On Submodularity of Quadratic Observation Selection in Constrained Networked Sensing Systems

Mahsa Ghasemi†, Abolfazl Hashemi†, Ufuk Topcu, and Haris Vikalo

***Abstract*— We study the problem of observation selection in a resource-constrained networked sensing system, where the objective is to select a small subset of observations from a large network to perform a state estimation task. When the measurements are gathered using nonlinear systems, majority of prior work resort to approximation techniques such as linearization of the measurement model to utilize the methods developed for linear models, e.g., (weak) submodular objectives and greedy selection schemes. In contrast, when the measurement model is quadratic, e.g., the range measurements in a radar system, by exploiting a connection to the classical Van Trees' inequality, we derive new optimality criteria without distorting the relational structure of the measurement model. We further show that under certain conditions these optimality criteria are monotone and (weak) submodular set functions. These results enable us to develop an efficient greedy observation selection algorithm uniquely tailored for constrained networked sensing systems following quadratic models and provide theoretical bounds on its achievable utility. Extensive numerical experiments demonstrate efficacy of the proposed framework.**

I. Introduction

The problem of state estimation in a network of sensing units that are capable of exchanging information arises in a variety of settings. An example is the task of multi-target tracking via a swarm of unmanned aerial vehicles (UAVs) where the UAVs sense a collection of objects with unknown locations through a nonlinear measurement model [1].

In addition to the challenge of nonlinearity, often due to constraints on computation, power and communication resources, instead of estimating the unknown states using information collected by the entire network, the fusion center (e.g., the control unit in the UAV example) typically queries a relatively small subset of the available observations.

For a variety of performance criteria and measurement models, finding an optimal subset of observations requires solving a computationally challenging combinatorial optimization problem, possibly using branch-and-bound search [2]. By reducing it to the set cover problem, the task of observation selection was in fact shown to be NP-hard [3]. Due to this hardness result, heuristic schemes that find a sub-optimal subset of observations are sought after. It has been shown that when the measurement model is linear, many of the performance criteria which are scalarizations of the error covariance matrix (e.g., $\log\det$ of its inverse or its trace), possess a diminishing return property known as submodularity or weak (a.k.a. approximate) submodularity. For such objectives, a simple greedy approximation scheme achieves near-optimal observation selection with provable performance guarantee [4], [5]. Examples of such greedy observation selection schemes which are developed in various problems in control systems, signal processing, and machine learning include sensor selection for Kalman filtering [6], [7], batch state estimation and stochastic process estimation [8], [9], minimal actuator placement [10], subset selection in machine learning [11], and sensor scheduling in wireless sensor networks [6], [12]. None of these schemes however consider the case of nonlinear measurement models as in these scenarios the error covariance matrix is in general unknown. Some important instances of nonlinearity are quadratic measurement models and inverse problems that occur in many natural phenomena and real-world applications. For instance, in object tracking and localization applications in robotics and autonomous systems, the range measurements gathered by the radar systems follow a quadratic relation [13], [14].

To arrive at a (weak) submodular objective in settings where the model is nonlinear, existing schemes resort to approximation techniques, e.g., linearizing the model (the so-called local optimality approach) prior to the actual observation selection step [15]–[22]. However, theoretical guarantees for the performance of greedy algorithms hold only for the linearized model, i.e., for the linear approximation of the actual nonlinear model, and hence the selected subset of observations is not necessarily the most informative collection of measurements.

To address these challenges, motivated by the multi-target tracking problem via a network of sensing units equipped with radar systems, we consider observation selection under models where the relation between unknown states and measurements (partially) follows a quadratic equation. By drawing a connection between the classical Van Trees' inequality [23] and alphabetical optimality criteria [15], we devise new objective functions that exploit the quadratic relation of the observation model. We further prove that these functions possess two appealing properties, namely, monotonicity and (weak) submodularity under mild conditions on the statistics of the problem and parameters of the model. These results allow us to develop a simple greedy scheme for observation selection with theoretical bounds on its achievable performance without requiring any a priori approximation step. To

Mahsa Ghasemi is with the Department of Mechanical Engineering, University of Texas at Austin, Austin, TX 78712 USA. Abolfazl Hashemi and Haris Vikalo are with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712 USA. Ufuk Topcu is with the Department of Aerospace Engineering and Engineering Mechanics, University of Texas at Austin, Austin, TX 78712 USA.

This work was supported in part by ONR N000141912054 and N000141812829, NSF 1809327, and DARPA D19AP00004.

†These authors contributed equally to the manuscript.

demonstrate efficacy of the proposed framework, we consider the task of multi-object tracking via swarm of UAVs and empirically verify that the subsets selected by the proposed greedy algorithm outperform approaches based on greedy selection of observations that rely on a linearized model.

Notation. We briefly summarize the notation used in the paper. Bold capital letters denote matrices while bold lowercase letters represent vectors. Sets are represented with calligraphic letters. $H_k(i,j)$ is the (i,j) entry of the time-varying matrix $\mathbf{H}_k$ at time k, $\mathbf{h}_{k,j}$ is the j^{th} row of $\mathbf{H}_k$, $\mathbf{H}_{k,\mathcal{S}}$ is a submatrix of $\mathbf{H}_k$ that consists of the rows of $\mathbf{H}_k$ indexed by the set $\mathcal{S}$, and $\lambda_{max}(\mathbf{H}_k)$ and $\lambda_{min}(\mathbf{H}_k)$ are the maximum and minimum eigenvalues of $\mathbf{H}_k$, respectively. $\mathbf{I}$ is the identity matrix and its dimension is determined from the context. Moreover, let $[n] := \{1, 2, \dots, n\}$. Finally, $\mathbf{1}_d$ and $\mathbf{0}_d$ denote all-one and all-zero vectors of dimension d.

