2019 IEEE 58th Conference on Decision and Control (CDC)
Palais des Congrès et des Expositions Nice Acropolis
Nice, France, December 11-13, 2019

Sensor Selection for Hypothesis Testing: Complexity and Greedy Algorithms

Lintao Ye and Shreyas Sundaram

Abstract— **In this paper, we consider sensor selection for (binary) hypothesis testing. Given a pair of hypotheses and a set of candidate sensors to measure (detect) the signals generated under the hypotheses, we aim to select a subset of the sensors (under a budget constraint) that yields the optimal signal detection performance. In particular, we consider the Neyman-Pearson detector based on measurements of the chosen sensors. The goal is to minimize (resp., maximize) the miss probability (resp., detection probability) of the Neyman-Pearson detector, while satisfying the budget constraint. We first show that the sensor selection for the Neyman-Pearson detector problem is NP-hard. We then characterize the performance of greedy algorithms for solving the sensor selection problem when we consider a surrogate to the miss probability as an optimization metric, which is based on the Kullback-Leibler distance. By leveraging the notion of submodularity ratio, we provide a bound on the performance of greedy algorithms.**

I. INTRODUCTION

Sensor selection problems arise in many different fields, including control system design and environment monitoring. In the general formulation of such problems, the system designers are faced with the situation where only a subset of candidate sensors can be installed on the system (under a budget constraint) to perform sensing tasks in order to achieve certain performance objectives.

While researchers have studied the sensor (or actuator) selection problem in control system design (e.g., [1]–[8]), and have provided complexity characterizations and algorithms, relatively less work has been done for the sensor selection problem in signal detection and hypothesis testing (e.g., [9]). In this paper, we consider the sensor selection problem for hypothesis testing based on the Neyman-Pearson detector [10]. We study the problem of choosing a subset of sensors (under a given budget constraint) to minimize the miss probability of the Neyman-Pearson detector such that the false-alarm probability is within a prescribed range. We further consider a surrogate (e.g., [9], [11], [12]), based on the Kullback-Leibler (KL) distance, to the miss probability of the Neyman-Pearson detector as an optimization metric. We summarize some related work as follows.

In [13], it was shown that the sensor selection problem for Kalman filtering is NP-hard and cannot be approximated within any constant factor (in polynomial time). Moreover, [13] showed that greedy algorithms can perform arbitrarily poorly. In contrast, in this paper we study the problem of Sensor Selection for the Neyman-Pearson detector (SSNP). We show that this problem is also NP-hard. We further provide performance bounds on greedy algorithms that depend on parameters of the SSNP instance, when we consider a certain surrogate to the objective function in SSNP.

This research was supported by NSF grant CMMI-1635014. Lintao Ye and Shreyas Sundaram are with the School of Electrical and Computer Engineering at Purdue University. Email: {ye159,sundara2}@purdue.edu.

The authors of [11] show that SSNP is NP-hard after they replace the miss probability with the surrogate, based on the KL distance, as the optimization metric. They further show that the objective function of the problem corresponding to the surrogate is not submodular in general. Here, we show that the original SSNP formulation (with the miss probability as the minimization metric) is NP-hard. When considering the surrogate (based on the KL distance) as the optimization metric, we provide bounds on the performance of greedy algorithms. We achieve this by leveraging the notion of submodularity ratio that is used to characterize how close a nonsubmodular set function is to being submodular.

In [12], the authors also study SSNP using surrogates to the original objective function. When studying SSNP using the KL distance as a surrogate, they first consider a special instance of SSNP and use a certain metric (different from the submodularity ratio) to characterize how close a nonsubmodular set function is to being submodular, which leads to a performance bound on the greedy algorithm. However, for more general instances of SSNP, the authors further consider submodular surrogates to the optimization metric corresponding to the KL distance. In this paper, we use the notion of submodularity ratio to provide a bound on the performance of greedy algorithms for solving different and potentially more useful instances of SSNP when directly using the optimization metric based on the KL distance.

Our contributions to this problem are as follows. First, we show that SSNP is NP-hard when we consider the miss probability of the Neyman-Pearson detector as the optimization metric. We then consider solving SSNP using an optimization criterion based on the KL distance. In such a case, we provide performance bounds on greedy algorithms that depend on the parameters of the SSNP problem.

A. Notation and terminology

The set of integers and real numbers are denoted as $\mathbb{Z}$ and $\mathbb{R}$, respectively. For $x \in \mathbb{R}$, denote $|x|$ as its absolute value. For a set $\mathcal{S}$, denote $|\mathcal{S}|$ as its cardinality. For a vector x of dimension n, denote x_i (or $(x)_i$) as its ith element, and let $\text{supp}(x) \triangleq \{i : x_i \neq 0\}$. Denote $\mathbf{1}_n$ as a column vector of dimension n with all of its elements equal to 1. For a matrix $P \in \mathbb{R}^{n\times n}$, let P^T, $\text{tr}(P)$ and $\det(P)$ be its transpose, trace and determinant, respectively. The eigenvalues of P are ordered with nondecreasing magnitude (i.e., $|\lambda_1(P)| \leq \cdots \leq$

978-1-7281-1398-2/19/$31.00 ©2019 IEEE

$|\lambda_n(P)|$). Denote P_{ij} as the element in the ith row and jth column of P. The set of n by n real positive semi-definite (resp., positive definite) matrices is denoted by $\mathbb{S}^n_+$ (resp., $\mathbb{S}^n_{++}$). The identity matrix with dimension n is denoted as I_n. For a random vector $X \in \mathbb{R}^n$, denote $\mathbb{E}[X] \in \mathbb{R}^n$ and $\text{Cov}(X) = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T] \in \mathbb{R}^{n\times n}$ as its mean vector and covariance, respectively. For two random vectors $X \in \mathbb{R}^{n_1}$ and $Y \in \mathbb{R}^{n_2}$, denote $\Sigma_{XY} = \text{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])^T] \in \mathbb{R}^{n_1\times n_2}$ as the cross-covariance between them.

