APPEARED IN
Proc. American Control Conference, 2017

Supermodular Mean Squared Error Minimization for Sensor Scheduling in Optimal Kalman Filtering

Prince Singh Min Chen* Luca Carlone Sertac Karaman Emilio Frazzoli David Hsu*

***Abstract*— We consider the problem of scheduling a set of sensors to observe the state of a discrete-time linear system subject to a limited energy budget. Our goal is to devise a sensor schedule that minimizes the mean squared error (MSE) of an optimal estimator (i.e., the Kalman Filter). Both the minimum-MSE and the minimum-cardinality optimal sensor scheduling problems are inherently combinatorial, and computationally intractable. We remedy the combinatorial complexity by using a greedy heuristic; the greedy heuristic is guaranteed to return near-optimal solutions when minimizing supermodular objectives. While it is known that the MSE is not supermodular (with counterexamples available in literature), we provide conditions on the prior information matrix and on the observation matrix under which supermodularity holds. More specifically, we require the prior information matrix to be a strictly-diagonally-dominant M-matrix (plus an extra technical requirement on its inverse). Empirical results confirm that random M-matrices lead to supermodular problems, while this is not the case for generic prior information matrices. M-matrices naturally arise in estimation problems over networks and we provide a practical application of our findings to an energy-constrained multi robot localization problem.**

I. Introduction

Sensor scheduling consists in the selection of a subset of available sensors to observe the state of a dynamical system. In many applications, including sensor networks and robotics, one may not use all available sensors due to energy, communication, or computation constraints [1]. Therefore, an *optimal sensor schedule* prescribes which fraction of the sensors to activate at each time-step in order to maximize a given performance metric in a resource-constrained fashion.

In this work, we consider sensor scheduling for the case where the system dynamics admits a discrete-time linear time-variant representation for both its process and measurement models. In this case, it is natural to estimate the state via a Kalman filter (KF), which is optimal in the minimum-mean-square-error sense, and sensor scheduling reduces to choosing the set of measurements to include in the filter update in order to maximize the estimation quality.

A natural function that characterizes the quality of the state estimate is the mean squared error (MSE), which is the trace of the estimation error covariance. The selection of the subset of sensors that minimizes the MSE subject to a given sensor budget is a combinatorial problem, which is inherently NP-hard and computationally intractable. One can hope to solve these problems using brute force for a small set of sensors by testing all possible subsets. However, such an approach quickly becomes computationally intractable since the number of subsets that need to be checked grows factorially with the number of available sensors. Hence, in this work, we resort to an efficient polynomial-time algorithm that attains a provably-close solution to the optimal solution by exploiting structural properties of the MSE metric.

Literature Review. The problem of sensor scheduling has re-gained much attention over the past decade. For instance, Jamali-Rad *et al.* [22] propose a distributed algorithm for solving the sensor scheduling problem with performance guarantees. Gupta *et al.* [2] propose a randomized sensor schedule that minimizes the error covariance for a network comprising of noisy sensors that communicate with each other to estimate a noisy process. More relevant to our discussion is the use of structural properties of the metric to be optimized, to enable formal performance guarantees. As discussed in the sequel, *submodularity* is an important property of set functions that facilitates combinatorial optimization in polynomial time via *greedy* algorithms with suboptimality guarantees [3]. Shamaiah *et al.* [4] have proven, over a single time-step, that the $\log\det$ of the error covariance of the Kalman filter is supermodular and monotone non-increasing. Additionally, the work [4] shows through empirical tests that the greedy heuristic outperforms the more computationally intensive semi-definite relaxation of Joshi and Boyd [5]. Further, Jawaid and Smith [6] prove that $\log\det$ of the error covariance of the Kalman filter is *sequence supermodular* and monotone non-increasing under restrictive assumptions on the system dynamics and parameters. Recently, Tzoumas *et al.* [7] have proven that the $\log\det$ of the error covariance from the Kalman filter is supermodular and monotone non-increasing over a finite observation interval. Similar effort has been invested in proving submodularity of other estimation error metrics. Krause *et al.* [20] have shown that the mutual information criterion is a submodular function, in the context of sensor placement. Jawaid and Smith [6] provide a counterexample showing that the trace of the error covariance is not, in general, a supermodular set function. Other counterexamples are given in [8], in the context of sensor selection. One of our goals here is to find specific classes of systems for which the trace of the error covariance can be proven to be supermodular.

P. Singh, L. Carlone, S. Karaman, E. Frazzoli are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA (e-mail: {prince1,lcarlone,sertac,frazzoli}@mit.edu).

* M. Chen and D. Hsu are with the Department of Computer Science, National University of Singapore, Singapore 117417, Singapore (e-mail: {chenmin,dyhsu}@comp.nus.edu.sg).

The first three authors equally contributed to this paper. This work is supported by the Singapore National Research Foundation through the SMART Future Urban Mobility project. Chen is supported by SMART graduate fellowship, Hsu is supported by SMART subawared agreement No. II-08-FM IRG.