II. System Model and Problem Formulation

Consider a networked sensing system where there are m sensing units in the network, sensing n objects with unknown locations. Sensing units are equipped with GPS and radar systems and can communicate with each other over locally established communication channels. Because of various practical restrictions such as power and communication constraints, only a subset of sensing units, known as leaders, can communicate to a control unit that surveys the environment via commanding the networked system. Each sensing unit acquires range and angular measurements of all the objects that are within the maximum radar detection range and transmits those measurements to its nearest leader.

Let $\mathbf{u}_k^i$ and $\mathbf{s}_k^j$ denote the location of i^{th} unit and j^{th} object at time k, respectively. Also, let $\mathbf{s}_k = [{\mathbf{s}_k^1}^\top, \dots, {\mathbf{s}_k^n}^\top]^\top \in \mathbb{R}^{3n}$ denote the collection of unknown states evolving according to the nonlinear state equation $\mathbf{s}_k = g(\mathbf{s}_{k-1}) + \mathbf{w}_k$, where $\mathbf{w}_k$ is the zero-mean white Gaussian process noise at time k with covariance $\mathbf{Q}_k$.

If j^{th} object is within the range of i^{th} unit, the range and angular measurements of the radar system at time k have the following forms:

$$r_{ij} = \frac{1}{2}\|\mathbf{u}_k^i - \mathbf{s}_k^j\|_2^2 + \nu_{ij}, \tag{1}$$

$$\phi_{ij} = \arcsin\frac{u_k^i(3) - s_k^j(3)}{\|\mathbf{u}_k^i - \mathbf{s}_k^j\|_2} + \zeta_{ij}, \tag{2}$$

$$\alpha_{ij} = \arctan\frac{u_k^i(1) - s_k^j(1)}{u_k^i(2) - s_k^j(2)} + \eta_{ij}, \tag{3}$$

where ν_{ij}, ζ_{ij} and η_{ij} are zero-mean white Gaussian observation noises.[1] We denote by $\mathcal{X}_r$, $\mathcal{X}_\phi$, and $\mathcal{X}_\alpha$ the corresponding subsets of all gathered range and angular measurements and further we define $\mathcal{X} := \mathcal{X}_r \cup \mathcal{X}_\phi \cup \mathcal{X}_\alpha$. Note that depending on the location of objects and units, $3n \le |\mathcal{X}| \le 3nm$.

Due to limitations on the rate of communication between the leaders and the control unit that mainly stems from power limitation, and to reduce delays in tracking from high computation, only a subset $\mathcal{S}_k \subset \mathcal{X}$ of the gathered measurements is communicated to the control unit such that $|\mathcal{S}_k| \le K$. In order to track the locations of the objects, the control unit employs extended Kalman filtering (EKF) using the received measurements. Hence, the selected subset by the unit leaders should be the one with lowest mean-square error of the EKF estimates of the objects' locations while satisfying the communication constraint.

To identify the most informative subset satisfying the communication constraint, existing locally optimal schemes (e.g., [17], [19]) linearize the measurement model in (1) – (3) around $\hat{\mathbf{s}}_{k-1}$, the estimate of objects' locations at time $k-1$, to obtain an approximate linearized measurement model $\mathbf{y}_k = \mathbf{H}_k\mathbf{s}_k + \mathbf{v}_k$, where $\mathbf{v}_k$ is the corresponding zero-mean white Gaussian observation noise with the diagonal covariance $\mathbf{R}_k = \text{diag}(\sigma_1^2, \dots \sigma_{|\mathcal{X}|}^2)$. Then, if for any subset of observations $\mathcal{S} \subseteq \mathcal{X}$, we consider the filtered error covariance of EKF,

$$\mathbf{P}_{k|k}(\mathcal{S}) = \left(\mathbf{P}_{k|k-1}^{-1} + \mathbf{H}_{k,\mathcal{S}}^\top \mathbf{R}_{k,\mathcal{S}}^{-1}\mathbf{H}_{k,\mathcal{S}}\right)^{-1}, \tag{4}$$

where $\mathbf{P}_{k|k-1}$ is the predicted error covariance of EKF, observation selection is performed at each time k by optimizing trace or $\log\det$ of inverse of $\mathbf{P}_{k|k}$. That is,

$$\mathcal{S}_k = \underset{|\mathcal{S}|\le K}{\arg\max} \quad \text{Tr}\left(\mathbf{P}_{k|k-1}\right) - \text{Tr}\left(\mathbf{P}_{k|k}(\mathcal{S})\right) \tag{5}$$

or

$$\mathcal{S}_k = \underset{|\mathcal{S}|\le K}{\arg\max} \quad \log\det\left(\mathbf{P}_{k|k}^{-1}(\mathcal{S})\right) - \log\det\left(\mathbf{P}_{k|k-1}^{-1}\right). \tag{6}$$

Both of the above optimization problems are NP-hard. Hence, existing schemes rely on greedy heuristics or convex relaxations to find a suboptimal subset $\mathcal{S}^g$.