II. Problem Formulation

We consider the classical binary hypothesis testing problem, where we assume that there are two possible hypotheses or states, denoted as H_0 and H_1, respectively. Denote $\mathcal{X} \triangleq \{1, 2, \dots, n\}$ as the set of all candidate sensors; each sensor is capable of providing a single measurement. Let $\boldsymbol{X} \triangleq \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}^T \in \mathbb{R}^n$ be the vector that collects the measurements for all the sensors in $\mathcal{X}$, where $X_k \in \mathbb{R}$ is the measurement from the kth sensor in $\mathcal{X}$ for all $k \in \mathcal{X}$. The measurements satisfy

$$\begin{aligned} H_0 &: X_k \sim p_k(x|H_0),\ k = 1, 2, \dots, n, \\ H_1 &: X_k \sim p_k(x|H_1),\ k = 1, 2, \dots, n, \end{aligned} \tag{1}$$

where $p_k(x|H_i)$ denotes the probability density function (pdf) of X_k conditioned on the state H_i, for $i = 0, 1$. Denote the pdf of $\boldsymbol{X}$ conditioned on H_i as $p(\boldsymbol{x}|H_i)$, for $i = 0, 1$.

We consider the scenario where we can only select a subset of sensors from $\mathcal{X}$ to deploy under a budget constraint. Specifically, sensor $k \in \mathcal{X}$ has a certain selection cost, denoted as $\omega_k \in \mathbb{R}_{\geq 0}$, for all $k \in \mathcal{X}$. Define $\omega = \begin{bmatrix} \omega_1 & \omega_2 & \cdots & \omega_n \end{bmatrix}^T$ as the sensor cost vector. We are given a total budget, denoted as $\Omega \in \mathbb{R}_{>0}$.

After a set of sensors is selected, we use their measurements to solve the hypothesis testing problem corresponding to Eq. (1). We define an indicator vector $\mu \in \{0,1\}^n$ indicating which sensors are selected, where $\mu_k = 1$ if sensor $k \in \mathcal{X}$ is selected, and $\mu_k = 0$ if otherwise. Given an indicator vector μ with $\text{supp}(\mu) = \{j_1, \dots, j_p\} \subseteq \{1, \dots, n\}$, we define $\boldsymbol{X}(\mu) = \begin{bmatrix} X_{j_1} & \cdots & X_{j_p} \end{bmatrix}^T$ as the vector that contains the measurements from the selected sensors indicated by μ. Denote the pdf of $\boldsymbol{X}(\mu)$ conditioned on H_i as $p(\boldsymbol{x}(\mu)|H_i)$, for $i = 0, 1$. We consider the Neyman-Pearson detector for hypothesis testing, which is based on the log-likelihood ratio of the two hypotheses [10]. The log-likelihood ratio is defined as

$$\log L(\boldsymbol{x}(\mu)) = \log \frac{p(\boldsymbol{x}(\mu)|H_1)}{p(\boldsymbol{x}(\mu)|H_0)}. \tag{2}$$

The Neyman-Pearson detector minimizes the miss probability (also known as Type II error) $P_m \triangleq P(H_0|H_1)$, where $P(H_0|H_1)$ is the conditional probability of deciding H_0 given that H_1 is true, such that the false-alarm probability (also known as Type I error) $P_f \triangleq P(H_1|H_0)$ is within a prescribed range, where $P(H_1|H_0)$ is the conditional probability of deciding H_1 given that H_0 is true. Moreover, the detection probability is defined as $P_D \triangleq 1 - P_m$. Given an indicator vector μ, denote $P_m(\mu)$, $P_D(\mu)$ and $P_f(\mu)$ as the miss probability, the detection probability, and the false-alarm probability obtained from the measurements of the sensors indicated by μ, respectively. For a given false-alarm rate $\alpha \in \mathbb{R}_{\geq 0}$ and an indicator vector μ, the Neyman-Pearson detector is of the form

$$\log L(\boldsymbol{x}(\mu)) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma(\mu), \tag{3}$$

where $\log L(\boldsymbol{x}(\mu))$ is as defined in Eq. (2) and $\gamma(\mu)$ is the threshold chosen such that $P_f(\mu) = \alpha$. The Sensor Selection for the Neyman-Pearson detector (SSNP) problem is given as follows.

Problem 1: (SSNP) Consider two possible states H_0 and H_1, a sensor measurement vector $\boldsymbol{X} \in \mathbb{R}^n$ where each element satisfies Eq. (1), a cost vector $\omega \in \mathbb{R}^n_{\geq 0}$, a budget $\Omega \in \mathbb{R}_{>0}$ and a prescribed false-alarm rate $\alpha \in \mathbb{R}_{\geq 0}$. The SSNP problem is to find a sensor selection μ that solves

$$\begin{aligned} &\min_{\mu \in \{0,1\}^n} \ P_m(\mu) \\ &s.t.\ \omega^T \mu \leq \Omega,\ P_f(\mu) \leq \alpha. \end{aligned}$$

□

III. Complexity Of SSNP

In this section, we will show that SSNP is NP-hard. To do this, we introduce the sensor selection problem for linear regression (e.g., [14]). Specifically, consider a predictor (random) variable $Z \in \mathbb{R}$ and observation (random) variables $Y_1, \dots, Y_m$. Denote $C \triangleq \text{Cov}(\boldsymbol{Y}) \in \mathbb{S}^m_{++}$, where $\boldsymbol{Y} \triangleq \begin{bmatrix} Y_1 & \cdots & Y_m \end{bmatrix}^T \in \mathbb{R}^m$. Define a covariance vector $b \in \mathbb{R}^m$ such that $b_i = \text{Cov}(Z, Y_i)$, $\forall i \in \{1, \dots, m\}$. Given an indicator vector ν with $\text{supp}(\nu) = \{i_1, \dots, i_q\} \subseteq \{1, \dots, m\}$, we denote $b(\nu) \triangleq \begin{bmatrix} b_{i_1} & \cdots & b_{i_q} \end{bmatrix}^T$ and $C(\nu) \in \mathbb{S}^q_{++}$ as the covariance vector and the covariance matrix corresponding to ν, respectively, where $C(\nu)$ is a submatrix of C that contains the rows and columns corresponding to $\text{supp}(\nu)$. Given $f_{LR}(\nu) \triangleq \big(b(\nu)\big)^T \big(C(\nu)\big)^{-1} b(\nu)$, the Sensor Selection for Linear Regression (SSLR) problem is defined as follows.

Problem 2: (SSLR) Consider a predictor variable Z and observation variables $Y_1, \dots, Y_m$ with the corresponding covariance matrix $C \in \mathbb{S}^m_{++}$ and covariance vector $b \in \mathbb{R}^m$, and an integer $s \in \mathbb{Z}_{>0}$. The SSLR problem is to find a sensor selection ν that solves

$$\begin{aligned} &\max_{\nu \in \{0,1\}^m} f_{LR}(\nu) \\ &s.t.\ |\text{supp}(\nu)| \leq s. \end{aligned}$$

□

We will use the following result (e.g., [15], [14] and [11]).