978-1-5090-5992-8/$31.00 ©2017 AACC

Contributions. The main contribution of this work is to provide conditions on the process and measurement model under which the trace of the a-posteriori covariance is a supermodular set function. These conditions mainly regard the structure of the prior information matrix, i.e., the inverse of the covariance resulting from the prediction phase of the KF, as well as the observation matrix. First, we derive technical conditions under which the trace of the covariance is supermodular. Then, we show that a class of prior information matrices (strictly-diagonally-dominant with ultrametric inverse) satisfies these conditions and leads to supermodular sensor scheduling problems. We currently consider monomial observation matrices (each sensor observes a single state of the system), while we believe that the results can be extended to more general observation models.

Our current results already enable interesting practical applications. The class of M-matrices naturally arises in estimation problems over networks. Indeed we provide an example in which a team of robots needs to localize from relative measurements and from GPS measurements; due to energy constraints, only a subset of robots can activate the GPS. We show that choosing this subset via an efficient greedy algorithm provides near-optimal solutions.

Interestingly, empirical evidence suggests that the diagonal dominance and ultrametric-inverse assumptions can be relaxed, since only the requirement of the prior information matrix being an M-matrix is sufficient for supermodularity. Relaxing these assumptions is subject of ongoing research.

For space reasons, all proofs are omitted from this paper; the interested reader can find them in the extended draft [9].

Outline. This paper is organized as follows. Section II provides a formal statement of the sensor scheduling problem. Section III provides preliminary notions on submodularity and presents our main contribution. Here we provide theoretical conditions under which the trace of the estimation covariance is supermodular. Section IV confirms our finding via numerical experiments on random systems and in a robotic example. Section V concludes the paper.

Notation. The set of natural numbers is denoted by $\mathbb{N}$, and the set of real numbers by $\mathbb{R}$. The cardinality of a set S is denoted by $|S|$. For a matrix A, its transpose is A^T. We write $A \succeq 0$ (resp. $A \succ 0$) to denote that A is symmetric and positive semi-definite (resp. positive definite). The symbols $0_{n\times n}, I_{n\times n}$ denote an $n \times n$ matrix of zeros and the identity of size n, respectively. For a random variable $x \in \mathbb{R}^n$, the expected value of x is $\mathbb{E}[x]$, and its covariance matrix is $\mathrm{Cov}(x) = \mathbb{E}\big[(x-\mathbb{E}[x])(x-\mathbb{E}[x])^T\big]$.

II. Problem Formulation

This section formalizes the sensor scheduling problem. Section II-A introduces the estimation problem over a linear system, tailoring the presentation to a networked system of sensors. Section II-B states the scheduling problem.

A. Network Dynamics

Consider a sensor network comprising of m sensors each of which produces scalar measurements and operates in discrete-time. Denote by $G \triangleq \{1, \dots, m\}$ the set of indices identifying each sensor. Assume that, at each time-step $t \in \mathbb{N}_{\geq 0}$, we can activate a time-varying subset $S_t \subseteq G$, with $|S_t| = r$ $(< m, \forall t)$, to observe a dynamical phenomenon.

The evolution of the dynamical phenomenon as well as the sensor measurements are modeled by the following discrete-time linear time-variant representation

$$\begin{aligned} x_{t+1} &= A_t x_t + u_t + w_t, \quad x_0 \sim \mathcal{N}(\hat{x}_0, P_{1|0}) \\ y_t &= C(S_t)x_t + v_t \end{aligned} \tag{1}$$

where $x_t \in \mathbb{R}^n$ is the state of the system at time t, $A_t \in \mathbb{R}^{n\times n}$ describes the system dynamics, $u_t \in \mathbb{R}^n$ is a known control input, $y_t \in \mathbb{R}^r$ is the measurement vector, and $C(S_t) \in \mathbb{R}^{r\times n}$ is a measurement matrix which is function of the set of sensors S_t that we activate at time t. The terms $w_t \sim \mathcal{N}(0, W_t)$ and $v_t \sim \mathcal{N}(0, V_t)$ are the process and the measurement noise, which are statistically independent zero-mean Gaussian noise processes of appropriate dimension, with covariances $W_t \succ 0$ and $V_t \succ 0$, respectively. The initial state of the system is assumed to be Normally distributed with given mean $(\hat{x}_0)$ and covariance $(P_{1|0} \succ 0)$.

