Randomized Greedy Sensor Selection: Leveraging Weak Submodularity

Abolfazl Hashemi, *Student Member, IEEE*, Mahsa Ghasemi, *Student Member, IEEE*, Haris Vikalo, *Senior Member, IEEE*, and Ufuk Topcu

***Abstract*—We study the problem of estimating a random process from the observations collected by a network of sensors that operate under resource constraints. When the dynamics of the process and sensor observations are described by a state-space model and the resource are unlimited, the conventional Kalman filter provides the minimum mean square error (MMSE) estimates. However, at any given time, restrictions on the available communications bandwidth and computational capabilities and/or power impose a limitation on the number of network nodes, whose observations can be used to compute the estimates. We formulate the problem of selecting the most informative subset of the sensors as a combinatorial problem of maximizing a monotone set function under a uniform matroid constraint. For the MMSE estimation criterion, we show that the maximum elementwise curvature of the objective function satisfies a certain upper-bound constraint and is, therefore, weak submodular. Building upon the work of Mirzasoleiman *et al.* on submodular maximization, we develop an efficient randomized greedy algorithm for sensor selection and establish guarantees on the estimator's performance in this setting. Extensive simulation results demonstrate the efficacy of the randomized greedy algorithm compared to state-of-the-art greedy and semidefinite programming relaxation methods.**

***Index Terms*—Kalman filtering, sensor networks, sensor selection, weak submodularity.**

Manuscript received August 12, 2019; revised January 6, 2020; accepted February 29, 2020. Date of publication March 16, 2020; date of current version December 24, 2020. This work was supported in part by the National Science Foundation under Grant 1652113 and Grant 1809327, in part by Sandia National Laboratories under Grant # 801KOB, in part by the Defense Advanced Research Projects Agency under Grant D19AP00004, and in part by the Office of Naval Research under Grant N00014161316, Grant N000141812829, Grant N000141712623, and Grant N000141912054. This paper was presented in part at the American Control Conference, Milwaukee, WI, USA, Jun. 2018 [1]. Recommended by Associate Editor F. Zhang. *(Corresponding author: Abolfazl Hashemi.)*

Abolfazl Hashemi and Haris Vikalo are with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712 USA (e-mail: abolfazl@utexas.edu; hvikalo@ece.utexas.edu).

Mahsa Ghasemi is with the Department of Mechanical Engineering, University of Texas at Austin, Austin, TX 78712 USA (e-mail: mahsa.ghasemi@utexas.edu).

Ufuk Topcu is with the Department of Aerospace Engineering and Engineering Mechanics, University of Texas at Austin, Austin, TX 78712 USA (e-mail: utopcu@utexas.edu).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2020.2980924

I. Introduction

MODERN sensor networks deploy a large number of nodes that either exchange their noisy and possibly processed observations of a random process or forward those observations to a data fusion center. Due to constraints on computation, power, and communication resources, instead of estimating the process using information collected by the entire network, the fusion center typically queries a relatively small subset of the available sensors. The problem of selecting the sensors that would acquire the most informative observations arises in a number of applications in control and signal processing systems, including sensor selection for Kalman filtering [3]–[5], batch state and stochastic process estimation [6], [7], minimal actuator placement [8], [9], voltage control and meter placement in power networks [10]–[12], sensor scheduling in wireless sensor networks [3], [13], and subset selection in machine learning [2].

For a variety of performance criteria, finding an optimal subset of sensors requires solving a computationally challenging combinatorial optimization problem, possibly using branch-and-bound search [14]. By reducing it to the set cover problem, sensor selection was, in fact, shown to be NP-hard [15]. This hardness result has motivated development of numerous heuristics and approximate algorithms. For instance, Joshi and Boyd [16] formulated the sensor selection problem as the maximization (minimization) of the $\log\det$ of the Fisher information matrix (error covariance matrix) and found a solution by relaxing the problem to a semidefinite program (SDP). The computational complexity of finding the solution to the SDP relaxation of the sensor selection problem is cubic in the total number of available sensors, which limits its practical feasibility in large-scale networks consisting of many sensing nodes. Moreover, the solution to the SDP relaxation comes with no performance guarantees. To overcome these drawbacks, Shamaiah *et al.* [4] proposed a greedy algorithm guaranteed to achieve at least $(1-1/e)$ of the optimal objective at a complexity lower than that of the SDP relaxation. The theoretical underpinnings of the greedy approach to the sensor selection problem in [4] are drawn from the area of submodular function optimization. In particular, these results stem from the fact that the logarithm of the determinant ($\log\det$) of the Fisher information matrix is a monotone submodular function. Nemhauser *et al.* [17] studied maximization of such a function subject to a uniform matroid constraint and showed that the greedy algorithm, which iteratively selects items providing maximum marginal gain, achieves $(1-1/e)$ approximation factor.

More recently, the authors of [5]–[7] and [9] have employed and analyzed greedy algorithms for finding approximate solutions to the $\log\det$ maximization problem in a number of practical settings.

Most of the existing work on greedy sensor selection has been focused on optimizing the $\log\det$ of the Fisher information matrix, an objective indicative of the volume of the confidence ellipsoid. However, this criterion does not explicitly relate to the mean square error (MSE), which is often a natural performance metric of interest in estimation problems. The MSE, i.e., the trace of the covariance matrix of the estimation error, is not supermodular [18]–[23]. Therefore, its negative value, which we would like to maximize, is not submodular. Consequently, the setting and results of [17] do not apply to the MSE minimization problem. Recently, Wang *et al.* [24] have analyzed the performance of the greedy algorithm in the general setting of maximizing a monotone nondecreasing objective function that is not necessarily submodular. They used a notion of the elemental curvature μ of the objective function to show that the greedy algorithm provides a $((1+\mu)^{-1})$ approximation under a matroid constraint. However, determining the elemental curvature defined in [24] is itself an NP-hard problem. Therefore, providing performance guarantees for the settings, where the objective function is not submodular or supermodular, such as the trace of the covariance matrix of the estimation error in the sensor selection problem, remains a challenge. On another note, processing massive amounts of data collected by modern large-scale networks may be challenging even for greedy algorithms. To further reduce the computational burden of maximizing a monotone increasing submodular function subject to cardinality constraints, the authors of [2] proposed a stochastic greedy algorithm that achieves $(1-1/e-\epsilon)$ approximation factor, where ϵ denotes a parameter that can be varied to explore the performance–complexity tradeoff. However, the results of [2] do not apply to the sensor selection problem under the (nonsubmodular) MSE objective.

The following are contributions of this article.

1) We formulate the task of selecting sensors in a large-scale network as the problem of maximizing a monotone nonsubmodular objective function directly related to the mean square estimation error.
2) By closely inspecting curvature of the objective function, we derive sufficient conditions under which the function is weak submodular. This enables us to argue that when the measurement vectors are Gaussian or Bernoulli, as frequently encountered in reduced-dimensionality Kalman filtering via random projections [25], the MSE objective is with high probability weak submodular.
3) We study the important setting, where the dynamics of the process and sensor observations is described by a state-space model and, building upon the work of Mirzasoleiman *et al.* [2], propose a randomized greedy algorithm for sensor selection and derive a bound on the MSE of the state estimate formed by the Kalman filter that uses the measurements of the selected sensors.
4) Our novel technique for the analysis of the randomized greedy algorithm provides results that improve over the existing performance guarantees of [2] for submodular maximization problems.
5) Extensive simulations demonstrate that the proposed randomized greedy sensor selection scheme significantly outperforms both greedy and SDP relaxation methods in terms of computational complexity, and hence runtime, while providing essentially the same or improved MSE.

Our preliminary work on randomized greedy sensor selection was presented at the 2018 American Control Conference [1]. The current article presents a significantly more thorough and detailed analysis of the proposed algorithmic framework. This includes the results in Theorem 3 and its corollary regarding performance guarantee of the randomized greedy algorithm for any instance of the sensor selection problem; prior results, summarized in Theorem 2, were limited to the guarantees of the expected performance. Moreover, we complement our theoretical results by presenting numerical evaluations for an application of multiobject tracking via UAVs.

The rest of this article is organized as follows. Section II presents a motivating example and sets up the system model. In Section III, we describe the novel formulation of the sensor selection problem and derive a bound on the curvature of the MSE-related objective function. In Section IV, we introduce the randomized greedy algorithm and analyze its performance. Section V presents the simulation results, Finally, Section VI concludes this article.

Notation: Bold capital letters denote matrices, while bold lowercase letters represent vectors. $H_k(i,j)$ is the (i,j) entry of the time-varying matrix $\mathbf{H}_k$ at time k, $\mathbf{h}_{k,j}$ is the jth row of $\mathbf{H}_k$, $\mathbf{H}_{k,S}$ is a submatrix of $\mathbf{H}_k$ that consists of the rows of $\mathbf{H}_k$ indexed by the set S, and $\lambda_{\max}(\mathbf{H}_k)$ and $\lambda_{\min}(\mathbf{H}_k)$ are the maximum and minimum eigenvalues of $\mathbf{H}_k$, respectively. Spectral (ℓ_2) norm of a matrix is denoted by $\|.\|$. $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ is the identity matrix. Moreover, let $[n] := \{1,2,\ldots,n\}$.

II. System Model and Problem Formulation

This section starts by a description of a motivating example of multiobject tracking under communication and power constraints. Then, we proceed to define the system model and mathematically formulate the sensor selection problem studied in this article.

A. Motivating Example: Accelerated Multiobject Tracking

Consider a tracking system, shown in Fig. 1, where a control unit surveys an area via a swarm of unmanned aerial vehicles (UAVs). The UAVs are equipped with GPS and radar systems and can communicate with each other over locally established communication channels. However, only a few of the UAVs known as *swarm leaders* are allowed to communicate to the control unit because of various practical restrictions such as power constraints. The UAVs patrol the area according to a predefined search pattern (i.e., a dynamic model) to gather information about the location of mobile objects of interest. That is, the UAVs move along an elliptically shaped path with constant speed. Each UAV, by using its radar system, acquires range and angular measurements of all the objects that are within the

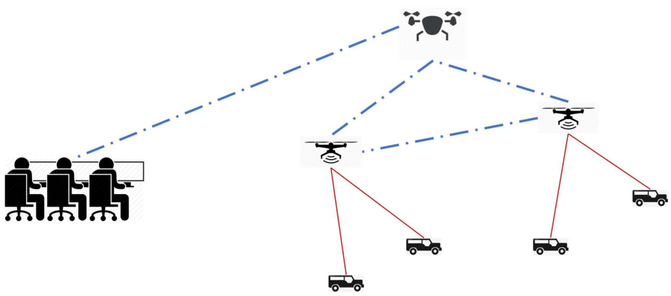


Fig. 1. Multiobject tracking via a swarm of UAVs. The UAVs can communicate with each other and are equipped with GPS and radar systems. The objective is to select a small subset of range and angular measurements gathered by the UAVs to communicate to the control unit.

maximum radar detection range and are capable of transmitting those measurements to the swarm leaders. Therefore, we assume that the detection probability is one for all the targets that are within the sensing range of each UAV. Additionally, we assume that there are a finite number of targets that are uniquely tagged in a way that the UAVs can exactly identify them in order to achieve error-free data association.

Due to limitations on the rate of communication between the swarm leaders and the control unit, and to reduce delays in tracking from high computation, only a subset of the gathered measurements is communicated to the control unit. In order to track the locations of the objects, the control unit employs Kalman filtering using the received measurements. Therefore, the goal of swarm leaders is to perform *sensor scheduling* and select a subset of range and angular measurements such that 1) the communication constraint is satisfied and 2) the MSE of the Kalman filter estimate of the objects' locations is minimized.