The major drawback of locally optimal approaches that are based on the linearized model is that the linearization step might distort the relational structure of the true nonlinear model severely. Hence, the selected subset of observations might not be the most informative collection of measurements. Although, a remedy for general and complex nonlinearities (such as angular measurements in (2) and (3)) seems rather infeasible to find, in the next section we develop a novel framework for quadratic models (such as range measurement in (1)). Our proposed framework builds upon the idea of optimizing alphabetical scalarizations of the Van Trees' bound [23] on the moment of a weakly biased estimator. The Van Trees' inequality is outlined in the following theorem.

Theorem 1. *Let $\boldsymbol{\theta}$ be a collection of random unknown parameters, and let $\mathbf{y}_\mathcal{S} = \{y_i\}_{i\in\mathcal{S}}$ denote the collection of measurements indexed by the subset $\mathcal{S}$. For any estimator $\hat{\boldsymbol{\theta}}_\mathcal{S}$ that satisfies*

$$\int_{-\infty}^{+\infty} \nabla_{\boldsymbol{\Theta}} \left(p_{\boldsymbol{\theta}}(\boldsymbol{\Theta})\, \mathbb{E}_{\mathbf{y};\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}_\mathcal{S} - \boldsymbol{\Theta}]\right) d\boldsymbol{\Theta} = \mathbf{0}, \tag{7}$$

it holds that

$$\mathbf{M}_\mathcal{S} \succeq \mathbb{E}_{\mathbf{y}_\mathcal{S},\boldsymbol{\theta}}\left[(\nabla_{\boldsymbol{\Theta}} \log q_{\boldsymbol{\theta}}(\boldsymbol{\Theta}))(\nabla_{\boldsymbol{\Theta}} \log q_{\boldsymbol{\theta}}(\boldsymbol{\Theta}))^\top\right]^{-1}, \tag{8}$$

[1] We occasionally omit the time index for simplicity of the notation.

where $\mathbf{M}_{\mathcal{S}} = \mathbb{E}[(\hat{\boldsymbol{\theta}}_{\mathcal{S}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{\mathcal{S}} - \boldsymbol{\theta})^{\top}]$ *is the so-called moment matrix associated with* $\hat{\boldsymbol{\theta}}_{\mathcal{S}}$*, and* $q_{\boldsymbol{\theta}}(\boldsymbol{\Theta}) = p_{\boldsymbol{\theta};\mathbf{y}_{\mathcal{S}}}(\boldsymbol{\Theta};\mathbf{y})$ *is the posterior distribution of* $\boldsymbol{\theta}$ *given* $\mathbf{y}_{\mathcal{S}}$.

The condition stated in Theorem 1 essentially quantifies to what extent the estimator is biased. Indeed, for an unbiased estimator satisfying $\mathbb{E}_{\mathbf{y};\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}_{\mathcal{S}}] = \boldsymbol{\theta}$, this condition is met.

The lower bound in the Van Trees' inequality which is essentially a lower bound on the achievable mean square error (MSE) cannot be computed in a closed-form for general nonlinear models. Nonetheless, as we show in the next section, the Van Trees' bound has a closed-form expression for the range measurements.

III. Quadratic Observation Selection

In this section, we devise a novel framework to select the most informative range measurements in a multi-object tracking sensing network. Throughout this section, we assume $\mathbb{E}[\mathbf{s}_k^j] = \hat{\mathbf{s}}_{k-1}^j$ for all $j \in [n]$. Admittedly, in the beginning of tracking, this assumption might not necessarily hold. Yet, as time passes the system generally improves the estimates of targets locations.

By defining $\tilde{\mathbf{s}}_k^j := \mathbf{s}_k^j - \hat{\mathbf{s}}_{k-1}^j$, $\mathbf{a}_{ij} := \hat{\mathbf{s}}_{k-1}^j - \mathbf{u}_k^i$, and $\tilde{r}_{ij} := r_{ij} - \frac{1}{2}\|\mathbf{a}_{ij}\|_2^2$, (1) can equivalently be written as

$$\tilde{r}_{ij} = \frac{1}{2}\|\tilde{\mathbf{s}}_k^j\|_2^2 + {\mathbf{a}_{ij}}^{\top}\tilde{\mathbf{s}}_k^j + \nu_{ij}. \tag{9}$$

The term $\mathbf{a}_{ij}$ can be thought of as the features or the design parameters. Let $\tilde{\mathbf{s}}_k = [{\tilde{\mathbf{s}}_k^1}^{\top}, \ldots, {\tilde{\mathbf{s}}_k^n}^{\top}]^{\top} \in \mathbb{R}^{3n}$, define $\mathbf{z}_{ij} := [\mathbf{0}_{3(j-1)}^{\top}, {\mathbf{a}_{ij}}^{\top}, \mathbf{0}_{3(n-j)}^{\top}]^{\top}$ and $\mathbf{X}_{ij} := \mathrm{diag}(\mathbf{0}_{3(j-1)}^{\top}, \mathbf{1}_3^{\top}, \mathbf{0}_{3(n-j)}^{\top})$. Then, (9) can be written in terms of the concatenated vector of all centralized unknowns $\tilde{\mathbf{s}}_k$ according to

$$\tilde{r}_{ij} = \frac{1}{2}\tilde{\mathbf{s}}_k^{\top}\mathbf{X}_{ij}\tilde{\mathbf{s}}_k + {\mathbf{z}_{ij}}^{\top}\tilde{\mathbf{s}}_k + \nu_{ij}. \tag{10}$$

Our first theoretical result, stated in the following theorem, demonstrates that for the quadratic model in (10) the Van Trees' bound has a closed-form expression.