From standard Kalman filtering theory, it follows that the error-covariance of an optimal estimator of the state in (1) can be computed recursively by the alternation of a prediction and an update step:

$$P_{t|t-1}(S_{t-1}) = A_{t-1}\, P_{t-1|t-1}(S_{t-1})\, A_{t-1}^T + W_{t-1} \tag{2}$$

$$P_{t|t}(S_t) = \big(P_{t|t-1}(S_{t-1})^{-1} + C(S_t)^T V_t(S_t)^{-1} C(S_t)\big)^{-1} \tag{3}$$

where $P_{t|t-1}(S_{t-1})$ is the *prior* covariance resulting from the prediction step, and $P_{t|t}(S_t)$ is the posterior covariance after the update. The covariance $P_{t|t}(S_t) \triangleq \mathrm{Cov}(x_t)$ and the estimate $\mathbb{E}[x_t]$ of x_t given the measurements collected till time t are characterized by the Kalman filter. Clearly, the covariances are functions of the set of sensors that we use for estimation, i.e., $P_{t|t} = P_{t|t}(S_t)$. Substituting (2) into (3) results in a Riccati equation that maps the posterior covariance at time $t-1$, namely $P_{t-1|t-1}(S_{t-1})$ into the posterior covariance at time t, namely $P_{t|t}(S_t)$.

In this paper we discuss how to select the set of sensors S_t so to "minimize" the estimation covariance. This notion is formalized in the following section.

B. Single time-step Optimal Sensor Scheduling Problems

A natural metric to evaluate the quality of a state estimator is the mean squared error (MSE), which corresponds to the trace of the estimation covariance. In this paper we look for a suitable choice of sensors S_t (with $|S_t| = r$) that induces the smallest MSE among all possible choices of r sensors. To be more formal, let us introduce the following set function:

$$g(S_t) \triangleq \mathrm{trace}(P_{t|t}(S_t)) - \mathrm{trace}(P_{t|t-1}(S_{t-1})). \tag{4}$$

which quantifies the change in the MSE after including the set of measurements S_t. Minimizing $g(S_t)$ is the same as minimizing the MSE (i.e., $\mathrm{trace}(P_{t|t}(S_t))$), since the second summand in (4) is a constant for the decision to be taken at time t. The second term is only used to make the cost (4)

normalized (see Definition 4). Therefore, our goal is to select an optimal set of sensors S_t at time t, that makes (4) minimum.

Since we minimize the MSE over a single time step, and to keep notation simple, in the following we drop the subscript t from S_t and we simply write $g(S)$. Additionally, since we already noted that $\mathrm{trace}(P_{t|t-1}(S_{t-1}))$ is a constant, we write $\mathrm{trace}(P_{t|t-1})$ henceforth.

Problem 1 (Cardinality-constrained single time-step optimal sensor scheduling)**.** *Given a sensor budget r $(< m) \in \mathbb{N}_{>0}$ on $|S|$, i.e., the number of sensors that can be activated at the current time, identify the optimal sensor schedule that minimizes $g(S)$:*

$$\begin{aligned} \underset{S \subseteq G}{\text{minimize}} \quad & g(S) \\ \text{subject to} \quad & |S| \leq r \end{aligned}$$

A related formulation requires to find the smallest set of sensors that meets a given MSE budget.

Problem 2 (MSE-constrained single time-step optimal sensor scheduling)**.** *Given an error budget $\xi \in \mathbb{R}$ on $g(S)$, find the minimal sensor set $S \subseteq G$ that meets the error budget:*

$$\begin{aligned} \underset{S \subseteq G}{\text{minimize}} \quad & |S| \\ \text{subject to} \quad & g(S) \leq \xi \end{aligned}$$

Both Problems 1 and 2 are hard combinatorial problems. In order to make the above problems computationally tractable, it will be crucial to prove that $g(S)$ is a supermodular and monotone non-increasing function with respect to the set of sensors S. To this end, we make the following assumption on the choice of the sensor matrix $C(S)$.

Assumption 1 (Monomial Measurement Matrix and Diagonal Measurement Covariance)**.** *For $S \subseteq G$, $C(S) \in \mathbb{R}^{r \times n}$ has at most one non-zero entry in each row. Moreover, the measurement covariance $V_t(S)$ is diagonal.*

In words, Assumption 1 requires that each sensor measures at most one entry of the state vector x_t and that all measurements are scalar and independent across sensors.

In the following we present a greedy algorithm to solve Problems 1 and 2, and we provide technical conditions under which the greedy algorithm produces near-optimal results.

III. Supermodularity in single time-step Optimal Sensor Scheduling

This section contains the key contributions of this paper. Before delving in the technical contribution, presented in Section III-C, we recall preliminary concepts about submodularity (Section III-A) and greedy algorithms (Section III-B).

A. Submodular functions

Submodularity is a property of *set functions*. Given a finite set G, a set function $g : 2^G \to \mathbb{R}$ assigns a real number to each subset of G. The following definitions and result will be useful in Section III-C.

Definition 1 (Modularity [10])**.** *For a given finite set G, a set function $g : 2^G \to \mathbb{R}$ is modular $\iff$ for any subset $A \subseteq G$, the set function can be expressed as $g(A) = \omega(\emptyset) + \sum_{a \in A} \omega(a)$, for some function $\omega : G \to \mathbb{R}$.*

A modular set function has the property that each element in G gives an independent contribution to the function value.

Definition 2 (Marginal gain)**.** *For a finite set G and a given set function $g : 2^G \to \mathbb{R}$, the marginal gain of g at a subset $A \subseteq G$ with respect to an element $a \in G \setminus A$ is*

$$\mathrm{MG}_g(a|A) \triangleq g(A \cup \{a\}) - g(A).$$

The above captures the change in the set function g when adding a new element a to the set A.