B. System Model

Consider a discrete-time, linear, time-varying state-space model described by

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \mathbf{w}_k$$
$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \tag{1}$$

where $\mathbf{x}_k \in \mathbb{R}^m$ is the state vector at time k that we aim to estimate, $\mathbf{y}_k \in \mathbb{R}^n$ is the measurement vector, $\mathbf{w}_k \in \mathbb{R}^m$ and $\mathbf{v}_k \in \mathbb{R}^n$ are zero-mean white Gaussian noise processes with covariances $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively, $\mathbf{A}_k \in \mathbb{R}^{m\times m}$ is the state transition matrix, and $\mathbf{H}_k \in \mathbb{R}^{n\times m}$ is the matrix, whose rows at time k are the measurement vectors $\mathbf{h}_{k,i} \in \mathbb{R}^m$, $1 \le i \le n$. We assume that the states $\mathbf{x}_k$ are uncorrelated with $\mathbf{w}_k$ and $\mathbf{v}_k$. Additionally, we assume that $\mathbf{x}_0 \sim \mathcal{N}(0, \mathbf{\Sigma}_x)$ with $\mathbf{\Sigma}_x \succ \mathbf{0}$, and $\mathbf{R}_k = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$. Note that, unlike the past work on greedy sensor selection in [4], [13], [26], and [27], this model does not restrict the measurement noise covariance matrix to be a multiple of identity.

Due to limited resources, the fusion center aims to select K out of n sensors and uses their measurements to estimate the state vector $\mathbf{x}_k$ such that the trace of the covariance matrix of the estimation error, i.e., the MSE of the estimator implemented using the Kalman filter, is minimized. Similar to prior work in [4], [13], and [16], we assume that the measurement vectors $\mathbf{h}_{k,i}$ are available at the fusion center. Let $\hat{\mathbf{x}}_{k|k-1}$ and $\hat{\mathbf{x}}_{k|k}$ denote the predicted and filtered linear minimum mean square error (LMMSE) estimators of $\mathbf{x}_k$, respectively. In other words, $\hat{\mathbf{x}}_{k|k-1}$ is the LMMSE estimator of $\mathbf{x}_k$ given $\{\mathbf{y}_{S_1}, \dots, \mathbf{y}_{S_{k-1}}\}$ and $\hat{\mathbf{x}}_{k|k}$ is the LMMSE estimator of $\mathbf{x}_k$ given $\{\mathbf{y}_{S_1}, \dots, \mathbf{y}_{S_k}\}$, where S_j denotes the set of sensors selected at time j and $\mathbf{y}_{S_j}$ denotes the vector of measurements collected by those sensors. Moreover, let $\mathbf{P}_{k|k-1}$ and $\mathbf{P}_{k|k}$ denote the predicted and filtered error covariance matrix of the Kalman filter at time instant k, respectively, i.e.,

$$\mathbf{P}_{k|k-1} = \mathbf{A}_k \mathbf{P}_{k-1|k-1} \mathbf{A}_k^\top + \mathbf{Q}_k$$
$$\mathbf{P}_{k|k} = \left(\mathbf{P}_{k|k-1}^{-1} + \mathbf{H}_{k,S_k}^\top \mathbf{R}_{k,S_k}^{-1} \mathbf{H}_{k,S_k}\right)^{-1}$$

where $P_{0|0} = \mathbf{\Sigma}_x$. Since $\mathbf{R}_k = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ and the measurements are uncorrelated across sensors, it holds that

$$\mathbf{P}_{k|k} = \left(\mathbf{P}_{k|k-1}^{-1} + \mathbf{H}_{k,S_k}^\top \text{diag}(\{\sigma_i^{-2}\}_{i\in S_k}) \mathbf{H}_{k,S_k}\right)^{-1}.$$

Furthermore, $\mathbf{F}_{S_k} = \mathbf{P}_{k|k}^{-1} = \mathbf{P}_{k|k-1}^{-1} + \sum_{i\in S_k} \sigma_i^{-2} \mathbf{h}_{k,i} \mathbf{h}_{k,i}^\top$ is the corresponding Fisher information matrix. In the information form, the filtered estimator of $\mathbf{x}_k$ is expressed as

$$\hat{\mathbf{x}}_{k|k} = \mathbf{F}_{S_k}^{-1} \mathbf{H}_{k,S_k}^\top \text{diag}(\{\sigma_i^{-2}\}_{i\in S_k}) \mathbf{y}_k. \tag{2}$$

The MSE of the estimate found in (2) is given by the trace of the filtered error covariance matrix $\mathbf{P}_{k|k}$

$$\text{MSE}_{S_k} = \mathbb{E}\left[\left\|\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}\right\|_2^2\right] = \text{Tr}\left(\mathbf{F}_{S_k}^{-1}\right). \tag{3}$$

To minimize (3), at each time step, the fusion center seeks a solution to the optimization problem

$$\min_S \quad \text{Tr}\left(\mathbf{F}_S^{-1}\right) \quad \text{s.t.} \quad S \subset [n], \quad |S| = K. \tag{4}$$

By a reduction to the well-known set cover problem, the combinatorial optimization (4) can be shown to be NP-hard [15], [28]. In principle, to find the optimal solution, one needs to exhaustively search over all schedules of K sensors. The techniques proposed in [16], albeit for an optimality criterion different from MSE and a simpler measurement model, find a subset of sensors that yields a suboptimal MSE performance while being computationally much more efficient than the exhaustive search. In particular, Joshi and Boyd [16] rely on finding the solution to the following SDP relaxation:

$$\begin{aligned} \min_{\mathbf{z}_k, \mathbf{Y}} \quad & \text{Tr}(\mathbf{Y}) \\ \text{s.t.} \quad & 0 \le z_{k,i} \le 1, \quad \forall i \in [n] \\ & \sum_{i=1}^n z_{k,i} = K \\ & \begin{bmatrix} \mathbf{Y} & \mathbf{I} \\ \mathbf{I} & \mathbf{P}_{k|k-1}^{-1} + \sum_{i=1}^n z_{k,i}\sigma_i^{-2} \mathbf{h}_{k,i}\mathbf{h}_{k,i}^\top \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \tag{5}$$

The complexity of the SDP algorithm scales as $\mathcal{O}(n^3)$, which is infeasible in many practical settings. Furthermore, there are no guarantees on the achievable MSE performance of the SDP relaxation. Note that when the number of sensors in a network and the size of the state vector $\mathbf{x}_k$ are relatively large, even the greedy algorithm proposed in [4] may be computationally prohibitive.

III. Sensor Selection via Optimizing a Weak Submodular Objective

Leveraging the idea of *weak submodularity*, in this section, we propose a new formulation of the sensor selection problem concerned with minimizing the MSE of the Kalman filter that relies on a subset of network nodes to track states of a hidden random process. We first overview concepts that are essential for the development of the proposed framework.

Definition 1: A set function $f : 2^X \rightarrow \mathbb{R}$ is monotone nondecreasing if $f(S) \leq f(T)$ for all $S \subseteq T \subseteq X$.

Definition 2: A set function $f : 2^X \rightarrow \mathbb{R}$ is submodular if

$$f(S \cup \{j\}) - f(S) \geq f(T \cup \{j\}) - f(T) \tag{6}$$

for all subsets $S \subseteq T \subset X$ and $j \in X \backslash T$.

A concept closely related to submodularity is the notion of curvature of a set function. The curvature quantifies how close the function is to being submodular. In particular, here, we state the definition of the elementwise curvature (also known as the approximate weak submodularity constant) [24], [26], [29], [30].

Definition 3: The elementwise curvature $\mathcal{C}_l$ of a monotone nondecreasing function f is defined as

$$\mathcal{C}_l = \max_{(S,T,i) \in \mathcal{X}_l} f_i(T)/f_i(S) \tag{7}$$

where $f_i(S) = f(S \cup \{i\}) - f(S)$ and $f_i(T) = f(T \cup \{i\}) - f(T)$ denote the marginal values of adding element i to sets S and T, respectively, and $\mathcal{X}_l = \{(S,T,i) | S \subset T \subset X, i \in X \backslash T, |T \backslash S| = l, |X| = n\}$. Furthermore, the maximum elementwise curvature is denoted by $\mathcal{C}_{\max} = \max_l \mathcal{C}_l$.

A set function is submodular if and only if $\mathcal{C}_{\max} \leq 1$. We refer to $f(S)$ as being weak submodular if its curvature $\mathcal{C}_{\max} > 1$ is bounded above. Note that while computing the elementwise curvature is NP-hard, the bound on the proposed sensor selection objective that we derive in Theorem 1 can be efficiently evaluated.

Definition 4: Let X be a finite set and let $\mathcal{I}$ denote a collection of subsets of X. The pair $\mathcal{M} = (X, \mathcal{I})$ is a matroid if the following two statements hold.

1) Hereditary property: If $T \in \mathcal{I}$, then $S \in \mathcal{I}$ for all $S \subseteq T$.
2) Augmentation property: If $S, T \in \mathcal{I}$ and $|S| < |T|$, then there exists $e \in T \backslash S$ such that $S \cup \{e\} \in \mathcal{I}$.

The collection $\mathcal{I}$ is called the set of independent sets of the matroid $\mathcal{M}$. A maximal independent set is a basis. It is easy to show that all the bases of a matroid have the same cardinality.

Given a monotone nondecreasing set function $f : 2^X \rightarrow \mathbb{R}$ with $f(\emptyset) = 0$, and a uniform matroid $\mathcal{M} = (X, \mathcal{I})$, we are interested in solving the combinatorial problem

$$\max_{S \in \mathcal{I}} f(S). \tag{8}$$

Recall that for Kalman filtering in the resource constrained scenario, if S_k is the set of sensors selected at time k, then the error covariance matrix of the filtered estimate is $\mathbf{P}_{k|k} = \mathbf{F}_{S_k}^{-1}$, the inverse of the corresponding Fisher information matrix. Let us define $f(S)$ as

$$f(S) = \operatorname{Tr}\left(\mathbf{P}_{k|k-1} - \mathbf{F}_S^{-1}\right).$$

Clearly, since $\mathbf{P}_{k|k-1}$ is known, there is a one-to-one correspondence between $f(S_k)$ computed for a given subset of sensors S_k and the MSE of the LMMSE estimator (i.e., filtered estimate of the Kalman filter) that uses measurements acquired by the sensors in S_k. Therefore, we can express the optimization problem (4) as

$$\max_{S} \quad f(S) \quad \text{s.t.} \quad S \subset [n], \quad |S| = K. \tag{9}$$

We now argue that (9) is indeed an instance of the general combinatorial problem (8). By defining $X = [n]$ and $\mathcal{I} = \{S \subset X || S| \leq K\}$, it is easy to see that $\mathcal{M} = (X, \mathcal{I})$ is a uniform matroid. In Proposition 1, we characterize important properties of $f(S)$ and develop a recursive scheme to efficiently compute the marginal gain of querying a sensor. The formula for the marginal gain of $f(S)$ is also of interest in our subsequent analysis of its weak submodularity properties.

Proposition 1: Let $f(S) = \operatorname{Tr}(\mathbf{P}_{k|k-1} - \mathbf{F}_S^{-1})$. Then, $f(S)$ is a monotonically increasing set function, $f(\emptyset) = 0$, and

$$f_j(S) = \frac{\mathbf{h}_{k,j}^\top \mathbf{F}_S^{-2} \mathbf{h}_{k,j}}{\sigma_j^2 + \mathbf{h}_{k,j}^\top \mathbf{F}_S^{-1} \mathbf{h}_{k,j}} \tag{10}$$

where upon adding element $j \in X \backslash S$ to S, $\mathbf{F}_S$ is updated according to

$$\mathbf{F}_{S \cup \{j\}}^{-1} = \mathbf{F}_S^{-1} - \frac{\mathbf{F}_S^{-1} \mathbf{h}_{k,j} \mathbf{h}_{k,j}^\top \mathbf{F}_S^{-1}}{\sigma_j^2 + \mathbf{h}_{k,j}^\top \mathbf{F}_S^{-1} \mathbf{h}_{k,j}}. \tag{11}$$

Proof: See Appendix A. ■

As stated in Section I, the MSE is not supermodular [18], [23]. Consequently, the proposed objective $f(S) = \operatorname{Tr}(\mathbf{P}_{k|k-1} - \mathbf{F}_S^{-1})$ is also not submodular. However, as we show in Theorem 1, under certain conditions, $f(S)$ is characterized by a bounded maximum elementwise curvature $\mathcal{C}_{\max}$. Theorem 1 also states a probabilistic theoretical upper bound on $\mathcal{C}_{\max}$ in scenarios, where, at each time step, the measurement vectors $\mathbf{h}_{k,j}$s are realizations of independent identically distributed (i.i.d.) random vectors drawn from a suitable distribution.