Theorem 2. *Let* $\mathbf{B}_{\mathcal{S}}$ *denote the lower bound in the Van Trees inequality for the quadratic model* (10)*. Then, for any subset* $\mathcal{S} \subseteq \mathcal{X}_r$ *it holds that*

$$\mathbf{B}_{\mathcal{S}} = \left(\sum_{ij\in\mathcal{S}} \frac{1}{\sigma_{ij}^2}\left(\mathbf{X}_{ij}\mathbf{P}_{k|k-1}\mathbf{X}_{ij}^{\top} + \mathbf{z}_{ij}\mathbf{z}_{ij}^{\top}\right) + \mathbf{P}_{k|k-1}^{-1}\right)^{-1}. \tag{11}$$

Proof. Let $\mathbf{r}$ denote the vector of all range measurements of the form (10), and $q_{\tilde{\mathbf{s}}_k}(\tilde{\mathbf{S}}) = p_{\mathbf{r}_{\mathcal{S}},\tilde{\mathbf{s}}_k}(\mathbf{r};\tilde{\mathbf{S}})$ denote the posterior distribution of $\tilde{\mathbf{s}}_k$ given $\mathbf{r}_{\mathcal{S}}$, and define

$$\boldsymbol{\mu}_{\mathcal{S}} = \mathrm{vec}(\{\frac{1}{2}\tilde{\mathbf{s}}_k^{\top}\mathbf{X}_{ij}\tilde{\mathbf{s}}_k + \mathbf{z}_{ij}^{\top}\tilde{\mathbf{s}}_k\}_{ij\in\mathcal{S}}).$$

Then the Van Trees' bound is found as

$$\begin{aligned}
\mathbf{B}_{\mathcal{S}}^{-1} &= \mathbb{E}_{\mathbf{r}_{\mathcal{S}},\tilde{\mathbf{s}}_k}[(\nabla_{\tilde{\mathbf{S}}}\log q_{\tilde{\mathbf{s}}_k}(\tilde{\mathbf{S}}))(\nabla_{\tilde{\mathbf{S}}}\log q_{\tilde{\mathbf{s}}_k}(\tilde{\mathbf{S}}))^{\top}] \\
&= \mathbb{E}_{\mathbf{r}_{\mathcal{S}},\tilde{\mathbf{s}}_k}[(\nabla_{\tilde{\mathbf{S}}}\log p_{\mathbf{r}_{\mathcal{S}};\tilde{\mathbf{s}}_k}(\mathbf{r};\tilde{\mathbf{S}})p_{\tilde{\mathbf{s}}_k}(\tilde{\mathbf{S}})) \\
&\qquad\qquad (\nabla_{\tilde{\mathbf{S}}}\log p_{\mathbf{r}_{\mathcal{S}};\tilde{\mathbf{s}}_k}(\mathbf{r};\tilde{\mathbf{S}})p_{\tilde{\mathbf{s}}_k}(\tilde{\mathbf{S}}))^{\top}] \qquad (12)\\
&= \mathbb{E}_{\mathbf{r}_{\mathcal{S}},\tilde{\mathbf{s}}_k}[(\nabla_{\tilde{\mathbf{S}}}\log p_{\mathbf{r}_{\mathcal{S}};\tilde{\mathbf{s}}_k}(\mathbf{r};\tilde{\mathbf{S}})) \\
&\qquad\qquad (\nabla_{\tilde{\mathbf{S}}}\log p_{\mathbf{r}_{\mathcal{S}};\tilde{\mathbf{s}}_k}(\mathbf{r};\tilde{\mathbf{S}}))^{\top}] + \mathbf{J}_x,
\end{aligned}$$

where

$$\mathbf{J}_x = \mathbb{E}_{\mathbf{r}_{\mathcal{S}},\tilde{\mathbf{s}}_k}\left[(\nabla_{\tilde{\mathbf{S}}}\log p_{\tilde{\mathbf{s}}_k}(\tilde{\mathbf{S}}))(\nabla_{\tilde{\mathbf{S}}}\log p_{\tilde{\mathbf{s}}_k}(\tilde{\mathbf{S}}))^{\top}\right]$$

is the prior Fisher information on $\tilde{\mathbf{s}}_k$. Since in EKF settings $p_{\tilde{\mathbf{s}}_k}(\tilde{\mathbf{S}}) = \mathcal{N}(\mathbf{0}, \mathbf{P}_{k|k-1})$, then $\mathbf{J}_x = \mathbf{P}_{k|k-1}^{-1}$. Note that the conditional distribution $p_{\mathbf{r}_{\mathcal{S}};\tilde{\mathbf{s}}_k}(\mathbf{r};\tilde{\mathbf{S}})$ is the normal distribution $\mathcal{N}(\boldsymbol{\mu}_{\tilde{\mathbf{s}}_k}, \mathbf{R}_{k,\mathcal{S}})$. Therefore,

$$\nabla_{\tilde{\mathbf{S}}}\log p_{\mathbf{r}_{\mathcal{S}};\tilde{\mathbf{s}}_k}(\mathbf{r};\tilde{\mathbf{S}}) = -(\nabla_{\tilde{\mathbf{S}}}\boldsymbol{\mu}_{\mathcal{S}})\mathbf{R}_{k,\mathcal{S}}^{-1}(\mathbf{r}_{\mathcal{S}} - \boldsymbol{\mu}_{\mathcal{S}}), \tag{13}$$