Definition 3 (Submodularity and Supermodularity)**.** *For a given finite set G, a set function $g : 2^G \to \mathbb{R}$ is* submodular *if for all subsets $A \subseteq B \subseteq G$ and $a \in G \backslash B$, it holds that*

$$\mathrm{MG}_g(a|A) \geq \mathrm{MG}_g(a|B), \tag{5}$$

A function g is said to be supermodular *if $-g$ is submodular.*

Submodularity captures the diminishing returns property: the contribution of adding a new element to a smaller set is higher w.r.t. adding the new element to a bigger set.

Definition 4 (Normalized and Monotone)**.** *For a given finite set G, a set function $g : 2^G \to \mathbb{R}$ is normalized if $g(\emptyset) = 0$, i.e., the empty set carries no value, and the set function is monotone non-increasing if for all subsets $A, B \subseteq G$ it holds*

$$A \subseteq B \Rightarrow g(A) \geq g(B),$$

Theorem 1 (Proposition 1.1 in [10])**.** *For a given finite set G, a set function $g : 2^G \to \mathbb{R}$ is submodular $\iff$ the derived set functions $g_a : 2^{G-\{a\}} \to \mathbb{R}$,*

$$g_a(A) \triangleq \mathrm{MG}_g(a|A) \triangleq g(A \cup \{a\}) - g(A)$$

are monotone decreasing for all $A \subseteq G$ and $a \in G \backslash A$.

Next, we outline greedy algorithms to approximately solve Problems 1 and 2, and present their associated performance guarantees in the context of supermodular minimization.

B. Greedy Algorithms

Solving Problems 1 and 2 is NP-hard [11]. Therefore, in this paper we are interested in approximation algorithms that can be implemented in polynomial time.

A popular approximation algorithm is the following greedy algorithm, applied to the minimization Problem 1.

Calling S^1_{greedy} the greedy solution resulting from Algorithm 1, it is well known that if g is a monotone supermodular (normalized) set function, then the greedy algorithm enjoys the following suboptimality guarantee:

$$g(S^1_{\text{greedy}}) \leq (1 - 1/e) g(S^1_*) \approx 0.63\ g(S^1_*), \tag{6}$$

where S^1_* is the optimal solution of Problem 1. In reading (6) we remark that the cost g is non-positive. Eq. (6) provides

Algorithm 1: Greedy algorithm for Problem 1.

Input: g, r;
Output: greedy solution S^1_{greedy};
$S_0 \leftarrow \emptyset$, $i \leftarrow 0$;
while $i < r$ **do**
  $S_{i+1} \leftarrow S_i \cup \{\arg\min_{a \in G \setminus S_i} \text{MG}_g(a|S_i)\}$;
  $i \leftarrow i + 1$;
end
return S_i;

the best approximation any polynomial-time algorithm can achieve [3]; the above is a worst-case bound and the greedy algorithm often performs much better in practice.

An analogous algorithm can be devised for Problem 2:

Algorithm 2: Greedy algorithm for Problem 2.

Input: g, ξ;
Output: greedy solution S^2_{greedy} ;
$S_0 \leftarrow \emptyset$, $i \leftarrow 0$;
while $g(S_i) > \xi$ **do**
  $S_{i+1} \leftarrow S_i \cup \{\arg\min_{a \in G \setminus S_i} \text{MG}_g(a|S_i)\}$;
  $i \leftarrow i + 1$;
end

When g is a monotone supermodular (normalized) set function, then the solution S^2_{greedy} of Algorithm 2 has suboptimality guarantees. In particular, if k is the smallest index such that $g(S_k) \leq \xi$, then, the following is true

$$\frac{|S^2_{\text{greedy}}|}{|S^2_*|} \leq 1 + \log_e \frac{g(G) - g(\emptyset)}{g(G) - g(S_{k-1})},$$

where S^2_* is the optimal solution of Problem 2. In words, the greedy algorithm does not select too many sensors to meet the error budget ξ, when compared to the optimal selection.

C. Supermodularity in MSE sensor scheduling

We provide conditions under which the MSE is a supermodular function of the sensor set. Related work [6] provides counterexamples showing that the trace of the covariance is not supermodular in general. Our route here is to devise a subset of linear systems for which supermodularity holds.

To further motivate our interest towards the MSE, we start our discussion with an alternative yet related cost function:

$$g_\Omega(S) \triangleq \text{trace}(P^{-1}_{t|t-1}) - \text{trace}(P_{t|t}(S)^{-1}). \tag{7}$$

The inverse of the covariance $P_{t|t}(S)^{-1}$ is called the *information matrix*. The following trivially follows from (3).

