Signal Processing 244 (2026) 110553

Contents lists available at ScienceDirect

Signal Processing

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

Greedy sensor selection for nonlinear models with performance guarantees

Jiaming Cui [a], Lingya Liu [b], Geert Leus [c], Yiyin Wang [a,*]

[a] State Key Laboratory of Submarine Geoscience, School of Automation and Sensing, Shanghai Jiao Tong University, Shanghai, 200240, China
[b] College of Information Engineering, Shanghai Maritime University, Shanghai, 200241, China
[c] Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Delft, 2628 CD, The Netherlands

ARTICLE INFO

Keywords:
Greedy sensor selection
Fisher information matrix
Nonlinear optimization
Submodularity

ABSTRACT

This paper focuses on developing a low-complexity greedy method to address the sensor selection problem with a generic nonlinear model. The optimal subset of sensors is chosen to satisfy specific performance constraints, which are functions of the Fisher information matrix (FIM) and depend on unknown parameters due to the nonlinear measurement model. Therefore, the sensor selection problem is a multi-constraint optimization problem, which is challenging to be addressed in a greedy way. In this paper, we design a performance metric, which is an average of truncated submodular functions, to ensure that all performance constraints are satisfied. An auxiliary term is introduced to the performance metric to make it well-defined. A greedy algorithm with optimality guarantees is proposed accordingly. In contrast to existing optimality guarantees, the proposed optimality guarantees are independent of the auxiliary term. Numerical results demonstrate the superiority of our proposed optimality guarantees and show that the proposed algorithm achieves similar performance to the convex methods in various scenarios with lower computational complexity.

1. Introduction

In wireless sensor networks (WSNs), a large number of spatially distributed sensors are deployed to collect measurements and extract relevant information [1,2]. To reduce the hardware cost and power consumption, it is essential to select the smallest subset of sensors from the candidate set. The selected sensor subset should ensure that the network performance satisfies the required criteria. This sensor selection problem arises in many applications, such as target localization [3–5], Internet of Things (IoT) systems [2,6], antenna selection [7], sound field control [8], and environmental monitoring [9]. Although numerous heuristic and approximate algorithms have been proposed [10–13], finding efficient solutions remains challenging even for moderate-sized WSNs.

1.1. Related prior works

Sensor selection is an intricate combinatorial optimization problem. Under a generic nonlinear measurement model, it is formulated as follows [3]:

$$\begin{aligned}\mathbf{P}_0: \quad & \min \quad \|\mathbf{w}\|_0 \\ & \text{s.t.} \quad J(\mathbf{w},\boldsymbol{\theta}) \geq \gamma, \forall \boldsymbol{\theta} \in \mathcal{U}_0, \\ & \qquad \mathbf{w} \in \{0,1\}^M,\end{aligned} \tag{1}$$

where $\mathbf{w} = [w_1, w_2, \ldots, w_M]^T$ is the binary selection vector. The entry $w_m = 1$ (or 0) indicates that the mth sensor is (or is not) selected. The unknown vector $\boldsymbol{\theta} \in \mathbb{R}^N$ within the parameter space $\mathcal{U}_0$ collects the parameters of interest, where N denotes its dimension. The function $J(\mathbf{w},\boldsymbol{\theta})$ is the performance metric, and the constant γ is a predefined threshold. For any unbiased estimator of $\boldsymbol{\theta}$, the error covariance matrix is lower bounded by the Cramér–Rao lower bound (CRLB), which depends on the Fisher information matrix (FIM). Therefore, a natural choice of $J(\mathbf{w},\boldsymbol{\theta})$ is a scalar metric function based on the FIM [2,3].

For linear measurement models with independent additive noise, the performance metric $J(\mathbf{w},\boldsymbol{\theta})$ reduces to $J(\mathbf{w})$ and is independent of $\boldsymbol{\theta}$. Consequently, $\mathbf{P}_0$ becomes a combinatorial optimization problem with a single performance constraint, i.e., $J(\mathbf{w}) \geq \gamma$. There are efficient convex [14–16] and greedy approaches [17–21] to deal with it. In contrast, under nonlinear measurement models, the performance metric $J(\mathbf{w},\boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$. The required performance should be satisfied for all $\boldsymbol{\theta} \in \mathcal{U}_0$, which yields an infinite number of constraints in $\mathbf{P}_0$. A common approach to address this issue is to discretize the parameter space of $\boldsymbol{\theta}$ with a prescribed resolution [3,5,7]. Let $\mathcal{D} = \{1, 2, \ldots, D\}$ denote the index set of the discretized points, where D is the number of grid points. The constraints in $\mathbf{P}_0$ are reformulated to be satisfied for all D discrete points in $\mathcal{U}_0$, i.e., $J(\mathbf{w},\boldsymbol{\theta}_d) \geq \gamma, \forall d \in \mathcal{D}$. The resulting problem is a combinatorial optimization with multiple performance constraints. It is difficult and nonconvex in general. A widely used approach relaxes

* Corresponding author.
E-mail address: yiyinwang@sjtu.edu.cn (Y. Wang).

https://doi.org/10.1016/j.sigpro.2026.110553
Received 23 October 2025; Received in revised form 7 February 2026; Accepted 10 February 2026
Available online 13 February 2026

the nonconvex parts to obtain a convex program [3,5,7,22]. Specifically, by exploiting the additive property of the FIM and relaxing $\mathbf{w} \in \{0,1\}^M$ to $\mathbf{w} \in [0,1]^M$, the sensor selection problem $\mathbf{P}_0$ is then approximated as a convex one. The relaxed solution is then rounded to produce a suboptimal Boolean selection. However, naive rounding can noticeably degrade performance. Accordingly, several refinement strategies have been proposed to trade computational cost for improved performance [15,21].

On the other hand, greedy algorithms are well-suited to sensor selection problems due to their low computational complexity [17–19,21]. They have an optimality guarantee of $(1-1/e)$ times the optimal performance metric for solving submodular optimization problems [23], where e is Euler's number. For nonlinear measurement models, greedy algorithms must handle multiple performance constraints associated with all D discrete points $\theta_d \in \mathcal{U}_0$. Let $\mathcal{S}$ and $f(\mathcal{S},\theta_d)$ denote the selected sensor set and the performance metric with respect to (w.r.t.) $\mathcal{S}$ and θ_d, respectively. As greedy algorithms require a scalar metric to determine the next sensor, it is natural to replace multiple constraints with a weighted average of them. For example, the weighted average frame potential (FP) and the weighted average log volume of the confidence ellipsoid (VCE) are employed in [24], which preserve submodularity. However, this averaging approach cannot guarantee the worst-case performance over all $d \in \mathcal{D}$. Alternatively, the minimum value among all performance metrics in the domain $\mathcal{U}_0$ is employed in [5], i.e., $\min_{d\in\mathcal{D}} f(\mathcal{S},\theta_d) \geq \gamma$. Although it ensures that all constraints are satisfied, the corresponding greedy algorithm may result in arbitrarily poor performance [25]. Because the function $\min_{d\in\mathcal{D}} f(\mathcal{S},\theta_d)$ is neither submodular nor convex, the greedy algorithm in [5] cannot provide an optimality guarantee. The robust submodular maximization problem related to $\mathbf{P}_0$ is addressed in [25] and [26], where greedy algorithms are developed to maximize the minimum value of multiple submodular functions under various constraints (e.g., cardinality or matroid constraints). A truncated submodular function is employed by the algorithm in [25] with an adaptive threshold, which is obtained by a binary search under an approximation of the cardinality constraint. The correlation between submodular functions is leveraged in [26]. However, the performance of the greedy algorithm in [26] may degrade when the curvature and correlation ratio of the submodular functions are low.

Sensor placement for target localization is one representative application scenario of the sensor selection problem $\mathbf{P}_0$. In this scenario, the finite candidate set for sensor selection can be obtained by discretizing the deployable region. This paper and the aforementioned related works primarily focus on the combinatorial formulation $\mathbf{P}_0$, which selects a subset of candidate locations from the finite set. In contrast, several sensor placement works [27–30] directly optimize continuous sensor positions. These works typically consider specific measurement modalities with corresponding noise assumptions, and design the deployment for a known or nominal target state. Therefore, they are conceptually related but differ from the sensor selection problem $\mathbf{P}_0$ in terms of optimization variables and modeling assumptions. Accordingly, this paper focuses on combinatorial sensor selection from a finite candidate set under a generic nonlinear measurement model. We enforce a worst-case requirement over $\mathcal{U}$, which differs from continuous placement designs that typically optimize deployments for a nominal target state.

1.2. Contributions

This paper develops a greedy algorithm to address the sensor selection problem under nonlinear measurement models. The main contributions are summarized as follows:

1. A performance metric is introduced to simultaneously maintain submodularity and ensure the performance constraints for all $\theta_d \in \mathcal{U}_0$ with $d \in \mathcal{D}$ are satisfied. This metric is defined as the average of truncated submodular functions, whose inputs are the predefined threshold γ and the values of submodular functions. In this work, the submodular functions are chosen as $\ln\det(\mathbf{F}(\mathcal{S},\theta_d)+\epsilon\mathbf{I})$, where $\mathbf{F}(\mathcal{S},\theta_d)$ is the FIM associated with the selected sensor set $\mathcal{S}$ and the discrete point θ_d. The auxiliary term $\epsilon\mathbf{I}$ helps to make $\ln\det(\mathbf{F}(\mathcal{S},\theta_d)+\epsilon\mathbf{I})$ well-defined.
2. As the performance metric is designed to be submodular, a greedy algorithm with optimality guarantees is proposed accordingly. Simulation results show that it achieves similar performance as its convex counterpart [3] with significantly reduced computational complexity.
3. The proposed sensor selection formulation belongs to the general class of submodular set covering problems [31]. However, the existing optimality guarantees on the number of selected sensors provided in [31] are overly conservative due to the auxiliary term $\epsilon\mathbf{I}$. To address this issue, novel ϵ-independent upper bounds on the number of selected sensors are derived. These bounds are also applicable to other greedy algorithms [17,19,21,24], where similar auxiliary terms are introduced.

The remainder of this paper is organized as follows. Section 2 reviews the sensor selection problem under a generic nonlinear measurement model. It also introduces the concept of the average of truncated submodular functions to address multiple constraints while preserving submodularity. Section 3 presents a greedy algorithm designed to solve the sensor selection problem with nonlinear measurement models. In Section 4, theoretical analyses and new optimality guarantees that are independent of the auxiliary parameter ϵ are provided. Section 5 presents numerical results, and Section 6 concludes the paper.

Notations: Upper case calligraphic and boldface upper (lower) case letters, e.g., $\mathcal{A}$ and $\mathbf{A}$ ($\mathbf{a}$), denote sets and matrices (column vectors), respectively. An identity matrix with proper dimensions is represented by $\mathbf{I}$. The operators $(\cdot)^T$, $\|\cdot\|$, $\det(\cdot)$, $\mathrm{tr}(\cdot)$, and $\mathrm{rank}(\cdot)$ correspond to the transpose, norm, determinant, trace, and rank operations, respectively. The function $|\cdot|$ denotes the cardinality of a set. The expression $\mathbf{A} \succ 0$ ($\mathbf{A} \succeq 0$) denotes that $\mathbf{A}$ is a positive (semi)definite matrix.

2. Problem analysis and statement

This section introduces the sensor selection problem with a generic nonlinear measurement model. The multi-constraint challenge is addressed by replacing the multiple inequality constraints in $\mathbf{P}_0$ with a single equality constraint. This equality enforces that the average of the truncated submodular metric functions equals a predefined threshold. The novel problem formulation helps to develop greedy algorithms with optimality guarantees, and ensures that all constraints are fulfilled accordingly.

2.1. Nonlinear measurement model

We explore the same generic nonlinear measurement model as [3]:

$$y_m = h_m(\theta, n_m),\ m = 1,2,\dots,M, \tag{2}$$

where the parameters of interest are collected in the vector $\theta \in \mathbb{R}^N$ with N denoting the length of θ. Moreover, the measurement y_m from the mth sensor is the output of the generic nonlinear function $h_m(\cdot)$, whose inputs are the unknown θ and the noise term n_m. Note that the noise n_m may depend on θ. Let us define $\mathbf{y} = [y_1, y_2, \dots, y_M]^T \in \mathbb{R}^M$ to encompass all measurements, and $p(\mathbf{y};\theta)$ to represent the probability density function (pdf) of the measurements parameterized by θ. The nonlinear model (2) fulfills the regularity condition [32] and satisfies the independent measurement assumption.

