $\mathcal{B} = \{1, 3, 5\}$. Because $\mathcal{A}$ is not a subset of $\mathcal{B}$, Algorithm 1 cannot obtain the optimal solution. However, in the proposed group greedy method, i.e. Algorithm 4, when $k = 2$ we reserve the other suboptimal solution $\{1, 5\}$ which is a subset of the optimal solution $\mathcal{B}$, and hence we can obtain the optimal solution. In Algorithm 4, if one of the group members is the optimal solution of the next step, we can obtain the optimal solution of (19).

Theorem 4: Denote the optimal solution of (19) by $\mathcal{S}^*_M$. Let $\mathbb{S}_k = \{\mathcal{S}_{k1}, \mathcal{S}_{k2}, ..., \mathcal{S}_{kL}\}$. Algorithm 4 can provide the optimal solution of (19) iif for every $k \in [1, M-1]$ there exists $i \in [1, L]$ such that $\mathcal{S}_{ki} \subset \mathcal{S}^*_{k+1}$.

Proof: See Appendix D. ■

Theorem 4 shows the necessary and sufficient condition that Algorithm 4 can find the optimal solution, i.e. one of the best

Algorithm 4: The General Group Greedy Method.

Input: $\tilde{\mathbf{\Phi}} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_N]^{\rm T} \in \mathbb{R}^{N \times n}$, $\mathcal{N} = \{1, 2, ..., N\}$, $\mathcal{L} = \{1, 2, ..., L\}$.

Output: M and $\mathcal{S}$.

1 **Initialization**: $k = 0$ and $\mathcal{S}_{0i} = \emptyset, i \in \mathcal{L}$;

2 **while** *constraint is not satisfied* **do**

3 $\quad \forall i \in \mathcal{L}$, $\mathcal{S}_{k+1i} = \mathcal{S}_{kl^*_i} \cup \{j^*_i\}$ where $\{l^*_i, j^*_i\} = \underset{l_i \in \mathcal{L}, j_i \in \mathcal{N} \setminus \mathcal{S}_{kl_i}}{\arg\ \rm rankde_{\it i}} \ f(\mathcal{S}_{kl_i} \cup \{j_i\})$;

4 $\quad k = k + 1$, $\mathcal{S} = \mathcal{S}_{k1}$, and $M = k$;

5 **end**

TABLE II
THE PROCEDURE OF SENSOR PLACEMENT FOR THE EXAMPLE IN FIG. 2 VIA THE GROUP GREEDY METHOD

$k = 1$	
potential choices	$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$
best 2 configurations	$\mathcal{S}_{11} = \{1\}$, $\mathcal{S}_{12} = \{5\}$
sensor selection	$\mathcal{S} = \mathcal{S}_{11} = \{1\}$
$k = 2$	
potential choices	$\{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{2,5\}, \{3,5\}, \{4,5\}$
best 2 configurations	$\mathcal{S}_{21} = \{1, 2\}$, $\mathcal{S}_{22} = \{1, 5\}$
sensor selection	$\mathcal{S} = \mathcal{S}_{21} = \{1, 2\}$
$k = 3$	
potential choices	$\{1,2,3\}, \{1,2,4\}, \{1,2,5\}, \{1,3,5\}, \{1,4,5\}$
best 2 configurations	$S_{31} = \{1, 3, 5\}$, $S_{32} = \{1, 2, 4\}$
sensor selection	$\mathcal{S} = \mathcal{S}_{31} = \{1, 3, 5\}$

Notes: The group size $L = 2$. For this example, using different criteria yields the same solution. From all the potential choices, we can obtain the top L sensor configurations in terms of an criterion. The top L configurations are applied to generate the potential choices of the next step. Finally, we select the best sensor configuration from the top L configurations.

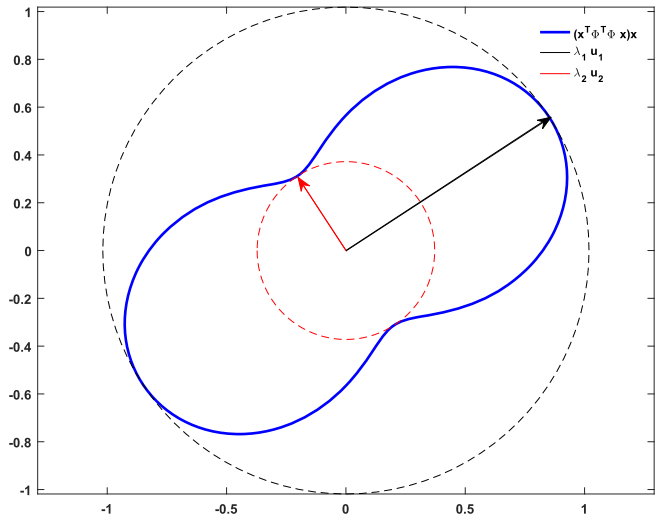


Fig. 3. A geometrical explanation of MSE, VCE, and WCEV based on the eigen-structure of the dual observation matrix. Following the example in Fig. 1, the blue curve consists of $(\mathbf{x}^{\mathrm{T}}\boldsymbol{\Psi}_{\mathcal{S}}\mathbf{x})\mathbf{x}$. Here $\|\mathbf{x}\|_2 = 1$, $\boldsymbol{\Psi}_{\mathcal{S}} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} + \epsilon\mathbf{I}$, and $\epsilon = 0$. The MSE corresponds to the harmonic mean of the lengths of the red vector and the black vector. To minimize the MSE, we should maximize the harmonic mean. The VCE corresponds to the geometrical mean of the two vectors. To minimize the VCE, we should maximize the area in the blue curve. The WCEV equals to the inverse of the length of the red vector. To minimize the WCEV, we should maximize the area in the red circle.

L sensor configurations is a subset of the optimal solution with one more sensor node. As mentioned before, if we can find the optimal solution of (19), we can also find the optimal solution of (6) using the same algorithm with a different terminating condition. Therefore, under such a condition, we can find the optimal solution for both (6) and (19).

Let $p_{\mathrm{groupgreedy}}$ denote the probability that a group of suboptimal sensor configurations include an element which is a subset of the optimal solution with one more sensor node. Let p_{greedy} denote the probability that the best one in the group is a subset of the optimal solution with one more sensor node. If the best solution of the current step is a subset of the optimal solution with one more sensor node, Algorithm 1 can find the optimal solution. Intuitively, Algorithm 4 has a higher potential to find the optimal solution than Algorithm 1 because it is always true that $p_{\mathrm{groupgreedy}} \geq p_{\mathrm{greedy}}$.

C. Influence of the Group Size

If we set the group size $L = 1$, the group greedy method will be exactly the same as the common greedy method.

In practice, we can also set L as a variable in terms of k. If we set $L(k) = \mathrm{C}_N^k$, Algorithm 4 will become the exhaustive search method, which can definitely find the optimal solution but at an unaffordable computational cost.