Lemma 1. *The function $g_\Omega(S)$ defined in* (7) *is a normalized, monotone non-increasing, modular function of the set S.*

One might argue that minimizing $\text{trace}(P_{t|t}(S))$ in (4) leads to similar choices of sensor sets compared to minimizing $-\text{trace}(P_{t|t}(S)^{-1})$ in (7), hence $g_\Omega(S)$ is a meaningful and easy to optimize cost function. The trace of the information matrix is indeed referred to as the T-optimality criterion in optimal experimental design, e.g., see [12].

Unfortunately, with a simple example one can realize that (7) may not lead to clever sensor schedules. Consider the case in which each sensor i measures the i-th state, i.e., $C(\{i\}) = e_i^T$, where e_i is the basis vector which is zero everywhere, except the i-th entry which is 1, and $V(\{i\}) = 1$. Then, using (3), the following equality holds:

$$g_\Omega(S) = \text{trace}(P^{-1}_{t|t-1}) - \text{trace}\left(P^{-1}_{t|t-1} + \sum_{i \in S} e_i e_i^T\right) = -|S|$$

Therefore, any set of given cardinality has the same cost using $g_\Omega(S)$, i.e., the function is "indifferent" to the selection of sensors. This motivates us to study the trace of the covariance (rather than the one of the information matrix).

We start by proving the following lemma.

Lemma 2 (Decrease in MSE)**.** *The set function $g(S)$ in* (4) *is normalized, non-positive, and monotonically non-increasing.*

While optimizing the trace of the information matrix is not a convenient choice, for our proofs it is still convenient to work in information form. In order to do this, we need the following notation.

Definition 5 (Information matrices in KF)**.** *The inverse of the prior covariance matrix $P_{t|t-1}$ is called the* prior information matrix*: $\Omega_{t|t-1} \triangleq P^{-1}_{t|t-1}$. Moreover, the inverse of the posterior covariance matrix $P_{t|t}(S)$ is called the* posterior information matrix*: $\Omega_{t|t}(S) \triangleq P_{t|t}(S)^{-1}$.*

Using Definition 5 it is possible to rewrite (3) (the update step) in information form:

$$\Omega_{t|t}(S) = \Omega_{t|t-1} + C(S)^T V_t(S)^{-1} C(S) \tag{8}$$

The information matrix $\Omega_{t|t}(S)$ is symmetric positive definite by construction with $\Omega_{t|t-1}$ comprising of the process contribution and $C(S)^T V_t(S)^{-1} C(S)$ comprising the contribution due to the measurements.

We note that the space of symmetric matrices is partially ordered by the semi-definite partial order and has the following useful properties.

Theorem 2 (Löwner-Heinz Inequality [13], 1934)**.** *For any two symmetric positive semi-definite matrices M, N:*

$$M \succeq N \succeq 0 \Rightarrow M^p \succeq N^p, \ \forall p \in [0, 1].$$

Theorem 2 succinctly captures the fact that among all non-negative exponents p, the power function $t \mapsto t^p$ is operator monotone (i.e., can be represented as a Pick function, see [13] for details) for $0 \leq p \leq 1$. In particular, the positive semi-definite partial ordering is not always preserved (i.e., the implication does not always hold) for integers $p > 1$.

Remark 1 (Additivity property of the information matrix $\Omega_{t-1:t}(S)$)**.** *For two disjoint subsets $S_1, S_2 \subseteq G$, we have*

$$\Omega_{t|t}(S_1 \cup S_2) = \Omega_{t|t}(S_1) + \Omega_{t|t}(S_2), \tag{9}$$

due to the additive structure of $\Omega_{t|t}(S)$ *in* (8).

We can now prove a first key result that relates supermodularity to the trace of a specific matrix product.

Proposition 1 (Trace Inequality $\Rightarrow$ Supermodularity)**.** *For a finite set* G*, the set function* g *in* (4) *is supermodular if, for any scalar* $\gamma \in [0,1]$ *and any subsets* $S_1 \subseteq S_2 \subseteq G$ *and element* $a \in G \setminus S_2$*, the following inequality holds:*

$$\operatorname{trace}(\Psi(\gamma)\ \mathcal{F}) \leq 0. \tag{10}$$

where the matrices $\Psi(\gamma)$ *and* $\mathcal{F}$ *are defined as follows:*

$$\Psi(\gamma) \triangleq \left(\Omega(\gamma) + \Delta_a\right)^{-2} - \Omega(\gamma)^{-2}, \tag{11}$$

$$\mathcal{F} \triangleq \Omega_{t|t}(S_2) - \Omega_{t|t}(S_1) \tag{12}$$

with $\Delta_a = C(a)^T V_t(a)^{-1} C(a)$ *and* $\Omega(\gamma) = \Omega_{t|t}(S_1) + \gamma(\Omega_{t|t}(S_2) - \Omega_{t|t}(S_1))$*. Moreover* $\mathcal{F}$ *is positive semidefinite.*

Proposition 1 provides conditions under which the MSE is supermodular. However, the conditions are still quite hard to grasp, since they cannot be easily related to the matrices defining the linear system (1). Moreover, verifying if the trace inequality (10) is satisfied still requires checking all possible subsets of G, which is intractable. This leads us to further develop (10) to get a more usable condition. Before doing so, two remarks are in order.