Before proceeding to Theorem 1 and its proof, we first state the matrix Bernstein inequality [31] and Weyl's inequality [32], which we will later use in the proof of Theorem 1.

Lemma 1 (Matrix Bernstein inequality [31]): Let $\{\mathbf{X}_\ell\}_{\ell=1}^n$ be a finite collection of independent, random, Hermitian matrices in $\mathbb{R}^{m \times m}$. Assume that for all $\ell \in [n]$, we have

$$\mathbb{E}\left[\mathbf{X}_\ell\right] = \mathbf{0}, \quad \lambda_{\max}(\mathbf{X}_\ell) \leq L. \tag{12}$$

Let $\mathbf{Y} = \sum_{\ell=1}^n \mathbf{X}_\ell$. Then, for all $q > 0$, it holds that

$$\Pr\{\lambda_{\max}(\mathbf{Y}) \geq q\} \leq m \exp\left(\frac{-q^2/2}{\|\mathbb{E}\left[\mathbf{Y}^2\right]\| + Lq/3}\right). \tag{13}$$

Lemma 2 (Weyl's inequality [32]): Let $\mathbf{A}$ and $\mathbf{B}$ be two $m \times m$ real positive-definite matrices. Then, it holds that

$$\lambda_l(\mathbf{A}) + \lambda_{\min}(\mathbf{B}) \leq \lambda_l(\mathbf{A}+\mathbf{B}) \leq \lambda_l(\mathbf{A}) + \lambda_{\max}(\mathbf{B}) \quad (14)$$

where $\lambda_l(\mathbf{A})$ denotes the lth largest eigenvalue of $\mathbf{A}$.

We now proceed to the statement and proof of Theorem 1.

Theorem 1: Let $\mathcal{C}_{\max}$ be the maximum elementwise curvature of $f(S)$, the objective function of the sensor selection problem. Assume that $\|\mathbf{h}_{k,j}\|_2^2 \leq C$ for all j and k. If

$$\lambda_{\max}(\mathbf{H}_k^\top \mathbf{H}_k) \leq \left(\frac{1}{\phi} - \frac{1}{\lambda_{\min}(\mathbf{P}_{k|k-1})}\right) \min_{j\in[n]} \sigma_j^2 \quad (15)$$

for some $0 < \phi < \lambda_{\min}(\mathbf{P}_{k|k-1})$, then it holds that

$$\mathcal{C}_{\max} \leq \max_{j\in[n]} \frac{\lambda_{\max}(\mathbf{P}_{k|k-1})^2(\sigma_j^2 + \lambda_{\max}(\mathbf{P}_{k|k-1})C)}{\phi^2(\sigma_j^2+\phi C)}. \quad (16)$$

Furthermore, if $\mathbf{h}_{k,j}$s are i.i.d. zero-mean random vectors with covariance matrix $\sigma_h^2\mathbf{I}_m$ such that $\sigma_h^2 < C$, then, for all $q > 0$, with probability

$$p \geq 1 - m\exp\left(\frac{-q^2/2}{(C-\sigma_h^2)(n\sigma_h^2+q/3)}\right) \quad (17)$$

it holds that

$$\phi = \min_{j\in[n]}\left(\frac{1}{\lambda_{\min}(\mathbf{P}_{k|k-1})} + \frac{n\sigma_h^2+q}{\sigma_j^2}\right)^{-1} > 0. \quad (18)$$

Proof: We prove the statement of the theorem by relying on the recursive expression for the marginal gain stated in Proposition 1. We first establish a sufficient condition for weak submodularity of $f(S)$. In particular, from the definition of the elementwise curvature and (10), for all $(S,T,j)\in\mathcal{X}_l$, we obtain

$$\begin{aligned}\mathcal{C}_l &= \max_{(S,T,j)\in\mathcal{X}_l} \frac{(\mathbf{h}_{k,j}^\top \mathbf{F}_T^{-2}\mathbf{h}_{k,j})(\sigma_j^2 + \mathbf{h}_{k,j}^\top \mathbf{F}_S^{-1}\mathbf{h}_{k,j})}{(\mathbf{h}_{k,j}^\top \mathbf{F}_S^{-2}\mathbf{h}_{k,j})(\sigma_j^2 + \mathbf{h}_{k,j}^\top \mathbf{F}_T^{-1}\mathbf{h}_{k,j})}\\
&\leq \max_{(S,T,j)\in\mathcal{X}_l} \frac{\lambda_{\max}(\mathbf{F}_T^{-2})(\sigma_j^2 + \lambda_{\max}(\mathbf{F}_S^{-1})\|\mathbf{h}_{k,j}\|_2^2)}{\lambda_{\min}(\mathbf{F}_S^{-2})(\sigma_j^2 + \lambda_{\min}(\mathbf{F}_T^{-1})\|\mathbf{h}_{k,j}\|_2^2)} \quad (19)\end{aligned}$$

where the inequality follows from the Courant–Fischer min–max theorem [32]. Notice that $\lambda_{\max}(\mathbf{F}_S^{-1}) = \lambda_{\min}(\mathbf{F}_S)^{-1}$ and $\lambda_{\min}(\mathbf{F}_T) \geq \lambda_{\min}(\mathbf{F}_S) \geq \lambda_{\min}(\mathbf{F}_\emptyset) = \lambda_{\min}(\mathbf{P}_{k|k-1}^{-1})$ by Lemma 2. This fact, along with the definition of $\mathcal{C}_{\max}$ implies

$$\begin{aligned}\mathcal{C}_{\max} &\leq \max_{j\in[n]} \frac{\lambda_{\max}(\mathbf{P}_{k|k-1})^2(\sigma_j^2 + \lambda_{\max}(\mathbf{P}_{k|k-1})\|\mathbf{h}_{k,j}\|_2^2)}{\lambda_{\max}(\mathbf{F}_S)^{-2}(\sigma_j^2 + \lambda_{\max}(\mathbf{F}_T)^{-1}\|\mathbf{h}_{k,j}\|_2^2)}\\
&\overset{(a)}{\leq} \max_{j\in[n]} \frac{\lambda_{\max}(\mathbf{P}_{k|k-1})^2(\sigma_j^2 + \lambda_{\max}(\mathbf{P}_{k|k-1})\|\mathbf{h}_{k,j}\|_2^2)}{\lambda_{\max}(\mathbf{F}_{[n]})^{-2}(\sigma_j^2 + \lambda_{\max}(\mathbf{F}_{[n]})^{-1}\|\mathbf{h}_{k,j}\|_2^2)}\\
&\overset{(b)}{\leq} \max_{j\in[n]} \frac{\lambda_{\max}(\mathbf{P}_{k|k-1})^2(\sigma_j^2 + \lambda_{\max}(\mathbf{P}_{k|k-1})C)}{\lambda_{\max}(\mathbf{F}_{[n]})^{-2}(\sigma_j^2 + \lambda_{\max}(\mathbf{F}_{[n]})^{-1}C)} \quad (20)\end{aligned}$$

where (a) follows from the fact that $\lambda_{\max}(\mathbf{F}_S) \leq \lambda_{\max}(\mathbf{F}_T) \leq \lambda_{\max}(\mathbf{F}_{[n]})$ and (b) holds since

$$g(x) = \frac{\sigma_j^2 + \lambda_{\max}(\mathbf{P}_{k|k-1})x}{\sigma_j^2 + \lambda_{\max}(\mathbf{F}_{[n]})^{-1}x} \quad (21)$$

is a monotonically increasing function for $x > 0$. Now, since the maximum eigenvalue of a positive-definite matrix satisfies the triangle inequality, we have

$$\begin{aligned}\lambda_{\max}(\mathbf{F}_{[n]}) &\leq \frac{1}{\lambda_{\min}(\mathbf{P}_{k|k-1})} + \lambda_{\max}\left(\sum_{j=1}^n \frac{1}{\sigma_j^2}\mathbf{h}_{k,j}\mathbf{h}_{k,j}^\top\right)\\
&\leq \frac{1}{\lambda_{\min}(\mathbf{P}_{k|k-1})} + \max_{j\in[n]}\frac{1}{\sigma_j^2}\lambda_{\max}(\mathbf{H}_k^\top\mathbf{H}_k). \quad (22)\end{aligned}$$

Therefore, by combining inequalities (15) and (20), we obtain the result in (16).

Next, to analyze the setting of i.i.d random measurement vectors, we bound $\lambda_{\max}(\mathbf{F}_{[n]})$ using Lemma 1. Let $\mathbf{X}_j = \mathbf{h}_{k,j}\mathbf{h}_{k,j}^\top - \sigma_h^2\mathbf{I}_m$ and $\mathbf{Y} = \sum_{j=1}^n \mathbf{X}_j$. To use the result of Lemma 1, one should first verify expressions in (12). To this end, note that

$$\begin{aligned}\mathbb{E}[\mathbf{X}_j] &= \mathbb{E}[\mathbf{h}_{k,j}\mathbf{h}_{k,j}^\top - \sigma_h^2\mathbf{I}_m]\\
&= \mathbb{E}[\mathbf{h}_{k,j}\mathbf{h}_{k,j}^\top] - \sigma_h^2\mathbf{I}_m = \mathbf{0}. \quad (23)\end{aligned}$$

This, in turn, implies that $\mathbb{E}[\mathbf{Y}] = \mathbf{0}$. Since $\mathbf{X}_j$s are independent, then

$$\|\mathbb{E}[\mathbf{Y}^2]\| = \|\mathbb{E}[\sum_{j=1}^n \mathbf{X}_j^2]\| \leq \sum_{j=1}^n \|\mathbb{E}[\mathbf{X}_j^2]\| \quad (24)$$

by the linearity of expectation and the triangle inequality. To proceed, we need to determine $\lambda_{\max}(\mathbf{X}_j)$ and $\mathbb{E}[\mathbf{X}_j^2]$. First, let us verify that $\mathbf{h}_{k,j}$ is an eigenvector of $\mathbf{X}_j$ by observing that

$$\begin{aligned}\mathbf{X}_j\mathbf{h}_j &= \left(\mathbf{h}_{k,j}\mathbf{h}_{k,j}^\top - \sigma_h^2\mathbf{I}_m\right)\mathbf{h}_{k,j}\\
&= \left(\|\mathbf{h}_{k,j}\|_2^2 - \sigma_h^2\right)\mathbf{h}_{k,j} \quad (25)\end{aligned}$$

where $\mathbf{h}_{k,j}\mathbf{h}_{k,j}^\top - \sigma_h^2\mathbf{I}_m$ is the corresponding eigenvalue. Since $\mathbf{h}_{k,j}\mathbf{h}_{k,j}^\top$ is a rank-1 matrix, other eigenvalues of $\mathbf{X}_j$ are all equal to $-\sigma_h^2$. Hence, we have

$$\lambda_{\max}(\mathbf{X}_j) \leq C - \sigma_h^2 \quad (26)$$

and we recall that $C - \sigma_h^2 > 0$. We can now establish an upper bound on $\mathbb{E}[\mathbf{X}_j^2]$ as

$$\begin{aligned}\mathbb{E}[\mathbf{X}_j^2] &= \mathbb{E}[\left(\mathbf{h}_{k,j}\mathbf{h}_{k,j}^\top - \sigma_h^2\mathbf{I}_m\right)\left(\mathbf{h}_{k,j}\mathbf{h}_{k,j}^\top - \sigma_h^2\mathbf{I}_m\right)]\\
&= \left(\|\mathbf{h}_{k,j}\|_2^2 - \sigma_h^2\right)\mathbb{E}[\mathbf{h}_{k,j}\mathbf{h}_{k,j}^\top]\\
&\quad - \sigma_h^2\mathbb{E}[\left(\mathbf{h}_{k,j}\mathbf{h}_{k,j}^\top - \sigma_h^2\mathbf{I}_m\right)]\\
&= \left(\|\mathbf{h}_{k,j}\|_2^2 - \sigma_h^2\right)\sigma_h^2\mathbf{I}_m \preceq (C-\sigma_h^2)\sigma_h^2\mathbf{I}_m \quad (27)\end{aligned}$$

where we have used the fact that $\mathbb{E}[\mathbf{X}_j] = \mathbf{0}$. Thus, $L = C - \sigma_h^2$ and $\|\mathbb{E}[\mathbf{Y}^2]\| \leq n(C-\sigma_h^2)\sigma_h^2$. Now, according to Lemma 1, for all $q > 0$, it holds that $\Pr\{\lambda_{\max}(\mathbf{Y}) \leq q\} \geq p$, where

$$p = 1 - m\exp\left(\frac{-q^2/2}{(C-\sigma_h^2)(n\sigma_h^2+q/3)}\right). \quad (28)$$

Therefore, we have

$$\lambda_{\max}(\mathbf{F}_{[n]}) \leq \frac{1}{\lambda_{\min}(\mathbf{P}_{k|k-1})} + \max_{j\in[n]}\frac{n\sigma_h^2+q}{\sigma_j^2} = \phi^{-1} \quad (29)$$

with probability p. This completes the proof. ■

Remark 1: The setting of i.i.d. random vectors described in Theorem 1 arises in scenarios where sketching techniques, such as random projections, are used to reduce dimensionality of the measurement equation (see [25] for more details). The following are often encountered examples of such settings.