Generally, for the proposed group greedy method, we set the group size $1 \leq L(k) < \mathrm{C}_N^k$. Considering Fig. 3, to optimize the three criteria, we hope that the areas in both the red circle and the blue curve, which corresponds to $(\mathbf{x}^{\mathrm{T}}\boldsymbol{\Psi}_{\mathcal{S}}\mathbf{x})\mathbf{x}$, are large enough. As we know $\mathbf{x}^{\mathrm{T}}\boldsymbol{\Psi}_{\mathcal{S}}\mathbf{x} = \sum_{k=1}^{|\mathcal{S}|}(\boldsymbol{\phi}_k^{\mathrm{T}}\mathbf{x})^2$, in certain step of the general greedy method, the sensor configurations associated with a small cost function f, which has a small contribution to the area in $(\mathbf{x}^{\mathrm{T}}\boldsymbol{\Psi}_{\mathcal{S}}\mathbf{x})\mathbf{x}$, is probably not a subset of the optimal solution with one more sensor, i.e. $\mathcal{S}^*(k+1)$. In the greedy method we only consider the best sensor configuration of each step while in the proposed group greedy method we consider the top L sensor configurations of each step, which helps find a better solution.

- Compared with the general greedy method, the group greedy strategy takes a more computational effort to achieve a better solution.
- Compared with the exhaustive search method, the group greedy method drops the 'bad' configurations in each step and only reserves the best L sensor configurations, which can greatly reduce the computational cost.

Apparently, the larger the group size L is, the better is the solution of the group greedy method. However, a larger L corresponds to larger computational cost. It is not easy to give a theoretical analysis to the balance between the computational cost and the solution performance. Intuitively, if the group size is large enough, the increase of the group size only has insignificant improvement on the solution. If the group size is large enough, we drop the 'bad' sensor configurations and all these 'bad' sensor configurations have insignificant influence on the potential sensor configurations with one more sensor, i.e. the 'bad' sensor configurations corresponds to a small area in the blue curve in Fig. 3. For this case, if we increase the group size, we actually reserve these 'bad' sensor configuration in each step, which has almost no help for the solution improvement. Therefore, we guess that if the increase of group size has only insignificant help for the solution performance, the solution almost has the same performance with the optimal solution. In the next section, we verify the guess via examples. Based on the results of the examples, we provide a practical way to find a reasonable group size with which we can obtain a solution that almost has the same performance as the optimal solution.

D. Computational Cost of the Group Greedy Method

The computational effort to solve the eigenvalues/inverse of an $n \times n$ matrix is $O(n^3)$. For the MSE pursuit algorithm, i.e. Algorithm 2, when determining k-th sensing location, we solve the inverse matrices of $N - k + 1$ matrices with the dimension $n \times n$. Hence, the computational cost of Algorithm 2 is $O(NMn^3)$. Similarly, the computational effort for VCE and WCEV via the greedy method is also $O(NMn^3)$, which can be found in [12] and [2], respectively. Hence, the computational cost of the general greedy method is $O(NMn^3)$.

For the fast MSE pursuit algorithm, to determine the k-th sensing location, the main computational cost is attributed to solving the optimization problem in line 3, which costs $O((N-k+1)n^2) = O(Nn^2)$. Hence, finding all the M sensing locations costs $O(NMn^2)$. Similarly, the computational cost for VCE and WCEV minimization via fast greedy methods shown in [12] and [2] is also $O(NMn^2)$.

For the proposed group greedy method, i.e. Algorithm 4, in the k-th step, the computational cost to solve the optimization problem in line 3 of Algorithm 4 is $O(NLn^3)$. Hence, the computational cost of the group greedy method is $O(NMLn^3)$.

TABLE III
THE COMPUTATIONAL COST OF THE MENTIONED GREEDY METHODS FOR SENSOR PLACEMENT

greedy method	fast greedy method	group greedy method	fast group greedy method
$O(NMn^3)$	$O(NMn^2)$	$O(NMLn^3)$	$O(NMLn^2)$

In practice, we can use the cost functions the same as those used in the fast greedy algorithms, and the computational cost will be $O(NMLn^2)$, which we call the fast group greedy method.

We summarize the computational costs of all the mentioned methods in Table III. It is clear that the computational cost of the proposed group greedy method is closely related to the group size L. Larger group size requires more computational cost.

V. THE EFFECTIVENESS OF THE GROUP GREEDY METHOD

In this section, we provide the following three examples to show the effectiveness of the proposed group greedy method.

Example 1: $\tilde{\mathbf{\Phi}} \in \mathbb{R}^{100\times 20}$ is a uniform random matrix with independent components ϕ_{ij} following the uniform distribution $\mathcal{U} \sim [0, 1]$. The measurement noise $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Example 2: $\tilde{\mathbf{\Phi}} \in \mathbb{R}^{20\times 5}$ is a uniform random matrix with independent components ϕ_{ij} following the uniform distribution $\mathcal{U} \sim [0, 1]$. The measurement noise $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Example 3: $\tilde{\mathbf{\Phi}} \in \mathbb{R}^{100\times 20}$ is a uniform random matrix with independent components ϕ_{ij} following the uniform distribution $\mathcal{U} \sim [0, 1]$. The measurement noise $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ where $\mathbf{C} = \mathbf{R}\mathbf{R}^{\mathrm{T}}$ and the entries of $\mathbf{R}$, $r_{ij} \sim \mathcal{U}(0, \tau)$.

We use the mean MSE, VCE, and WCEV indices of 100 Monte-Carlo simulation run results of the three examples to assess the performance of different sensor placement algorithms. For the i-th ($1 \le i \le 100$) simulation run, we determine the observation matrix $\mathbf{\Phi}_k^{(i)}$ based on the random matrix $\tilde{\mathbf{\Phi}}^{(i)}$, and obtain the following MSE VCE and WCEV indices from sensor placement methods:

$$\mathrm{MSE}_k^{(i)} = \mathrm{tr}((\mathbf{\Psi}_k^{(i)})^{-1}) \tag{22}$$

$$\mathrm{VCE}_k^{(i)} = -\log\det(\mathbf{\Psi}_k^{(i)}) \tag{23}$$

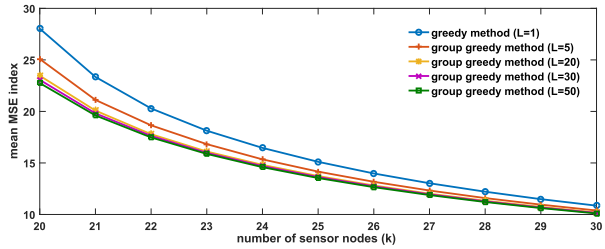
$$\mathrm{WCEV}_k^{(i)} = \lambda_{\max}((\mathbf{\Psi}_k^{(i)})^{-1}) \tag{24}$$

where k represents the number of sensor nodes, $\mathbf{\Psi}_k^{(i)} = (\mathbf{\Phi}_k^{(i)})^{\mathrm{T}}\mathbf{\Phi}_k^{(i)}$, and we ignore the constant β in (4) for the VCE index. Then, the mean MSE, VCE and WCEV indices of the 100 Monte-Carlo simulation run results are given by

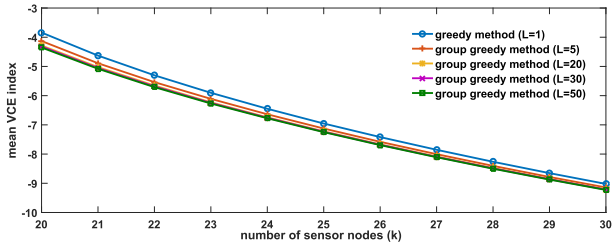
$$\overline{\mathrm{MSE}}_k = \frac{1}{100}\sum_{i=1}^{100}\mathrm{MSE}_k^{(i)} \tag{25}$$

$$\overline{\mathrm{VCE}}_k = \frac{1}{100}\sum_{i=1}^{100}\mathrm{VCE}_k^{(i)} \tag{26}$$

$$\overline{\mathrm{WCEV}}_k = \frac{1}{100}\sum_{i=1}^{100}\mathrm{WCEV}_k^{(i)} \tag{27}$$



(a) The mean MSE index, i.e. $\overline{\mathrm{MSE}}_k$ in (25).