Remark 2. *It may be tempting to believe that, since* $\Delta_a \succeq 0$*:*

$$\begin{aligned}\Omega(\gamma) + \Delta_a \succeq \Omega(\gamma) \overset{(?)}{\Rightarrow} (\Omega(\gamma) + \Delta_a)^2 \ \succeq \Omega(\gamma)^2 \\ \Rightarrow (\Omega(\gamma) + \Delta_a)^{-2} \preceq \Omega(\gamma)^{-2}\end{aligned} \tag{13}$$

and thus claim that $\Psi(\gamma) \preceq 0$ *in* (11)*. A similar logic was undertaken in [14] to prove Theorem 5 in turn claiming that their set function is submodular and monotone increasing irrespectively of the system dynamics or choice of sensor set; however, this is not generally the case as the first implication* (13) *(labeled with the question mark) is not true in general when the exponent* $p > 1$ *(see Theorem 2). This is further substantiated by the counterexample in [15, Proposition 2.4].*

A sufficient condition that guarantees satisfaction of Proposition 1 is to require $\Psi(\gamma) \preceq 0$. In the following corollary we show that this requirement is indeed too strict to be practically useful.

Corollary 1. *The trace inequality* (10) *in Proposition 1 is satisfied if the matrix* $\Psi(\gamma)$ *is negative semidefinite; moreover,* $\Psi(\gamma) \preceq 0$ *if and only if* $C(a)$ *is either zero or is an eigenvector of* $\Omega(\gamma)$*.*

Corollary 1 suggests that asking $\Psi(\gamma) \preceq 0$ is too restrictive. Thus, in the following we leverage our assumption on the measurement matrix (Assumption 1) to get more usable conditions for the trace inequality (10) to be satisfied.

Corollary 2. *If Assumption 1 holds true, the trace inequality* (10) *in Proposition 1 is satisfied if all the diagonal entries of* $\Psi(\gamma)$ *are non-positive.*

In the following, we first derive conditions on the inverse of $\Omega(\gamma)$ that make the diagonal elements of $\Psi(\gamma)$ negative, satisfying Corollary 2 and in turn the trace inequality (10). Then we translate these conditions into conditions on the prior information matrix $\Omega_{t|t-1}$.

Proposition 2. *If Assumption 1 holds true, the trace inequality* (10) *in Proposition 1 is satisfied if all the columns of:*

$$P(\gamma) \triangleq \Omega(\gamma)^{-1} \tag{14}$$

satisfy the following inequality,

$$\left\| P(\gamma)_{:i} - \left(\frac{P(\gamma)_{ij}}{\bar{c}^{-2} + P(\gamma)_{jj}} \right) P(\gamma)_{:j} \right\|_2 \leq \|P(\gamma)_{:i}\|_2, \tag{15}$$

where $P(\gamma)_{:i}$ *denotes the* i*-th column of the matrix* $P(\gamma)$*,* $P(\gamma)_{ij}$ *denotes the entry of* $P(\gamma)$ *in position* (i,j)*, and* $\bar{c}$ *is the only nonzero entry of the matrix* $C(a)^T V_t(a)^{-1} C(a)$*.*

Next, we specify two classes of matrices which leads the inequality in (15) to satisfaction.

Definition 6 (Strictly-diagonally dominant M-matrix [16])**.** *A matrix* $M \in \mathbb{R}^{n \times n}$ *is said to be an M-matrix if it has non-positive off-diagonal entries (*$M_{ij} \leq 0$*) and its eigenvalue have positive real parts. An M-matrix* M *is said to be strictly diagonally dominant if the following inequalities hold:*

$$|M_{ii}| > \sum_{j \neq i} |M_{ij}|, \quad \forall i = 1, \ldots, n. \tag{16}$$

Definition 7 (Strictly ultrametric matrix [17])**.** *A matrix* $U \in \mathbb{R}^{n \times n}$ *with elements* $[U_{ij}]$ *is a strictly ultrametric matrix if*

1) U *is symmetric with non-negative entries*
2) $U_{ij} \geq \min\{U_{ik}, U_{kj}\}$*,* $\forall (i,k,j) \in \{1, \ldots, n\}$
3) $U_{ii} > U_{ik}$*,* $\forall (i \neq k) \in \{1, \ldots, n\}$

Strictly ultrametric matrices have been studied as covariance matrices of random energy models in statistical physics, e.g., see [18], as generalizations of the diagonal case. The following theorem formalizes the fact that ultrametric matrices imply supermodularity in MSE minimization.

Theorem 3. *If Assumption 1 holds true, the trace inequality* (10) *in Proposition 1 is satisfied if the matrix* $P(\gamma)$*, defined in* (14)*, is a strictly ultrametric matrix.*

With the support of Definition 6 we present our final results. This result translates the conditions in Theorem 3, which are not easy to check in practice (the matrix $P(\gamma)$ depends on γ, S_1, and S_2), into conditions on the prior information matrix $\Omega_{t|t-1}$; the latter conditions can be readily verified by inspection.