1) *Multivariate Gaussian measurement vectors:* Let $\mathbf{h}_{k,j} \sim \mathcal{N}(0, \frac{1}{m}\mathbf{I}_m)$ for all j. It is straightforward to show that $\mathbb{E}[\|\mathbf{h}_{k,j}\|_2^2] = 1$. Furthermore, it can be shown that $\|\mathbf{h}_{k,j}\|_2^2$ is with high probability concentrated around its expected value. Therefore, for this case, $\sigma_h^2 = \frac{1}{m}$ and $C = 1$.
2) *Centered Bernoulli measurement vectors:* Let each entry of $\mathbf{h}_{k,j}$ be $\pm\frac{1}{\sqrt{m}}$ with equal probability. Therefore, $\|\mathbf{h}_{k,j}\|_2^2 = 1 = C$. Additionally, $\sigma_h^2 = \frac{1}{m}$, since the entries of $\mathbf{h}_{k,j}$ are i.i.d. zero-mean random variables with variance $\frac{1}{m}$.

We can interpret the conditions stated in Theorem 1 as requirements on the condition number of $\mathbf{P}_{k|k-1}$ as argued next. For a sufficiently large m and $\sigma_h^2 = \frac{1}{m}$, it holds that $C \approx 1$. Assume that $\phi \geq \lambda_{\max}(\mathbf{P}_{k|k-1})/\Delta$ for some $\Delta > 1$, and $\sigma_j^2 = \sigma^2$ for all $i \in [n]$. Define

$$\text{SNR} = \frac{\lambda_{\max}(\mathbf{P}_{k|k-1})}{\sigma^2} \tag{30}$$

and let

$$\kappa = \frac{\lambda_{\max}(\mathbf{P}_{k|k-1})}{\lambda_{\min}(\mathbf{P}_{k|k-1})} \geq 1 \tag{31}$$

be the condition number of $\mathbf{P}_{k|k-1}$. Then, following some elementary numerical approximations, we obtain the following corollary.

Corollary 1.1: Let

$$\Delta \geq \kappa + c_1 \frac{n}{m}\text{SNR} \tag{32}$$

for some $c_1 > 1$. Then, with probability

$$p \geq 1 - m \exp\left(-\frac{n}{m}c_2\right) \tag{33}$$

it holds that $\mathcal{C}_{\max} \leq \Delta^3$ for some $c_2 > 0$.

Informally, Theorem 1 states that for a well-conditioned $\mathbf{P}_{k|k-1}$, the curvature of $f(S)$ is small, which implies weak submodularity of $f(S)$. Furthermore, the probability of such an event exponentially increases with the number of available measurements.

IV. Randomized Greedy Sensor Selection

The complexity of SDP relaxation and greedy algorithms for sensor selection becomes prohibitive in large-scale systems. Motivated by the need for practically feasible schemes, we present a randomized greedy algorithm for finding an approximate solution to (9) and derive its performance guarantees. In particular, inspired by the technique in [2] proposed in the context of optimizing submodular objective functions, we develop a computationally efficient randomized greedy algorithm (see Algorithm 1) that finds an approximate solution to (9) with a guarantee on the achievable MSE performance of the Kalman filter that uses only the observations of the selected sensors. Algorithm 1 performs the task of sensor scheduling in the following way. At each iteration of the algorithm, a subset R of size s is sampled uniformly at random and without replacement from the set of available sensors. The marginal gain provided by each of these s sensors to the objective function is computed using (10), and the one yielding the highest marginal gain is added to the set of selected sensors. Then, the efficient recursive formula in (11) is used to update $\mathbf{F}_S^{-1}$ so it can be analyzed when making the selection in the next iteration. This procedure is repeated K times.

Algorithm 1: Randomized Greedy Sensor Scheduling.

Input: $\mathbf{P}_{k|k-1}$, $\mathbf{H}_k$, K, ϵ.
Output: Subset $S_k \subseteq [n]$ with $|S_k| = K$.
Initialize $S_k^{(0)} = \emptyset$, $\mathbf{F}^{-1}_{S_k^{(0)}} = \mathbf{P}_{k|k-1}$.
for $i = 0, \ldots, K-1$ **do**
  Choose R by sampling $s = \frac{n}{K}\log(1/\epsilon)$ indices uniformly at random from $[n]\backslash S_k^{(i)}$.
  $i_s = \text{argmax}_{j\in R} \frac{\mathbf{h}_{k,j}^\top \mathbf{F}^{-2}_{S_k^{(i)}} \mathbf{h}_{k,j}}{\sigma_j^2 + \mathbf{h}_{k,j}^\top \mathbf{F}^{-1}_{S_k^{(i)}} \mathbf{h}_{k,j}}$.
  Set $S_k^{(i+1)} = S_k^{(i)} \cup \{i_s\}$.
  $\mathbf{F}^{-1}_{S_k^{(i+1)}} = \mathbf{F}^{-1}_{S_k^{(i)}} - \frac{\mathbf{F}^{-1}_{S_k^{(i)}} \mathbf{h}_{k,i_s} \mathbf{h}_{k,i_s}^\top \mathbf{F}^{-1}_{S_k^{(i)}}}{\sigma_j^2 + \mathbf{h}_{k,i_s}^\top \mathbf{F}^{-1}_{S_k^{(i)}} \mathbf{h}_{k,i_s}}$
end for
return $S_k = S_k^{(K)}$.

Remark 2: The parameter ϵ in Algorithm 1, $e^{-K} \leq \epsilon < 1$, is a predefined constant that is chosen to strike a desired balance between performance and complexity. When $\epsilon = e^{-K}$, each iteration includes all of the nonselected sensors in R, and Algorithm 1 coincides with the conventional greedy scheme. However, as ϵ approaches 1, $|R|$ and, thus, the overall computational complexity decrease.

A. Performance Analysis of the Proposed Scheme

In this section, we analyze Algorithm 1 and in Theorem 2 provide a bound on the performance of the proposed randomized greedy scheme when applied to finding an approximate solution to maximization problem (9).

Before deriving the main result, we first provide two lemmas. Lemma 3 from [24] states an upper bound on the difference between the values of the objective function corresponding to two sets having different cardinalities, while Lemma 4 provides a lower bound on the expected marginal gain.

Lemma 3: Let $\{\mathcal{C}_l\}_{l=1}^{n-1}$ denote the elementwise curvatures of $f(S)$. Let S and T be any subsets of sensors such that $S \subset T \subseteq [n]$ with $|T\backslash S| = r$. Then, it holds that

$$f(T) - f(S) \leq C(r) \sum_{j\in T\backslash S} f_j(S) \tag{34}$$

where $C(r) = \frac{1}{r}(1 + \sum_{l=1}^{r-1} \mathcal{C}_l)$.

Proof: See Appendix B. ∎

Lemma 4: Let $S_k^{(i)}$ be the set of selected sensors at the end of the ith iteration of Algorithm 1. Then, we have

$$\mathbb{E}\left[f_{(i+1)_s}(S_k^{(i)})|S_k^{(i)}\right] \geq \frac{1-\epsilon^\beta}{K} \sum_{j\in O_k\backslash S_k^{(i)}} f_j(S_k^{(i)}) \qquad (35)$$

where O_k is the set of optimal sensors at time k, $(i+1)_s$ is the index of the selected sensor at the $(i+1)$th iteration, $\beta = 1+\max\{0, \frac{s}{2n} - \frac{1}{2(n-s)}\}$, and $s = \frac{n}{K}\log(1/\epsilon)$.

Proof: See Appendix C. ∎

Theorem 2 specifies how accurate the approximate solution to the sensor selection problem found by Algorithm 1 is. In particular, if $f(S)$ is characterized by a bounded maximum elementwise curvature, Algorithm 1 returns a subset of sensors yielding an objective that is, on average, within a multiplicative factor of the objective achieved by the optimal schedule.

Theorem 2: Let $\mathcal{C}_{\max}$ be the maximum elementwise curvature of $f(S)$, i.e., the objective function of sensor scheduling problem in (9). Let S_k denote the subset of sensors selected by Algorithm 1 at time k, and let O_k be the optimum solution to (9) such that $|O_k| = K$. Then, $f(S_k)$ is on expectation a multiplicative factor away from $f(O_k)$, i.e.,

$$\mathbb{E}\left[f(S_k)\right] \geq \left(1 - e^{-\frac{1}{c}} - \frac{\epsilon^\beta}{c}\right) f(O_k) \qquad (36)$$

where $c = \max\{\mathcal{C}_{\max}, 1\}$, $e^{-K} \leq \epsilon < 1$, and $\beta = 1+\max\{0, \frac{s}{2n} - \frac{1}{2(n-s)}\}$. Furthermore, the computational complexity of Algorithm 1 is $\mathcal{O}(nm^2\log(\frac{1}{\epsilon}))$, where n is the total number of sensors and m is the dimension of $\mathbf{x}_k$.

Proof: Consider $S_k^{(i)}$, the set generated by the end of the ith iteration of Algorithm 1. Employing Lemma 3 with $S = S_k^{(i)}$ and $T = O_k \cup S_k^{(i)}$, and using monotonicity of f, yields

$$\begin{aligned}\frac{f(O_k) - f(S_k^{(i)})}{\frac{1}{r}\left(1+\sum_{l=1}^{r-1}\mathcal{C}_l\right)} &\leq \frac{f(O_k\cup S_k^{(i)}) - f(S_k^{(i)})}{\frac{1}{r}\left(1+\sum_{l=1}^{r-1}\mathcal{C}_l\right)} \\ &\leq \sum_{j\in O_k\backslash S_k^{(i)}} f_j(S_k^{(i)}) \qquad (37)\end{aligned}$$

where $|O_k\backslash S_k^{(i)}| = r$. Now, using Lemma 4, we obtain

$$\mathbb{E}\left[f_{(i+1)_s}(S_k^{(i)})|S_k^{(i)}\right] \geq \left(1-\epsilon^\beta\right)\frac{f(O_k) - f(S_k^{(i)})}{\frac{K}{r}\left(1+\sum_{l=1}^{r-1}\mathcal{C}_l\right)}. \qquad (38)$$

Applying the law of total expectation yields

$$\begin{aligned}\mathbb{E}\left[f_{(i+1)_s}(S_k^{(i)})\right] &= \mathbb{E}\left[f(S_k^{(i+1)}) - f(S_k^{(i)})\right] \\ &\geq \left(1-\epsilon^\beta\right)\frac{f(O_k) - \mathbb{E}\left[f(S_k^{(i)})\right]}{\frac{K}{r}\left(1+\sum_{l=1}^{r-1}\mathcal{C}_l\right)}. \qquad (39)\end{aligned}$$

Using the definition of the maximum elementwise curvature, we obtain

$$\frac{1}{r}\left(1+\sum_{l=1}^{r-1}\mathcal{C}_l\right) \leq \frac{1}{r}(1+(r-1)\mathcal{C}_{\max}) = g(r). \qquad (40)$$