The covariance of any unbiased estimate of θ based on $\mathbf{y}$ is lower bounded by the inverse of the FIM, where the FIM is given by [3]:

$$\mathbf{F}(\theta) = \sum_{m=1}^{M} \mathbf{F}_m(\theta) = \sum_{m=1}^{M} \mathbb{E}\left\{ \left(\frac{\partial \ln p(y_m;\theta)}{\partial \theta}\right)\left(\frac{\partial \ln p(y_m;\theta)}{\partial \theta}\right)^T \right\}. \tag{3}$$

The Fisher information exhibits the additive property for independent observations. The additive property helps to develop a submodular performance metric. Note that the individual FIM $\mathbf{F}_m(\theta)$ of size $N \times N$ is related to the mth sensor. It may not be full rank. For example, when the noise n_m is zero-mean additive white Gaussian noise (AWGN) with variance σ_m^2 independent of θ, the individual FIM $\mathbf{F}_m(\theta)$ in (3) simplifies to

$$\mathbf{F}_m(\theta) = \frac{1}{\sigma_m^2}\left(\frac{\partial h_m(\theta)}{\partial \theta}\right)\left(\frac{\partial h_m(\theta)}{\partial \theta}\right)^T, \tag{4}$$

which is of rank 1. As the individual FIM $\mathbf{F}_m(\theta)$ may not be of full rank, it may introduce a rank-deficiency issue, which will be discussed in detail in the following subsection.

2.2. Performance metric

Let $\mathcal{M}$ represent the set containing all M sensors. By the additive property of the FIM, the FIM related to the selected subset $\mathcal{S}$ at the point $\theta \in \mathcal{U}_0$ is given by

$$\mathbf{F}(\mathcal{S}, \theta) = \sum_{m\in\mathcal{S}} \mathbf{F}_m(\theta). \tag{5}$$

Since the individual FIMs $\mathbf{F}_m(\theta)$ may not be full rank, the overall FIM $\mathbf{F}(\mathcal{S}, \theta)$ can also be singular. This rank deficiency may render some FIM-based performance metrics ill-defined, as discussed below. Three typical scalar forms of the FIM are widely used in the literature [2,3,7,21]:

- $\mathrm{tr}\left((\mathbf{F}(\mathcal{S}, \theta) + \varepsilon\mathbf{I})^{-1}\right)$: the trace of the inverse of the FIM.
- $\ln\det(\mathbf{F}(\mathcal{S}, \theta) + \varepsilon\mathbf{I})$: the logarithm of the determinant of the FIM.
- $\lambda_{\min}(\mathbf{F}(\mathcal{S}, \theta))$: the minimum eigenvalue of the FIM.

These three criteria are commonly known as the A-, D-, and E-optimality criteria, respectively. The auxiliary term $\varepsilon\mathbf{I}$ is introduced in the first two metrics for two main reasons. First, when $\mathcal{S} = \emptyset$, we have $\mathbf{F}(\emptyset, \theta) = \mathbf{0}$. In this case, the inverse and $\ln\det(\cdot)$ operations are undefined. The term $\varepsilon\mathbf{I}$ avoids this issue. Second, $\mathbf{F}(\mathcal{S}, \theta)$ may be singular even when $\mathcal{S} \neq \emptyset$. The term $\varepsilon\mathbf{I}$ also helps mitigate the rank-deficiency issue. Thus, the inclusion of $\varepsilon\mathbf{I}$ is necessary to ensure that the first two scalar measures are well-defined. When ε is properly chosen, it does not change the optimal solution of the sensor selection problem [19]. However, it can influence the optimality guarantees of greedy algorithms. This effect is discussed in detail in Section 4.

Among the aforementioned three scalar forms, only the performance metric $\ln\det(\mathbf{F}(\mathcal{S}, \theta) + \varepsilon\mathbf{I})$ is a monotone submodular function w.r.t $\mathcal{S}$ [7,17,19]. Its submodularity and nondecreasing nature have been proved in [7, Theorem VI.2]. These properties enable the use of a greedy algorithm with a $(1 - 1/e)$ optimality guarantee [23]. Appendix A provides the definition of submodularity and the $(1 - 1/e)$ optimality guarantee. The scalar function $\mathrm{tr}\left((\mathbf{F}(\mathcal{S}, \theta) + \varepsilon\mathbf{I})^{-1}\right)$ w.r.t. $\mathcal{S}$ in general is neither submodular nor supermodular [33]. The proof of Theorem 2 in [21] claims that $\mathrm{tr}\left((\mathbf{F}(\mathcal{S}, \theta) + \varepsilon\mathbf{I})^{-1}\right)$ is a submodular function w.r.t. $\mathcal{S}$. However, it implicitly uses the incorrect lemma that if $\mathbf{A} \preceq \mathbf{B}$ then $\mathbf{A}^2 \preceq \mathbf{B}^2$ for positive semidefinite $\mathbf{A}$ and $\mathbf{B}$. Counterexamples can be found in [34]. Therefore, we focus on the submodular function $\ln\det(\mathbf{F}(\mathcal{S}, \theta) + \varepsilon\mathbf{I})$. Accordingly, the performance constraint based on it is given by

$$f(\mathcal{S}, \theta) = \ln(\det(\mathbf{F}(\mathcal{S}, \theta) + \varepsilon\mathbf{I})) \geq \gamma, \text{ with } \gamma = 2N\ln\left(\sqrt{\xi}/\bar{R}_e\right), \tag{6}$$

where $\bar{R}_e$ is the geometric mean of the semi-axes of the confidence ellipsoid with an error probability P_e, and $\xi = F_{\chi_N^2}^{-1}(P_e)$, where $F_{\chi_N^2}(\cdot)$ denotes the cumulative distribution function of a Chi-squared random variable with N degrees of freedom [3]. The values of $\bar{R}_e$ and P_e specify the predefined threshold γ, and are assumed to be known.

As the performance metric depends on the unknown parameter θ, we discretize the uncertainty region $\mathcal{U}_0$ with a prescribed resolution. This discretization yields D points. These points are collected into the set $\mathcal{U} = \{\theta_1, \theta_2, \ldots, \theta_D\} \subset \mathcal{U}_0$. Its cardinality is $|\mathcal{U}| = D$. Moreover, different θ_d may correspond to different predefined thresholds γ_d. Consequently, the performance constraint $f(\mathcal{S}, \theta_d) \geq \gamma_d$ must hold for all $d \in \mathcal{D}$. This requirement leads to D constraints in $\mathbf{P}_0$. However, greedy algorithms optimize a single scalar objective at each selection step. Therefore, nonlinear measurements motivate the development of a novel performance metric. The metric should ensure that $f(\mathcal{S}, \theta_d) \geq \gamma_d$ holds for all $d \in \mathcal{D}$. It should also preserve submodularity. To this end, we define a truncated metric function as

$$g(\mathcal{S}, \theta_d, \gamma_d) = \min\left(f(\mathcal{S}, \theta_d), \gamma_d\right), \tag{7}$$

where $\min(a, b)$ returns the smaller of a and b. For any fixed $\gamma_d, \forall d \in \mathcal{D}$, the following equivalence holds:

$$f(\mathcal{S}, \theta_d) \geq \gamma_d, \forall d \in \mathcal{D} \Leftrightarrow \frac{1}{D}\sum_{d=1}^{D} g(\mathcal{S}, \theta_d, \gamma_d) = \frac{1}{D}\sum_{d=1}^{D}\gamma_d = \bar{\gamma}, \tag{8}$$

where $\Leftrightarrow$ denotes equivalence. Hence, all inequality constraints $f(\mathcal{S}, \theta_d) \geq \gamma_d, \forall d \in \mathcal{D}$ can be replaced by a single equality constraint:

$$\frac{1}{D}\sum_{d=1}^{D} g(\mathcal{S}, \theta_d, \gamma_d) = \bar{\gamma}. \tag{9}$$

The truncation operation preserves both the monotonicity and submodular properties of functions [35]. Thus, $g(\mathcal{S}, \theta_d, \gamma_d)$ remains submodular and nondecreasing w.r.t. $\mathcal{S}$. Furthermore, a positive linear combination of submodular functions is also submodular [23]. Therefore, the average of the truncated submodular functions in (9) remains nondecreasing and submodular for $\gamma_d \geq 0$, $\forall d \in \mathcal{D}$.

2.3. Problem statement

Consequently, the sensor selection problem using a single constraint to ensure the estimation performance for all $\theta_d \in \mathcal{U}$ can be formulated as:

$$\begin{aligned} \mathbf{P}_1: \quad & \min_{\mathcal{S}\subseteq\mathcal{M}} \quad |\mathcal{S}| \\ & \text{s.t.} \quad \frac{1}{D}\sum_{d=1}^{D} g(\mathcal{S}, \theta_d, \gamma_d) = \bar{\gamma}. \end{aligned} \tag{10}$$

Note that the auxiliary term $\varepsilon\mathbf{I}$ is introduced in the performance metric $g(\mathcal{S}, \theta_d, \gamma_d)$. When ε is properly designed, the resulting problem shares the same optimal solution as the case without this auxiliary term [19].

3. Greedy sensor selection with worst performance guarantee

This section presents a greedy algorithm to solve $\mathbf{P}_1$ with low computational complexity and optimality guarantees. Prior to sensor selection, it is essential to verify the feasibility of $\mathbf{P}_1$. The following proposition provides the corresponding feasibility condition.

Proposition 1. *The problem $\mathbf{P}_1$ is feasible if and only if the following equality condition holds:*

$$\sum_{d=1}^{D} g(\mathcal{M}, \theta_d, \gamma_d)/D = \bar{\gamma}. \tag{11}$$

Proof. We first demonstrate the sufficiency of the equality condition (11). If this condition holds, a feasible solution that satisfies the performance constraint always exists with $\mathcal{S} = \mathcal{M}$. Next, we show the necessity of the equality condition (11). If problem $\mathbf{P}_1$ is feasible, there exists a subset $\mathcal{S} \subseteq \mathcal{M}$ such that $\sum_{d=1}^{D} g(\mathcal{S}, \theta_d, \gamma_d)/D = \bar{\gamma}$. Because the performance metric is nondecreasing, the equality condition in (11) also holds. □

After $\mathbf{P}_1$ is verified to be feasible, sensors are selected one by one. Let $\mathcal{S}^{(k)} = \{s_1, s_2, \ldots, s_k\}$ denote the subset of sensors selected at the kth step $(1 \leq k \leq M)$, where $\mathcal{S}^{(0)} = \emptyset$. The performance improvement resulting from the addition of sensor $j \in \mathcal{M}\backslash\mathcal{S}^{(k-1)}$ to $\mathcal{S}^{(k-1)}$ is given by

$$\delta_j\left(\mathcal{S}^{(k-1)}\right) = \frac{1}{D}\sum_{d=1}^{D}\left(g\left(\mathcal{S}^{(k-1)} \cup \{j\}, \theta_d, \gamma_d\right) - g\left(\mathcal{S}^{(k-1)}, \theta_d, \gamma_d\right)\right). \tag{12}$$

At the kth step, given $\mathcal{S}^{(k-1)}$, the next sensor is selected according to

$$s_k = \underset{j\in\mathcal{M}\setminus\mathcal{S}^{(k-1)}}{\arg\max}\ \delta_j\left(\mathcal{S}^{(k-1)}\right). \tag{13}$$

The corresponding performance increment is defined as

$$\beta_k = \delta_{s_k}\left(\mathcal{S}^{(k-1)}\right). \tag{14}$$

The selection process can be delineated into two stages.