Theorem 4 (Condition for supermodularity)**.** *If the prior information matrix* $\Omega_{t|t-1}$ *in Definition 5 is a strictly-diagonally-dominant M-matrix with strictly ultrametric inverse, then,* $P(\gamma)$ *in* (14) *is strictly ultrametric. Further, under Assumption 1, the trace inequality* (10) *in Proposition 1 is satisfied and the cost* $g(S)$ *is supermodular w.r.t.* S*.*

The conditions of Theorem 4 can be readily checked from the prior information matrix and its inverse by applying the

Definitions 6-7. We note that a symmetric positive definite M-matrix is also referred to as a *Stieltjes* matrix [16]. The class of Stieltjes matrices is well studied and a celebrated result says that the inverse of a strictly ultrametric matrix is a s.d.d. Stieltjes matrix [17], while the reverse is not necessarily true, for $n \geq 2$.

The satisfaction of Proposition 1 under the conditions of Theorem 4 ensures that the corresponding MSE minimization is supermodular, hence a greedy algorithm produces near-optimal solutions. We conclude the theoretical contribution of this paper with a remark about the tightness of Theorem 4.

Remark 3 (Tightness of Theorem 4)**.** *One might wonder if the conditions (on the prior information matrix and its inverse) required by Theorem 4 are also necessary for supermodularity. Empirical evidence presented in Section IV suggests that the assumptions can be loosened. In particular, the experiments suggest that whenever the prior information matrix is an M-matrix, supermodularity holds. This observation is useful since M-matrices naturally arise in estimation problem over networks (Section IV-B). Ongoing work involves relaxing some of the assumptions in Theorem 4.*

IV. Numerical Experiment

In this section, we validate our results in two different setups. For the first setup, we consider random linear systems and empirically show that some conditions in Theorem 4 can be relaxed, as the requirement of the prior information matrix being an M-matrix is sufficient for supermodularity to hold. The second setup provides an example of the applicability of our results to a multi robot localization problem.

A. Random Linear Systems

In this experiment, we empirically show that if the prior information matrix $\Omega_{t|t-1}$ is an M-matrix, then supermodularity holds. The numerical setup is as follows: we generate 1000 random systems ($n = 6$) with $\Omega_{t|t-1}$ being an M-matrix, and another 1000 random systems with $\Omega_{t|t-1}$ being a positive definite matrix. The latter one is a generic prior information matrix resulting from the Kalman filter. We assume $C(i) = e_i^T$, i.e., sensor i measures the i-th entry of the state vector. To test supermodularity, we check Definition 4 for all choices of subsets $S_1 \subseteq S_2 \subseteq G$, and $a \in G \setminus S_2$. Supermodularity holds if the marginal gain is monotonically increasing, i.e., $\delta(a, S_1, S_2) = \mathrm{MG}_g(a|S_1) - \mathrm{MG}_g(a|S_2) \leq 0$; this brute-force check is feasible since the dimension of the state space is small ($n = 6$). For each system we store the maximum value of $\delta(a, S_1, S_2)$ across the tests, which is denoted as $\delta_m(a, S_1, S_2)$. If $\delta_m(a, S_1, S_2) \leq 0$, supermodularity holds numerically for the corresponding random system.

Fig. 1 shows the results of the supermodularity tests. When the prior information matrix is a general positive definite matrix, Fig. 1(b), then supermodularity fails and $\delta_m(a, S_1, S_2)$ becomes positive; this agrees with the counterexamples in the literature. However, when the prior information matrix is an M-matrix, then supermodularity empirically holds, Fig. 1(a).

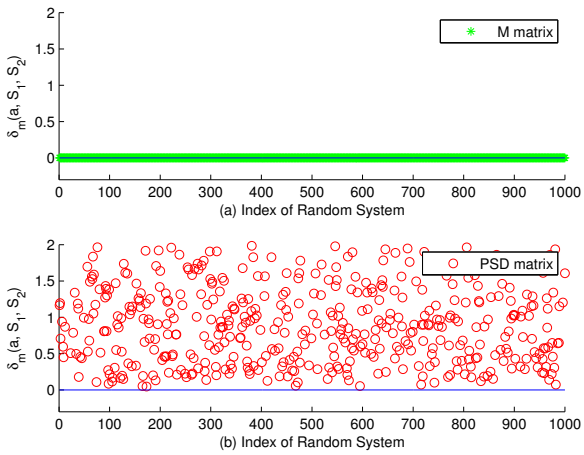

Fig. 1: Supermodularity tests on random systems. $\delta_m(a, S_1, S_2) \leq 0$ indicates that supermodularity holds, while $\delta_m(a, S_1, S_2) > 0$ corresponds to failure in the problem being supermodular.