where $[\nabla_{\tilde{\mathbf{S}}}\boldsymbol{\mu}_{\mathcal{S}}]_{ij} = \mathbf{X}_{ij}\tilde{\mathbf{s}}_k + \mathbf{z}_{ij}$. Using this result and applying the law of total expectation we obtain

$$\mathbf{B}_{\mathcal{S}}^{-1} = \sum_{ij\in\mathcal{S}} \frac{1}{\sigma_{ij}^2}\left(\mathbf{X}_{ij}\mathbf{P}_{k|k-1}\mathbf{X}_{ij}^{\top} + \mathbf{z}_{ij}\mathbf{z}_{ij}^{\top}\right) + \mathbf{P}_{k|k-1}^{-1}. \tag{14}$$

Inverting the last line that consists of an invertible positive definite matrix establishes the stated result and completes the proof. ∎

Relying on the result of Theorem 2, we propose to use the trace and $\log\det$ of inverse of $\mathbf{B}_{\mathcal{S}}$ as the objective functions in the observation selection task (effectively replacing $\mathbf{P}_{k|k}(\mathcal{S})$ with $\mathbf{B}_{\mathcal{S}}$ in (5) and (6)). That is, instead of linearizing the range measurements we propose to select the most informative range measurements according to one of the following optimization problems:

$$\mathcal{S}_k = \underset{|\mathcal{S}|\leq K}{\arg\max} \quad \mathrm{Tr}\left(\mathbf{P}_{k|k-1}\right) - \mathrm{Tr}\left(\mathbf{B}_{\mathcal{S}}\right), \tag{15}$$

$$\mathcal{S}_k = \underset{|\mathcal{S}|\leq K}{\arg\max} \quad \log\det\left(\mathbf{B}_{\mathcal{S}}^{-1}\right) - \log\det\left(\mathbf{P}_{k|k-1}^{-1}\right), \tag{16}$$

which are computationally challenging and NP-hard [3]. Theorem 2 opens a new avenue in the task of observation selection for quadratic models which, as we see in our simulation results, enables selection of observations leading to lower estimation error (i.e., higher information) as compared to the locally optimal approximation methods based on linearization [17], [19]. We note that the Van Trees' lower bound is asymptotically tight, i.e., it is tight in the high signal-to-noise ratio settings or in the case of sufficiently large number of observations. Hence, we expect to select a near-optimal subset by using the proposed selection criteria in such settings. In the next section, we further demonstrate monotonicity and weak submodularity of the proposed optimality criteria which in turn enables us to devise a novel greedy observation selection scheme with theoretical performance guarantee.

IV. A Near-Optimal Greedy Approach

In this section, we show that the proposed objective functions are monotone and (weak) submodular, two important concepts that we will later exploit to develop a novel greedy observation selection scheme. These concepts along with their properties are overviewed next.

A. Combinatorial optimization background

Definition 1. *Set function $f : 2^{\mathcal{X}} \to \mathbb{R}$ is monotone non-decreasing if $f(\mathcal{S}) \leq f(\mathcal{T})$ for all $\mathcal{S} \subseteq \mathcal{T} \subseteq \mathcal{X}$.*

Definition 2. *Set function $f : 2^{\mathcal{X}} \to \mathbb{R}$ is submodular if*

$$f(\mathcal{S} \cup \{j\}) - f(\mathcal{S}) \geq f(\mathcal{T} \cup \{j\}) - f(\mathcal{T}) \tag{17}$$

for all subsets $\mathcal{S} \subseteq \mathcal{T} \subset \mathcal{X}$ and $j \in \mathcal{X} \backslash \mathcal{T}$. The term $f_j(\mathcal{S}) = f(\mathcal{S} \cup \{j\}) - f(\mathcal{S})$ is the marginal value of adding element j to set $\mathcal{S}$.

Definition 3. *The multiplicative curvature of a monotone non-decreasing function f is defined as*

$$c_f = \max_{(\mathcal{S},\mathcal{T},i) \in \tilde{\mathcal{X}}} f_i(\mathcal{T}) / f_i(\mathcal{S}), \tag{18}$$

where $\tilde{\mathcal{X}} = \{(\mathcal{S}, \mathcal{T}, i) | \mathcal{S} \subseteq \mathcal{T} \subset \mathcal{X}, i \in \mathcal{X} \backslash \mathcal{T}\}$.

The multiplicative curvature [24], [25] is a closely related concept to submodularity and essentially quantifies how close the set function is to being submodular. A set function with bounded curvature is called weak submodular. It is worth noting that a set function $f(\mathcal{S})$ is submodular if and only if its multiplicative curvature satisfies $c_f \leq 1$ [26]–[28].

A similar notion of weak submodularity is the additive curvature defined below [24].

Definition 4. *The additive curvature of a monotone non-decreasing function f is defined as*

$$\epsilon_f = \max_{(\mathcal{S},\mathcal{T},i) \in \tilde{\mathcal{X}}} f_i(\mathcal{T}) - f_i(\mathcal{S}), \tag{19}$$

where $\tilde{\mathcal{X}} = \{(\mathcal{S}, \mathcal{T}, i) | \mathcal{S} \subseteq \mathcal{T} \subset \mathcal{X}, i \in \mathcal{X} \backslash \mathcal{T}\}$.