(b) The mean VCE index, i.e. $\overline{\mathrm{VCE}}_k$ in (26).

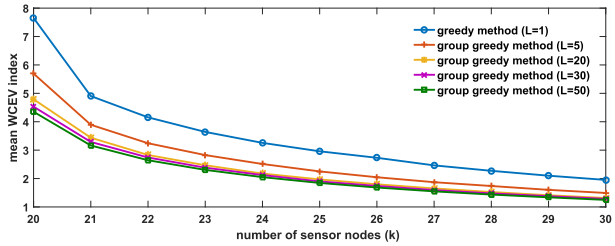
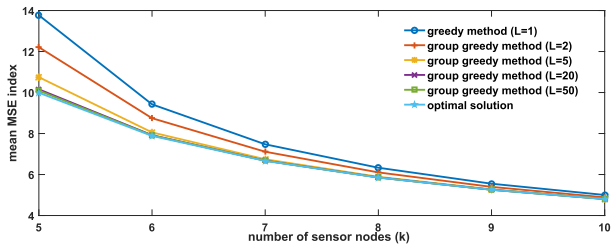


Fig. 4. The mean MSE, VCE, and WCEV indices of the solutions of *Example* 1 obtained from the group greedy method, i.e., Algorithm 4, with different group size L. If $L = 1$, the group greedy method is identical with the greedy method, i.e., Algorithm 1. When the potential sensor configurations less than the group size, we set the group size as the number of potential sensor configurations. For example, if $L = 50$, when we determining the first sensor configurations, only 20 potential sensing locations are available and we set the group size as 20. When we determining the second sensing location, the number potential sensing configurations will be more than 50 and we set the group size as 50.

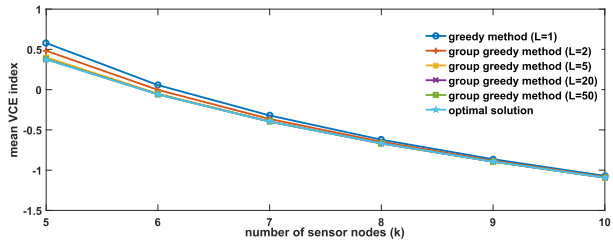
A. Performance of the Proposed Group Greedy Method

If $L = 1$, the group greedy method is the general greedy method. The results of *Example* 1 are given in Fig. 4. The findings from Fig. 4 are detailed as follows.

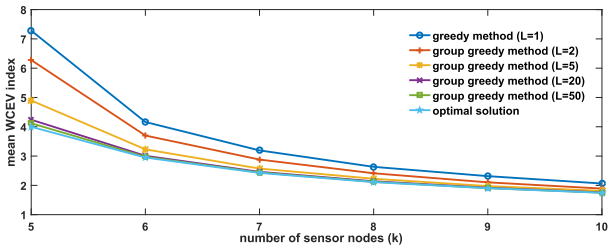
1) With more sensor nodes, the performance will be better. It is because all the three cost functions are monotonic functions w.r.t. the number of sensor nodes (M).
2) The proposed group greedy method outperforms the general greedy method ($L = 1$) in the sense of minimizing all the three metrics.
3) With larger group size L, the performance will be better.
4) All the three figures show that, compared with the greedy method ($L = 1$), the group greedy method can save sensor nodes. In other words, given the same pre-defined performance threshold, i.e. $\gamma_{\mathrm{A}}, \gamma_{\mathrm{D}}$, and γ_{E}, the solution of group greedy method is with less number of sensor nodes. For example, see Fig. 4(a), $\overline{\mathrm{MSE}}_{k-1}(L = 20) \le \overline{\mathrm{MSE}}_k$ $(L = 1)$.
5) With more sensor nodes, the differences between the results of different group size L decrease. Therefore, if the number of sensor nodes is large enough, the sensing



(a) The mean MSE index, i.e. $\overline{\mathrm{MSE}}_k$ in (25).



(b) The mean VCE index, i.e. $\overline{\mathrm{VCE}}_k$ in (26).



(c) The mean WCEV index, i.e. $\overline{\mathrm{WCEV}}_k$ in (27).

Fig. 5. The performance of the proposed group greedy method for *Example* 2. The optimal solution is obtained from the exhaustive search method.

locations are not so critical. Fig. 4(a)–4(b) shows that when the number of sensor nodes is 30, different group size almost provide the same mean MSE index and mean VCE index.

6) With larger group size L, the group greedy method has a higher capacity to find the critical sensing locations. The number of sensor nodes is smaller, the performance improvement w.r.t. the increase of group size is larger.
7) If the group size is large enough, the increase of group size makes insignificant improvement of the performance indices. If $L \geq 20$, the increase of L almost has no help for the improvement of the three performance indices but it greatly increases the computational cost.
8) If the group size is large enough, e.g., $L(k) = \mathrm{C}_N^k$, the solution will be the optimal solution. Considering the facts of 3) and 7), we guess that if we increase the group size and the new solution has insignificant improvement on the performance, the new solution would have almost the same performance as the optimal solution. It is not easy to obtain the optimal solution of *Example* 1 in which $N = 100$ and $n = 20$. To verify our conjecture, next, we apply the proposed methods for a small scale problem, i.e. *Example* 2.

The results of *Example* 2 are shown in Fig. 5 from which we can obtain the same findings, i.e. the items 1)-7). In addition, Fig. 5 provides the performances of the optimal solution of *Example* 2. We find that if the group size (L) is large enough (for this example $L \geq 20$), the increase of group size almost has no help for the improvement of the performance indices. Meanwhile, the performance of the solution is almost the same as the optimal solution, which is obtained from the exhaustive search method.

All the three sub-figures in Fig. 5 verify our previous conjecture: if we increase the group size but the solution has insignificant improvement, we can view the solution has almost the same performance as the optimal solution. Therefore, in practice, we can test two different group size, if the results associated with the large group size almost have no improvement from the results associated with the small one, we can view the group size is large enough, and the performance of the solution is almost the same as that of the optimal solution.

B. The Case of Correlated Measurement Noises

In aforementioned parts, we mainly consider the cases of uncorrelated measurement noises. Here we show that for the case of correlated measurement noises, the proposed group greedy method still outperforms the general greedy methods.