Stage i): At the kth step, no selection can satisfy the performance constraint for any discrete point in the set $\mathcal{U}$, i.e., $f\left(\mathcal{S}^{(k-1)}\cup\{j\},\theta_d\right) < \gamma_d, \forall d\in\mathcal{D}$ and $\forall j\in\mathcal{M}\setminus\mathcal{S}^{(k-1)}$. In this case, we have $g\left(\mathcal{S}^{(k-1)}\cup\{j\},\theta_d,\gamma_d\right) = f\left(\mathcal{S}^{(k-1)}\cup\{j\},\theta_d\right)$, and $g\left(\mathcal{S}^{(k-1)},\theta_d,\gamma_d\right) = f\left(\mathcal{S}^{(k-1)},\theta_d\right)$. Thus, according to (13), the kth sensor is actually selected as

$$s_k = \underset{j\in\mathcal{M}\setminus\mathcal{S}^{(k-1)}}{\arg\max}\ \frac{1}{D}\sum_{d=1}^{D}\left(f\left(\mathcal{S}^{(k-1)}\cup\{j\},\theta_d\right) - f\left(\mathcal{S}^{(k-1)},\theta_d\right)\right). \tag{15}$$

Stage ii): At the kth step, at least one candidate sensor can satisfy the performance constraint for at least one discrete point in $\mathcal{U}$, i.e., $\exists j\in\mathcal{M}\setminus\mathcal{S}^{(k-1)}, d\in\mathcal{D}$, such that $f\left(\mathcal{S}^{(k-1)}\cup\{j\},\theta_d\right)\geq\gamma_d$. Define the index subset $\bar{\mathcal{D}}^{(k)}\subset\mathcal{D}$ to include the indices of discrete points whose estimation accuracy with the current subset $\mathcal{S}^{(k)}$ does not yet meet the performance constraint, i.e., $f\left(\mathcal{S}^{(k)},\theta_d\right)<\gamma_d, \forall d\in\bar{\mathcal{D}}^{(k)}$. When $\left|\bar{\mathcal{D}}^{(k)}\right| = D$, the algorithm corresponds to *Stage i)*. Thus, we obtain

$$\sum_{d\in\mathcal{D}\setminus\bar{\mathcal{D}}^{(k-1)}}\left(g\left(\mathcal{S}^{(k-1)}\cup\{j\},\theta_d,\gamma_d\right) - g\left(\mathcal{S}^{(k-1)},\theta_d,\gamma_d\right)\right) = 0, \tag{16}$$

where $g\left(\mathcal{S}^{(k-1)},\theta_d,\gamma_d\right)=\gamma_d$ for $d\in\mathcal{D}\setminus\bar{\mathcal{D}}^{(k-1)}$ based on the definition of $\bar{\mathcal{D}}^{(k-1)}$, and $g\left(\mathcal{S}^{(k-1)}\cup\{j\},\theta_d,\gamma_d\right)=\gamma_d$ for $d\in\mathcal{D}\setminus\bar{\mathcal{D}}^{(k-1)}$ due to the non-decreasing property of $f(\mathcal{S},\theta_d)$. Therefore, the sensor selection problem at the kth step given by (13) can be expressed as

$$s_k = \underset{j\in\mathcal{M}\setminus\mathcal{S}^{(k-1)}}{\arg\max}\ \frac{1}{D}\sum_{d\in\bar{\mathcal{D}}^{(k-1)}}\left(g\left(\mathcal{S}^{(k-1)}\cup\{j\},\theta_d,\gamma_d\right) - g\left(\mathcal{S}^{(k-1)},\theta_d,\gamma_d\right)\right). \tag{17}$$

This formulation indicates that, in this case, the proposed algorithm only considers the discrete points whose performance constraints remain unsatisfied after $(k-1)$ selections. The selection continues until $\sum_{d=1}^{D} g\left(\mathcal{S}^{(K)},\theta_d,\gamma_d\right)/D=\bar{\gamma}$, where K denotes the total number of selected sensors. The proposed greedy algorithm, referred to as `Avg.-Trunc.`, is summarized in Algorithm 1.

Algorithm 1: `Avg.-Trunc.` algorithm.

```
Input: H = {h_m(θ_d, n_m)}_{m=1,d=1}^{M,D}, ε, U = {θ_1, θ_2, …, θ_D},
       {γ_d}_{d=1}^{D};
Output: S^(K) and K;
Obtain the FIMs F_m(θ_d) by (3), m = 1,2,…,M, d = 1,2,…,D;
Initialize: S^(0) = ∅, k = 0;
if Σ_{d=1}^{D} g(M, θ_d, γ_d)/D ≠ γ̄ then
    The problem has no solution.
else
    while Σ_{d=1}^{D} g(S^(k), θ_d, γ_d)/D ≠ γ̄ do
        k = k + 1;
        s_k = arg max_{j∈M\S^(k−1)} {δ_j(S^(k−1))};
        S^(k) = S^(k−1) ∪ {s_k};
    end
    Set K = k.
end
```

Note that the parameter ε should be sufficiently small to ensure that it does not significantly affect the sensor selection. Therefore, based on the first sensor selection, ε should at least fulfill the following condition

$$\varepsilon \ll \min_{m\in\mathcal{M},d\in\mathcal{D}} \lambda_{\min}^{+}\left(\mathbf{F}_m(\theta_d)\right), \tag{18}$$

where $\lambda_{\min}^{+}\left(\mathbf{F}_m(\theta_d)\right)$ denotes the minimum non-zero eigenvalue of $\mathbf{F}_m(\theta_d)$.

The computational complexity of the proposed `Avg.-Trunc.` algorithm is $O(MN^3\sum_{k=0}^{K-1}|\bar{\mathcal{D}}^{(k)}|)$. This is because the aggregated marginal gain is evaluated only over the currently unsatisfied discretized points. In the worst case, $|\bar{\mathcal{D}}^{(k)}| = D$ for all k. Then, the above complexity reduces to $O\big(DKMN^3\big)$, which is of the same order as the greedy algorithm in [24]. However, by definition, $|\bar{\mathcal{D}}^{(k)}|$ often decreases with k. This yields a lower effective complexity in practice. Under the assumption that $M\gg N$, the proposed `Avg.-Trunc.` algorithm has substantial computational advantages compared to the convex approximation algorithm based on the ℓ_1-norm in [3]. The latter exhibits a complexity of $O(DM^3)$ per iteration when employing the projected Newton's algorithm. In addition, the robust submodular maximization algorithm in [26] incurs a complexity of $O(DKM^2N^3)$. This complexity is higher than that of the proposed method. In general, greedy algorithms exhibit a trade-off between performance and computational cost. Group selection techniques can further improve empirical performance at the price of higher per-iteration cost [21]. For large-scale deployments, advanced techniques can be incorporated into the proposed algorithm to further reduce the complexity. Examples include subsampling-based randomized greedy [36], streaming frameworks [37], and parallelizable greedy schemes [38]. They can be incorporated as enhancements that mainly reduce the runtime of candidate evaluation and selection, without changing the proposed performance metric. For example, subsampling reduces the per-iteration search from M candidates to $M_s\ll M$ candidates, where M_s is the subsample size. This leads to an effective cost $O(M_sN^3\sum_{k=0}^{K-1}|\bar{\mathcal{D}}^{(k)}|)$. These techniques are complementary to the proposed greedy algorithm.

4. Theoretical performance analysis

In this section, we analyze the optimality guarantees of the proposed `Avg.-Trunc.` algorithm. We start with the optimality guarantees on the number of selected sensors proposed in [31]. We show that the introduction of the auxiliary term $\varepsilon\mathbf{I}$ leads to the excessive looseness of the optimality guarantees in [31] for greedy algorithms. To address this, we propose novel optimality guarantees that are independent of the auxiliary number ε. The proposed guarantees can also be extended to other greedy algorithms employing similar auxiliary terms, such as those in [17,19,21,24].

4.1. State-of-the-art optimality guarantees

Problem $\mathbf{P}_1$ falls within the class of general submodular set covering problems, whose optimality guarantees have been thoroughly analyzed in [31]. Thus, the results presented therein can be applied to our proposed algorithm. Note that $\mathcal{S}^{(K)}$ is the solution obtained by the proposed algorithm and $K=|\mathcal{S}^{(K)}|$. Let $\mathcal{S}_{\text{opt}}$ represent the optimal solution of $\mathbf{P}_1$ with $K_{\text{opt}}=|\mathcal{S}_{\text{opt}}|$. Theorem 1 in [31] is restated below.

Theorem 1. *If the greedy algorithm is applied to $\mathbf{P}_1$, the solution K always satisfies [31]:*

$$1)\quad K\leq K_{\text{opt}}\left(1+\ln\left(\max_{j\in\mathcal{M},\,p\in\mathcal{K}}\left\{\frac{\delta_j(\mathcal{S}^{(0)})}{\delta_j(\mathcal{S}^{(p)})}\right\}\right)\right), \tag{19}$$

$$2)\quad K\leq K_{\text{opt}}\left(1+\ln\left(\frac{\beta_1}{\beta_K}\right)\right), \tag{20}$$

$$3)\quad K\leq K_{\text{opt}}\left(1+\ln\left(\frac{\delta(\mathcal{M},\mathcal{S}^{(0)})}{\delta(\mathcal{M},\mathcal{S}^{(K-1)})}\right)\right), \tag{21}$$

where $\mathcal{K}=\{1,\ldots,K\}$, and the performance difference between the full set $\mathcal{M}$ and a subset $\mathcal{S}$ is defined as $\delta(\mathcal{M},\mathcal{S})=\sum_{d=1}^{D}\left(g(\mathcal{M},\theta_d,\gamma_d)-g(\mathcal{S},\theta_d,\gamma_d)\right)/D$. Let αK_{opt} denote the upper bound given by Theorem 1, where $\alpha\geq 1$ denotes the right hand side *(r.h.s.) term of (19)–(21) excluding $K_{\rm opt}$. Accordingly, (19)–(21) can be summarized as:*

$$K \leq \alpha K_{\rm opt}. \tag{22}$$

Theorem 1 provides upper bounds on the number of sensors selected by the proposed algorithm. These bounds follow from the diminishing returns property of submodular functions. A detailed derivation can be found in [31]. However, the optimality guarantees in Theorem 1 become excessively loose due to the introduction of the auxiliary term $\varepsilon\mathbf{I}$. This issue is not unique to the proposed algorithm. It also appears in existing studies such as [17,19,21,24], where similar auxiliary terms are introduced to ensure that the performance metrics are well-defined. Although these works are accompanied by the $(1-1/e)$ optimality guarantee, which is reviewed as Theorem 3 in Appendix A. Both Theorem 1 and the $(1-1/e)$ optimality guarantee are derived from the diminishing returns property of submodular functions, and both are too loose in the presence of auxiliary terms. For example, the term $\varepsilon\mathbf{I}$ causes the parameter β_1 in (20) to become excessively large. This leads to a significant increase in the upper bound (20). A similar impact is imposed on the other bounds in Theorem 1 as well.

4.2. Optimality guarantees independent of ε

In this subsection, we establish novel optimality guarantees that are independent of the auxiliary number ε, thereby addressing the aforementioned issue. Moreover, there always exists a range of ε for which the proposed guarantees are tighter than those established in Theorem 1. We first introduce the following definition.

Definition 1. Let N^* denote the first step of the greedy algorithm at which the FIM $\mathbf{F}\big(\mathcal{S}^{(N^*)},\theta_d\big)$ attains full rank for every $\theta_d \in \mathcal{U}$, i.e.,

$$N^* = \min\left\{ k \,|\, \mathrm{rank}\left(\mathbf{F}\big(\mathcal{S}^{(k)},\theta_d\big)\right) = N, \forall d \in \mathcal{D}\right\}. \tag{23}$$

Note that the greedy algorithm adds sensors sequentially. At the kth selection, the rank of $\mathbf{F}(\mathcal{S}^{(k)},\theta_d), \forall d \in \mathcal{D}$ is checked. The parameter N^* is determined accordingly. Based on the definition of N^*, we have $\mathrm{rank}(\mathbf{F}(\mathcal{S}^{(N^*)},\theta_d)) = N$ for all $d\in\mathcal{D}$, while there exists at least one $d \in \mathcal{D}$ such that $\mathrm{rank}(\mathbf{F}(\mathcal{S}^{(N^*-1)},\theta_d)) < N$. Recall from Section 2.1 that the individual FIM $\mathbf{F}_m(\theta_d)$ may not be full rank. Furthermore, the individual FIM may have different ranks due to heterogeneous sensor measurements. If $\mathbf{F}(\mathcal{S}^{(1)},\theta_d)$ is full rank for every $d\in\mathcal{D}$, then $N^*=1$; otherwise, $N^*>1$. For instance, we have $\mathrm{rank}(\mathbf{F}_m(\theta_d)) = 1$ for all $m\in\mathcal{M}$ and $d \in \mathcal{D}$, when the measurement noise in (2) is AWGN with the variance independent of θ. Therefore, it follows that $N^*>1$ as long as $N>1$. Moreover, the estimation performance constraint $f(\mathcal{S}^{(K)},\theta_d)\geq\gamma_d, \forall d\in\mathcal{D}$ can only be satisfied when $\mathbf{F}(\mathcal{S}^{(K)},\theta_d)$ is full rank for all $d\in\mathcal{D}$. As a result, we have $1\leq N^*\leq K$.

In order to derive tighter bounds, we propose to divide the selection process into two parts based on N^*, namely, the first N^* steps and the remaining $K-N^*$ ones. The iterations of the greedy algorithm to solve $\mathbf{P}_1$ are divided into two segments.

1) $1\leq k\leq N^*$: The first segment corresponds to one of the following two cases: i) the FIM $\mathbf{F}(\mathcal{S}^{(1)},\theta_d)$ has full rank for all $d\in\mathcal{D}$, yielding $N^*=1$; or ii) there exists at least one $d\in\mathcal{D}$ such that the FIM $\mathbf{F}(\mathcal{S}^{(k-1)},\theta_d)$ is singular, leading to $N^*>1$. At the N^*th iteration, the selected sensor set is $\mathcal{S}^{(N^*)}$.