Lemma 1: SSLR is NP-hard even when $b = \mathbf{1}_m$. □

We will relate SSNP to SSLR and prove the following.

Theorem 1: The SSNP problem is NP-hard. □

Proof: We show the result by giving a reduction from SSLR to SSNP. Consider any instance of SSLR with the predictor variable Z, observation variables $Y_1, \dots, Y_m$ and the positive integer s, where the covariance matrix is $C \in$

$\mathbb{S}^m_{++}$ and the covariance vector is $b = \mathbf{1}_m$. We construct an instance of SSNP as follows. The measurement vector $\boldsymbol{X}$ is Gaussian distributed conditioned on H_i for $i = 0, 1$, i.e.,

$$H_0 : \boldsymbol{X} \sim \mathcal{N}(\theta_0, C),$$
$$H_1 : \boldsymbol{X} \sim \mathcal{N}(\theta_1, C),$$

where $\mathcal{N}(\theta_i, C)$ denotes the pdf of a multivariate Gaussian with mean $\theta_i \in \mathbb{R}^m$ and covariance $C \in \mathbb{S}^m_{++}$. Moreover, we set $\theta_0 = \mathbf{0}$ and $\theta_1 = b = \mathbf{1}_m$. We set the cost vector as $\omega = \mathbf{1}_m$ and the budget as $\Omega = s$. The required false-alarm rate for the Neyman-Pearson detector is set as $\alpha = \frac{1}{2}$.

Considering any sensor selection $\mu \in \{0,1\}^m$ with its support denoted as $\text{supp}(\mu) = \{j_1, \ldots, j_p\}$, where $p \leq s$ and $\{j_1, \ldots, j_p\} \subseteq \{1, \ldots, m\}$, we define $\theta_i(\mu) \triangleq \begin{bmatrix}(\theta_i)_{j_1} & \cdots & (\theta_i)_{j_p}\end{bmatrix}^T$, for $i = 0, 1$. We have from Eq. (2):

$$\begin{aligned}\log L(\boldsymbol{x}(\mu)) &= \big(\theta_1(\mu)\big)^T \big(C(\mu)\big)^{-1} \boldsymbol{x}(\mu) \\ &\quad - \frac{1}{2}\big(\theta_1(\mu)\big)^T \big(C(\mu)\big)^{-1} \theta_1(\mu).\end{aligned} \tag{4}$$

Let $T(\mu) \triangleq \big(\theta_1(\mu)\big)^T \big(C(\mu)\big)^{-1} \boldsymbol{x}(\mu) \in \mathbb{R}$, where the pdf of $T(\mu)$ conditioned on H_i, for $i = 0, 1$, is given as

$$H_0 : T(\mu) \sim \mathcal{N}(0, \sigma(\mu)),$$
$$H_1 : T(\mu) \sim \mathcal{N}(\sigma(\mu), \sigma(\mu)),$$

where $\sigma(\mu) \triangleq \big(\theta_1(\mu)\big)^T \big(C(\mu)\big)^{-1} \theta_1(\mu) > 0$ for all $\mu \neq \mathbf{0}$, since $\big(C(\nu)\big)^{-1}$ is positive definite and $\theta_1 = \mathbf{1}_m$. We have from (3) and (4) that the Neyman-Pearson detector is of the form

$$T(\mu) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma'(\mu), \tag{5}$$

where $\gamma'(\mu) \triangleq \gamma(\mu) + \frac{1}{2}\sigma(\mu)$. We then know from [10] (Case III.B.2) that $\gamma'(\mu)$ satisfies

$$\gamma'(\mu) = \sqrt{\sigma(\mu)}\Phi^{-1}(1-\alpha) = \sqrt{\sigma(\mu)}\Phi^{-1}(\frac{1}{2}) = 0,$$

where $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard normal distribution, and $\Phi^{-1}(\cdot)$ is the inverse of $\Phi(\cdot)$. The corresponding detection probability is given by

$$\begin{aligned}P_D(\mu) &= P(T(\mu) > \gamma'(\mu) | H_1) \\ &= 1 - \Phi\Big(\frac{\gamma'(\mu) - \sigma(\mu)}{\sqrt{\sigma(\mu)}}\Big) = 1 - \Phi\big(-\sqrt{\sigma(\mu)}\big) \\ &= \Phi\Big(\sqrt{\big(\theta_1(\mu)\big)^T \big(C(\mu)\big)^{-1} \theta_1(\mu)}\Big),\end{aligned} \tag{6}$$

where $P(T(\mu) > \gamma'(\mu) | H_1)$ is the conditional probability of $T(\mu) > \gamma'(\mu)$ given that H_1 is true. Noting that $\Phi(x)$ is monotonically nondecreasing on $x \in \mathbb{R}$, Eq. (6) then yields that in order to maximize $P_D(\mu)$ over sensor selections μ that satisfy the budget constraint, we have to maximize $\big(\theta_1(\mu)\big)^T \big(C(\mu)\big)^{-1} \theta_1(\mu)$. By our construction of the SSNP instance, it follows that a sensor selection μ for SSLR is optimal if and only if μ is optimal for the corresponding SSNP instance that we construct. Since we know from Lemma 1 that SSLR is NP-hard, the SSNP problem is also NP-hard, which completes the proof of the theorem. ■

Remark 1: Theorem 1 shows that SSNP is NP-hard even when the measurement vector is Gaussian distributed and all the sensors have the same selection cost. Our direct proof of complexity complements the existing result in [11], which optimized a surrogate (which we shall discuss in the next section) to the objective function as defined in Eq. (1), and then showed that the modified SSNP is NP-hard. □