If the measurement noises are correlated, i.e. the variance $\mathbf{C}$ is not a diagonal matrix, the MVUE of $\boldsymbol{\alpha}$ is

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{\Phi}^{\mathrm{T}}(\mathbf{H}\mathbf{C}\mathbf{H}^{\mathrm{T}})^{-1}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}(\mathbf{H}\mathbf{C}\mathbf{H}^{\mathrm{T}})^{-1}\mathbf{y},$$

and the covariance of $\hat{\boldsymbol{\alpha}}$ is $\boldsymbol{\Sigma} = (\boldsymbol{\Phi}^{\mathrm{T}}(\mathbf{H}\mathbf{C}\mathbf{H}^{\mathrm{T}})^{-1}\boldsymbol{\Phi})^{-1}$ [39, p. 186]. Correspondingly, we can obtain the metrics to assess the estimation performance:

$$\mathrm{MSE}(\hat{\boldsymbol{\alpha}}) = \mathrm{tr}\left((\boldsymbol{\Phi}^{\mathrm{T}}(\mathbf{H}\mathbf{C}\mathbf{H}^{\mathrm{T}})^{-1}\boldsymbol{\Phi})^{-1}\right), \tag{28}$$

$$\mathrm{VCE}(\hat{\boldsymbol{\alpha}}) = \beta - \frac{1}{2}\log\det\left(\boldsymbol{\Phi}^{\mathrm{T}}(\mathbf{H}\mathbf{C}\mathbf{H}^{\mathrm{T}})^{-1}\boldsymbol{\Phi}\right), \tag{29}$$

$$\mathrm{WCEV}(\hat{\boldsymbol{\alpha}}) = \lambda_{\max}\left((\boldsymbol{\Phi}^{\mathrm{T}}(\mathbf{H}\mathbf{C}\mathbf{H}^{\mathrm{T}})^{-1}\boldsymbol{\Phi})^{-1}\right). \tag{30}$$

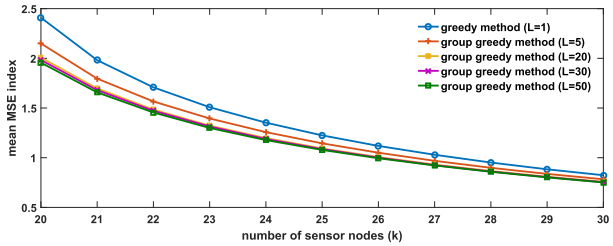
The fast greedy algorithm is not suitable for the cases of correlated measurement noises. However, the greedy method, i.e. Algorithm 1, still works well. In practice, we only replace the corresponding cost functions and constraints in Algorithm 1 to be:

$$f_{\mathrm{MSE}}(\mathcal{S}) = -\mathrm{tr}\left((\boldsymbol{\Phi}^{\mathrm{T}}(\mathbf{H}\mathbf{C}\mathbf{H}^{\mathrm{T}})^{-1}\boldsymbol{\Phi})^{-1}\right) \geq -\gamma_{\mathrm{A}},$$

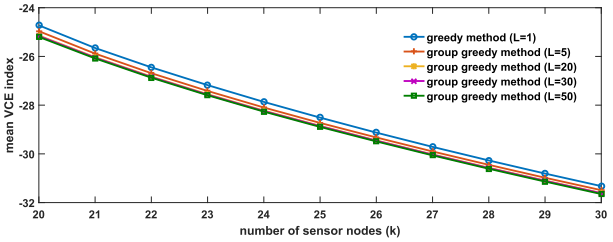
$$f_{\mathrm{VCE}}(\mathcal{S}) = \log\det\left(\boldsymbol{\Phi}^{\mathrm{T}}(\mathbf{H}\mathbf{C}\mathbf{H}^{\mathrm{T}})^{-1}\boldsymbol{\Phi}\right) \geq \gamma_{\mathrm{D}},$$

$$f_{\mathrm{WCEV}}(\mathcal{S}) = \lambda_{\min}\left(\boldsymbol{\Phi}^{\mathrm{T}}(\mathbf{H}\mathbf{C}\mathbf{H}^{\mathrm{T}})^{-1}\boldsymbol{\Phi}\right) \geq \frac{1}{\gamma_{\mathrm{E}}}.$$

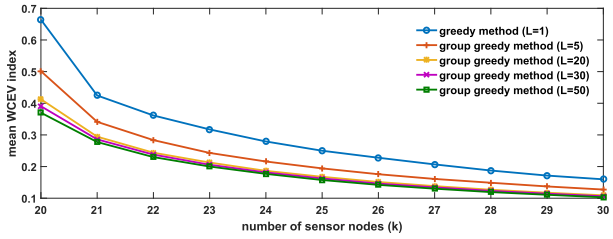
Like the case of uncorrelated measurement noises, the proposed group greedy strategy can be applied for the case of correlated measurement noises. We applied the group greedy method for *Example* 3 in which the variance of the measurement noise $\mathbf{C}$ is not a diagonal matrix. In this example, the parameter τ represents the noise level. We set $\tau = 0.1$, and the corresponding simulation results are given in Fig. 6, which shows that the proposed group greedy method outperforms the general greedy methods. Compared Fig. 6 with Fig. 4, we can obtain the similar conclusions for the case of the correlated measurement noises.



(a) The mean MSE index in (25).



(b) The mean VCE index in (26).



(c) The mean WCEV index in (27).

Fig. 6. The performance of the proposed group greedy method for *Example* 3 in which the measurement noises are correlated. The noise level $\tau = 0.1$. For this correlated noise case, in equations (25)–(27) and the corresponding equations (22)–(24), the matrix $\mathbf{\Psi}_k^{(i)}$ is instead of $(\mathbf{\Phi}_k^{(i)})^{\mathrm{T}} \left(\mathbf{H}_k^{(i)} \mathbf{C} (\mathbf{H}_k^{(i)})^{\mathrm{T}}\right)^{-1} \mathbf{\Phi}_k^{(i)}$.

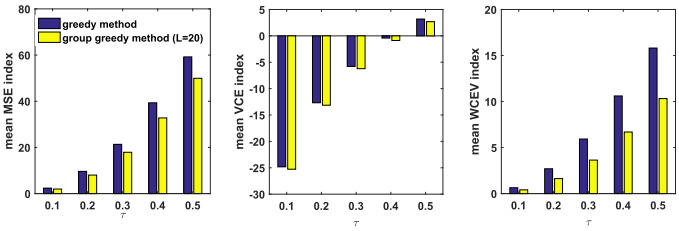
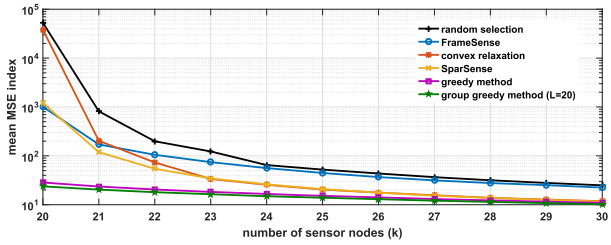
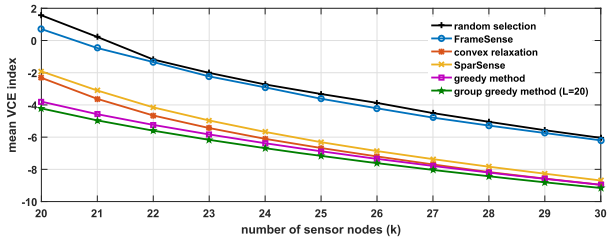


Fig. 7. Simulation results for *Example* 3 with different noise level τ. The number of selected sensor nodes is 20, i.e. $M = 20$.