2) $N^*<k\leq K$: Following the first segment of iterations, the second segment actually solves the following problem:

$$\begin{aligned}\mathbf{P}_2:\quad &\min_{\mathcal{S}\subseteq\mathcal{M}\setminus\mathcal{S}^{(N^*)}} \quad |\mathcal{S}| \\ &\text{s.t.} \quad \frac{1}{D}\sum_{d=1}^{D} g(\mathcal{S}\cup\mathcal{S}^{(N^*)},\theta_d,\gamma_d) = \bar{\gamma}.\end{aligned} \tag{24}$$

Let $\mathcal{S}_2$ denote the output of the proposed algorithm applied to $\mathbf{P}_2$, and $K_2 = |\mathcal{S}_2|$. Note that $\mathcal{S}^{(N^*)}\cup\mathcal{S}_2 = \mathcal{S}^{(K)}$ and $K = N^* + K_2$. According to Theorem 1, we have

$$K_2 \leq \tilde{\alpha} K_{{\rm opt},2}, \tag{25}$$

where $K_{{\rm opt},2}$ represents the number of selected sensors corresponding to the optimal solution $\mathcal{S}_{{\rm opt},2}$ of $\mathbf{P}_2$. The parameter $\tilde{\alpha}>1$ can be determined using any of the following expressions:

$$1)\quad \tilde{\alpha} = 1 + \ln\left(\max_{j\in\mathcal{M}\setminus\mathcal{S}^{(N^*)},\, p\in\mathcal{K}\setminus\mathcal{N}^*}\left\{\frac{\delta_j(\mathcal{S}^{(N^*)})}{\delta_j(\mathcal{S}^{(p)})}\right\}\right), \tag{26}$$

$$2)\quad \tilde{\alpha} = 1 + \ln\left(\frac{\beta_{N^*+1}}{\beta_K}\right), \tag{27}$$

$$3)\quad \tilde{\alpha} = 1 + \ln\left(\frac{\delta(\mathcal{M},\mathcal{S}^{(N^*)})}{\delta(\mathcal{M},\mathcal{S}^{(K-1)})}\right), \tag{28}$$

where $\mathcal{N}^* = \{1,\ldots,N^*\}$. We propose the following optimality guarantees.

Theorem 2. *If the greedy algorithm is applied to $\mathbf{P}_1$, the number of selected sensors K satisfies*

$$K \leq \bar{\alpha} K_{\rm opt}, \tag{29}$$

where $\bar{\alpha} = \tilde{\alpha} + N^*/K_{\rm opt}$, and $\tilde{\alpha}$ can be chosen from (26)–(28).

Proof. We first prove $K_{{\rm opt},2}\leq K_{\rm opt}$ using the method of contradiction. Suppose $K_{{\rm opt},2} > K_{\rm opt}$, then there exists a better solution $\mathcal{S}_{\rm opt}\setminus\mathcal{S}^{(N^*)}$ for $\mathbf{P}_2$. This contradicts the optimality of $\mathcal{S}_{{\rm opt},2}$, thereby proving that $K_{{\rm opt},2}\leq K_{\rm opt}$. Adding N^* to both sides of (25) yields $K_2+N^*=K\leq \tilde{\alpha}K_{{\rm opt},2}+N^*$. Using $K_{{\rm opt},2}\leq K_{\rm opt}$, we obtain $K\leq\tilde{\alpha}K_{\rm opt}+N^*$. To be consistent with (22), define $\bar{\alpha}=\tilde{\alpha}+N^*/K_{\rm opt}$, leading to (29). □

Remark 1. (Independence of the proposed optimality guarantees in Theorem 2 from ε): The proposed optimality guarantees are composed of two parts, i.e., $\tilde{\alpha}$ and $N^*/K_{\rm opt}$. As shown in (23), N^* corresponding to the case $k<N^*$ does not depend on ε, and thus the term $N^*/K_{\rm opt}$ is independent of ε. When $k\geq N^*$ and ε satisfies (18), the performance metric in (6) can be simplified as $f(\mathcal{S}^{(k)},\theta)\approx\ln\left(\det\left(\mathbf{F}(\mathcal{S}^{(k)},\theta)\right)\right)$, since the FIM $\mathbf{F}(\mathcal{S}^{(N^*)},\theta_d)$ is already full rank. Under these conditions, the performance metric $g(\mathcal{S}^{(k)},\theta_d,\gamma_d)$ is effectively independent of the auxiliary term $\varepsilon\mathbf{I}$. Consequently, the influence of $\varepsilon\mathbf{I}$ on $\delta_j\big(\mathcal{S}^{(k)}\big)$ in (26), $\beta_{k+1}=\delta_{s_{k+1}}\big(\mathcal{S}^{(k)}\big)$ in (27), and $\delta(\mathcal{M},\mathcal{S}^{(k)})$ in (28) is negligible. Therefore, the impact of the auxiliary term $\varepsilon\mathbf{I}$ on $\tilde{\alpha}$ can also be regarded as negligible. In summary, the proposed optimality guarantees are independent of the auxiliary number ε.

Remark 2. (Applicability of the proposed optimality guarantees in Theorem 2 w.r.t. N^*):

Note that a larger N^* indicates a larger number of iterations of the first segment, and the improvement of the proposed bound w.r.t. the ones in Theorem 1 becomes less. In the worst case where $N^*=K\neq 1$, the FIM $\mathbf{F}\big(\mathcal{S}^{(N^*)},\theta_d\big)$ is full rank for all $d\in\mathcal{D}$ at the Kth step, which is also the last step. Thus, the second part does not exist. The influence of the auxiliary term $\varepsilon\mathbf{I}$ is distributed across every step of the greedy algorithm, rendering the process inseparable. In the aforementioned worst case, the proposed optimality guarantees in Theorem 2 become inapplicable, and the bounds in Theorem 1 remain valid. Conversely, when $N^*=K=1$, the constraints in (10) are satisfied at the first step, and the proposed `Avg.-Trunc.` algorithm provides the optimal solution to $\mathbf{P}_1$. There is no need for optimality guarantees. Hence, the proposed optimality guarantees are designed for cases where $1\leq N^*<K$ and $K>1$, which commonly occur in practice.

To further compare the optimality bounds proposed in Theorem 2 with the ones in Theorem 1, we first present the following corollary.

Corollary 1. *For the upper bounds presented in Theorems 1 and 2, the inequality $\tilde{\alpha}\leq\alpha$ holds.*

Proof. See Appendix B. □

Note that $\tilde{\alpha}$ contains an additive term N^*/K_{opt}. We next show that there always exists a range of ε such that the $\tilde{\alpha}$ associated with $\tilde{\alpha}$ in (27) is smaller than α in (20). The existence of such a range of ε, which ensures that the other bounds in Theorem 2 are tighter than those in Theorem 1, can be established similarly. Moreover, the range of ε depends on the specific form of the performance metric $f(\mathcal{S},\boldsymbol{\theta})$. We take the metric in (6) as an illustrative example. For simplicity, we have the following assumption:

AS1 For all $d \in \mathcal{D}$, $f\left(\mathcal{S}^{(1)},\boldsymbol{\theta}_d\right) \le \gamma_d$, which means that selecting a single sensor cannot satisfy any performance constraint.

Let $\lambda_i(\mathbf{A})$ denote the ith eigenvalue of the matrix $\mathbf{A}$, and $\lambda(\mathbf{F}(\mathcal{S},\boldsymbol{\theta}_d)) = \left[\lambda_1\left(\mathbf{F}(\mathcal{S},\boldsymbol{\theta}_d)\right),\ldots,\lambda_N\left(\mathbf{F}(\mathcal{S},\boldsymbol{\theta}_d)\right)\right]^T$ denotes the eigenvalue vector of the FIM $\mathbf{F}(\mathcal{S},\boldsymbol{\theta}_d)$, arranged in descending order. The following proposition is then established.

Proposition 2. *Using the performance metric in (6), the upper bound of Theorem 2 obtained with $\tilde{\alpha}$ in (27) is tighter than the upper bound in (20) of Theorem 1, i.e., $\alpha > \tilde{\alpha}$, if the auxiliary number ε satisfies*

$$\varepsilon^{\sum_{n=1}^{D}\tilde{n}_d} < \prod_{d=1}^{D}\left(\prod_{n=1}^{\tilde{n}_d}\lambda_n\left(\mathbf{F}\left(\mathcal{S}^{(1)},\boldsymbol{\theta}_d\right)\right)\left(\frac{1}{\prod_{n=1}^{N}\left(1+\kappa_{d,n}\right)}\right)^{e^{N^*/K_{\text{opt}}}}\right), \tag{30}$$

where $\tilde{n}_d$ denotes the rank of $\mathbf{F}\left(\mathcal{S}^{(1)},\boldsymbol{\theta}_d\right)$, and $\kappa_{d,n} = \lambda_n((\mathbf{F}(\mathcal{S}^{(N^*)},\boldsymbol{\theta}_d))^{-1}\mathbf{F}_{s_{N^*+1}}(\boldsymbol{\theta}_d)) \ge 0$, since each individual FIM $\mathbf{F}_m(\boldsymbol{\theta})$ for all $d \in \mathcal{D}$ and $m \in \mathcal{M}$ is positive semidefinite.

Proof. See Appendix C. □

It follows from (30) that a sufficiently small ε always exists to satisfy the condition. The condition in (30) offers a range of ε that ensures $\tilde{\alpha}$, with $\tilde{\alpha}$ in (27), is less than α in (20). Hence, Proposition 2 establishes the theoretical existence of the range of ε, whereas the exact numerical range of ε is not of primary concern as it has no impact on our proposed bound. In practice, it is sufficient to choose ε such that the condition (18) is satisfied [17,19,21,24]. The performance guarantees provided by Theorems 1 and 2 can be computed individually, and the tighter one can be adopted. Further reducing ε does not change the upper bounds in Theorem 2, but only loosens the upper bounds in Theorem 1.

Moreover, the optimality guarantees proposed in Theorem 2 are general enough and can also be applied to other greedy algorithms [17,19,21,24] by considering their respective performance metrics, where similar auxiliary terms are introduced. The existence of a range of ε can be proven by replacing $f(\mathcal{S},\boldsymbol{\theta})$ in (C.6) in Appendix C with the considered nondecreasing submodular performance metric, which introduces auxiliary terms. To illustrate the generality of our proposed optimality guarantees, we apply them to the algorithm proposed in [24] as an example in the following subsection. In summary, compared to the upper bounds in Theorem 1 that depend on the auxiliary number ε, Theorem 2 provides novel upper bounds that are independent of ε. Moreover, Proposition 2 proves the existence of a range of ε, within which the proposed optimality guarantees are tighter than those in Theorem 1.

4.3. Optimality guarantee simulations

In this subsection, we evaluate the optimality guarantees provided by Theorems 1 and 2, as depicted in Figs. 1 and 2. We consider a sensor placement problem for localization with a terrestrial WSN. The scenario and simulation parameters are identical to those described in Section 5.1. To assess the effect of ε on the optimality guarantees, we vary the value of the auxiliary number ε while ensuring that (18) holds. The benchmark is a straight line with a constant value of 1.

Fig. 1 illustrates the optimality guarantees provided by Theorems 1 and 2 for the proposed algorithm. The simulation results indicate that the bound (20) and its counterpart (27) are tightest for Theorems 1 and 2 w.r.t. $\mathbf{P}_1$, respectively. Therefore, Fig. 1 only compares these tightest bounds. In the simulations, we determine N^* by checking whether

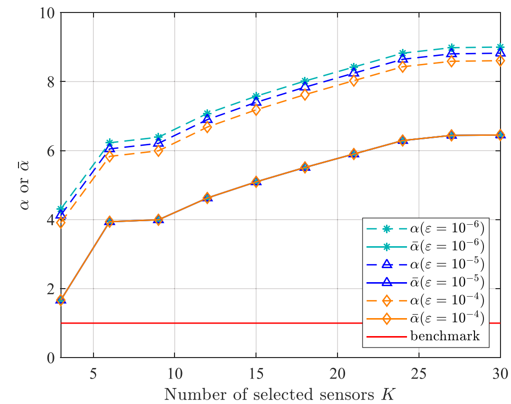

Fig. 1. The optimality bounds provided by Theorems 1 and 2 for the `Avg.-Trunc.` algorithm to solve $\mathbf{P}_1$.

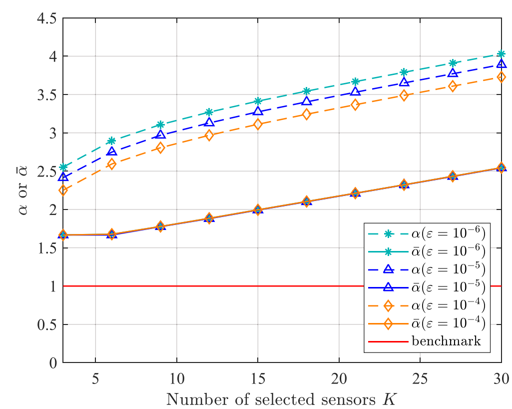

Fig. 2. The optimality bounds provided by Theorems 1 and 2 for the greedy algorithm in [24] to solve $\mathbf{P}_{\text{avg}}$.