Note that when $f(\mathcal{S})$ is submodular, its additive curvature satisfies $\epsilon_f \leq 0$. Multiplicative and additive curvatures are closely related to submodularity ratio [26].

For the above additive and multiplicative curvatures, we have the following proposition [4], [7], [24], [29], [30].

Proposition 1. *Let c_f and ϵ_f be the multiplicative and additive curvatures of $f(\mathcal{S})$, a monotone non-decreasing function with $f(\emptyset) = 0$. Let $\mathcal{S}$ and $\mathcal{T}$ be any subsets such that $\mathcal{S} \subset \mathcal{T} \subseteq \mathcal{X}$ with $|\mathcal{T} \backslash \mathcal{S}| = r$. Then, it holds that*

$$f(\mathcal{T}) - f(\mathcal{S}) \leq \frac{1}{r} \left(1 + (r-1)c_f\right) \sum_{j \in \mathcal{T} \backslash \mathcal{S}} f_j(\mathcal{S}), \tag{20}$$

and

$$f(\mathcal{T}) - f(\mathcal{S}) \leq (r-1)\epsilon_f + \sum_{j \in \mathcal{T} \backslash \mathcal{S}} f_j(\mathcal{S}). \tag{21}$$

Algorithm 1 Greedy Observation Selection

1: **Input:** Utility function $f(\mathcal{S})$, set of all observations $\mathcal{X}$, number of selected observations K.
2: **Output:** Subset $\mathcal{S}^g \subseteq \mathcal{X}$ with $|\mathcal{S}^g| = K$.
3: Initialize $\mathcal{S}^g = \emptyset$
4: **for** $i = 0, \ldots, K-1$ **do**
5: $\quad j_s = \text{argmax}_{j \in \mathcal{X} \backslash \mathcal{S}^g} f_j(\mathcal{S}^g)$
6: $\quad \mathcal{S}^g \leftarrow \mathcal{S}^g \cup \{j_s\}$
7: **end for**
8: **return** $\mathcal{S}^g$.

Given a monotone non-decreasing set function $f : 2^{\mathcal{X}} \to \mathbb{R}$ with $f(\emptyset) = 0$, we are interested in solving the combinatorial problem

$$\max_{\mathcal{S} \subseteq \mathcal{X}, |\mathcal{S}| \leq K} f(\mathcal{S}). \tag{22}$$

It has been shown that finding an optimal solution to (22) is generally NP-hard [3]. To this end, efficient heuristic approaches that rely on a simple greedy search (see Algorithm 1) are developed. If the set function $f(\mathcal{S})$ is monotone, Algorithm 1 has a guaranteed lower bound on its achievable performance as stated in Proposition 2 [4], [7], [24], [29], [30].

Proposition 2. *Let c_f and ϵ_f be the multiplicative and additive curvatures of $f(\mathcal{S})$, a monotone non-decreasing function with $f(\emptyset) = 0$. Let $\mathcal{S}^g \subseteq \mathcal{X}$ with $|\mathcal{S}^g| \leq K$ be the subset selected when maximizing $f(\mathcal{S})$ subject to a cardinality constraint via the greedy observation selection scheme, and let $\mathcal{S}^\star$ denote the optimal subset. Then*

$$f(\mathcal{S}^g) \geq \left(1 - e^{-\frac{1}{c}}\right) f(\mathcal{S}^\star), \tag{23}$$

where $c = \max\{c_f, 1\}$ and

$$f(\mathcal{S}^g) \geq \left(1 - \frac{1}{e}\right) \left(f(\mathcal{S}^\star) - (k-1)\epsilon_f\right). \tag{24}$$

The results of Propositions 1 and 2 imply that if the objective function (22) is monotone and (weak) submodular, the greedy selection scheme that in each iteration selects an observation with the highest marginal value satisfies the approximation bounds given in Proposition 2.

B. Greedy selection of range observations

In the following theorems, we consider trace and $\log\det$ scalarizations of the Van Trees' bound $\mathbf{B}_S$ defined in Theorem 2 and show that they are monotonically non-decreasing as well as either submodular, or weak submodular. These results illustrate not only that the proposed objective functions deal with the quadratic model of range measurements without resorting to any approximations, but also that one can use the greedy observation selection method of Algorithm 1 to find a near-optimal subset of observations with performance guarantees established in Proposition 2. Proofs of the subsequent results are established by employing tools from linear algebra and matrix analysis such as Weyl's inequality,

Sylvester's determinant identity, matrix inversion lemma, and Courant–Fischer min-max theorem [31].

Theorem 3. *Instate the notation and hypothesis of Theorem 2. The D-optimality of the Van Trees' bound, i.e.,*

$$f^D(\mathcal{S}) = \log\det\left(\mathbf{B}_{\mathcal{S}}^{-1}\right) - \log\det\left(\mathbf{P}_{k|k-1}^{-1}\right), \tag{25}$$

is monotone and submodular.