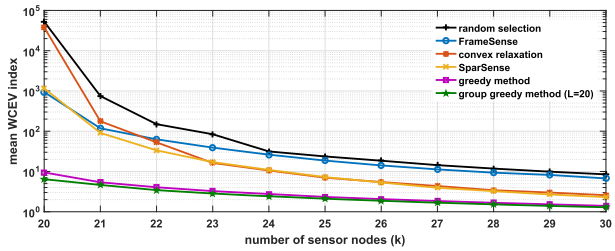
If the variance matrix $\mathbf{C} = \sigma^2 \mathbf{I}$, the signal-to-noise ratio has no influence on the solution of the sensor placement. As shown in (3)–(5), the noise level (σ or β) and the sensing location involved parameter ($\mathbf{\Phi}$) are uncorrelated. If the measurement noises are correlated, as can be seen in (28)–(30), the variance matrix $\mathbf{C}$ is correlated with the observation matrix $\mathbf{\Phi}$, which depends on the selection of sensing locations. To show the impact of signal-to-noise ratio, we show the results of *Example* 3 in which the noise level τ is set to be different values, in Fig. 7. Obviously, increasing noise level increases the three metric indices, and the proposed group greedy method outperforms the greedy method for all noise levels. In addition, Fig. 7 shows that with a higher noise level, the proposed method has a larger performance improvement.



(a) The mean MSE index, i.e. $\overline{\mathrm{MSE}}_k$ in (25).



(b) The mean VCE index, i.e. $\overline{\mathrm{VCE}}_k$ in (26).



(c) The mean WCEV index, i.e. $\overline{\mathrm{WCEV}}_k$ in (27).

Fig. 8. Performance comparison between the proposed group greedy method and the state-of-the-art for *Example* 1. The results of convex relaxation are without the operation of the so called local optimization in [13]. The SparSense refers to the centralized SparSense algorithm in [14] and [26].

C. Comparison With the State-of-the-Art

We apply the random selection, the FrameSense [1], the convex relaxation [13], the SparSense [14], [26], the greedy methods [2], [12], and the proposed group greedy method for *Example* 1. The simulation results are shown in Fig. 8. Here, the random selection means randomly select the sensing location one-by-one from the set of potential sensing locations.

In practice, these state-of-the-art approaches have their own advantages for some special cases. The computational cost of the random selection is lowest. FrameSense can obtain very good solution when the matrix $\tilde{\mathbf{\Phi}}$ corresponds to a tight frame, i.e. all the rows of $\tilde{\mathbf{\Phi}}$ have the same Euclidean norm [1]. With an operation of local optimization for the solution of convex relaxation, the performance can be greatly improved [13]. The SparSense can be easily extended for the distributed sensor placement [14], [26] while all the other mentioned approaches cannot. However, in a general centralized context, the proposed group greedy method can provide the best solution, as can be seen from Fig. 8.

VI. Conclusions

Sensor placement for linear inverse problems is a challenging combinatorial problem. The optimal solution can be obtained from the exhaustive search method but with extremely high computational cost. The greedy method can find near-optimal solutions with moderate computational cost. In this paper, we proposed a group greedy strategy which can improve the greedy method.

We perform Monte-Carlo simulations to show the effectiveness of the proposed group greedy method. Based on the simulation results, we conclude that:

- The group greedy method outperforms the state-of-the-art approaches including the common greedy method, the FrameSense, the SparSense, and the convex relaxation method.
- To meet certain performance requirement, compared with the greedy method, the proposed group greedy method has a potential to save sensor nodes.
- Larger group size leads to a better solution. If the group size is large enough, the group greedy method can obtain the optimal solution.
- If the group size is larger than a certain value, the increase of group size almost has no help for the improvement of performance indices, and hence the solution almost has the same performance as the optimal solution.
- The group greedy method is also effective for the case of correlated measurement noises, and with higher noise level, the proposed method has a larger performance improvement from the corresponding greedy method.

Appendix A

We provide some *a prior* results in linear algebra which are used in this paper.

Theorem 5 (Courant-Fischer Minimax Theorem): If $\mathbf{A} \in \mathbb{R}^{n\times n}$ is symmetric and the eigenvalues $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$, then for $i = 1 : n$,

$$\lambda_i(\mathbf{A}) = \max_{\dim(\mathbb{U})=i} \min_{\mathbf{0}\neq\mathbf{x}\in\mathbb{U}} \frac{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}} \tag{31}$$

Proof: See the proof of Theorem 8.1.2 in [40]. ■

Lemma 6: Suppose $\mathbf{B} = \mathbf{A} + \mathbf{c}\mathbf{c}^{\mathrm{T}}$ where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is symmetric, and $\mathbf{c} \in \mathbb{R}^n$ is a non-zero vector. Then,

$$\lambda_1(\mathbf{B}) \geq \lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{B}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{B}) \geq \lambda_n(\mathbf{A})$$

Proof: See [41] and the Theorem 8.1.8 in [40]. ■

Lemma 7 (Sherman-Morrison Formula): Suppose $\mathbf{A} \in \mathbb{R}^{n\times n}$ is invertible and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. If $\mathbf{A} + \mathbf{u}\mathbf{v}^{\mathrm{T}}$ is invertible,

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^{\mathrm{T}})^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}}{1 + \mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{u}}. \tag{32}$$

Proof: See [42].

The following rules can be found from [43]:

$$\frac{\mathrm{d}\mathrm{tr}(\mathbf{A}(x))}{\mathrm{d}x} = \mathrm{tr}\left(\frac{\mathrm{d}\mathbf{A}(x)}{\mathrm{d}x}\right), \tag{33}$$

$$\frac{\mathrm{d}\mathbf{A}(x)^{-1}}{\mathrm{d}x} = -\mathbf{A}(x)^{-1}\frac{\mathrm{d}\mathbf{A}(x)}{\mathrm{d}x}\mathbf{A}(x)^{-1}, \tag{34}$$

$$\mathrm{tr}(\mathbf{A}\mathbf{B}) = \mathrm{tr}(\mathbf{B}\mathbf{A}), \tag{35}$$

$$\begin{aligned}\det(\mathbf{A} + \mathbf{u}\mathbf{v}^{\mathrm{T}}) &= \det(\mathbf{A})\det(\mathbf{I} + (\mathbf{A}^{-1}\mathbf{u})\mathbf{v}^{\mathrm{T}}) \\ &= \det(\mathbf{A})(1 + \mathbf{u}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{v})\end{aligned} \tag{36}$$

■

Lemma 8: $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$, $\mathbf{A} \succeq \mathbf{0}$ and $\mathbf{B} \preceq \mathbf{0}$, $\mathrm{tr}(\mathbf{A}\mathbf{B}) \leq 0$.