$\mathbf{F}(\mathcal{S}^{(k)},\boldsymbol{\theta}_d)$ is full rank for all $d \in \mathcal{D}$ after the kth selection. We observe that $N^* = N$ for all tested values of K. It can be seen that the optimality guarantee provided by Theorem 2 ($\tilde{\alpha}$ denoted by solid lines) is much tighter than the one in Theorem 1 (α denoted by dashed lines). Note that the performance guarantee provided by Theorem 1 improves as the auxiliary number ε increases. However, it remains considerably weaker than the guarantee proposed in Theorem 2. Moreover, the proposed novel optimality guarantee on the number of selected sensors is independent of the auxiliary number ε, as long as the condition (18) is satisfied. Increasing ε affects the sensor selection process and leads to a degraded algorithm performance.

Fig. 2 illustrates the application of Theorems 1 and 2 to the greedy algorithm [24] for solving the sensor placement problem $\mathbf{P}_{\text{avg}}$. The problem $\mathbf{P}_{\text{avg}}$ is obtained from $\mathbf{P}_1$ by replacing its performance constraint with $\sum_{d=1}^{D} f(\mathcal{S},\boldsymbol{\theta}_d) \ge \bar{\gamma}$. As the simulation results show that the bound (21) and its counterpart using $\tilde{\alpha}$ in (28) are tightest for Theorems 1 and 2 w.r.t. $\mathbf{P}_{\text{avg}}$, respectively, Fig. 2 illustrates only these tightest bounds. In this scenario, we observe that $N^* = N$ for all tested values of K. The proposed optimality guarantee in Theorem 2 ($\tilde{\alpha}$ denoted by solid lines) also has a significant improvement compared to that in Theorem 1 (α denoted by dashed lines). Note that α and $\tilde{\alpha}$ in Fig. 2 are generally smaller than those in Fig. 1. However, the greedy algorithm in [24] for $\mathbf{P}_{\text{avg}}$ cannot guarantee the performance constraints for all grid points.

5. Applications: Sensor selection for localization

Target localization is a fundamental task in WSNs. Sensor placement is critical to localization system design. Unfavorable sensor-target geometries can severely degrade accuracy and may cause ambiguity or

Table 1
Summary of the sensor selection algorithms in the simulation.

Algorithm	Performance constraint	Select the sensor in the kth step	Complexity
`Avg.-Trunc.` (proposed)	$\sum_{d=1}^{D} g(\mathcal{S},\boldsymbol{\theta}_d,\gamma_d)/D=\bar{\gamma}$	$s_k=\arg\max_{j\in\mathcal{M}\setminus\mathcal{S}^{(k-1)}}\sum_{d=1}^{D} g\left(\mathcal{S}^{(k-1)}\cup\{j\},\boldsymbol{\theta}_d,\gamma_d\right)$	$O\left(\sum_{k=0}^{K-1}\lvert\bar{\mathcal{D}}^{(k)}\rvert MN^3\right)$
`Avg.` [24] (modified)	$\sum_{d=1}^{D} f\left(\mathcal{S},\boldsymbol{\theta}_d\right)/D\ge\bar{\gamma}$	$s_k=\arg\max_{j\in\mathcal{M}\setminus\mathcal{S}^{(k-1)}}\sum_{d=1}^{D} f\left(\mathcal{S}^{(k-1)}\cup\{j\},\boldsymbol{\theta}_d\right)$	$O(DKMN^3)$
`Min.` [5]	$\min_{d\in\mathcal{D}} f\left(\mathcal{S},\boldsymbol{\theta}_d\right)\ge\bar{\gamma}$	$s_k=\arg\max_{j\in\mathcal{M}\setminus\mathcal{S}^{(k-1)}}\min_{d\in\mathcal{D}} f\left(\mathcal{S}^{(k-1)}\cup\{j\},\boldsymbol{\theta}_d\right)$	$O(DKMN^3)$
`Robust` [26] (modified)	$\min_{d\in\mathcal{D}} f\left(\mathcal{S},\boldsymbol{\theta}_d\right)\ge\bar{\gamma}$	$s_k=\arg\max_{j\in\mathcal{M}\setminus\mathcal{S}^{(k-1)}}\min_{d\in\mathcal{D}}\frac{f(\mathcal{S}^{(k-1)}\cup\{j\},\boldsymbol{\theta}_d)-f(\mathcal{S}^{(k-1)},\boldsymbol{\theta}_d)}{f(\mathcal{S}^{(k-1)}\cup\{j_{d,k}^*\},\boldsymbol{\theta}_d)-f(\mathcal{S}^{(k-1)},\boldsymbol{\theta}_d)}$	$O(DKM^2N^3)$
`Convex-`ℓ_1 [3]	Performance constraint: $\ln\det\left(\sum_{m=1}^{M} w_m\mathbf{F}_m\left(\boldsymbol{\theta}_d\right)\right)\ge\gamma_d,\forall d\in\mathcal{D}$, and $w\in[0,1]^M$		$O(DM^3)$

even loss of observability. When the deployable region of sensors is discretized into a set of candidate positions, the sensor placement problem reduces to selecting the most informative subset of sensor positions. This formulation is mathematically equivalent to a sensor selection problem. In this section, we compare the proposed algorithm with several greedy and convex algorithms under two typical nonlinear measurement models for sensor placement. The selected sensor subset is optimized to guarantee the localization performance within the surveillance area $\mathcal{U}$. As long as the target moves within $\mathcal{U}$, the selected sensors can maintain the prescribed localization performance along its trajectory. This is because a target trajectory can be viewed as a sequence of target locations within $\mathcal{U}$. Moreover, the sensor availability or the sensing conditions may vary over time. In such cases, the proposed sensor selection can be re-executed with the updated candidate set or model parameters. More detailed extensions for dynamic scenarios are left for future work.

Table 1 summarizes the algorithms involved in the simulations. For a fair comparison, the parameters are set as $\gamma_1=\cdots=\gamma_D=\bar{\gamma}$. The evaluated algorithms are described as follows:

- `Avg.-Trunc.` (proposed): the proposed algorithm in Algorithm 1.
- `Avg.` (modified): the greedy algorithm [24] that optimizes the average of the weighted performance metrics $\sum_{d=1}^{D}\pi_d f(\mathcal{S},\boldsymbol{\theta}_d)/D$. Here, we set the weights $\pi_d=1$. To facilitate a fair comparison with the proposed `Avg.-Trunc.` algorithm, we modify the termination condition of the greedy algorithm [24] to $\min_{d\in\mathcal{D}} f(\mathcal{S},\boldsymbol{\theta}_d)\ge\bar{\gamma}$.
- `Convex-`ℓ_1: the convex approximation algorithm based on the ℓ_1-norm with the randomized rounding technique [3].
- `Min.`: the greedy algorithm [5] that optimizes the minimum performance metrics.
- `Robust` (modified): the robust submodular maximization algorithm [26] that leverages correlation between the submodular functions, where $j_{d,k}^*=\arg\max_{j\in\mathcal{M}\setminus\mathcal{S}^{(k-1)}}\left\{f\left(\mathcal{S}^{(k-1)}\cup\{j\},\boldsymbol{\theta}_d\right)-f\left(\mathcal{S}^{(k-1)},\boldsymbol{\theta}_d\right)\right\}$ is employed in Table 1. We modify the termination condition of the algorithm in [26] to $\min_{d\in\mathcal{D}} f(\mathcal{S},\boldsymbol{\theta}_d)\ge\bar{\gamma}$ for fair comparisons.

5.1. Terrestrial WSNs

We consider a WSN deployed in a two-dimensional plane as shown in Fig. 3. A target is located in the target region $\mathcal{U}$, and we can only place sensors on grid points in the set $\mathcal{M}$, where $M=|\mathcal{M}|=80$. The absolute positions of the grid points are known. Hence, the sensors under consideration are often referred to as anchor nodes. Let $\boldsymbol{\theta}=\left[\theta_1,\theta_2\right]^T$ and $\mathbf{a}_m=\left[a_{m,1},a_{m,2}\right]^T$ denote the coordinates of the target and the mth anchor node, respectively, where $\boldsymbol{\theta}$ is unknown but lies within $\mathcal{U}$. To avoid infinitely many constraints, the surveillance area is uniformly discretized with a resolution of 2.5 m along both horizontal and vertical directions, resulting in $D=|\mathcal{U}|=81$ grid points within a $20\,\text{m}\times 20\,\text{m}$ area. This setup is representative of typical target localization scenarios and has also been adopted in [3,5,24]. Simulations are conducted under the distance measurement model. Let $d_m(\boldsymbol{\theta})=\|\boldsymbol{\theta}-\mathbf{a}_m\|_2$ denote the Euclidean distance between the target and the mth anchor. The range

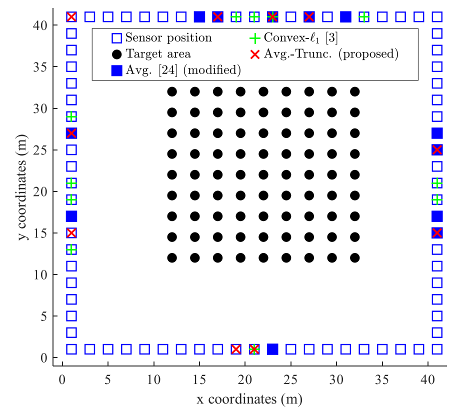

Fig. 3. Sensor selection for target localization. .

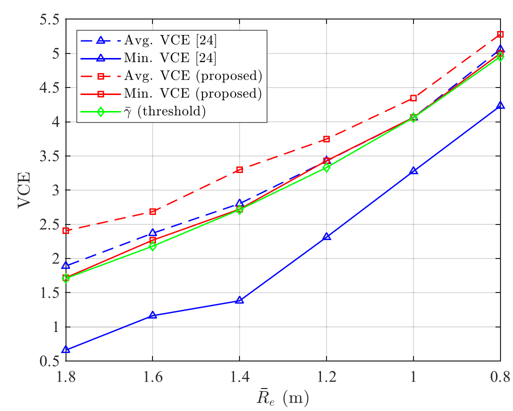

Fig. 4. Minimum and average VCE of the sensor selection performances. .

measurements follow a nonlinear observation model expressed as

$$y_m=d_m(\boldsymbol{\theta})+n_m,\ \ m=1,2,\ldots,M, \tag{31}$$

where $n_m\sim\mathcal{N}(0,\sigma_m^2)$ is AWGN with $\sigma_m^2=\sigma^2 d_m^2(\boldsymbol{\theta})$, and σ^2 denotes the nominal noise variance. The local FIM $\mathbf{F}_m(\boldsymbol{\theta})$ is given by

$$\mathbf{F}_m(\boldsymbol{\theta})=\frac{\left(\boldsymbol{\theta}-\mathbf{a}_m\right)\left(\boldsymbol{\theta}-\mathbf{a}_m\right)^T}{\sigma_m^2||\boldsymbol{\theta}-\mathbf{a}_m||_2^2}. \tag{32}$$

The dependence of σ_m^2 on $\boldsymbol{\theta}$ is neglected in the derivation of $\mathbf{F}_m(\boldsymbol{\theta})$ because its variation is insignificant. Each local FIM $\mathbf{F}_m(\boldsymbol{\theta})\in\mathbb{R}^{N\times N}$ has rank 1. Unless otherwise specified, the simulations use $\sigma^2=2\times10^{-3}$, $\varepsilon=1\times10^{-5}$, and $P_e=0.9$.

Fig. 3 also shows the sensor positions selected by the different algorithms under a specific threshold setting. The threshold $\bar{\gamma}$ is computed using $\bar{R}_e=1.2$ m. The sensor positions selected by the modified `Avg.` algorithm [24] are concentrated on the upper and right sides of the

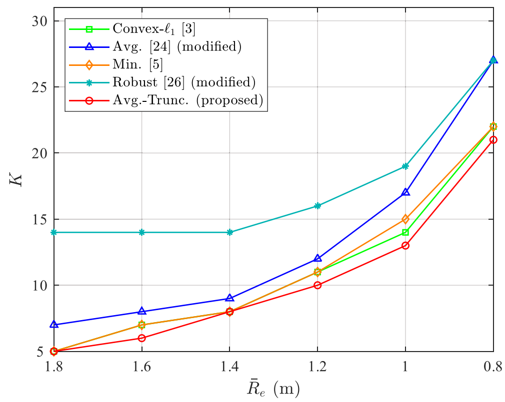

Fig. 5. The number of selected sensors K for different values of $\bar{R}_e$.