Theorem 4. *Instate the notation and hypothesis of Theorem 2. The A-optimality of the Van Trees' bound, i.e.,*

$$f^A(\mathcal{S}) = \mathrm{Tr}\left(\mathbf{P}_{k|k-1}\right) - \mathrm{Tr}\left(\mathbf{B}_{\mathcal{S}}\right), \tag{26}$$

is monotone and weak submodular and its additive and multiplicative curvatures satisfy

$$c_{f^A} \leq \max_{ij\in\mathcal{X}_r} \frac{\lambda_{\max}(\mathbf{B}_{\mathcal{X}_r}^{-1} + \mathbf{B}_{\mathcal{X}_r}^{-1}\mathbf{J}_{ij}\mathbf{B}_{\mathcal{X}_r}^{-1})}{\lambda_{\min}(\mathbf{P}_{k|k-1}^{-1} + \mathbf{P}_{k|k-1}^{-1}\mathbf{J}_{ij}\mathbf{P}_{k|k-1}^{-1})}, \tag{27}$$

$$\epsilon_{f^A} \leq \max_{ij\in\mathcal{X}_r} \lambda_{\max}(\mathbf{B}_{\mathcal{X}_r}^{-1} + \mathbf{B}_{\mathcal{X}_r}^{-1}\mathbf{J}_{ij}\mathbf{B}_{\mathcal{X}_r}^{-1}) - \lambda_{\min}(\mathbf{P}_{k|k-1}^{-1} + \mathbf{P}_{k|k-1}^{-1}\mathbf{J}_{ij}\mathbf{P}_{k|k-1}^{-1}), \tag{28}$$

where $\mathbf{J}_{ij} = \frac{1}{\sigma_{ij}^2}\left(\mathbf{X}_{ij}\mathbf{P}_{k|k-1}\mathbf{X}_{ij}^\top + \mathbf{z}_{ij}\mathbf{z}_{ij}^\top\right)$*, for all* $ij \in \mathcal{X}_r$.

The term $\mathbf{J}_{ij}$ is reflective of the amount of *information* captured by the ij^{th} observation. In this regard, Theorem 4 states that if the difference between the minimum and maximum information of individual observations is small, the objective in (15) is nearly submodular. Hence, the greedy observation selection scheme is expected to find a good (informative) subset.

Theorems 3 and 4 establish monotonicity and (weak) submodularity of the proposed objective functions in (15) and (16). Hence, a suboptimal subset of range observations found by the greedy observation selection scheme (Algorithm 1) satisfies the performance bounds given in Proposition 2.

V. Simulation Results

In this section, we test the efficacy of the proposed quadratic observation selection objectives in a multi-object tracking application via UAV swarm using radar measurements (Fig. 1(a)) and compare their performance with those of random and locally optimal (linearization-based) schemes.

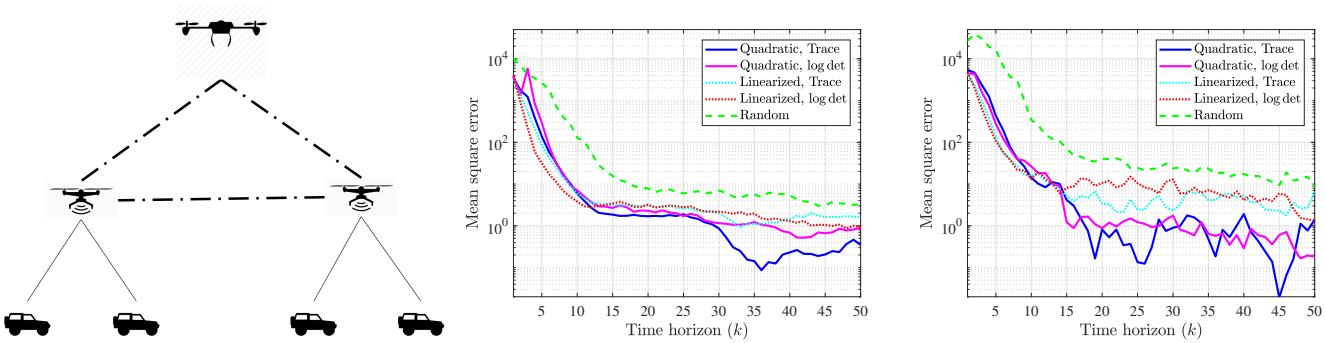
We consider a Monte Carlo simulation with 50 independent instances where 10 moving objects are initially uniformly distributed in a 5×10 area. At each time instance, the objects move in a random direction with a constant velocity set to 0.2. The swarm consists of 10 UAVs, equidistantly spread over the area, that move according to a periodic *parallel-path* search pattern [1]. 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 130-170 range and angular measurements. The communication bandwidth constraints limit the number of measurements transmitted to the control unit to 10% of the gathered measurements. For the proposed scheme, we select the range measurements using the proposed quadratic observation selection scheme while for angular measurements, we follow the locally optimal approach of [17], [19], i.e., linearization around the prior estimates. Performance of different schemes is assessed using the MSE of the EKF estimates of objects' locations. We consider two noise models: in the first scenario, the noise terms are i.i.d. Gaussian with $\sigma_{ij} = 0.01$ while in the second scenario, we logarithmically space the interval $(0.001, 0.01)$ to generate 10 points and select σ_i for each measurement uniformly at random from one of these 10 numbers.

The results for the first noise model are illustrated in Fig. 1(b). There, at the beginning of tracking all schemes have relatively high error. However, since the observations selected by the proposed schemes are chosen according to the exact range model, as time passes the MSE of the proposed schemes becomes significantly lower than those of locally optimal and random selection methods (especially under the A-optimality criterion). Fig. 1(b) also depicts that the MSE of estimates formed from observations selected by the proposed quadratic observation selection scheme using A-optimality is lower than the MSE achieved by selecting the observations via D-optimality. The explanation of this phenomenon is that if the estimator (here the EKF) is a minimum variance unbiased estimator attaining (8) with equality, the A-optimality scalarization of the Van Trees' bound becomes equivalent to the MSE, the performance measure shown in Fig. 1(b). Therefore, intuitively, one expects to achieve lower MSE using the A-optimality scalarization of the Van Trees' bound, which is the case in this simulation.