Proof:

$$\mathrm{tr}(\mathbf{A}\mathbf{B}) = \mathrm{tr}(\sqrt{\mathbf{A}}\sqrt{\mathbf{A}}\mathbf{B}) \overset{(35)}{=} \mathrm{tr}(\sqrt{\mathbf{A}}\mathbf{B}\sqrt{\mathbf{A}}) \leq 0 \tag{37}$$

Appendix B
Proof of Theorem 2

If $\mathcal{S} = \emptyset$, $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$. Substituting these zero eigenvalues into (7) yields that $f_{\mathrm{MSE}}(\emptyset) = 0$.

We denote the observation matrix and the dual observation matrix corresponding to the sensor configuration $\mathcal{S}$ by $\mathbf{\Phi}_{\mathcal{S}}$ and $\mathbf{\Psi}_{\mathcal{S}}$, respectively. We find that

$$\begin{aligned}\mathbf{\Psi}_{\mathcal{S}\cup\{j\}} &= \begin{bmatrix}\mathbf{\Phi}_{\mathcal{S}}^{\mathrm{T}} & \boldsymbol{\phi}_j\end{bmatrix}\begin{bmatrix}\mathbf{\Phi}_{\mathcal{S}}^{\mathrm{T}} & \boldsymbol{\phi}_j\end{bmatrix}^{\mathrm{T}} + \epsilon\mathbf{I} \\ &= \mathbf{\Psi}_{\mathcal{S}} + \boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}}.\end{aligned} \tag{38}$$

Then we can obtain from Lemma 6 that

$$\begin{aligned}\lambda_1(\mathbf{\Psi}_{\mathcal{S}\cup\{j\}}) \geq \lambda_1(\mathbf{\Psi}_{\mathcal{S}}) \geq \lambda_2(\mathbf{\Psi}_{\mathcal{S}\cup\{j\}}) \geq \lambda_2(\mathbf{\Psi}_{\mathcal{S}}) \geq \\ \cdots \geq \lambda_n(\mathbf{\Psi}_{\mathcal{S}\cup\{j\}}) \geq \lambda_n(\mathbf{\Psi}_{\mathcal{S}})\end{aligned} \tag{39}$$

Obviously, both $\mathbf{\Psi}_{\mathcal{S}}$ and $\mathbf{\Psi}_{\mathcal{S}\cup\{j\}}$ are positive semi-define matrices. From (39), we can obtain that

$$\sum_{i=1}^{n}\frac{1}{(\lambda_i(\mathbf{\Psi}_{\mathcal{S}\cup\{j\}}))} \leq \sum_{i=1}^{n}\frac{1}{(\lambda_i(\mathbf{\Psi}_{\mathcal{S}}))}$$

from which we can obtain

$$f_{\mathrm{MSE}}(\mathcal{S}\cup\{j\}) \geq f_{\mathrm{MSE}}(\mathcal{S}).$$

Hence, the cost function f_{MSE} is a nondecreasing function.

Considering Theorem 1, if $f_{\mathrm{MSE}}(\mathcal{S}), \mathcal{S} \subseteq \mathcal{N}$ is a submodular function, (12) holds. Next, we show that $f_{\mathrm{MSE}}(\mathcal{S})$ is a submodular function, i.e. $\forall\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{N}$ and $j \in \mathcal{N}\setminus\mathcal{B}$,

$$\begin{aligned}h = f_{\mathrm{MSE}}(\mathcal{A}\cup\{j\}) - f_{\mathrm{MSE}}(\mathcal{A}) - f_{\mathrm{MSE}}(\mathcal{B}\cup\{j\}) \\ + f_{\mathrm{MSE}}(\mathcal{B}) \geq 0.\end{aligned}$$

Considering (7), we can obtain that

$$\begin{aligned}f_{\mathrm{MSE}}(\mathcal{A}\cup\{j\}) - f_{\mathrm{MSE}}(\mathcal{A}) &= \mathrm{tr}\left(\mathbf{\Psi}_{\mathcal{A}}^{-1}\right) - \mathrm{tr}\left(\mathbf{\Psi}_{\mathcal{A}\cup\{j\}}^{-1}\right) \\ &= \mathrm{tr}\left(\mathbf{\Psi}_{\mathcal{A}}^{-1}\right) - \mathrm{tr}\left((\mathbf{\Psi}_{\mathcal{A}} + \boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}})^{-1}\right).\end{aligned}$$

Let

$$g(x) = \mathrm{tr}\left((\mathbf{\Psi}_{\mathcal{A}} + x\mathbf{\Omega})^{-1}\right) - \mathrm{tr}\left((\mathbf{\Psi}_{\mathcal{A}} + x\mathbf{\Omega} + \boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}})^{-1}\right)$$

where $x \in [0,1]$ and $\mathbf{\Omega} = \mathbf{\Psi}_{\mathcal{B}} - \mathbf{\Psi}_{\mathcal{A}} \succeq \mathbf{0}$. It is obvious that

$$g(0) = f_{\mathrm{MSE}}(\mathcal{A} \cup \{j\}) - f_{\mathrm{MSE}}(\mathcal{A}),$$
$$g(1) = f_{\mathrm{MSE}}(\mathcal{B} \cup \{j\}) - f_{\mathrm{MSE}}(\mathcal{B}),$$

and $h = g(0) - g(1)$.

Taking the derivative of $g(x)$ with respect to x yields

$$\begin{aligned}\frac{\mathrm{d}g(x)}{\mathrm{d}x} &= \frac{\mathrm{dtr}\left((\mathbf{\Psi}_{\mathcal{A}} + x\mathbf{\Omega})^{-1}\right)}{\mathrm{d}x} \\ &\quad - \frac{\mathrm{dtr}\left((\mathbf{\Psi}_{\mathcal{A}} + x\mathbf{\Omega} + \boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}})^{-1}\right)}{\mathrm{d}x} \\ &\overset{(33)}{=} \mathrm{tr}\left(\frac{\mathrm{d}(\mathbf{\Psi}_{\mathcal{A}} + x\mathbf{\Omega})^{-1}}{\mathrm{d}x}\right. \\ &\quad \left. - \frac{\mathrm{d}(\mathbf{\Psi}_{\mathcal{A}} + x\mathbf{\Omega} + \boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}})^{-1}}{\mathrm{d}x}\right) \\ &\overset{(34)}{=} \mathrm{tr}\left((\mathbf{\Psi}_{\mathcal{A}} + x\mathbf{\Omega} + \boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}})^{-1}\mathbf{\Omega}(\mathbf{\Psi}_{\mathcal{A}} + x\mathbf{\Omega}\right. \\ &\quad \left. + \boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}})^{-1} - (\mathbf{\Psi}_{\mathcal{A}} + x\mathbf{\Omega})^{-1}\mathbf{\Omega}(\mathbf{\Psi}_{\mathcal{A}} + x\mathbf{\Omega})^{-1}\right) \\ &\overset{(35)}{=} \mathrm{tr}\left(\mathbf{\Omega}\left((\mathbf{\Psi}_{\mathcal{A}} + x\mathbf{\Omega} + \boldsymbol{\phi}_j\boldsymbol{\phi}_j^{\mathrm{T}})^{-2} - (\mathbf{\Psi}_{\mathcal{A}} + x\mathbf{\Omega})^{-2}\right)\right) \\ &\overset{(37)}{\leq} 0\end{aligned}$$

Hence, $g(0) \geq g(1)$ and $h = g(0) - g(1) \geq 0$.