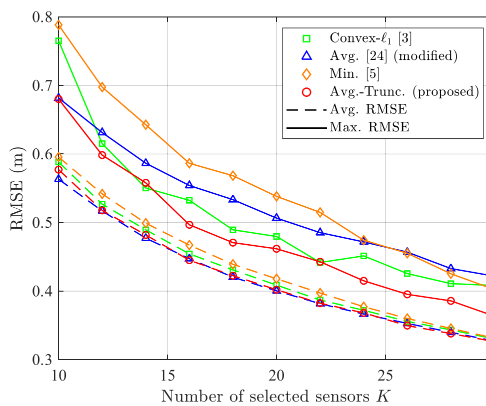

Fig. 6. Comparison of the RMSE using K sensors selected by different algorithms. .

candidate region, close to the target area. In contrast, the sensor positions selected by the `Convex-`ℓ_1 algorithm [3] and the proposed `Avg.-Trunc.` algorithm are more evenly distributed. Furthermore, the proposed `Avg.-Trunc.` algorithm achieves the same localization accuracy with only 10 sensors, fewer than those required by the modified `Avg.` algorithm (12 sensors) and the `Convex-`ℓ_1 algorithm (11 sensors).

As the determinant constraint is adopted, Fig. 4 displays the VCE performance of sensors selected by the proposed `Avg.-Trunc.` algorithm and the `Avg.` algorithm [24] under the termination condition $\sum_{d\in D} f(S,\theta_d)/D \geq \bar{\gamma}$ for various $\bar{R}_e$. The Avg. VCE (or Min. VCE) indicates the average (or minimum) logarithm of the determinant of the FIM over different target positions. As shown in Fig. 4, the `Avg.-Trunc.` algorithm ensures the measurement performance (the Min. VCE) for all $\theta_d \in \mathcal{U}$. In contrast, the `Avg.` algorithm [24] with the termination condition of $\sum_{d\in D} f(S,\theta_d)/D \geq \bar{\gamma}$ fails to provide such assurance, as its Min. VCE is lower than the threshold $\bar{\gamma}$.

Fig. 5 compares the number of sensors selected by different algorithms to achieve the same localization accuracy. The proposed `Avg.-Trunc.` algorithm selects the fewest sensors across different values of $\bar{R}_e$. The modified `Robust` algorithm [26] and the modified `Avg.` algorithm [24] require more sensors than all other algorithms. The poor performance of the modified `Robust` algorithm [26] is because the introduction of $\epsilon\mathbf{I}$ significantly reduces the curvature of the performance metric functions, and leads to a looser lower bound of the correlation ratio of the submodular metric functions. Thus, the modified `Robust` algorithm [26] is not suitable for the sensor selection problem in this scenario.

Figs. 3–5 show the outputs of our proposed `Avg.-Trunc.` algorithm, i.e., $S^{(K)}$, $\sum_{d=1}^{D} \ln\det(\mathbf{F}(S^{(K)},\theta_d))/D$, and K, respectively. The performance benefits of the proposed algorithm are evidenced by them. Furthermore, the impact of the selected sensors on the root mean square error (RMSE) of target localization is evaluated, as the localization RMSE is also of interest. Fig. 6 provides a localization performance comparison of the average RMSE (Avg. RMSE) and maximum RMSE (Max. RMSE) using the selected K sensors. Due to the poor performance of the modified `Robust` algorithm [26], it is excluded from Fig. 6. The Avg. RMSE is computed as $\sqrt{\sum_{mc=1}^{Q}\sum_{d=1}^{D}||\hat{\theta}_d^{(mc)}-\theta_d||_2^2/QD}$, and the Max. RMSE is defined as $\sqrt{\sum_{mc=1}^{Q}\max_{d\in D}||\hat{\theta}_d^{(mc)}-\theta_d||_2^2/Q}$. The total number Q of Monte-Carlo (MC) runs is 10000. We employ a weighted least squares estimator for localization and solve it iteratively using Gauss-Newton's method with 10 iterations [32]. The weight matrix is diagonal, and its diagonal elements are given by $1/(y_m^2\sigma^2)$ corresponding to different sensor positions. The weight $1/(y_m^2\sigma^2)$ is an approximation of $1/\sigma_m^2 = 1/(d_m^2\sigma^2)$ due to the unknown value of d_m. As the noise variance increases, the accuracy of the localization result diminishes.

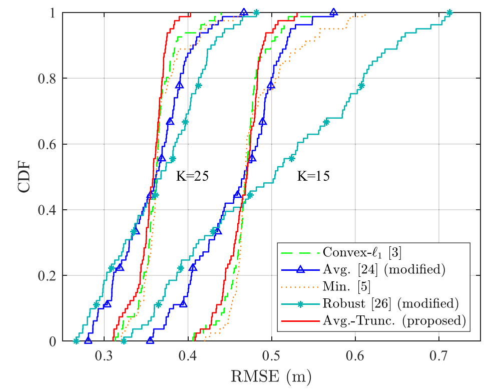

Fig. 7. CDF curves of the RMSE using the K sensors.

As shown in Fig. 6, the proposed `Avg.-Trunc.` algorithm achieves the minimum Max. RMSE for most values of K, and significantly improves the worst-case localization accuracy in the surveillance area. The `Convex-`ℓ_1 algorithm [3], which satisfies the performance constraint by optimizing the worst VCE, is in the second place in terms of Max. RMSE. However, it has substantially higher computational complexity compared to the proposed `Avg.-Trunc.` algorithm. Although the modified `Min.` algorithm [5] also optimizes the worst VCE, its Max. RMSE exceeds that of the modified `Avg.` algorithm [24] when $K \leq 24$. Moreover, the modified `Min.` algorithm [5] lacks an optimality guarantee, and its performance can be arbitrarily poor in scenarios where the increment of the performance metric of a given sensor j, i.e., $f(S^{(k-1)}\cup\{j\},\theta_d) - f(S^{(k-1)},\theta_d)$, varies significantly across different values of θ_d. This observation is consistent with the results in [25, Table 1]. In terms of the Avg. RMSE, the proposed `Avg.-Trunc.` algorithm achieves comparable performance to the modified `Avg.` algorithm [24], and consistently outperforms all the other competitors. It indicates that the proposed `Avg.-Trunc.` algorithm can reduce the Max. RMSE without significantly increasing the Avg. RMSE.

To show more details about the localization RMSE, Fig. 7 illustrates the cumulative distribution function (CDF) curves of the localization RMSE obtained with the K selected sensors, where $K = 15$ and $K = 25$, respectively. Although the modified `Avg.` algorithm [24] and the modified `Robust` algorithm [26] achieve a smaller minimum RMSE in some MC realizations, their RMSE distributions shown in Fig. 7 exhibit heavier right tails. Thus, they have a larger Max. RMSE. By contrast, the proposed `Avg.-Trunc.` algorithm achieves the smallest Max. RMSE in Fig. 7. These results indicate that the proposed algorithm is optimal in terms of the worst-case localization error (Max. RMSE) among the compared algorithms. It also aligns with the proposed metric. Moreover, the proposed `Avg.-Trunc.` algorithm attains a smaller variance of the RMSE than the modified `Avg.` algorithm [24] and the modified `Robust`

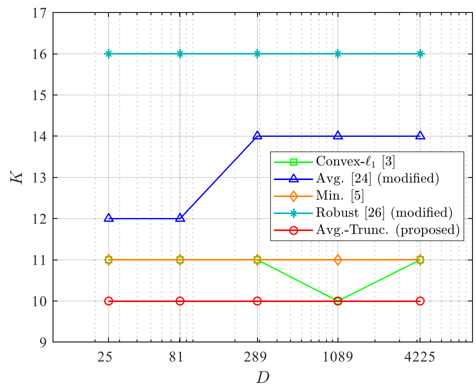

Fig. 8. The number of selected sensors K for different values of D.

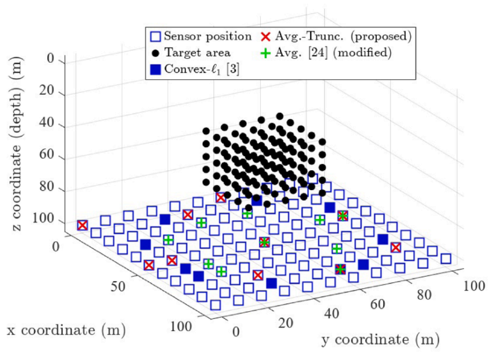

Fig. 9. Sensor selection for target localization. The threshold $\bar{\gamma}$ is computed using $\bar{R}_e = 3\,\mathrm{m}$. .

algorithm [26]. Accordingly, the localization accuracy using the sensors selected by the proposed `Avg.-Trunc.` algorithm is more reliable.

In the above simulations, the surveillance area is uniformly discretized into $D = 81$ grid points. We next evaluate the sensitivity of the algorithms to the resolution of the discretization. The total number of grid points D is chosen from the set $\{25, 81, 289, 1089, 4225\}$, which corresponds to the set of grid spacings $\{5\text{ m}, 2.5\text{ m}, 1.25\text{ m}, 0.625\text{ m}, 0.3125\text{ m}\}$ over the $20\text{ m} \times 20\text{ m}$ region. Since our objective is to minimize the number of selected sensors, Fig. 8 illustrates the number of selected sensors K versus D. The threshold $\bar{\gamma}$ is computed using $\bar{R}_e = 1.2$ m. All algorithms except the modified `Avg.` algorithm [24] exhibit weak sensitivity to D in this setup. The small fluctuations of `Convex-`ℓ_1 [3] are mainly due to the randomized rounding used to obtain a binary selection. For the modified `Avg.` algorithm [24], changing D is equivalent to changing the relative weights of different locations in the aggregated performance metric. Consequently, the resulting K is more sensitive to the discretization resolution. Moreover, the proposed `Avg.-Trunc.` consistently selects the smallest number of sensors across all tested values of D. In practice, the choice of D introduces a trade-off between the coverage of $\mathcal{U}_0$ and the computational cost.

5.2. Underwater WSNs

In this subsection, we consider a more intricate three-dimensional underwater positioning scenario. As illustrated in Fig. 9, sensors are deployed over a seabed area of $100\,\mathrm{m} \times 100\,\mathrm{m}$ with a grid resolution of $10\,\mathrm{m}$, resulting in $M = 121$ potential sensor locations. The target is located within a cubic region of side length $32\,\mathrm{m}$, centered at $[50\,\mathrm{m}, 50\,\mathrm{m}, 50\,\mathrm{m}]^T$. The target space is discretized into $D = 125$ grid points with a resolution of $8\,\mathrm{m}$. The sound speed profile (SSP) $C(z)$ depends solely on the depth z [39] and is expressed as $C(z) = b + az$, where b denotes the sound speed at the surface and a is an environment-dependent constant. The underwater wireless sensor network (UWSN) estimates the target position using time-of-flight (ToF) measurements. Let $\mathbf{y} = [y_1, y_2, \dots, y_M]^T$ represent the ToF measurements between the target node and the sensor nodes, given by $\mathbf{y} = \mathbf{t}(\mathbf{x}) + \mathbf{n}$, where $\mathbf{t}(\mathbf{x}) = [t_1, t_2, \dots, t_M]^T$ represents the true ToFs w.r.t. the target location $\mathbf{x} = [x_1, x_2, x_3]^T$, calculated as [39]:

$$t_m = -\frac{1}{a}\left(\log\frac{1+\sin\psi^{\mathrm{G}}}{\cos\psi^{\mathrm{G}}} - \log\frac{1+\sin\psi^{\mathrm{A}}}{\cos\psi^{\mathrm{A}}}\right), \tag{33}$$

with ψ^{G} and ψ^{A} denoting the angles of underwater acoustic radiation at the target and sensor node locations, respectively. The additive noise $\mathbf{n}$ follows a Gaussian distribution with $\mathbb{E}(\mathbf{n}\mathbf{n}^T) = \mathrm{diag}\left(\sigma_1^2, \sigma_2^2, \dots, \sigma_M^2\right)$, where $\sigma_m^2 = K_E A(l_m, \omega)$ [39] with $K_E = -30\,\mathrm{dB}$, and $A(l_m, \omega) = \left(l_m/l_0\right)^{\zeta}(L(\omega))^{l_m - l_0}$ represents the overall path loss [40], with l_m being the propagation distance referenced to $l_0 = 1000\,\mathrm{m}$. The path loss exponent is $\zeta = 2$, and the absorption coefficient $L(\omega) = 1\,\mathrm{dB/km}$ is obtained using an empirical formula [40]. The specific expression of the FIM is shown in [39, eq. (30)]. The simulation parameters are set as $\varepsilon = 1 \times 10^{-3}$, $P_e = 0.9$, $b = 1480\,\mathrm{m/s}$, $a = 0.1\,\mathrm{m/s}$. Fig. 9 reveals that compared to the modified `Avg.` algorithm [24], the sensors selected by the `Avg.-Trunc.` algorithm or `Convex-`ℓ_1 algorithm [3] are closer to the boundary of the seabed area. Moreover, they select the same number of sensors to achieve the same localization accuracy.