IV. Greedy algorithms

We now turn to a modified version of SSNP, and analyze the performance of greedy algorithms for the modified problem. Specifically, we use the KL distance between two distributions as an alternate metric in the SSNP problem, due to the fact that P_m and P_f may not yield known closed-form expressions in general. We refer interested readers to, e.g., [11], [9], [16] and [10], for more detailed justification and explanation of using this metric as a surrogate for hypothesis testing problems in the Neyman-Pearson setting. Given the conditional pdfs of $\boldsymbol{X}(\mu)$ on H_0 and H_1, denoted as $p(\boldsymbol{x}(\mu)|H_0)$ and $p(\boldsymbol{x}(\mu)|H_1)$, respectively, we denote the KL distance between these two distributions as $f_{KL}\big(p(\boldsymbol{x}(\mu)|H_1)\big|\big|p(\boldsymbol{x}(\mu)|H_0)\big)$ ($f_{KL}(\mu)$ for simplicity).[1] Furthermore, we focus on cases, as studied in [11] and [12], when the measurement vector $\boldsymbol{X} \in \mathbb{R}^n$ is Gaussian distributed given both hypotheses H_0 and H_1, i.e.,

$$\begin{aligned}H_0 : \boldsymbol{X} \sim \mathcal{N}(\theta_0, \Sigma_0), \\ H_1 : \boldsymbol{X} \sim \mathcal{N}(\theta_1, \Sigma_1),\end{aligned} \tag{7}$$

where $\theta_0, \theta_1 \in \mathbb{R}^n$ and $\Sigma_0, \Sigma_1 \in \mathbb{S}^n_{++}$. Given a sensor selection $\mu \in \{0,1\}^n$ with its support denoted as $\text{supp}(\mu) = \{j_1, \ldots, j_p\}$, where $\{j_1, \ldots, j_p\} \subseteq \{1, \ldots, n\}$, we define $\theta_i(\mu) = \begin{bmatrix}(\theta_i)_{j_1} & \cdots & (\theta_i)_{j_p}\end{bmatrix}^T$ and $\Sigma_i(\mu)$ as the submatrix of Σ_i that contains the rows and columns corresponding to $\text{supp}(\mu)$, for $i = 0, 1$. The KL distance $f_{KL}(\mu)$ is then given by (e.g., [11])

$$\begin{aligned}f_{KL}(\mu) = \frac{1}{2}\Big(&\text{tr}\big(\tilde{\Sigma}_0^{-1}\tilde{\Sigma}_1\big) + \big(\tilde{\theta}_1 - \tilde{\theta}_0\big)^T \tilde{\Sigma}_0^{-1}\big(\tilde{\theta}_1 - \tilde{\theta}_0\big) \\ &+ \log\big(\frac{\det(\tilde{\Sigma}_0)}{\det(\tilde{\Sigma}_1)}\big) - |\text{supp}(\mu)|\Big),\end{aligned} \tag{8}$$

where $\tilde{\theta}_i \triangleq \theta_i(\mu)$ and $\tilde{\Sigma}_i \triangleq \Sigma_i(\mu)$, for $i = 0, 1$.

Remark 2: Noting from Eq. (8) that $f_{KL}(\mu)$ depends on $\theta_0(\mu)$ and $\theta_1(\mu)$ through $\theta_1(\mu) - \theta_0(\mu)$ only, we can assume without loss of generality that $\theta_0 = \mathbf{0}$ in what follows. □

The Modified Sensor Selection for the Neyman-Pearson detector (MSSNP) problem, as studied in [11] and [12], is then given as follows.

Problem 3: (MSSNP) Consider two possible states H_0 and H_1, a sensor measurement vector $\boldsymbol{X} \in \mathbb{R}^n$ satisfying Eq. (7), and a budget $\Omega \in \mathbb{Z}_{\geq 1}$. The MSSNP problem is to find a sensor selection μ that solves

$$\begin{aligned}&\max_{\mu \in \{0,1\}^n} f_{KL}(\mu) \\ &s.t.\ |\text{supp}(\mu)| \leq \Omega,\end{aligned} \tag{9}$$

[1] Noting that the KL distance is always nonnegative [16], we have that $f_{KL}(\mu) \geq 0$ for all μ.

where $f_{KL}(\mu)$ is given in Eq. (8) if $\mu \neq \mathbf{0}$ and $f_{KL}(\mathbf{0}) = 0$. □

Since it has been shown in [11] that $f_{KL}(\cdot)$ is not a submodular function with respect to the subsets of candidate sensors in general, we first introduce the following concept from [14] and [17], which characterizes how close a nonsubmodular function $f(\cdot)$ is to being submodular, where $f(\cdot)$ is nonnegative and $f(\emptyset) = 0$.

Definition 1: (Submodularity ratio) Given a set $\mathcal{V}$, the submodularity ratio of a nonnegative set function $f : 2^{\mathcal{V}} \to \mathbb{R}_{\geq 0}$ is the largest $\gamma \in \mathbb{R}$ that satisfies

$$\sum_{a \in A \setminus B} \big(f(\{a\} \cup B) - f(B)\big) \geq \gamma\big(f(A \cup B) - f(B)\big), \quad (10)$$

for all $A, B \subseteq \mathcal{V}$. □

Given such a set function as described in Definition 1 and a positive integer K, a greedy algorithm for the problem

$$\max_{A \subseteq \mathcal{V}, |A| \leq K} f(A) \quad (11)$$

starts with a set $\mathcal{S} = \emptyset$ and iteratively adds an element $i \in \mathcal{V} \setminus \mathcal{S}$ to $\mathcal{S}$ such that $f(\{i\} \cup \mathcal{S})$ is maximized. The algorithm then returns $\mathcal{S}$ after K iterations.

We will use the following result from [14] and [17].

Lemma 2: If $f(\cdot)$ is a nonnegative and (monotonically) nondecreasing set function with submodularity ratio $\gamma \geq 0$, a greedy algorithm for solving problem (11) yields

$$f(\mathcal{S}) \geq (1 - e^{-\gamma}) f(A^*), \quad (12)$$

where $\mathcal{S}$ is the solution returned by the greedy algorithm and A^* is the optimal solution of problem (11). □

Remark 3: For a nonnegative and nondecreasing function $f(\cdot)$ with submodularity ratio γ, we have $\gamma \in [0, 1]$ [17]. Moreover, $f(\cdot)$ is submodular if and only if $\gamma = 1$ [17]. We then know from inequality (12) that to characterize the performance of the greedy algorithm for solving problem (11), we can give a lower bound on γ. □

Remark 4: If $\Sigma_0 = \Sigma_1$, the expression for $f_{KL}(\mu)$ in Eq. (8) simplifies to $f_{KL}(\mu) = \frac{1}{2}\big(\theta_1(\mu)\big)^T \big(\Sigma_0(\mu)\big)^{-1} \theta_1(\mu)$. If we let $b = \theta_1$, $C = \Sigma_1$ and $s = \Omega$ in the SSLR problem (Problem 2), we can solve such an instance of MSSNP by solving the corresponding instance of SSLR. Thus, the analysis in [14] can be applied when $\Sigma_0 = \Sigma_1$. □

Our goal in this section is to give a (lower) bound on the submodularity ratio of $f_{KL}(\cdot)$, when Σ_0 and Σ_1 are not necessarily equal. In the proceedings, we write $f_{KL}(\mu)$ as $f_{KL}(A)$ if $A = \text{supp}(\mu) \subseteq \mathcal{X}$, where $\mathcal{X}$ is the set of all candidate sensors. Similarly, we use $\boldsymbol{X}(A)$, and $\Sigma_i(A)$ and $\theta_i(A)$ for $i = 0, 1$. We will use the following result and give a sketched proof in the Appendix.