The results for the second noise model are illustrated in Fig. 1(c) where we again observe superiority of the proposed quadratic framework to select a subset of observations with the lowest estimation error. Compared to Fig. 1(b), since the noise terms here are random, the MSE curves in Fig. 1(c) are not as smooth as those in Fig. 1(b).

VI. Conclusion

In this paper, we considered the problem of state estimation in time-varying networked sensing systems with resource constraints, following a (partially) quadratic measurement models. For this setting, we derived new optimality criteria by relying on the Van Trees' inequality and proved that they are monotone and (weak) submodular set functions. In particular, we showed that the $\log\det$ of the inverse of the Van Trees' bound is submodular while its trace is weak submodular under certain conditions on the unknown states, noise statistics, and the parameters of the model. Following these results, we developed an efficient greedy observation selection algorithm for networked sensing systems with theoretical bounds on its achievable utility that efficiently exploits the quadratic structure of the measurement model in its selection criteria. Our experimental studies demonstrated efficacy of the proposed optimality criteria in estimating the



(a) tracking via a swarm of UAVs (b) tracking under identical noise powers (c) tracking under random noise powers

Fig. 1: Comparison of MSEs for random, linearized, and quadratic observation selection schemes in the multi-target tracking application.

location of unknown targets with minimal mean square error in a multi-object tracking application with UAV swarms.

REFERENCES

[1] P. Vincent and I. Rubin, "A framework and analysis for cooperative search using UAV swarms," in *Proc. ACM symposium on Applied computing*, pp. 79–86, ACM, 2004.

[2] 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.

[3] D. P. Williamson and D. B. Shmoys, *The design of approximation algorithms*. Cambridge university press, 2011.

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

[5] A. Krause and D. Golovin, *Submodular function maximization*. Cambridge University Press, 2014.

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

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

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

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

[10] V. Tzoumas, M. A. Rahimian, G. J. Pappas, and A. Jadbabaie, "Minimal actuator placement with optimal control constraints," in *Proc. American Control Conference*, pp. 2081–2086, IEEE, 2015.

[11] B. Mirzasoleiman, A. Badanidiyuru, A. Karbasi, J. Vondrak, and A. Krause, "Lazier than lazy greedy," in *Proc. AAAI Conference on Artificial Intelligence*, AAAI, 2015.

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

[13] M. I. Skolnik, *Radar handbook*. McGraw-Hill Education, 1970.

[14] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," *Computer*, vol. 34, no. 8, pp. 57–66, 2001.

[15] K. Chaloner and I. Verdinelli, "Bayesian experimental design: A review," *Statistical Science*, pp. 273–304, 1995.

[18] A. Krause, H. B. McMahan, C. Guestrin, and A. Gupta, "Robust submodular observation selection," *Journal of Machine Learning Research*, vol. 9, no. Dec, pp. 2761–2801, 2008.

[16] P. Sebastiani and H. P. Wynn, "Maximum entropy sampling and optimal Bayesian experimental design," *Journal of the Royal Statistical Society: Series B*, vol. 62, no. 1, pp. 145–157, 2000.

[17] P. Flaherty, A. Arkin, and M. I. Jordan, "Robust design of biological experiments," in *Proc. Advances in neural information processing systems*, pp. 363–370, 2006.

[19] S. Rao, S. P. Chepuri, and G. Leus, "Greedy sensor selection for non-linear models," in *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, pp. 241–244, IEEE, 2015.

[20] X. Shen, S. Liu, and P. K. Varshney, "Sensor selection for nonlinear systems in large sensor networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 4, pp. 2664–2678, 2014.

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

[22] M. Davidian, *Nonlinear models for repeated measurement data*. Routledge, 2017.

[23] H. L. Van Trees, *Detection, estimation, and modulation theory, part I: Detection, estimation, and linear modulation theory*. John Wiley & Sons, 2004.

[24] L. Chamon and A. Ribeiro, "Approximate supermodularity bounds for experimental design," in *Proc. Advances in Neural Information Processing Systems*, pp. 5409–5418, 2017.

[25] M. Sviridenko, J. Vondrák, and J. Ward, "Optimal approximation for submodular and supermodular optimization with bounded curvature," *Mathematics of Operations Research*, vol. 42, no. 4, pp. 1197–1218, 2017.

[26] A. Das and D. Kempe, "Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection," in *Proc. International Conference on Machine Learning*, pp. 1057–1064, 2011.

[27] E. R. Elenberg, R. Khanna, A. G. Dimakis, S. Negahban, *et al.*, "Restricted strong convexity implies weak submodularity," *The Annals of Statistics*, vol. 46, no. 6B, pp. 3539–3568, 2018.

[28] T. Horel and Y. Singer, "Maximization of approximately submodular functions," in *Proc. Advances in Neural Information Processing Systems*, pp. 3045–3053, 2016.

[29] A. Krause and C. Guestrin, "Near-optimal observation selection using submodular functions," in *Proc. national conference on Artificial intelligence*, vol. 7, pp. 1650–1654, AAAI Press, 2007.

[30] 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. American Control Conference*, pp. 1027–1032, IEEE, 2018.

[31] R. Bellman, *Introduction to matrix analysis*. SIAM, 1997.