It is easy to verify, e.g., by taking the derivative, that $g(r)$ is decreasing (increasing) with respect to r if $\mathcal{C}_{\max} < 1$ ($\mathcal{C}_{\max} > 1$). Let $c = \max\{\mathcal{C}_{\max}, 1\}$. Then, we have

$$\frac{1}{r}\left(1+\sum_{l=1}^{r-1}\mathcal{C}_l\right) \leq \frac{1}{r}(1+(r-1)\mathcal{C}_{\max}) \leq c. \qquad (41)$$

Hence, we have

$$\mathbb{E}\left[f(S_k^{(i+1)}) - f(S_k^{(i)})\right] \geq \frac{1-\epsilon^\beta}{Kc}\left(f(O_k) - \mathbb{E}\left[f(S_k^{(i)})\right]\right). \qquad (42)$$

Using an inductive argument and due to the fact that $f(\emptyset) = 0$, we obtain

$$\mathbb{E}[f(S_k)] \geq \left(1 - \left(1 - \frac{1-\epsilon^\beta}{Kc}\right)^K\right) f(O_k). \qquad (43)$$

Finally, using the fact that $(1+x)^y \leq e^{xy}$ for $y > 0$ and the easily verifiable fact that $e^{ax} \leq 1 + axe^a$ for $0 < x < 1$, we obtain

$$\begin{aligned}\mathbb{E}[f(S_k)] &\geq \left(1 - e^{-\frac{1-\epsilon^\beta}{c}}\right) f(O_k) \\ &\geq \left(1 - e^{-\frac{1}{c}} - \frac{\epsilon^\beta}{c}\right) f(O_k). \qquad (44)\end{aligned}$$

To take a closer look at computational complexity of Algorithm 1, note that step 6 costs $\mathcal{O}(\frac{n}{K}m^2\log(\frac{1}{\epsilon}))$, since one needs to compute $\frac{n}{K}\log(\frac{1}{\epsilon})$ marginal gains, each requiring $\mathcal{O}(m^2)$ operations. Furthermore, step 8 requires $\mathcal{O}(m^2)$ arithmetic operations. Since there are K such iterations, the running time of Algorithm 1 is $\mathcal{O}(nm^2\log(\frac{1}{\epsilon}))$. This completes the proof. ∎

Using the definition of $f(S)$, we obtain Corollary 2.1 stating that, at each time step, the achievable MSE in (3) obtained by forming an estimate using sensors selected by the randomized greedy algorithm is within a factor of the optimal MSE.

Corollary 2.1: Consider the notation and assumptions of Theorem 2 and introduce $\alpha = 1 - e^{-\frac{1}{c}} - \frac{\epsilon^\beta}{c}$. Let MSE_{S_k} denote the mean square estimation error obtained by forming an estimate using information provided by the sensors selected by Algorithm 1 at time k, and let MSE_o be the optimal MSE formed using information collected by the sensors specified by the optimum solution of (9). Then, the expected MSE_{S_k} is bounded as

$$\mathbb{E}\left[\mathrm{MSE}_{S_k}\right] \leq \alpha\mathrm{MSE}_o + (1-\alpha)\mathrm{Tr}(\mathbf{P}_{k|k-1}). \qquad (45)$$

Remark 3: Since the proposed sensor selection scheme is a randomized algorithm, the analysis of its *expected* MSE, as provided by Theorem 2 and Corollary 2.1, is a meaningful performance characterization. Notice that, as expected, α is decreasing in both c and ϵ. If $f(S)$ is characterized by a small curvature, then $f(S)$ is nearly submodular, and the randomized greedy algorithm delivers a near-optimal sensor scheduling. As we decrease ϵ, α increases, which, in turn, leads to a better approximation factor. Moreover, by following an argument similar to that of the classical analysis in [17], one can show that the approximation factor for the greedy algorithm is given by $\alpha_g = 1 - e^{-\frac{1}{c}}$ (see also [22] and [27]). Therefore, the term $\frac{\epsilon}{c}$ in α denotes the difference between the approximation factors of the proposed randomized greedy algorithm and the conventional greedy scheme.

Remark 4: The computational complexity of the greedy method for sensor selection that finds marginal gains via the efficient recursion given in Proposition 1 is $\mathcal{O}(Knm^2)$. Hence, our proposed scheme provides a reduction in complexity by $K/\log(\frac{1}{\epsilon})$, which may be particularly beneficial in large-scale networks, as illustrated in our simulation results.

Remark 5: In contrast to the results of [2] derived in the context of maximizing monotone submodular functions, Theorem 2 relaxes the submodularity assumption and states that the randomized greedy algorithm does not require submodularity to achieve near-optimal performance. Rather, if the set function is *weak submodular*, Algorithm 1 still selects a subset of sensors that provide an MSE near that achieved by the optimal subset of sensors. In addition, even if the function is submodular (e.g., if we use the $\log\det$ objective instead of the MSE), the results of Theorem 2 offer an improvement over the theoretical results of [2] due to a tighter approximation bound stemming from the analysis presented in the proof of Theorem 2. Moreover, a major assumption in [2] is that R is constructed by sampling with replacement. Clearly, this contradicts the fact that a sensor selected in one iteration will not be in R in the subsequent iteration with probability 1. On the contrary, we assume that R is constructed by sampling without replacement and carry out the analysis in this setting that matches the actual randomized greedy sensor selection strategy.

The randomized selection step of Algorithm 1 can be interpreted as an approximation of the marginal gains of the selected sensors using a greedy scheme [4]. More specifically, for the ith iteration, it holds that $f_{j_{rg}}(S_k^{(i)}) = \eta_k^{(i)} f_{j_g}(S_k^{(i)})$, where subscripts rg and g refer to the sensors selected by the randomized greedy algorithm (see Algorithm 1) and the greedy algorithm, respectively, and $\{\eta_k^{(i)}\}_{i=1}^K$ are random variables with mean $\mu_i(\epsilon)$ that satisfy $0 < \ell_i(\epsilon) \leq \eta_k^{(i)} \leq 1$ for all $i \in [K]$.[1] In view of this argument, we obtain Theorem 3, which states that if $f(S)$ is characterized by a bounded maximum elementwise curvature and $\{\eta_k^{(i)}\}_{i=1}^K$ are independent random variables, Algorithm 1 returns a subset of sensors yielding an objective that, with high probability, is only a multiplicative factor away from the objective achieved by the optimal schedule.

Theorem 3: Instate the notation and assumptions of Theorem 2. Let $\{\eta_k^{(i)}\}_{i=1}^K$ denote a collection of random variables such that $0 < \ell_i(\epsilon) \leq \eta_k^{(i)} \leq 1$, and $\mathbb{E}[\eta_k^{(i)}] = \mu_i(\epsilon)$ for all i and k. Let $\ell_{\min}(\epsilon) = \min_{i,k}\{\ell_i(\epsilon)\}$ and $\mu_{\min}(\epsilon) = \min_{i,k}\{\mu_i(\epsilon)\}$. Then, we have

$$f(S_k) \geq \left(1 - e^{-\frac{\ell_{\min}(\epsilon)}{c}}\right) f(O_k). \tag{46}$$

Furthermore, if $\{\eta_k^{(i)}\}_{i=1}^K$ are independent, then for all $0 < q < 1$ with probability at least $1 - e^{-CK}$, it holds that

$$f(S_k) \geq \left(1 - e^{-\frac{(1-q)\mu_{\min}(\epsilon)}{c}}\right) f(O_k) \tag{47}$$

for some $C > 0$.

Proof: Consider $S_k^{(i)}$, the set generated by the end of the ith iteration of Algorithm 1 and let $(i+1)_g$ and $(i+1)_{rg}$ denote the sensors selected by the greedy and randomized greedy algorithm in the ith iteration, respectively. Let $c = \max\{\mathcal{C}_{\max}, 1\}$. Employing Lemma 3 with $S = S_k^{(i)}$ and $T = O_k \cup S_k^{(i)}$, and using monotonicity of f, yields

$$\begin{aligned} f(O_k) - f(S_k^{(i)}) &\leq f(O_k \cup S_k^{(i)}) - f(S_k^{(i)}) \\ &\leq c \sum_{j \in O_k \setminus S_k^{(i)}} f_j(S_k^{(i)}). \end{aligned} \tag{48}$$

Using the fact that

$$f_j(S_k^{(i)}) \leq f_{(i+1)_{rg}}(S_k^{(i)}) \leq f_{(i+1)_g}(S_k^{(i)}) \tag{49}$$

for all j, we obtain

$$f(O_k) - f(S_k^{(i)}) \leq cK f_{(i+1)_g}(S_k^{(i)}). \tag{50}$$

On the other hand, we have

$$\begin{aligned} f(S_k^{(i+1)}) - f(S_k^{(i)}) &= f_{(i+1)_{rg}}(S_k^{(i)}) \\ &= \eta_k^{(i+1)} f_{(i+1)_g}(S_k^{(i)}). \end{aligned} \tag{51}$$

Combining (50) and (51) yields

$$f(S_k^{(i+1)}) - f(S_k^{(i)}) \geq \frac{\eta_k^{(i+1)}}{Kc}\left(f(O_k) - f(S_k^{(i)})\right). \tag{52}$$

Using an inductive argument similar to the one in the proof of Theorem 2, and noting that $f(\emptyset) = 0$, we have

$$\begin{aligned} f(S_k) &\geq \left(1 - \left(1 - \sum_{i=1}^K \frac{\eta_k^{(i)}}{Kc}\right)\right) f(O_k) \\ &\overset{(a)}{\geq} \left(1 - e^{-\sum_{i=1}^K \frac{\eta_k^{(i)}}{Kc}}\right) f(O_k) \end{aligned} \tag{53}$$

where to obtain (a), we use the fact that $(1+x)^y \leq e^{xy}$ for $y > 0$. Therefore, since, by assumption, $\ell_{\min}(\epsilon) \leq \ell_i(\epsilon) \leq \eta_k^{(i)} \leq 1$, we establish (46).

To show the second statement, i.e., to prove that (47) holds in the setting of independent $\{\eta_k^{(i)}\}_{i=1}^K$, we apply the Bernstein's inequality [33] to the sum of independent random variables $\sum_{i=1}^K \eta_k^{(i)}$. Since $\{\eta_k^{(i)}\}$ are bounded random variables, from Popoviciu's inequality [33] for all $i \in [K]$, it follows that

$$\mathrm{Var}[\eta_k^{(i)}] \leq \frac{1}{4}(1 - \ell_i(\epsilon))^2. \tag{54}$$

[1] Notice that $\ell_i(\epsilon)$ and $\mu_i(\epsilon)$ are time-varying quantities, where the time index is omitted for the simplicity of notation.

Hence, based on the Bernstein's inequality, for all $0 < q < 1$, we have

$$\Pr\left\{\sum_{i=1}^{K}\eta_k^{(i)} < (1-q)\sum_{i=1}^{K}\mu_i\right\} < p \tag{55}$$

where

$$\begin{aligned} p &= \exp\left(-\frac{(1-q)^2(\sum_{i=1}^{K}\mu_i(\epsilon))^2}{\frac{1-q}{3}\sum_{i=1}^{K}\mu_i(\epsilon)+\frac{1}{4}\sum_{i=1}^{K}(1-\ell_i(\epsilon))^2}\right)\\ &\overset{(b)}{\leq} \exp\left(-\frac{K(1-q)^2\mu_{\min}^2(\epsilon)}{\frac{1-q}{3}\mu_{\min}(\epsilon)+\frac{1}{4}(1-\ell_{\min}(\epsilon))^2}\right)\\ &= e^{-C(\epsilon,q)K} \end{aligned} \tag{56}$$

where (b) follows because p increases as we replace $\mu_i(\epsilon)$ and $\ell_i(\epsilon)$ by their lower bounds. Finally, substituting this result into (53) yields

$$f(S_k) \geq \left(1 - e^{-\frac{(1-q)\mu_{\min}(\epsilon)}{c}}\right) f(O_k) \tag{57}$$

with probability at least $1 - e^{C(\epsilon,q)K}$. This completes the proof. ■

Our simulation studies in Section V empirically confirm the results of Theorems 2 and 3 and illustrate that Algorithm 1 performs favorably compared to the competing schemes both on average as well as for each individual sensor scheduling task.