Lemma 3: Consider $\boldsymbol{X} = \begin{bmatrix} X_1 & \dots & X_n \end{bmatrix}^T \in \mathbb{R}^n$ that collects the measurements of all candidate sensors. Suppose that $\boldsymbol{X}$ has mean $\theta \in \mathbb{R}^n$ and covariance $\Sigma \in \mathbb{S}^n_{++}$. For all $p < n$ ($p \in \mathbb{Z}_{>0}$), consider the sensor selection $\mu \in \{0, 1\}^n$ with $\text{supp}(\mu) = \{1, \dots, p\}$. Denote $\boldsymbol{X}(\mu^c) \triangleq \begin{bmatrix} X_{p+1} & \cdots & X_n \end{bmatrix}^T$ with the corresponding covariance denoted as $\Sigma(\mu^c)$. Partitioning Σ as

$$\Sigma = \begin{bmatrix} \Sigma(\mu) & \Sigma_{\boldsymbol{X}(\mu)\boldsymbol{X}(\mu^c)} \\ \Sigma^T_{\boldsymbol{X}(\mu)\boldsymbol{X}(\mu^c)} & \Sigma(\mu^c) \end{bmatrix},$$

the following holds

$$\lambda_1(\Sigma) \leq \lambda_1\big(\Sigma(\mu) - \Sigma_{\boldsymbol{X}(\mu)\boldsymbol{X}(\mu^c)} (\Sigma(\mu^c))^{-1} \Sigma^T_{\boldsymbol{X}(\mu)\boldsymbol{X}(\mu^c)}\big).$$

□

Assumption 1: Let $\Sigma_0 = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$, where $\sigma_i \in \mathbb{R}_{>0}$ for all $i \in \{1, \dots, n\}$. Let Σ_1 be any matrix in $\mathbb{S}^n_{++}$.

Lemma 4: The set function $f_{KL} : 2^{\mathcal{X}} \to \mathbb{R}$ under Assumption 1 has the following properties.

(a) $f_{KL}(\cdot)$ is nonnegative and nondecreasing, i.e., for all $A \subseteq B \subseteq \mathcal{X}$, $0 \leq f_{KL}(A) \leq f_{KL}(B)$ holds.

(b) The submodularity ratio of $f_{KL}(\cdot)$, denoted as γ_{KL}, satisfies

$$\gamma_{KL} \geq \frac{\underline{r}_1^2 + r - 1}{\bar{r}_1^2 + r_2 - 1 + \log \frac{\lambda_n(\Sigma_0)}{\lambda_1(\Sigma_1)}}, \quad (13)$$

where

$$r \triangleq \min_{1 \leq i \leq n} \Big(\frac{(\Sigma_1)_{ii}}{\sigma_i^2} - \log \frac{(\Sigma_1)_{ii}}{\sigma_i^2}\Big), r_2 \triangleq \max_{1 \leq i \leq n} \frac{(\Sigma_1)_{ii}}{\sigma_i^2},$$

$$\underline{r}_1 \triangleq \min_{1 \leq i \leq n} \Big|\frac{(\theta_1)_i}{\sigma_i}\Big|, \text{ and } \bar{r}_1 \triangleq \max_{1 \leq i \leq n} \Big|\frac{(\theta_1)_i}{\sigma_i}\Big|.$$

□

Proof of (a): Note that $f_{KL}(B) \geq f_{KL}(\emptyset) = 0, \forall B \subseteq \mathcal{X}$. To prove that $f_{KL}(\cdot)$ is nondecreasing, it is then sufficient to show that for all $B \subseteq \mathcal{X}$ ($B \neq \emptyset$ and $B \neq \mathcal{X}$) and for all $a \in \mathcal{X} \setminus B$, $f_{KL}(\{a\} \cup B) - f_{KL}(B) \geq 0$ holds. Recall from Remark 2 that we set, without loss of generality, $\theta_0 = \mathbf{0}$. Denote $\bar{B} \triangleq \{a\} \cup B$. Denote $\Sigma_{Ba} \triangleq \text{Cov}(\boldsymbol{X}(B), \boldsymbol{X}(a))$ as the cross-covariance between $\boldsymbol{X}(B)$ and $\boldsymbol{X}(a)$ under hypothesis H_1, where we note that $\Sigma_{Ba} \in \mathbb{R}^{|B|}$. Note that $\Sigma_i(\bar{B}) \in \mathbb{S}^{|B|+1}_{++}$, and $\Sigma_i(B) \in \mathbb{S}^{|B|}_{++}$, for $i = 0, 1$. We have from Eq. (8) the following:

$$\begin{aligned} &2\big(f_{KL}(\{a\} \cup B) - f_{KL}(B)\big) \\ &= \frac{\Sigma_1(a)}{\Sigma_0(a)} + \frac{\big(\theta_1(a)\big)^2}{\Sigma_0(a)} + \log \frac{\det\big(\Sigma_1(B)\big)}{\det\big(\Sigma_1(\bar{B})\big)} + \log \Sigma_0(a) - 1 \quad (14) \\ &= \frac{\big(\theta_1(a)\big)^2}{\Sigma_0(a)} + \frac{\Sigma_1(a)}{\Sigma_0(a)} + \log \Sigma_0(a) \\ &\quad - \log\big(\Sigma_1(a) - \Sigma^T_{Ba}\big(\Sigma_1(B)\big)^{-1}\Sigma_{Ba}\big) - 1 \quad (15) \\ &\geq \frac{\big(\theta_1(a)\big)^2}{\Sigma_0(a)} + \frac{\Sigma_1(a)}{\Sigma_0(a)} + \log \Sigma_0(a) - \log \Sigma_1(a) - 1 \quad (16) \\ &= \frac{\big(\theta_1(a)\big)^2}{\Sigma_0(a)} + \frac{\Sigma_1(a)}{\Sigma_0(a)} - \log \frac{\Sigma_1(a)}{\Sigma_0(a)} - 1 \geq 0, \quad (17) \end{aligned}$$

where Eq. (14) follows from the assumption that Σ_0 is diagonal and invertible. To obtain Eq. (15) we use the following identity [18]:

$$\begin{aligned} \det\big(\Sigma_1(\bar{B})\big) &= \det \begin{bmatrix} \Sigma_1(B) & \Sigma_{Ba} \\ \Sigma^T_{Ba} & \Sigma_1(a) \end{bmatrix} \\ &= \det\big(\Sigma_1(B)\big) \det\big(\Sigma_1(a) - \Sigma^T_{Ba}\big(\Sigma_1(B)\big)^{-1}\Sigma_{Ba}\big), \quad (18) \end{aligned}$$

where $\big(\Sigma_1(a) - \Sigma_{Ba}^T\big(\Sigma_1(B)\big)^{-1}\Sigma_{Ba}\big)$ is a scalar. To obtain inequality (16), we use the fact $0 < \Sigma_1(a) - \Sigma_{Ba}^T\big(\Sigma_1(B)\big)^{-1}\Sigma_{Ba}^T \le \Sigma_1(a)$. Noting that the function $h(x) \triangleq x - \log x$ achieves its unique minimum on $x > 0$ at $x = 1$ with $h(1) = 1$, the inequality in (17) then follows.