APPENDIX C
PROOF OF THEOREM 3

If Algorithm 1 can find the optimal solution of (19), we can easily find that $\mathcal{A} \subset \mathcal{B}$.

Next, we use the mathematical induction technique to show that if $\mathcal{A} \subset \mathcal{B}$, Algorithm 1 can find the optimal solution. We introduce a set sequence $\mathcal{S}_k^* \subseteq \mathcal{N}$, $k = |\mathcal{S}_k^*|$, and

$$f(\mathcal{S}_k^*) = \underset{|\mathcal{S}|=k, \mathcal{S}\subseteq\mathcal{N}}{\arg\max} f(\mathcal{S})$$

Obviously, $\mathcal{S}_1^* = \{i^*\}$ and

$$\boldsymbol{\phi}_{i^*} = \underset{\boldsymbol{\phi}_i, i\in\mathcal{N}}{\arg\max} \|\boldsymbol{\phi}_i\|_2 \tag{40}$$

no matter which cost function (i.e. f_{MSE}, f_{VCE} or f_{WCEV}) is selected. Meanwhile, the first sensing location obtained from Algorithm 1 is exactly i^* for all the three different cost functions.

$\forall k < N$ if $\mathcal{S}_k^*$ is obtained from Algorithm 1, the $(k+1)$-th sensing location obtained from Algorithm 1 is

$$s^* = \underset{j\in\mathcal{N}\setminus\mathcal{S}_k^*}{\arg\max} f(\mathcal{S}_k^* \cup \{j\}).$$

Since $\mathcal{S}_k^* \subset \mathcal{S}_{k+1}^*$, we can find that

$$f(\mathcal{S}_{k+1}^*) = \max_{|\mathcal{S}|=k+1, \mathcal{S}\subseteq\mathcal{N}} f(\mathcal{S}) = \max_{j\in\mathcal{N}\setminus\mathcal{C}_k} f(\mathcal{S}_k^* \cup \{j\}).$$

which implies that $\mathcal{S}_{k+1}^* = \mathcal{S}_k^* \cup \{s^*\}$, i.e. we can obtain the optimal solution $\mathcal{S}_{k+1}^*$ from Algorithm 1. Therefore, if $\forall M_{\mathcal{A}} < M_{\mathcal{B}}$, $\mathcal{A} \subset \mathcal{B}$, then Algorithm 1 can find the optimal solution of (19).

APPENDIX D
PROOF OF THEOREM 4

We use the mathematical induction technique to show if $\forall k \in [1, M-1]$ there exists $i \in [1, L]$ such that $\mathcal{S}_{ki} \subset \mathcal{S}_{k+1}^*$, Algorithm 4 can find the optimal solution.

No matter which cost function (i.e. f_{MSE}, f_{VCE} or f_{WCEV}) is selected, Algorithm 4 can obtain the optimal solution for $M = 1$, i.e. $\mathcal{S}_1^* = \{i^*\}$ where i^* is given in (40).

$\forall k < N$ we assume that $\mathcal{S}_k^*$ is obtained from Algorithm 4. Let $M = k+1$ and we can obtain from Algorithm 4 that

$$\{j^*, l^*\} = \underset{l\in\mathcal{L}, j\in\mathcal{N}\setminus\mathcal{S}_{kl}}{\arg\max} f(\mathcal{S}_{kl} \cup \{j\})$$

and $\mathcal{S}_{(k+1)1} = \mathcal{S}_{kl^*} \cup \{j^*\}$.

Since $\exists l \in \mathcal{L}, \mathcal{S}_{kl} \subset \mathcal{S}_{k+1}^*$, we have

$$f(\mathcal{S}_{k+1}^*) = \max_{|\mathcal{S}|=k+1, \mathcal{S}\subseteq\mathcal{N}} f(\mathcal{S}) = \max_{l\in\mathcal{L}, j\in\mathcal{N}\setminus\mathcal{S}_{kl}} f(\mathcal{S}_{kl} \cup \{j\})$$

from which we can find the optimal solution $\mathcal{S}_{k+1}^* = \mathcal{S}_{kl^*} \cup \{j^*\}$. Therefore, if for every $k \in [1, M-1]$ there exists $i \in [1, L]$ such that $\mathcal{S}_{ki} \subset \mathcal{S}_{k+1}^*$, then Algorithm 4 can provide the optimal solution of (19).

REFERENCES

[1] 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.

[2] 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.

[3] K. Cohen, S. Siegel, and T. McLaughlin, "A heuristic approach to effective sensor placement for modeling of a cylinder wake," *Comput. Fluids*, vol. 35, no. 1, pp. 103–120, 2006.

[4] K. Willcox, "Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition," *Comput. Fluids*, vol. 35, no. 2, pp. 208–226, 2006.

[5] B. Yildirim, C. Chryssostomidis, and G. Karniadakis, "Efficient sensor placement for ocean measurements using low-dimensional concepts," *Ocean Model.*, vol. 27, no. 3, pp. 160–173, 2009.

[6] C. Jiang, Y. C. Soh, and H. Li, "Sensor and CFD data fusion for airflow field estimation," *Appl. Thermal Eng.*, vol. 92, pp. 149–161, 2016.

[7] P. Astrid, S. Weiland, K. Willcox, and T. Backx, "Missing point estimation in models described by proper orthogonal decomposition," *IEEE Trans. Autom. Control*, vol. 53, no. 10, pp. 2237–2251, Nov. 2008.

[8] C. Jiang, Y. C. Soh, and H. Li, "Two-stage indoor physical field reconstruction from sparse sensor observations," *Energy Build.*, vol. 151, pp. 548–563, 2017.

[9] C. Jiang, Y. C. Soh, H. Li, and H. Zhou, "Physical field estimation from CFD database and sparse sensor observations," in *Proc. IEEE Int. Conf. Autom. Sci. Eng.*, 2015, pp. 1294–1299.

[10] S. Liu, N. Cao, and P. K. Varshney, "Sensor placement for field estimation via Poisson disk sampling," in *Proc. IEEE Global Conf. Signal Inf. Process.*, 2016, pp. 520–524.

[11] S. Liu, A. Vempaty, M. Fardad, E. Masazade, and P. K. Varshney, "Energy-aware sensor selection in field reconstruction," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1476–1480, 2014.

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

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

[14] 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.

[15] M. Naeem, S. Xue, and D. Lee, "Cross-entropy optimization for sensor selection problems," in *Proc. Int. Symp. Commun. Inf. Tech.*, 2009, pp. 396–401.