Similar to Corollary 2.1, we can now obtain a probabilistic bound on the MSE (3) achievable at each time step using the proposed randomized greedy algorithm. This result is stated in Corollary 3.1.

Corollary 3.1: Consider the notation and assumptions of Corollary 2.1 and Theorem 3. Let $0 < q < 1$ and define $\alpha = 1 - \exp(-\frac{(1-q)\mu_{\min}(\epsilon)}{c})$. Then, with probability at least $1 - e^{-CK}$, it holds that

$$\mathrm{MSE}_{S_k} \leq \alpha \mathrm{MSE}_o + (1-\alpha)\mathrm{Tr}(\mathbf{P}_{k|k-1}) \tag{58}$$

for some $C > 0$.

V. Simulation Results

To test the performance of the proposed randomized greedy algorithm, we compare it with the classic greedy algorithm and the SDP relaxation in a variety of settings, as detailed next. We implemented the greedy and randomized greedy algorithms in MATLAB and the SDP relaxation scheme via CVX [34]. All simulations were run on a laptop with 2.0-GHz Intel Core i7-4510 U CPU and 8.00 GB of RAM.

A. Kalman Filtering in Random Sensor Networks

We first consider the problem of state estimation in a linear time-varying system via Kalman filtering. For simplicity, we assume the state transition matrix to be identity, i.e., $\mathbf{A}_k = \mathbf{I}_m$. At each time step, the measurement vectors, i.e., the rows of the measurement matrix $\mathbf{H}_k$, are drawn according to $\mathcal{N} \sim (0, \frac{1}{m}\mathbf{I}_m)$. The initial state is a zero-mean Gaussian random

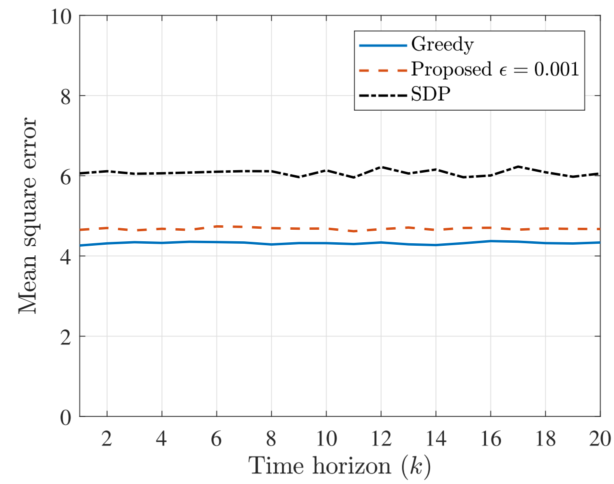


Fig. 2. MSE comparison of randomized greedy, greedy, and SDP relaxation sensor selection schemes employed in Kalman filtering.

TABLE I
Running Time Comparison of the Randomized Greedy, Greedy, and SDP Relaxation Sensor Selection Schemes ($m = 50$, $n = 400$, $K = 55$, and $\epsilon = 0.001$)

Randomized Greedy	Greedy	SDP Relaxation
0.20 s	0.38 s	249.86 s

vector with covariance $\mathbf{\Sigma_x} = \mathbf{I}_m$, and the process and measurement noise are zero-mean Gaussian with covariance matrices $\mathbf{Q} = 0.05\mathbf{I}_m$ and $\mathbf{R} = 0.05\mathbf{I}_n$, respectively.

The MSE of the filtered estimator and running time of each scheme is averaged over 100 Monte Carlo simulations. The time horizon for each run is $T = 10$ s.

We first consider a system having state dimension $m = 50$ and the total number of sensors $n = 400$. We set a constraint on the number of sensors allowed to be queried at each time step to $K = 55$ and compare the MSE achieved by each sensor selection method over the time horizon of interest. For the randomized greedy algorithm, we set $\epsilon = 0.001$. Fig. 2 shows that the greedy method consistently yields the lowest estimation MSE, while the MSE provided by the randomized greedy algorithm is slightly higher. The MSE performance achieved by solving the SDP relaxation is considerably larger than those of the greedy and randomized greedy algorithms. The time it takes each method to select K sensors is given in Table I. Both the greedy algorithm and the randomized greedy algorithm are much faster than the SDP formulation. Moreover, the randomized greedy scheme is nearly two times faster than the greedy method.

Note that, in this example, in each iteration of the sensor selection procedure, the randomized scheme only computes the marginal gain for a sampled subset of size 50. In contrast, the classic greedy approach computes the marginal gain for all 400 sensors. In summary, the greedy method yields slightly lower MSE but is much slower than the proposed randomized greedy algorithm.

To study the effect of the number of selected sensors on the MSE performance, we vary K from 55 to 115 with increments

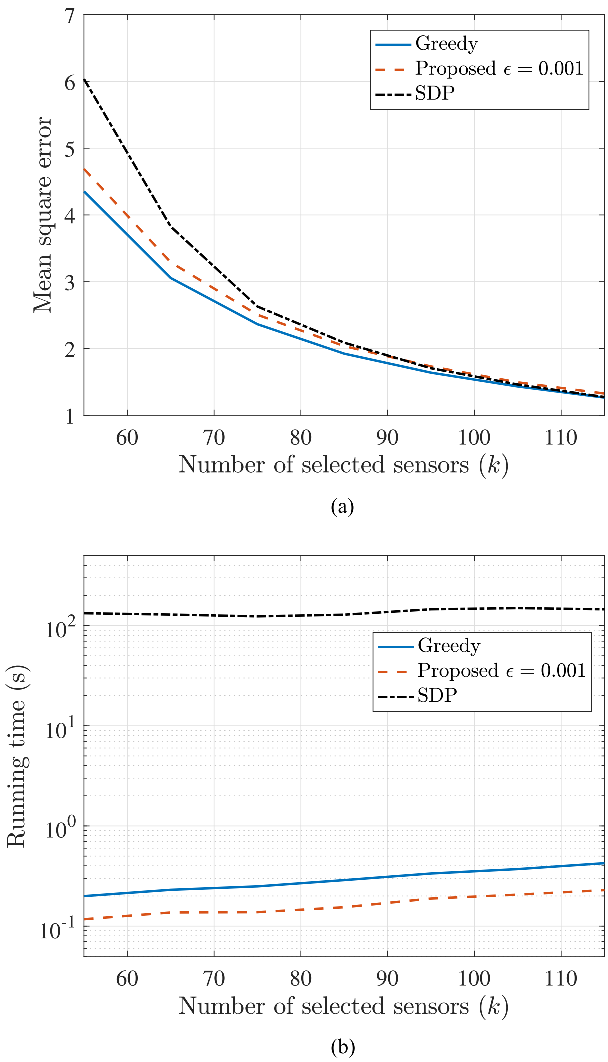


Fig. 3. Comparison of randomized greedy, greedy, and SDP relaxation schemes as the number of selected sensors increases. (a) Comparing MSE performance of different schemes. (b) Running time comparison.

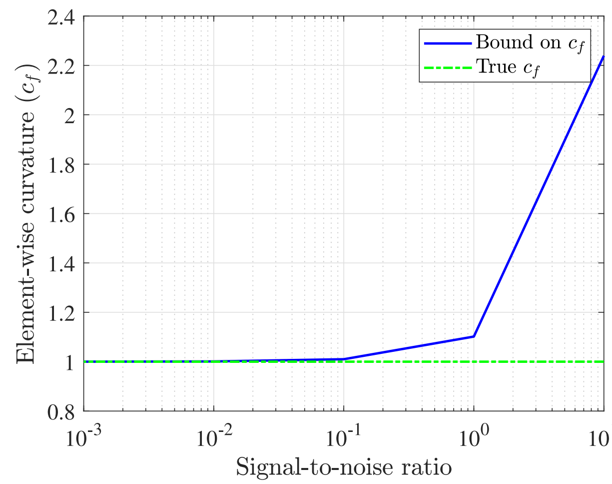


Fig. 4. Evaluation of theoretical results in Theorem 1 for a sensor network with $m = 3$ and $n = 12$.

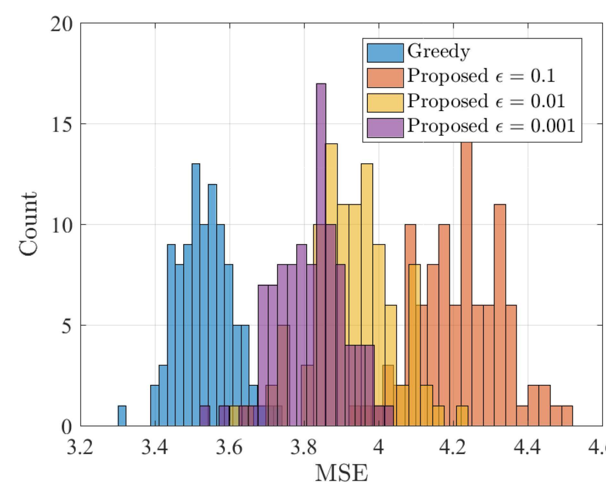


Fig. 5. Histogram of MSE values for 100 independent realization of a sensor scheduling task for a sensor network with $m = 50$, $K = 60$, and $n = 400$.

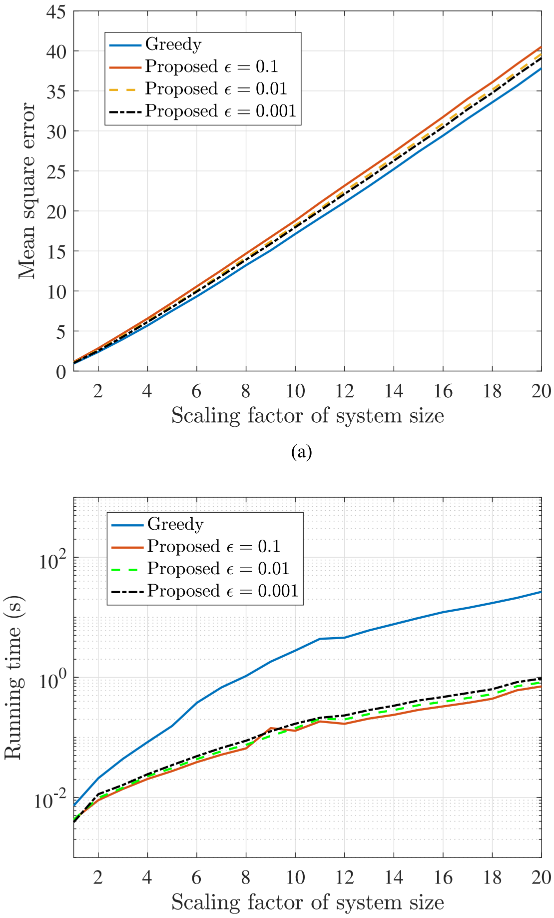
of 10. The MSE values at the last time step for each algorithm are shown in Fig. 3(a). As the number of selected sensors increases, the estimation becomes more accurate, as reflected by the MSE of the estimates provided by each algorithm. Moreover, the differences between the MSE values achieved by different schemes monotonically decrease as more sensors are selected. The sensor selection running times shown in Fig. 3(b) indicate that the randomized greedy scheme is nearly twice as fast as the greedy method, while the SDP method is orders of magnitude slower than both greedy and randomized greedy algorithms.

To test the tightness of the bound established in Theorem 1, we empirically study a sensor selection problem with $n = 12$ Gaussian observations. Fig. 4 shows the true values of the maximum elementwise curvature found via exhaustive search as well as the bound stated in Theorem 1. As can be seen in the figure, the gap between the two is negligible at small SNR but becomes relatively loose at high SNR.

Finally, to empirically verify the results of Theorem 3, in Fig. 5, we compare histograms of MSE achieved by the greedy and the proposed randomized greedy sensor selection schemes with various choices of ϵ when $K = 60$. As shown in Fig. 5, the MSE of sets selected by the proposed scheme is relatively close to that selected by the state-of-the-art greedy algorithm. In addition, as ϵ decreases, the MSE of the randomized greedy algorithm approaches that of the greedy algorithm. These empirical observations coincide with our theoretical results in Theorem 3. That is, the proposed algorithm, although a randomized scheme, returns a near-optimal subset of sensors for each individual sensor selection task.