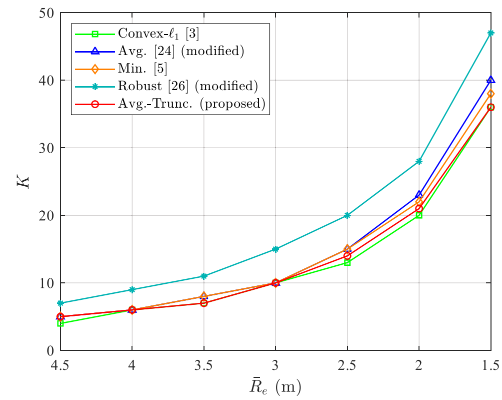

Fig. 10. The number of selected sensors K for different values of $\bar{R}_e$.

Fig. 10 compares the number of sensors selected by the different algorithms under various $\bar{R}_e$. The modified `Robust` algorithm [26] always chooses the greatest number of sensors. The number of sensors selected by the proposed `Avg.-Trunc.` algorithm is always no greater than the ones chosen by the modified `Avg.` [24] and the `Min.` [5] algorithms, but also no less than the one picked by the `Convex-`ℓ_1 algorithm [3]. Therefore, the proposed `Avg.-Trunc.` algorithm achieves a good balance between performance and computational cost.

In the above simulations, we compare the proposed algorithm with several greedy and convex baselines under two typical nonlinear measurement models. In practical applications, such as WSN/IoT deployments, the proposed method can be applied to select a small set of sensing locations from a feasible candidate pool. This pool accounts for site constraints such as power supply, accessibility, and communication coverage. The performance metric can be instantiated using a measurement model for a specific application (e.g., TDOA/FDOA/RSS/AoA). Nominal model parameters and noise statistics can be obtained from calibration or historical measurements. Potential model mismatches can be accommodated by adopting conservative uncertainty and noise modeling, e.g., using inflated noise levels or enlarged uncertainty regions. These treatments integrate naturally into the proposed greedy sensor selection framework and help maintain reliable performance over the region of interest under imperfect modeling.

6. Conclusions

The sensor selection problem with a general nonlinear model can be formulated as an optimization problem with multiple constraints. In this work, we introduce a performance metric defined as the average of truncated submodular functions to consolidate multiple inequality constraints into a single equality constraint. This formulation not only ensures satisfactory performance for each θ_d within the domain $\mathcal{U}$ but also preserves the submodular optimality. The proposed `Avg.-Trunc.` algorithm achieves comparable performance to the `Convex-`ℓ_1 method [3], while substantially reducing computational complexity, making it well-suited for large-scale sensor selection problems. Furthermore, we present a performance analysis of the proposed greedy algorithm and establish new optimality guarantees that are independent of the auxiliary parameter ε, thereby addressing the limitations introduced by the small auxiliary term $\varepsilon\mathbf{I}$. The proposed guarantees can also be extended to other sensor selection problems with different performance metrics.

CRediT authorship contribution statement

Jiaming Cui: Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization; **Lingya Liu:** Writing – review & editing, Resources, Investigation, Formal analysis, Conceptualization; **Geert Leus:** Writing – review & editing, Investigation, Formal analysis; **Yiyin Wang:** Writing – review & editing, Writing – original draft, Resources, Investigation, Formal analysis, Conceptualization.

Data availability

No data was used for the research described in the article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Background on Submodularity

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

$$q(\mathcal{A} \cup \{j\}) - q(\mathcal{A}) \geq q(\mathcal{B} \cup \{j\}) - q(\mathcal{B}). \tag{A.1}$$

Theorem 3 (The $(1-1/e)$ optimality guarantee). *Let $q : 2^{\mathcal{M}} \to \mathbb{R}$ be a nondecreasing submodular function with $q(\emptyset) = 0$ on a finite set $\mathcal{M}$. Let $\mathcal{S}_{\text{opt}}$ be the optimal solution of the following combinatorial optimization problem:*

$$\max_{\mathcal{S}\subseteq\mathcal{M}} q(\mathcal{S}), \quad \text{s.t. } |\mathcal{S}| = K.$$

Let $\mathcal{S}^{(K)}$ be the subset chosen from $\mathcal{M}$ by the greedy method. Due to the submodularity of $q(\mathcal{S})$, the solution $\mathcal{S}^{(K)}$ has the $(1-1/e)$ optimality guarantee [23] as follows

$$q\left(\mathcal{S}^{(K)}\right) \geq \left(1 - \frac{1}{e}\right) q(\mathcal{S}_{\text{opt}}), \tag{A.2}$$

where e is Euler's number.

Proof. See [23]. □

Remark 3. (Impact of ε on the $(1-1/e)$ optimality guarantee): If the performance metric in (6), which introduces the auxiliary term $\varepsilon\mathbf{I}$, is used, substituting it into the $(1-1/e)$ optimality guarantee in (A.2) yields

$$f\left(\mathcal{S}^{(K)}, \boldsymbol{\theta}\right) \geq \left(1 - \frac{1}{e}\right) f\left(\mathcal{S}_{\text{opt}}, \boldsymbol{\theta}\right) + \frac{1}{e} N \ln \varepsilon. \tag{A.3}$$

The additional term $(N \ln \varepsilon)/e$ in (A.3), resulting from the small number ε, makes the $(1-1/e)$ optimality guarantee overly loose.

Appendix B. Proof of Corollary 1

We provide the following proof based on the specific form of each bound.

1. For the bounds in (19) and (26), the proof of $\tilde{\alpha} \leq \alpha$ is trivial. According to the diminishing returns property of submodular functions, we obtain

$$\max_{j\in\mathcal{M}\backslash\mathcal{S}^{(N^*)},\, p\in\mathcal{K}\backslash\mathcal{N}^*} \left\{ \frac{\delta_j\left(\mathcal{S}^{(N^*)}\right)}{\delta_j\left(\mathcal{S}^{(p)}\right)} \right\} \leq \max_{j\in\mathcal{M},\, p\in\mathcal{K}} \left\{ \frac{\delta_j\left(\mathcal{S}^{(N^*)}\right)}{\delta_j\left(\mathcal{S}^{(p)}\right)} \right\} \leq \max_{j\in\mathcal{M},\, p\in\mathcal{K}} \left\{ \frac{\delta_j\left(\mathcal{S}^{(0)}\right)}{\delta_j\left(\mathcal{S}^{(p)}\right)} \right\}. \tag{B.1}$$

2. Taking the specific expressions of the bounds in (20) and (27) into $\alpha - \tilde{\alpha}$, we obtain that $\alpha - \tilde{\alpha} = \ln \beta_1/\beta_{N^*+1}$. First of all, we have $\beta_1 = \delta_{s_1}\left(\mathcal{S}^{(0)}\right) \geq \delta_{s_{N^*+1}}\left(\mathcal{S}^{(0)}\right)$ based on the definition of β_k in (14). According to the diminishing returns property of submodular functions, we further obtain $\delta_{s_{N^*+1}}\left(\mathcal{S}^{(0)}\right) \geq \delta_{s_{N^*+1}}\left(\mathcal{S}^{(N^*)}\right) = \beta_{N^*+1}$. As a result, we arrive at $\beta_1 \geq \beta_{N^*+1}$ and $\alpha - \tilde{\alpha} = \ln(\beta_1/\beta_{N^*+1}) \geq 0$.
3. Taking the specific expressions of the bounds in (21) and (28) into $\alpha - \tilde{\alpha}$, we obtain that $\alpha - \tilde{\alpha} = \ln\left(\delta(\mathcal{M},\emptyset)/\delta\left(\mathcal{M},\mathcal{S}^{(N^*)}\right)\right) \geq 0$. Because the performance metric is a nondecreasing set function, $g\left(\mathcal{S}^{(N^*)},\theta_d,\gamma_d\right) \geq g(\emptyset,\theta_d,\gamma_d)$. As a result, $\delta(\mathcal{M},\emptyset) \geq \delta\left(\mathcal{M},\mathcal{S}^{(N^*)}\right)$.

This completes the proof of Corollary 1.

Appendix C. Proof of Proposition 2

Taking (20), (27) and (29) into account, we need to prove that

$$\alpha - \bar{\alpha} = \ln\frac{\beta_1}{\beta_K} - \left(\ln\frac{\beta_{N^*+1}}{\beta_K} + \frac{N^*}{K_{\text{opt}}}\right) = \ln\frac{\beta_1}{\beta_{N^*+1}} - \frac{N^*}{K_{\text{opt}}} > 0, \tag{C.1}$$

if the auxiliary number ε satisfies (30). Substituting the definitions of β_k in (14) and $\delta_j\left(\mathcal{S}^{(k-1)}\right)$ in (12) into (C.1) yields

$$\frac{\beta_1}{\beta_{N^*+1}} = \frac{\delta_{s_1}\left(\mathcal{S}^{(0)}\right)}{\delta_{s_{N^*+1}}\left(\mathcal{S}^{(N^*)}\right)} = \frac{\sum_{d=1}^{D}\left(g\left(\mathcal{S}^{(1)},\theta_d,\gamma_d\right) - g\left(\mathcal{S}^{(0)},\theta_d,\gamma_d\right)\right)}{\sum_{d=1}^{D}\left(g\left(\mathcal{S}^{(N^*+1)},\theta_d,\gamma_d\right) - g\left(\mathcal{S}^{(N^*)},\theta_d,\gamma_d\right)\right)}. \tag{C.2}$$

According to the definition of $\bar{\mathcal{D}}^{(k)}$, we derive the numerator of (C.2) as

$$\begin{aligned}&\sum_{d=1}^{D}\left(g\left(\mathcal{S}^{(1)},\theta_d,\gamma_d\right) - g\left(\mathcal{S}^{(0)},\theta_d,\gamma_d\right)\right)\\ &= \sum_{d\in\bar{\mathcal{D}}^{(1)}}\left(f\left(\mathcal{S}^{(1)},\theta_d\right) - f(\emptyset,\theta_d)\right) + \sum_{d\in\mathcal{D}\backslash\bar{\mathcal{D}}^{(1)}}\left(\gamma_d - f(\emptyset,\theta_d)\right).\end{aligned} \tag{C.3}$$

Recall that $f\left(\mathcal{S}^{(k)},\theta_d\right)$ is nondecreasing If $f\left(\mathcal{S}^{(1)},\theta_d\right) \geq \gamma_d$, then both $f\left(\mathcal{S}^{(N^*)},\theta_d\right)$ and $f\left(\mathcal{S}^{(N^*+1)},\theta_d\right)$ are greater than or equal to γ_d. In this case, $g\left(\mathcal{S}^{(N^*+1)},\theta_d,\gamma_d\right) - g\left(\mathcal{S}^{(N^*)},\theta_d,\gamma_d\right) = \gamma_d - \gamma_d = 0$. Moreover, if $g\left(\mathcal{S}^{(N^*+1)},\theta_d,\gamma_d\right) = \gamma_d$ and $g\left(\mathcal{S}^{(N^*)},\theta_d,\gamma_d\right) \neq \gamma_d$, then, according to the definition of $g\left(\mathcal{S}^{(k)},\theta_d,\gamma_d\right)$ in (7), we have $g\left(\mathcal{S}^{(N^*+1)},\theta_d,\gamma_d\right) - g\left(\mathcal{S}^{(N^*)},\theta_d,\gamma_d\right) = \gamma_d - f\left(\mathcal{S}^{(N^*)},\theta_d\right) \leq f\left(\mathcal{S}^{(N^*+1)},\theta_d\right) - f\left(\mathcal{S}^{(N^*)},\theta_d\right)$. Consequently, we obtain

$$g\left(\mathcal{S}^{(N^*+1)},\theta_d,\gamma_d\right) - g\left(\mathcal{S}^{(N^*)},\theta_d,\gamma_d\right) \leq f\left(\mathcal{S}^{(N^*+1)},\theta_d\right) - f\left(\mathcal{S}^{(N^*)},\theta_d\right). \tag{C.4}$$

Using (C.3) and (C.4), we arrive at

$$\frac{\beta_1}{\beta_{N^*+1}} \geq \frac{\sum_{d\in\bar{\mathcal{D}}^{(1)}}\left(f\left(\mathcal{S}^{(1)},\theta_d\right) - f(\emptyset,\theta_d)\right) + \sum_{d\in\mathcal{D}\backslash\bar{\mathcal{D}}^{(1)}}\left(\gamma_d - f(\emptyset,\theta_d)\right)}{\sum_{d\in\bar{\mathcal{D}}^{(1)}}\left(f\left(\mathcal{S}^{(N^*+1)},\theta_v\right) - f\left(\mathcal{S}^{(N^*)},\theta_v\right)\right)}, \tag{C.5}$$

and if

$$\frac{\sum_{d\in\bar{\mathcal{D}}^{(1)}}\left(f\left(\mathcal{S}^{(1)},\theta_d\right) - f(\emptyset,\theta_d)\right) + \sum_{d\in\mathcal{D}\backslash\bar{\mathcal{D}}^{(1)}}\left(\gamma_d - f(\emptyset,\theta_d)\right)}{\sum_{d\in\bar{\mathcal{D}}^{(1)}}\left(f\left(\mathcal{S}^{(N^*+1)},\theta_v\right) - f\left(\mathcal{S}^{(N^*)},\theta_v\right)\right)} > e^{\frac{N^*}{K_{\text{opt}}}}, \tag{C.6}$$

then $\ln(\beta_1/\beta_{N^*+1}) > N^*/K_{\text{opt}}$, which implies $\alpha > \bar{\alpha}$.