[16] W. J. Welch, "Branch-and-bound search for experimental designs based on d optimality and other criteria," *Technometrics*, vol. 24, no. 1, pp. 41–48, 1982.
[17] L. Yao, W. A. Sethares, and D. C. Kammer, "Sensor placement for on-orbit modal identification via a genetic algorithm," *AIAA J.*, vol. 31, no. 10, pp. 1922–1928, 1993.
[18] S. Lau, R. Eichardt, L. Di Rienzo, and J. Haueisen, "Tabu search optimization of magnetic sensor systems for magnetocardiography," *IEEE Trans. Magn.*, vol. 44, no. 6, pp. 1442–1445, Jun. 2008.
[19] S. P. Chepuri and G. Leus, "Continuous sensor placement," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 544–548, May 2015.
[20] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for nonlinear measurement models," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 684–698, Feb. 2015.
[21] 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.
[22] 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, 2014.
[23] 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.
[24] H. Godrich, A. P. Petropulu, and H. V. Poor, "Sensor selection in distributed multiple-radar architectures for localization: A knapsack problem formulation," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 247–260, Jan. 2012.
[25] S. Liu, S. P. Chepuri, G. Leus, and A. O. Hero, "Distributed sensor selection for field estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2017, pp. 4257–4261.
[26] H. Jamali-Rad, A. Simonetto, X. Ma, and G. Leus, "Distributed sparsity-aware sensor selsection," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 5951–5964, Nov. 2015.
[27] S. Liu, S. P. Chepuri, M. Fardad, E. Maşazade, G. Leus, and P. K. Varshney, "Sensor selection for estimation with correlated measurement noise," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3509–3522, Jul. 2016.
[28] X. Shen, S. Liu, and P. Varshney, "Sensor selection for nonlinear systems in large sensor networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 4, pp. 2664–2678, Oct. 2014.
[29] S. Liu, M. Fardad, E. Masazade, and P. K. Varshney, "Optimal periodic sensor scheduling in networks of dynamical systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3055–3068, Jun. 2014.
[30] S. P. Chepuri and G. Leus, "Sparsity-promoting adaptive sensor selection for non-linear filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2014, pp. 5080–5084.
[31] V. Roy, A. Simonetto, and G. Leus, "Spatio-temporal field estimation using Kriged Kalman filter (KKF) with sparsity-enforcing sensor placement," *Sensors*, vol. 18, no. 6, 2018, Art. no. 1778.
[32] S. Liu, E. Masazade, M. Fardad, and P. K. Varshney, "Sparsity-aware field estimation via ordinary kriging," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2014, pp. 3948–3952.
[33] H. Wang, K. Yao, G. Pottie, and D. Estrin, "Entropy-based sensor selection heuristic for target localization," in *Proc. Int. Symp. Inf. Process. Sens. Netw.*, 2004, pp. 36–45.
[34] D. MacKay, "Information-based objective functions for active data selection," *Neural Comput.*, vol. 4, no. 4, pp. 590–604, 1992.
[35] C. Guestrin, A. Krause, and A. P. Singh, "Near-optimal sensor placements in Gaussian processes," in *Proc. Int. Conf. Mach. Learn.*, 2005, pp. 265–272.
[36] C. Tillmann and H. Ney, "Word reordering and a dynamic programming beam search algorithm for statistical machine translation," *Comput. Linguistics*, vol. 29, no. 1, pp. 97–133, 2003.
[37] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions–I," *Math. Program.*, vol. 14, no. 1, pp. 265–294, 1978.
[38] A. Krause and D. Golovin, "Submodular function maximization," *Tractability: Practical Approaches to Hard Problems*, vol. 3, pp. 71–104, 2014.
[39] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
[40] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: JHU Press, 2013.
[41] R. Thompson, "The behavior of eigenvalues and singular values under perturbations of restricted rank," *Linear Algebra Appl.*, vol. 13, no. 1, pp. 69–78, 1976.
[42] M. S. Bartlett, "An inverse matrix adjustment arising in discriminant analysis," *Annals Math. Statist.*, vol. 22, no. 1, pp. 107–111, 1951.
[43] H. Lütkepohl, *Handbook of Matrices*. Singapore: Wiley, 1996.

Chaoyang Jiang (S'15–M'17) received the B.Eng. degree in electrical engineering and automation from the China University of Mining and Technology, Xuzhou, China in 2009, the M.Eng. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China in 2011, and the Ph.D. degree in electrical and electronic engineering from Nanyang Technological University (NTU), Singapore, in 2017. He was a Research Fellow in NTU before he joined Beijing Institute of Technology in 2018, where he is currently an Associate Professor with the School of Mechanical Engineering. His current research interests include statistical signal processing, machine learning, sparse sensing, and intelligent vehicles.

Zhenghua Chen received the B.Eng. degree in mechatronics engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2011, and the Ph.D. degree in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2017. Currently, he is a scientist at the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. His research interests include data analytics in smart buildings, ubiquitous computing, Internet of things, machine learning, and deep learning.

Rong Su (M'11–SM'14) received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 1997, and the M.A.Sc. and Ph.D. degrees from the University of Toronto, Toronto, ON, Canada, in 2000 and 2004, respectively. He was with the University of Waterloo and the Technical University of Eindhoven before he joined Nanyang Technological University, Singapore in 2010, where he is currently an Associate Professor with the School of Electrical and Electronic Engineering. He has been involved as a PI or a Co-PI in many projects sponsored by the Singapore National Research Foundation; Singapore Agency of Science, Technology, and Research; Singapore Ministry of Education, Singapore Civil Aviation Authority; and Singapore Economic Development Board. He has authored and coauthored more than 160 journal and conference publications, one granted U.S. Patent, and one filed Singapore patent. His research interests include multi-agent systems, discrete-event system theory, model-based fault diagnosis, control and optimization of complex networks with applications in smart manufacturing, intelligent transportation, human robot interface, power management, and green buildings. Dr Su is an Associate Editor for *Automatica*, the *Journal of Discrete Event Dynamic Systems: Theory and Applications*, and the *Journal of Control and Decision*. He is also the Chair of the Technical Committee on Smart Cities in IEEE Control Systems Society.

Yeng Chai Soh (M'87–SM'06) received the B.Eng. (Hons. I) degree in electrical and electronic engineering from the University of Canterbury, Christchurch, New Zealand, and the Ph.D. degree in electrical engineering from the University of Newcastle, Callaghan, Australia. He joined the Nanyang Technological University, Singapore, after his Ph.D. study and is currently a Professor with the School of Electrical and Electronic Engineering. He has served as the Head of the Control and Instrumentation Division, the Associate Dean (Research and Graduate Studies) and the Associate Dean (Research) at the College of Engineering. His research interests have been in robust control and applications, robust estimation and filtering, optical signal processing, and energy efficient systems. He has authored and coauthored more than 260 refereed journal papers in these areas. His most recent research projects and activities are in sensor networks, sensor fusion, distributed control and optimization, and control and optimization of ACMV systems. Dr. Soh has served as panel members of several national grants and scholarships evaluation and awards committees.