Proof of (b): First, suppose that $A \setminus B \neq \emptyset$ and $B \neq \emptyset$. We begin by providing a lower bound on $\sum_{a\in A\setminus B}\big(f_{KL}(\{a\}\cup B) - f_{KL}(B)\big)$. We follow inequality (16) and obtain:

$$\begin{aligned} &2\sum_{a\in A\setminus B}\big(f_{KL}(\{a\}\cup B) - f_{KL}(B)\big)\\ &\ge \sum_{a\in A\setminus B}\Big(\frac{\big(\theta_1(a)\big)^2}{\Sigma_0(a)} + \frac{\Sigma_1(a)}{\Sigma_0(a)} - \log\frac{\Sigma_1(a)}{\Sigma_0(a)} - 1\Big)\\ &\ge |A\setminus B|(\underline{r}_1^2 + r - 1), \end{aligned} \tag{19}$$

where inequality (19) follows from the definitions of $\underline{r}_1$ and r. Note that $r - 1 \ge 0$.

We then give an upper bound on $f_{KL}(A\cup B) - f_{KL}(B)$. Denote $\tilde{B} \triangleq A \cup B$ and $\tilde{A} \triangleq A \setminus B$. Further denote $\Sigma_{B\tilde{A}} \triangleq \text{Cov}(\boldsymbol{X}(B), \boldsymbol{X}(\tilde{A}))$ as the cross-covariance between $\boldsymbol{X}(B)$ and $\boldsymbol{X}(\tilde{A})$ under hypothesis H_1, where $\Sigma_{B\tilde{A}} \in \mathbb{R}^{|B|\times|A\setminus B|}$. Note that $\Sigma_i(\tilde{B}) \in \mathbb{S}_{++}^{|A\cup B|}$ and $\Sigma_i(\tilde{A}) \in \mathbb{S}_{++}^{|A\setminus B|}$, for $i = 0, 1$. We have from Eq. (8) the following:

$$\begin{aligned} &2\big(f_{KL}(\tilde{B}) - f_{KL}(B)\big)\\ &= \Big\{\sum_{a\in A\setminus B}\Big(\frac{\big(\theta_1(a)\big)^2}{\Sigma_0(a)} + \frac{\Sigma_1(a)}{\Sigma_0(a)} - 1\Big)\Big\}\\ &\qquad + \log\frac{\big(\prod_{a\in A\setminus B}\Sigma_0(a)\big)\det\big(\Sigma_1(B)\big)}{\det\big(\Sigma_1(\tilde{B})\big)} \quad (20)\\ &= \Big\{\sum_{a\in A\setminus B}\Big(\frac{\big(\theta_1(a)\big)^2}{\Sigma_0(a)} + \frac{\Sigma_1(a)}{\Sigma_0(a)} - 1\Big)\Big\}\\ &\qquad + \log\frac{\prod_{a\in A\setminus B}\Sigma_0(a)}{\det\big(\Sigma_1(\tilde{A}) - \Sigma_{B\tilde{A}}^T\big(\Sigma_1(B)\big)^{-1}\Sigma_{B\tilde{A}}\big)} \quad (21)\\ &\le \Big\{\sum_{a\in A\setminus B}\Big(\frac{\big(\theta_1(a)\big)^2}{\Sigma_0(a)} + \frac{\Sigma_1(a)}{\Sigma_0(a)} - 1\Big)\Big\} + \log\frac{\prod_{a\in A\setminus B}\Sigma_0(a)}{\big(\lambda_1(\Sigma_1)\big)^{|A\setminus B|}} \quad (22)\\ &\le |A\setminus B|\big(\bar{r}_1^2 + r_2 - 1 + \log\frac{\lambda_n(\Sigma_0)}{\lambda_1(\Sigma_1)}\big), \quad (23) \end{aligned}$$

where Eq. (20) follows from the assumption that Σ_0 is diagonal and invertible. To obtain Eq. (21), we use the following identity [18]:

$$\begin{aligned} \det\big(\Sigma_1(\tilde{B})\big) &= \det\begin{bmatrix}\Sigma_1(B) & \Sigma_{B\tilde{A}}\\ \Sigma_{B\tilde{A}}^T & \Sigma_1(\tilde{A})\end{bmatrix} \quad (24)\\ &= \det\big(\Sigma_1(B)\big)\det(\Sigma_1(\tilde{A}) - \Sigma_{B\tilde{A}}^T\big(\Sigma_1(B)\big)^{-1}\Sigma_{B\tilde{A}}\big). \end{aligned}$$

To obtain inequality (22), we first note that $\Sigma_{B\tilde{A}} = \Sigma_{\tilde{A}B}^T$, where $\Sigma_{\tilde{A}B} \triangleq \text{Cov}(\boldsymbol{X}(\tilde{A}), \boldsymbol{X}(B))$. This leads to a partition of $\Sigma_1(\tilde{B})$ as

$$\Sigma_1(\tilde{B}) = \begin{bmatrix}\Sigma_1(\tilde{A}) & \Sigma_{\tilde{A}B}\\ \Sigma_{\tilde{A}B}^T & \Sigma_1(B)\end{bmatrix}, \tag{25}$$

where we obtain $\Sigma_1(\tilde{B})$ of the form in Eq. (25) by appropriate permutations of rows and columns of $\Sigma_1(\tilde{B})$ from Eq. (24), which do not change the eigenvalues of $\Sigma_1(\tilde{B})$. It then follows from Lemma 3 that

$$\lambda_1\big(\Sigma_1(\tilde{A}) - \Sigma_{\tilde{A}B}(\Sigma_1(B))^{-1}\Sigma_{\tilde{A}B}^T\big) \ge \lambda_1\big(\Sigma_1(\tilde{B})\big) \ge \lambda_1(\Sigma_1),$$