Since we assume that $f(\mathcal{S}^{(1)}, \theta_d) \leq \gamma_d$, it follows that $\bar{\mathcal{D}}^{(1)} = \mathcal{D}$. By substituting the performance metric into (6) and $f(\emptyset, \theta_d) = N \ln \varepsilon$ in (C.6), the condition on ε can be further derived and as

$$\varepsilon^{DN} < \prod_{d=1}^{D} \left(\det\left(\mathbf{F}(\mathcal{S}^{(1)}, \theta_d) + \varepsilon\mathbf{I}\right) \left(\frac{\det\left(\mathbf{F}(\mathcal{S}^{(N^*)}, \theta_d) + \varepsilon\mathbf{I}\right)}{\det\left(\mathbf{F}(\mathcal{S}^{(N^*+1)}, \theta_d) + \varepsilon\mathbf{I}\right)} \right)^{e^{N^*/K_{\text{opt}}}} \right), \tag{C.7}$$

Since the FIM $\mathbf{F}(\mathcal{S}^{(N^*)}, \theta_d)$ is full rank, we can decompose

$$\begin{aligned}
&\det\left(\mathbf{F}\left(\mathcal{S}^{(N^*+1)}, \theta_d\right) + \varepsilon\mathbf{I}\right) \\
&= \det\left(\mathbf{F}\left(\mathcal{S}^{(N^*)}, \theta_d\right) + \varepsilon\mathbf{I}\right) \det\left(\mathbf{I} + \left(\mathbf{F}\left(\mathcal{S}^{(N^*)}, \theta_d\right) + \varepsilon\mathbf{I}\right)^{-1} \mathbf{F}_{s_{N^*+1}}(\theta_d)\right) \\
&\leq \det\left(\mathbf{F}\left(\mathcal{S}^{(N^*)}, \theta_d\right) + \varepsilon\mathbf{I}\right) \det\left(\mathbf{I} + \left(\mathbf{F}\left(\mathcal{S}^{(N^*)}, \theta_d\right)\right)^{-1} \mathbf{F}_{s_{N^*+1}}(\theta_d)\right) \\
&= \det\left(\mathbf{F}\left(\mathcal{S}^{(N^*)}, \theta_d\right) + \varepsilon\mathbf{I}\right) \prod_{n=1}^{N} \left(1 + \kappa_{d,n}\right),
\end{aligned} \tag{C.8}$$

where $\kappa_{d,n} = \lambda_n((\mathbf{F}(\mathcal{S}^{(N^*)}, \theta_d))^{-1}\mathbf{F}_{s_{N^*+1}}(\theta_d))$. Therefore, the condition on ε in (C.7) is fulfilled if

$$\varepsilon^{DN} < \prod_{d=1}^{D} \left(\det\left(\mathbf{F}(\mathcal{S}^{(1)}, \theta_d)\right) \left(\frac{1}{\prod_{n=1}^{N}\left(1 + \kappa_{d,n}\right)} \right)^{e^{N^*/K_{\text{opt}}}} \right). \tag{C.9}$$

Let $\tilde{n}_d$ denotes the rank of the FIM $\mathbf{F}(\mathcal{S}^{(1)}, \theta_d)$, we have

$$\det\left(\mathbf{F}(\mathcal{S}^{(1)}, \theta_d)\right) = \left(\prod_{n=1}^{\tilde{n}_d} \lambda_n\left(\mathbf{F}(\mathcal{S}^{(1)}, \theta_d)\right) \right) \varepsilon^{N-\tilde{n}_d}. \tag{C.10}$$

Substituting (C.10) into (C.9), we then can obtain (30).

Supplementary material

Supplementary material associated with this article can be found in the online version at 10.1016/j.sigpro.2026.110553.

References

[1] A. Srivastava, P.K. Mishra, A survey on WSN issues with its heuristics and meta-heuristics solutions, Wireless Pers. Commun. 121 (1) (2021) 745–814.
[2] S. Xu, L. Wu, K. Doğançay, M.R. Bhavani Shankar, K.C. Ho, X. Wu, Optimal sensor placement for target localization in IoT systems: a Cramér-Rao bound perspective, IEEE Internet Things Mag. 8 (5) (2025) 120–126.
[3] S.P. Chepuri, G. Leus, Sparsity-promoting sensor selection for non-linear measurement models, IEEE Trans. Signal Process. 63 (3) (2015) 684–698.
[4] Y. Zhao, Z. Li, B. Hao, J. Shi, Sensor selection for TDOA-based localization in wireless sensor networks with non-line-of-sight condition, IEEE Trans. Veh. Technol. 68 (10) (2019) 9935–9950.
[5] W. Ma, X. Dang, Q. Cheng, H. Zhu, Sensor selection based on sparse sensing in the presence of sensor position error, IEEE Trans. Aerosp. Electron. Syst. 59 (6) (2023) 8915–8930.
[6] C. Di, F. Li, S. Li, Sensor deployment for wireless sensor networks: a conjugate learning automata-based energy-efficient approach, IEEE Wireless Commun. 27 (5) (2020) 80–87.
[7] E. Tohidi, M. Coutino, S.P. Chepuri, H. Behroozi, M.M. Nayebi, G. Leus, Sparse antenna and pulse placement for colocated MIMO radar, IEEE Trans. Signal Process. 67 (3) (2018) 579–593.
[8] S. Koyama, G. Chardon, L. Daudet, Optimizing source and sensor placement for sound field control: an overview, IEEE/ACM Trans. Audio Speech Lang. Process. 28 (2020) 696–714.
[9] X. Yu, K. Ergun, L. Cherkasova, T.S. Rosing, Optimizing sensor deployment and maintenance costs for large-scale environmental monitoring, IEEE Trans. Comput.-Aided Design Integr. Circuits Syst. 39 (11) (2020) 3918–3930.
[10] M. Younis, K. Akkaya, Strategies and techniques for node placement in wireless sensor networks: a survey, Ad Hoc Netw. 6 (4) (2008) 621–655.
[11] R. Asorey-Cacheda, A.J. Garcia-Sanchez, F. García-Sánchez, J. García-Haro, A survey on non-linear optimization problems in wireless sensor networks, J. Netw. Comput. Appl. 82 (2017) 1–20.
[12] A. Krause, D. Golovin, Submodular Function Maximization, Cambridge University Press, 2014, pp. 71–104.
[13] E. Tohidi, R. Amiri, M. Coutino, D. Gesbert, G. Leus, A. Karbasi, Submodularity in action: from machine learning to signal processing applications, IEEE Signal Process. Mag. 37 (5) (2020) 120–133.
[14] S. Joshi, S. Boyd, Sensor selection via convex optimization, IEEE Trans. Signal Process. 57 (2) (2009) 451–462.
[15] C. Rusu, J. Thompson, N.M. Robertson, Sensor scheduling with time, energy, and communication constraints, IEEE Trans. Signal Process. 66 (2) (2017) 528–539.
[16] V. Kekatos, G.B. Giannakis, B. Wollenberg, Optimal placement of phasor measurement units via convex relaxation, IEEE Trans. Power Syst. 27 (3) (2012) 1521–1530.
[17] M. Shamaiah, S. Banerjee, H. Vikalo, Greedy sensor selection: leveraging submodularity, in: Proc. IEEE Conf. Decis. Control, Atlanta, GA, USA, 2010, pp. 2572–2577.
[18] J. Ranieri, A. Chebira, M. Vetterli, Near-optimal sensor placement for linear inverse problems, IEEE Trans. Signal Process. 62 (5) (2014) 1135–1146.
[19] L. Liu, C. Hua, J. Xu, G. Leus, Y. Wang, Greedy sensor selection: leveraging submodularity based on volume ratio of information ellipsoid, IEEE Trans. Signal Process. 71 (2023) 2391–2406.
[20] C. Jiang, Y.C. Soh, H. Li, Sensor placement by maximal projection on minimum eigenspace for linear inverse problems, IEEE Trans. Signal Process. 64 (21) (2016) 5595–5610.
[21] C. Jiang, Z. Chen, R. Su, Y.C. Soh, Group greedy method for sensor placement, IEEE Trans. Signal Process. 67 (9) (2019) 2249–2262.
[22] Y. Lin, Y. Li, D. Song, W. Wang, Sensor selection for TOA-Based multi-target localization with non-shared sensors, IEEE Trans. Aerosp. Electron. Syst. 61 (2) (2025) 2383–2396.
[23] G.L. Nemhauser, L.A. Wolsey, M.L. Fisher, An analysis of approximations for maximizing submodular set functions i, Math. Prog. 14 (1978) 265–294.
[24] S. Rao, S.P. Chepuri, G. Leus, Greedy sensor selection for non-linear models, in: Proc. IEEE CAMSAP, Cancun, Mexico, 2015, pp. 241–244.
[25] A. Krause, H.B. McMahan, C. Guestrin, A. Gupta, Robust submodular observation selection, J. Mach. Learn. Res. 9 (12) (2008) 2761–2801.
[26] Q. Hou, A. Clark, Robust maximization of correlated submodular functions, in: Proc. IEEE Conf. Decis. Contr., Nice, France, 2019, pp. 7177–7183.
[27] S. Xu, K. Doğançay, Optimal sensor placement for 3-D angle-of-arrival target localization, IEEE Trans. Aerosp. Electron. Syst. 53 (3) (2017) 1196–1211.
[28] N. Sahu, L. Wu, P. Babu, M.R. Bhavani Shankar, B. Ottersten, Optimal sensor placement for source localization: a unified ADMM approach, IEEE Trans. Veh. Technol. 71 (4) (2022) 4359–4372.
[29] Y. Yang, J. Zheng, H. Liu, K.C. Ho, Z. Yang, S. Gao, Optimal sensor placement and velocity configuration for TDOA-FDOA localization and tracking of a moving source, IEEE Trans. Aerosp. Electron. Syst. 60 (6) (2024) 8255–8272.
[30] Y. Yang, J. Zheng, H. Liu, K.C. Ho, Y. Chen, Z. Yang, Optimal sensor placement for source tracking under synchronization offsets and sensor location errors with distance-dependent noises, Signal Process. 193 (2022) 108399.
[31] L.A. Wolsey, An analysis of the greedy algorithm for the submodular set covering problem, Combinatorica 2 (4) (1982) 385–393.
[32] S.M. Kay, Fundamentals of Statistical Signal Processing: Estimation theory, Prentice-Hall, Inc., 1993.
[33] L.F.O. Chamon, G.J. Pappas, A. Ribeiro, The mean square error in Kalman filtering sensor selection is approximately supermodular, in: Proc. IEEE Conf. Decis. Control, Melbourne, VIC, Australia, 2017, pp. 343–350.
[34] A. Olshevsky, On (non) supermodularity of average control energy, IEEE Trans. Control Netw. Syst. 5 (3) (2017) 1177–1181.
[35] T. Fujito, Approximation algorithms for submodular set cover with applications, IEICE Trans. Inf. Syst. 83 (3) (2000) 480–487.
[36] B. Mirzasoleiman, A. Badanidiyuru, A. Karbasi, J. Vondrák, A. Krause, Lazier than lazy greedy, in: Proc. AAAI, Austin, Texas, 2015, pp. 1812–1818.
[37] A. Badanidiyuru, B. Mirzasoleiman, A. Karbasi, A. Krause, Streaming submodular maximization: massive data summarization on the fly, in: Proc. ACM KDD, New York, NY, USA, 2014, pp. 671–680.
[38] C. Chekuri, K. Quanrud, Parallelizing greedy for submodular set function maximization in matroids and beyond, in: Proc. ACM STOC, New York, NY, USA, 2019, pp. 78–89.
[39] H. Ramezani, H. Jamali-Rad, G. Leus, Target localization and tracking for an isogradient sound speed profile, IEEE Trans. Signal Process. 61 (6) (2012) 1434–1446.
[40] M. Stojanovic, J. Preisig, Underwater acoustic communication channels: propagation models and statistical characterization, IEEE Commun. Mag. 47 (1) (2009) 84–89.