B. State Estimation in Large-Scale Networks

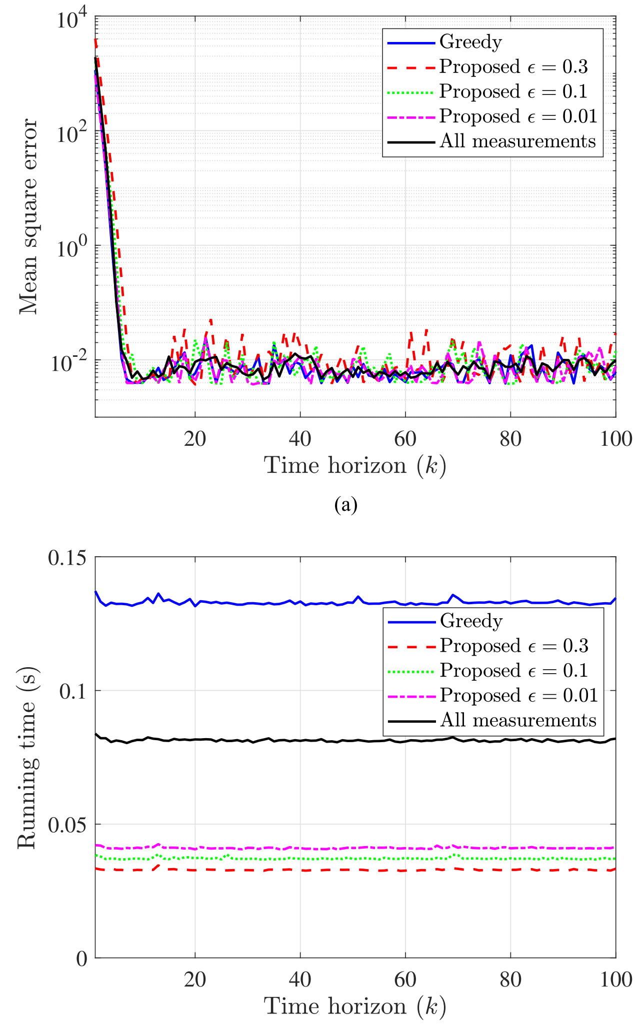
Next, we compare the performance of the randomized greedy algorithm to that of the greedy algorithm as the size of the system increases. We run both methods for 20 different system dimensions. The initial dimensions are set to $m = 20$, $n = 200$, and $K = 25$, and all three parameters are scaled by γ, where γ varies from 1 to 20. In addition, to evaluate the effect of ϵ on the



(a)

(b)

Fig. 6. Comparison of the randomized greedy and greedy algorithms for varied network size. (a) Comparing MSE performance of different schemes. (b) Running time comparison.



(a)

(b)

Fig. 7. Comparison of the randomized greedy and greedy algorithms for a multiobject tracking application. (a) Comparing MSE performance of different schemes. (b) Running time comparison.

performance and runtime of the randomized greedy approach, we repeat experiments for $\epsilon \in \{0.1, 0.01, 0.001\}$. Note that the computational complexity of the SDP relaxation scheme is prohibitive in this setting, and hence, it is omitted. Fig. 6(a) illustrates the MSE comparison of the greedy and randomized greedy schemes. It shows that the difference between the MSEs is negligible. The running time is plotted in Fig. 6(b). As the figure illustrates, the gap between the running times grows with the size of the system, and the randomized greedy algorithm performs nearly 28 times faster than the greedy method for the largest network. Fig. 6 shows that using a smaller ϵ results in a lower MSE, while it slightly increases the running time. These results suggest that, for large systems, the randomized greedy provides almost the same MSE while being much faster than the greedy algorithm.

C. Accelerated Multiobject Tracking

Finally, we study the multiobject tracking application introduced in Section II-A. Specifically, we consider a scenario, where 20 moving objects are initially uniformly distributed in a 5×10 area. At each time instance, each object moves in a random direction with a constant speed set to 0.2. Twenty UAVs, equidistantly spread over the area, move according to a periodic *parallel-path* search pattern [35]. The initial phases of the UAVs' motions are uniformly distributed to provide a better coverage of the area. The UAVs can acquire range and angular measurements of the objects that are within the maximum radar detection range. The maximum radar detection range is set such that at each time step, the UAVs together collect approximately 600 range and angular measurements. The communication bandwidth constraints limit the number of measurements transmitted to the control unit to $K = 100$. Note that since the radar measurement model is nonlinear, the control unit tracks objects via the extended Kalman filter. Fig. 7 shows a comparison in terms of the MSE and running time between the greedy and randomized greedy schemes for various values of ϵ. In Fig. 7, we show the performance of the scheme that ignores communications constraints and uses all the available measurements gathered by the UAVs. As Fig. 7(a) illustrates, the MSE performance of the greedy and proposed schemes are relatively close and similar to the performance of the scheme that uses all the measurements. However, a closer look at the running time comparison shown in Fig. 7(b) reveals that the combined runtime of randomized greedy sensor selection and Kalman filtering tasks is approximately two times faster than the runtime of the Kalman filter that uses all the measurements, and approximately four times faster than the combined runtime of the classical greedy sensor selection and Kalman filtering. Therefore, the proposed scheme not only satisfies the communication constraint and performs nearly as well as using all the measurements, but also significantly reduces the time needed to perform sensor selection and process the selected measurements in extended Kalman filtering.

VI. Conclusion

In this article, we studied the problem of state estimation in large-scale linear time-varying dynamical systems. We proposed a randomized greedy algorithm for selecting sensors to query such that their choice minimizes the estimator's MSE at each time step. We established the performance guarantee for the proposed algorithm and analyzed its computational complexity. To our knowledge, the proposed scheme is the first randomized algorithm for sensor scheduling with an explicit bound on its achievable MSE. In addition, we provided a probabilistic theoretical bound on the elementwise curvature of the objective function. Furthermore, in several simulated settings, we demonstrated that the proposed algorithm is superior to the classical greedy and SDP relaxation methods in terms of running time while providing the same or better utility.

As a future work, it is of interest to extend this approach to nonlinear dynamical systems and obtain theoretical guarantees on the quality of the resulting approximate solution found by the randomized greedy algorithm. Moreover, it would be of interest to extend the framework established in this article to related problems, such as the minimal actuator placement.

Appendix A
Proof of Proposition 1

First, note that

$$f(\emptyset) = \operatorname{Tr}\left(\mathbf{P}_{k|k-1} - \mathbf{F}_\emptyset^{-1}\right) = \operatorname{Tr}\left(\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\right) = 0. \tag{59}$$

Now, for $j \in [n]\backslash S$, it holds that

$$\begin{aligned} f_j(S) &= f(S\cup\{j\}) - f(S) \\ &= \operatorname{Tr}\left(\mathbf{P}_{k|k-1} - \mathbf{F}_{S\cup\{j\}}^{-1}\right) - \operatorname{Tr}\left(\mathbf{P}_{k|k-1} - \mathbf{F}_S^{-1}\right) \\ &= \operatorname{Tr}\left(\mathbf{F}_S^{-1}\right) - \operatorname{Tr}\left(\mathbf{F}_{S\cup\{j\}}^{-1}\right) \\ &= \operatorname{Tr}\left(\mathbf{F}_S^{-1}\right) - \operatorname{Tr}\left(\left(\mathbf{F}_S + \sigma_j^{-2}\mathbf{h}_{k,j}\mathbf{h}_{k,j}^\top\right)^{-1}\right) \\ &\overset{(a)}{=} \operatorname{Tr}\left(\frac{\mathbf{F}_S^{-1}\mathbf{h}_{k,j}\mathbf{h}_{k,j}^\top\mathbf{F}_S^{-1}}{\sigma_j^2 + \mathbf{h}_{k,j}^\top\mathbf{F}_S^{-1}\mathbf{h}_{k,j}}\right) \\ &\overset{(b)}{=} \frac{\mathbf{h}_{k,j}^\top\mathbf{F}_S^{-2}\mathbf{h}_{k,j}}{\sigma_j^2 + \mathbf{h}_{k,j}^\top\mathbf{F}_S^{-1}\mathbf{h}_{k,j}} \end{aligned} \tag{60}$$

where (a) is obtained by applying matrix inversion lemma (Sherman–Morrison formula) [32] to $(\mathbf{F}_S + \sigma_j^{-2}\mathbf{h}_{k,j}\mathbf{h}_{k,j}^\top)^{-1}$, and (b) follows from the properties of the matrix trace operator. Finally, since $\mathbf{F}_S$ is a symmetric positive-definite matrix, $f_j(S) > 0$, which, in turn, implies monotonicity.

Appendix B
Proof of Lemma 3

Let $S \subset T$ and $T\backslash S = \{j_1,\dots,j_r\}$. Therefore, we have

$$\begin{aligned} f(T) - f(S) &= f(S \cup \{j_1,\dots,j_r\}) - f(S) \\ &= f_{j_1}(S) + f_{j_2}(S\cup\{j_1\}) + \dots \\ &\quad + f_{j_r}(S\cup\{j_1,\dots,j_{r-1}\}). \end{aligned} \tag{61}$$

Definition of the elementwise curvature implies that

$$\begin{aligned} f(T) - f(S) &\le f_{j_1}(S) + \mathcal{C}_1 f_{j_2}(S) + \cdots + \mathcal{C}_{r-1} f_{j_r}(S) \\ &= f_{j_1}(S) + \sum_{l=1}^{r-1}\mathcal{C}_l f_{j_{l+1}}(S). \end{aligned} \tag{62}$$

Note that (62) is established for a specific ordering of elements in $T\backslash S$. Given an ordering $\{j_1,\dots,j_r\}$, one can form a set $P = \{\mathcal{P}_1,\dots,\mathcal{P}_r\}$ of r permutations (e.g., by defining the right circular-shift operator $\mathcal{P}_t(\{j_1,\dots,j_r\}) = \{j_{r-t+1},\dots,j_1,\dots\}$ for $1 \le t \le r$) such that $\mathcal{P}_p(j) \neq \mathcal{P}_q(j)$ for $p \neq q$ and $\forall j \in T\backslash S$; (62) holds for each such permutation. By summing the corresponding r inequalities, we obtain

$$r(f(T) - f(S)) \le \left(1 + \sum_{l=1}^{r-1}\mathcal{C}_l\right)\sum_{j\in T\backslash S} f_j(S). \tag{63}$$

Rearranging (63) yields the desired result.