where the second inequality follows from the Cauchy interlacing theorem for positive definite matrices [18]. Inequality (22) then follows from the fact that $\det(\Sigma_1(\tilde{A})) = \prod_{i=1}^{|\tilde{A}|}\lambda_i(\Sigma_1(\tilde{A}))$. We then obtain inequality (23) from the definitions of $\bar{r}_1$ and r_2. We note that if $\bar{r}_1^2 + r_2 - 1 + \log\frac{\lambda_n(\Sigma_0)}{\lambda_1(\Sigma_1)} = 0$, we have $f_{KL}(A\cup B) - f_{KL}(B) = 0$ for all $A, B \subseteq \mathcal{V}$, where $A\setminus B \neq \emptyset$ and $B \neq \emptyset$. If we set B as a singleton, we then see that the (optimal) solution for MSSNP can be achieved via a sensor selection μ with supp$(\mu) = 1$. Thus, we consider $\bar{r}_1^2 + r_2 - 1 + \log\frac{\lambda_n(\Sigma_0)}{\lambda_1(\Sigma_1)} > 0$.

Next, suppose that $A \setminus B \neq \emptyset$ and $B = \emptyset$. Following similar arguments as above, one can show that

$$2\sum_{a\in A\setminus B}\big(f_{KL}(\{a\}\cup B) - f_{KL}(B)\big) \ge |A|(\underline{r}_1^2 + r - 1), \tag{26}$$

and

$$2\big(f_{KL}(A\cup B) - f_{KL}(B)\big) \le |A|(\bar{r}_1^2 + r_2 - 1 + \log\frac{\lambda_n(\Sigma_0)}{\lambda_1(\Sigma_1)}). \tag{27}$$

Again, we note that if $\bar{r}_1^2 + r_2 - 1 + \log\frac{\lambda_n(\Sigma_0)}{\lambda_1(\Sigma_1)} = 0$, we have $f_{KL}(A) = 0$ for all $A \subseteq \mathcal{V}$. Therefore, we consider $\bar{r}_1^2 + r_2 - 1 + \log\frac{\lambda_n(\Sigma_0)}{\lambda_1(\Sigma_1)} > 0$.

Combining inequalities (19), (23), (26) and (27), we obtain inequality (13) from Definition 1. ■

We then summarize the above arguments as follows.

Theorem 2: Consider an instance of the MSSNP problem. Suppose that the measurement vector $\boldsymbol{X} \in \mathbb{R}^n$ satisfies

$$\begin{aligned} H_0 &: \boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, \Sigma_0),\\ H_1 &: \boldsymbol{X} \sim \mathcal{N}(\theta_1, \Sigma_1), \end{aligned}$$

where $\Sigma_0 = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ with $\sigma_i \in \mathbb{R}_{>0}$, $\forall i \in \{1, \ldots, n\}$, and $\Sigma_1 \in \mathbb{S}_{++}^n$. The greedy algorithm, when applied to such instances of MSSNP, yields

$$f_{KL}(\mu_{gre}) \ge (1 - e^{-\gamma^*})f_{KL}(\mu^*), \tag{28}$$

where μ_{gre} is the solution returned by the greedy algorithm, μ^* is the optimal solution to the problem and $\gamma^* \triangleq \frac{\underline{r}_1^2 + r - 1}{\bar{r}_1^2 + r_2 - 1 + \log\frac{\lambda_n(\Sigma_0)}{\lambda_1(\Sigma_1)}}$. □

Remark 5: Note that the bound on the performance of the greedy algorithm depends on the parameters of the MSSNP instance. We know from Eq. (13) and Eq. (28) that the bound on the performance of the greedy algorithm becomes tighter when the lower bound on γ_{KL} increases. For example, considering the instances with $\Sigma_0 = \sigma^2 I_n$ and $\theta_1 = \theta\mathbf{1}_n$, where $\sigma, \theta \in \mathbb{R}$, the bound in inequality (13) simplifies into

$$\gamma_{KL} \ge \frac{\frac{\theta^2}{\sigma^2} + \min_{1\le i\le n}\big(\frac{(\Sigma_1)_{ii}}{\sigma^2} - \log\frac{(\Sigma_1)_{ii}}{\sigma^2}\big) - 1}{\frac{\theta^2}{\sigma^2} + \max_{1\le i\le n}\frac{(\Sigma_1)_{ii}}{\sigma^2} - 1 + \log\frac{\sigma^2}{\lambda_1(\Sigma_1)}},$$

It follows that

$$\gamma_{KL} \geq \frac{\frac{\theta^2}{\sigma^2} + \min_{1\leq i\leq n}\left(\frac{(\Sigma_1)_{ii}}{\sigma^2} - \log\frac{(\Sigma_1)_{ii}}{\sigma^2}\right) - 1}{\frac{\theta^2}{\sigma^2} + \frac{\lambda_n(\Sigma_1)}{\sigma^2} - \log\frac{\lambda_1(\Sigma_1)}{\sigma^2} - 1}, \quad (29)$$

where we use the fact that $\lambda_n(\Sigma_1) \geq \max_{i\leq i\leq n}(\Sigma_1)_{ii}$ [18]. Note that $\frac{\lambda_n(\Sigma_1)}{\sigma^2} - \log\frac{\lambda_1(\Sigma_1)}{\sigma^2} - 1 \geq 0$. Supposing that σ and Σ_1 are fixed, we obtain from inequality (29) that the lower bound on γ_{KL} increases if θ^2 increases. If we further consider that Σ_1 is diagonal and $\lambda_1(\Sigma_1) \geq \sigma^2$, inequality (29) becomes

$$\gamma_{KL} \geq \frac{\frac{\theta^2}{\sigma^2} + \frac{\lambda_1(\Sigma_1)}{\sigma^2} - \log\frac{\lambda_1(\Sigma_1)}{\sigma^2} - 1}{\frac{\theta^2}{\sigma^2} + \frac{\lambda_n(\Sigma_1)}{\sigma^2} - \log\frac{\lambda_1(\Sigma_1)}{\sigma^2} - 1}, \quad (30)$$

where we use the fact that the function $h(x) = x - \log x$ is nondecreasing on $x \geq 1$. Now supposing that θ, $\lambda_1(\Sigma_1)$ and σ are fixed, we have from inequality (30) that the lower bound on γ_{KL} increases as $\lambda_n(\Sigma_1)$ decreases, and the lower bound on γ_{KL} tends to zero if $\lambda_n(\Sigma_1)$ goes to infinity. In summary, a higher value of $\frac{\theta^2}{\sigma^2}$ or a smaller gap between $\lambda_1(\Sigma_1)$ and $\lambda_n(\Sigma_1)$ can potentially improve the performance of the greedy algorithm. Note that when Σ_1 is diagonal, $\lambda_1(\Sigma_1)$ and $\lambda_n(\Sigma_1)$ correspond to the smallest variance and the largest variance among the measurements of the sensors under H_1, respectively. □