Appendix C
Proof of Lemma 4

First, we aim to bound the probability of an event that a random set R contains at least one index from the optimal set of sensors, which is a necessary condition to reach the optimal MSE. Let us consider $S_t^{(i)}$, the set of sensors selected by the end of the ith iteration of Algorithm 1 and let $\Phi = R \cap (O_k\backslash S_t^{(i)})$. It holds that[2]

$$\begin{aligned} \Pr\{\Phi = \emptyset\} &= \prod_{l=0}^{s-1}\left(1 - \frac{|O_k\backslash S_k^{(i)}|}{|[n]\backslash S_k^{(i)}| - l}\right) \\ &\overset{(a)}{\le} \left(1 - \frac{|O_k\backslash S_k^{(i)}|}{s}\sum_{l=0}^{s-1}\frac{1}{|[n]\backslash S_k^{(i)}| - l}\right)^s \\ &\overset{(b)}{\le} \left(1 - \frac{|O_k\backslash S_k^{(i)}|}{s}\sum_{l=0}^{s-1}\frac{1}{n-l}\right)^s \end{aligned} \tag{64}$$

where (a) holds due to the inequality of arithmetic and geometric means, and (b) holds since $|[n]\backslash S_i| \le n$. Now, recall that for any

[2] Without loss of generality, we assume that s is an integer.

integer p,

$$H_p = \sum_{l=1}^{p} \frac{1}{p} = \log p + \gamma + \zeta_p \quad (65)$$

where H_p is the pth harmonic number, γ is the Euler–Mascheroni constant, and $\zeta_p = \frac{1}{2p} - \mathcal{O}_{\|}(\frac{1}{p^4})$ is a monotonically decreasing sequence related to the Hurwitz zeta function [37]. Therefore, using the identity (65) we obtain

$$\begin{aligned}\Pr\{\Phi = \emptyset\} &\leq \left(1 - \frac{|O_k \backslash S_k^{(i)}|}{s}(H_n - H_{n-s})\right)^s \\ &= \left(1 - \frac{|O_k \backslash S_k^{(i)}|}{s}\left(\log\left(\frac{n}{n-s}\right) + \zeta_n - \zeta_{n-s}\right)\right)^s \\ &\overset{(c)}{\leq} \left(1 - \frac{|O_k \backslash S_k^{(i)}|}{s}\left(\log\left(\frac{n}{n-s}\right) - \frac{s}{2n(n-s)}\right)\right)^s \\ &\overset{(d)}{\leq} \left(\left(1 - \frac{s}{n}\right) e^{\frac{s}{2n(n-s)}}\right)^{|O_k \backslash S_k^{(i)}|} \quad (66)\end{aligned}$$

where (c) follows since $\zeta_n - \zeta_{n-s} = \frac{1}{2n} - \frac{1}{2(n-s)} + \mathcal{O}_{\|}(\frac{1}{(n-s)^4})$, and (d) is due to the fact that $(1+x)^y \leq e^{xy}$ for any real number $y \geq 1$. Next, the fact that $\log(1-x) \leq -x - \frac{x^2}{2}$ for $0 < x < 1$ yields

$$(1 - \frac{s}{n})e^{\frac{s}{2n(n-s)}} \leq e^{-\frac{\beta_1 s}{n}} \quad (67)$$

where $\beta_1 = 1 + (\frac{s}{2n} - \frac{1}{2(n-s)})$. On the other hand, we can also upper-bound $\Pr\{\Phi = \emptyset\}$ as

$$\begin{aligned}\Pr\{\Phi = \emptyset\} &\leq \left(1 - \frac{|O_k \backslash S_k^{(i)}|}{s} \sum_{l=0}^{s-1} \frac{1}{n-l}\right)^s \\ &\leq \left(1 - \frac{|O_k \backslash S_k^{(i)}|}{n}\right)^s \\ &\leq e^{-\frac{s}{n}|O_k \backslash S_k^{(i)}|} \quad (68)\end{aligned}$$

where we again employed the inequality $(1+x)^y \leq e^{xy}$. Let us denote $\beta = \max\{1, \beta_1\}$. Then, we have

$$\Pr\{\Phi \neq \emptyset\} \geq 1 - e^{-\frac{\beta s}{n}|O_k \backslash S_k^{(i)}|} \geq \frac{1 - \epsilon^\beta}{K}(|O_k \backslash S_k^{(i)}|) \quad (69)$$

by the definition of s and the fact that $1 - e^{-\frac{\beta s}{n}x}$ is a concave function. Finally, according to [2, Lemma 2], we have

$$\mathbb{E}[f_{(i+1)_s}(S_k^{(i)})|S_k^{(i)}] \geq \frac{\Pr\{\Phi \neq \emptyset\}}{|O_k \backslash S_k^{(i)}|} \sum_{j \in O_k \backslash S_k^{(i)}} f_o(S_k^{(i)}). \quad (70)$$

Combining (69) and (70) yields the stated results.

References

[1] A. Hashemi, M. Ghasemi, H. Vikalo, and U. Topcu, "A randomized greedy algorithm for near-optimal sensor scheduling in large-scale sensor networks," in *Proc. IEEE Amer. Control Conf.*, Jun. 2018, pp. 1027–1032.

[2] B. Mirzasoleiman, A. Badanidiyuru, A. Karbasi, J. Vondrak, and A. Krause, "Lazier than lazy greedy," in *Proc. 29th AAAI Conf. Artif. Intell.*, Jan. 2015, pp. 1812–1818.

[3] A. Nordio, A. Tarable, F. Dabbene, and R. Tempo, "Sensor selection and precoding strategies for wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 63, no. 16, pp. 4411–4421, Jun. 2015.

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

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

[6] V. Tzoumas, N. A. Atanasov, A. Jadbabaie, and G. J. Pappas, "Scheduling nonlinear sensors for stochastic process estimation," in *Proc. IEEE Amer. Control Conf.*, Jul. 2017, pp. 580–585.

[7] V. Tzoumas, A. Jadbabaie, and G. J. Pappas, "Near-optimal sensor scheduling for batch state estimation: Complexity, algorithms, and limits," in *Proc. IEEE Conf. Decis. Control*, Dec. 2016, pp. 2695–2702.

[8] T. H. Summers, F. L. Cortesi, and J. Lygeros, "On submodularity and controllability in complex dynamical networks," *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 1, pp. 91–101, Mar. 2016.

[9] V. Tzoumas, M. A. Rahimian, G. J. Pappas, and A. Jadbabaie, "Minimal actuator placement with optimal control constraints," in *Proc. Amer. Control Conf.*, Jul. 2015, pp. 2081–2086.

[10] M. G. Damavandi, V. Krishnamurthy, and J. R. Martí, "Robust meter placement for state estimation in active distribution systems," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1972–1982, Feb. 2015.

[11] N. Gensollen, V. Gauthier, M. Marot, and M. Becker, "Submodular optimization for control of prosumer networks," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, Nov. 2016, pp. 180–185.

[12] Z. Liu, A. Clark, P. Lee, L. Bushnell, D. Kirschen, and R. Poovendran, "Towards scalable voltage control in smart grid: A submodular optimization approach," in *Proc. ACM/IEEE 7th Int. Conf. Cyber-Phys. Syst.*, Apr. 2016, pp. 1–10.

[13] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection under channel uncertainty," *IEEE Wireless Commun. Lett.*, vol. 1, no. 4, pp. 376–379, Aug. 2012.

[14] W. J. Welch, "Branch-and-bound search for experimental designs based on D optimality and other criteria," *Technometrics*, vol. 24, no. 1, pp. 41–48, 1982.

[15] D. P. Williamson and D. B. Shmoys, *The Design of Approximation Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

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

[17] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions I," *Math. Program.*, vol. 14, no. 1, pp. 265–294, Dec. 1978.

[18] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, pp. 235–284, Feb. 2008.

[19] S. T. Jawaid and S. L. Smith, "Submodularity and greedy algorithms in sensor scheduling for linear dynamical systems," *Automatica*, vol. 61, pp. 282–288, Nov. 2015.

[20] H. Zhang, R. Ayoub, and S. Sundaram, "Sensor selection for optimal filtering of linear dynamical systems: Complexity and approximation," in *Proc. Conf. Decis. Control*, Dec. 2015, pp. 5002–5007.

[21] P. Singh, S. Z. Yong, and E. Frazzoli, "Supermodular batch state estimation in optimal sensor scheduling," *IEEE Control Syst. Lett.*, vol. 1, no. 2, pp. 292–297, Oct. 2017.

[22] P. Singh, M. Chen, L. Carlone, S. Karaman, E. Frazzoli, and D. Hsu, "Supermodular mean squared error minimization for sensor scheduling in optimal Kalman filtering," in *Proc. Amer. Control Conf.*, 2017, pp. 5787–5794.

[23] A. Olshevsky, "On (non) supermodularity of average control energy," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1177–1181, Sep. 2018.

[24] Z. Wang, B. Moran, X. Wang, and Q. Pan, "Approximation for maximizing monotone non-decreasing set functions with a greedy method," *J. Combinatorial Optim.*, vol. 31, no. 1, pp. 29–43, Jan. 2016.

[25] D. Berberidis and G. B. Giannakis, "Data sketching for large-scale Kalman filtering," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3688–3701, Jul. 2017.

[26] L. Chamon and A. Ribeiro, "Approximate supermodularity bounds for experimental design," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, Dec. 2017, pp. 5409–5418.

[27] L. F. Chamon, G. J. Pappas, and A. Ribeiro, "The mean square error in Kalman filtering sensor selection is approximately supermodular," in *Proc. IEEE Conf. Decis. Control*, Dec. 2017, pp. 343–350.

[28] U. Feige, "A threshold of ln n for approximating set cover," *J. ACM*, vol. 45, no. 4, pp. 634–652, Jul. 1998.

[29] H. Zhang and Y. Vorobeychik, "Submodular optimization with routing constraints," in *Proc. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 819–825.

[30] M. Sviridenko, J. Vondrák, and J. Ward, "Optimal approximation for submodular and supermodular optimization with bounded curvature," *Math. Oper. Res.*, vol. 42, no. 4, pp. 1197–1218, 2017.

[31] J. A. Tropp, "An introduction to matrix concentration inequalities," *Found. Trends Mach. Learn.*, vol. 8, nos. 1/2, pp. 1–230, May 2015.

[32] R. Bellman, *Introduction to Matrix Analysis*. Philadelphia, PA, USA: SIAM, 1997.

[33] R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics*. Upper Saddle River, NJ, USA: Prentice-Hall, 1995.

[34] M. Grant, S. Boyd, and Y. Ye, *CVX: Matlab Software for Disciplined Convex Programming*. Austin, TX, USA: CVX Research, Inc., 2008.

[35] P. Vincent and I. Rubin, "A framework and analysis for cooperative search using UAV swarms," in *Proc. ACM Symp. Appl. Comput.*, 2004, pp. 79–86.

[36] S. Liu, S. P. Chepuri, M. Fardad, E. Masazade, G. Leus, and P. K. Varshney, "Sensor selection for estimation with correlated measurement noise," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3509–3522, Jul. 2016.

[37] S. Lang, *Algebraic Number Theory*, vol. 110. New York, NY, USA: Springer, 2013.

Abolfazl Hashemi (Student Member, IEEE) received the B.S. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2014, and the M.S.E. degree in electrical engineering, in 2016 from the University of Texas at Austin, Austin, TX, USA, where he is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering.

He was a Visiting Undergraduate Researcher with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, in the summer of 2013. His research interests include machine learning, signal processing, and control.

Mr. Hashemi was the recipient of the Iranian National Elite Foundation Fellowship and a Best Student Paper Award finalist at the 2018 American Control Conference.

Mahsa Ghasemi (Student Member, IEEE) received the B.S. degree in mechanical engineering from the Sharif University of Technology, Tehran, Iran, in 2014, and the M.S.E. degree in 2017 from the University of Texas at Austin, Austin, TX, USA, where she is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering.

Her research interests include control systems, formal methods, robotics, and artificial intelligence.

Ms. Ghasemi was the recipient of the Iranian National Elite Foundation Fellowship (2010–2014) and the Cockrell School of Engineering Fellowship at the University of Texas at Austin (2014–2015). She is a member of the American Society of Mechanical Engineers and the Society for Industrial and Applied Mathematics.

Haris Vikalo (Senior Member, IEEE) received the B.S. degree from the University of Zagreb, Zagreb, Croatia, in 1995, the M.S. degree from Lehigh University, Bethlehem, PA, USA, in 1997, and the Ph.D. degree from Stanford University, Stanford, CA, USA, in 2003, all in electrical engineering.

He held a short-term appointment with Bell Laboratories, Murray Hill, NJ, USA, in the summer of 1999. He was a Postdoctoral Researcher from January 2003 to July 2003 and an Associate Scientist from July 2003 to August 2007 with the California Institute of Technology, CA, USA. Since September 2007, he has been with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, USA, where he is currently an Associate Professor. His research interests include signal processing, machine learning, communications, and bioinformatics.

Dr. Vikalo is a recipient of the 2009 National Science Foundation Career Award.

Ufuk Topcu received the Ph.D. degree from the University of California at Berkeley, Berkeley, CA, USA, in 2008.

He joined the Department of Aerospace Engineering, The University of Texas at Austin, Austin, TX, USA, as an Assistant Professor in Fall 2015. He held research positions with the University of Pennsylvania and the California Institute of Technology. His research interests include theoretical, algorithmic, and computational aspects of design and verification of autonomous systems through novel connections between formal methods, learning theory, and controls.