Remark 6: The result in Theorem 2 complements the results in [11] and [12]. Specifically, in [11], the authors only show that the objective function of MSSNP is not submodular without further analyzing the submodularity ratio of the objective function. Furthermore, they propose a selection algorithm without providing any theoretical guarantee on the performance. Under the assumption that $\Sigma_0 = \Sigma_1$, the authors in [12] consider a different metric to characterize how close a nonsubmodular function is to being submodular, which leads to a bound on the performance ratio of the greedy algorithm, defined as $\frac{f_{KL}(\mu^*)}{f_{KL}(\mu_{gre})}$, that depends on the value of $f_{KL}(\mu_{gre})$. For general instances of MSSNP, where $\theta_0 \neq \theta_2$ or $\Sigma_0 \neq \Sigma_1$ (or both), the authors in [12] instead consider solving the MSSNP problem with submodular surrogates to $f_{KL}(\mu)$. □

V. Conclusions

In this paper, we studied the sensor selection for the Neyman-Pearson detector problem in hypothesis testing. We first showed that the sensor selection for the Neyman-Pearson detector problem is NP-hard in general. We related our problem to the subset selection for linear regression to achieve this. We then characterized the performance of greedy algorithms for solving the sensor selection problem when we considered a surrogate (based on the KL distance) to the miss probability of the Neyman-Pearson detector. By using the notion of submodularity ratio to characterize how close a nonsubmodular set function is to being submodular, we provided a bound on the performance of greedy algorithms that depends on the parameters of the problem. Future studies on other types of detectors and hypothesis testing in networks are of interest.

Appendix

Proof of Lemma 3 (sketch):

Note that since $\Sigma \in \mathbb{S}^n_{++}$, we have that $\Sigma(\mu)$ and $\Sigma(\mu^c)$ are positive definite for all $\mu \in \{0,1\}^n$ with $\text{supp}(\mu) = \{1,\dots,p\}$, where $p < n$ ($p \in \mathbb{Z}_{>0}$). Denote $\Sigma' = \Sigma(\mu) - \Sigma_{\boldsymbol{X}(\mu)\boldsymbol{X}(\mu^c)}(\Sigma(\mu^c))^{-1}\Sigma^T_{\boldsymbol{X}(\mu)\boldsymbol{X}(\mu^c)}$. Let e_1 be an eigenvector of the eigenvalue $\lambda_1(\Sigma') > 0$, i.e., $\Sigma' e_1 = \lambda_1(\Sigma')e_1$, and define $e_0 = \begin{bmatrix} e_1 \\ -(\Sigma(\mu^c))^{-1}\Sigma^T_{\boldsymbol{X}(\mu)\boldsymbol{X}(\mu^c)}e_1\end{bmatrix}$. Using Eq. (0.8.5.3) in [18], one can show that $\Sigma e_0 = \begin{bmatrix}\Sigma' e_1 \\ \mathbf{0}\end{bmatrix}$. It follows that

$$e_0^T\lambda_1(\Sigma)e_0 \leq e_0^T\Sigma e_0 = e_1^T\Sigma' e_1 = e_1^T\lambda_1(\Sigma')e_1, \quad (31)$$

where the first inequality follows from the fact that $\big(\Sigma - \lambda_1(\Sigma)I_n\big)$ is positive semi-definite [18]. Noting that $e_0^Te_0 \geq e_1^Te_1 > 0$, we have from (31) that $\lambda_1(\Sigma) \leq \lambda_1\big(\Sigma'\big)$. ■

References

[1] M. Van De Wal and B. De Jager, "A review of methods for input/output selection," *Automatica*, vol. 37, no. 4, pp. 487–510, 2001.

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

[3] S. Pequito, S. Kar, and A. P. Aguiar, "Minimum cost input/output design for large-scale linear structural systems," *Automatica*, vol. 68, pp. 384–391, 2016.

[4] 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.

[5] C. Yang, J. Wu, X. Ren, W. Yang, H. Shi, and L. Shi, "Deterministic sensor selection for centralized state estimation under limited communication resource," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2336–2348, 2015.

[6] V. Tzoumas, A. Jadbabaie, and G. J. Pappas, "Sensor placement for optimal Kalman filtering: Fundamental limits, submodularity, and algorithms," in *Proc. American Control Conference*, 2016, pp. 191–196.

[7] H. Zhang, R. Ayoub, and S. Sundaram, "Sensor selection for Kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms," *Automatica*, vol. 78, pp. 202–210, 2017.

[8] L. Ye, S. Roy, and S. Sundaram, "Optimal sensor placement for Kalman filtering in stochastically forced consensus network," in *Proc. IEEE Conference on Decision and Control*, 2018, pp. 6686–6691.

[9] S. P. Chepuri and G. Leus, "Sparse sensing for distributed detection," *IEEE Trans. Signal Processing*, vol. 64, no. 6, pp. 1446–1460, 2016.

[10] H. V. Poor, *An introduction to signal detection and estimation*. Springer, 1994.

[11] D. Bajovic, B. Sinopoli, and J. Xavier, "Sensor selection for event detection in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4938–4953, 2011.

[12] M. Coutino, S. P. Chepuri, and G. Leus, "Submodular sparse sensing for gaussian detection with correlated observations," *IEEE Transactions on Signal Processing*, vol. 66, no. 15, pp. 4025–4039, 2018.

[13] L. Ye, S. Roy, and S. Sundaram, "On the complexity and approximability of optimal sensor selection for Kalman filtering," in *Proc. American Control Conference*, 2018, pp. 5049–5054.

[14] A. Das and D. Kempe, "Approximate submodularity and its applications: subset selection, sparse approximation and dictionary selection," *JMLR*, vol. 19, no. 1, pp. 74–107, 2018.

[15] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive approximation*, vol. 13, no. 1, pp. 57–98, 1997.

[16] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[17] A. A. Bian, J. M. Buhmann, A. Krause, and S. Tschiatschek, "Guarantees for greedy maximization of non-submodular functions with applications," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, 2017, pp. 498–507.

[18] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.