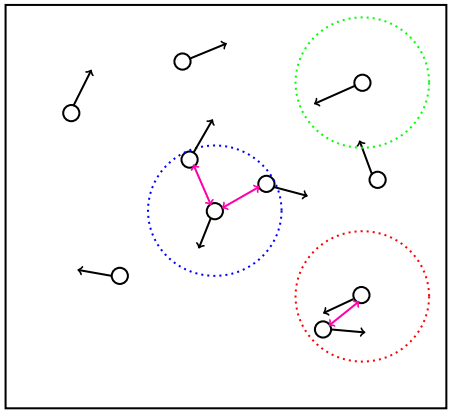

Fig. 2: Multi robot localization. Each solid circle denotes a robot, moving in the direction defined by the corresponding arrow. The dotted circles picture the sensing range of the robots. Each robot can communicate and take relative measurements of teammates within its sensing range. The (bi-directional) magenta arrows represent pairs of robots that can communicate with each other. Due to energy constraints, only a subset of the robots can take GPS measurements.

B. Resource-Constrained Multi Robot Localization

Consider a swarm of n robots moving in the 2D plane according to the following linear dynamics (see Fig. 2):

$$x_{t+1} = A_t x_t + u_t + w_t \tag{17}$$

where $x_t \triangleq [x_t^{(1)} \; x_t^{(2)} \; \dots \; x_t^{(n)}] \in \mathbb{R}^{2n}$ denotes the position of all robots in the swarm at time t, and $x_t^{(i)} \in \mathbb{R}^2$ denotes the 2D position of robot i. We assume $A_t = I_{2n\times 2n}$ meaning that each robot controls its displacement independently through its local control action $u_t^{(i)}$ (with $u_t \triangleq [u_t^{(1)} \; u_t^{(2)} \; \dots \; u_t^{(n)}] \in \mathbb{R}^{2n}$); as in the rest of the paper, $w_t \sim \mathcal{N}(0, W_t)$, for all $t \geq 0$. For our tests we assume that the control actions are randomly chosen at each time-step, resulting in each robot performing a random walk.

We tackle a multi robot localization problem in which we estimate the absolute positions x_t of the robots from on-board sensor measurements. We assume that each robot

has two sensors: a laser scanner, that allows measuring the relative position of other robots within a sensing radius, and a GPS, that measures the absolute position of the robot. While a robot can always acquire relative measurements of neighboring robots, we consider the case in which energy constraints prevent all robots to collect GPS data, hence only a subset of the robots can use the GPS at each time step and our task is to compute the most convenient subset of those robots, so to minimize the mean squared error of the global positions of all robots. For this purpose, we show that for this localization problem the prior information (before GPS measurements are included) is an M-matrix, and that the greedy algorithm produces near-optimal results.

Mathematically, the measurement model for the relative measurements from the laser scanner is as follows:

$$y^r_{t,ij} = (x^j_t - x^i_t) + v^r_{t,ij} \tag{18}$$

where each pair of robots $(i, j) \in \{1, \dots, n\}$, within sensing radius, measures the relative state $x^j_t - x^i_t$ of the sensed robot, plus zero-mean additive Gaussian noise $v^r_{t,ij}$. For the sake of simplicity we assume $v^r_{t,ij} \sim \mathcal{N}(0, I_{2\times 2})$. Stacking all measurements $y^r_{t,ij} \in \mathbb{R}^2$ in (19) into a single vector y^r_t we get a more compact measurement model:

$$y^r_t = C^r_t x_t + v^r_t \tag{19}$$

It is well known [19] that $C^r_t \in \mathbb{R}^{2n_e \times 2n}$ is a close relative of the incidence matrix of the graph underlying the problem; such graph has a vertex for each robot and edges connecting robots within communication radius (we denote with n_e the number of edges). In particular, C^r_t includes n_e block-rows of size $2 \times 2n$, where each block row is zero everywhere and equal to $I_{2\times 2}$ and $-I_{2\times 2}$ for the entries corresponding to the position of robot j and i, respectively.

The GPS measurement model for robot i is:

$$y^a_{t,i} = x^{(i)}_t + v^a_{t,i}, \tag{20}$$

i.e., the GPS measures the absolute position $x^{(i)}_t$ of robot i, plus zero mean Gaussian noise $v^a_{t,i}$. Assuming that $G = \{1, \dots, n\}$ is the set of all robots, and assuming that we can only acquire GPS measurements from a subset $S \subseteq G$ of r robots (i.e., $|S| = r$), then we can express model (21) in matrix form as:

$$y^a_t = C(S)x_t + v^a_t, \tag{21}$$

We note that $C(S)$ is a monomial measurement matrix.

This setup naturally falls within the framework of the cardinality-constrained single time-step optimal sensor scheduling problem outlined in Problem 1. Note that we only choose the subset of GPS measurements, while we always collect the relative measurements when possible.

The prior information matrix, before including GPS measurements, is given by:

$$\Omega_{t|t-1} = (P_{t-1|t-1} + W_{t-1})^{-1} + (C^r_t)^T (C^r_t) \tag{22}$$

where we used our assumptions about the matrix A_t and the measurement covariance being the identity. Assuming that

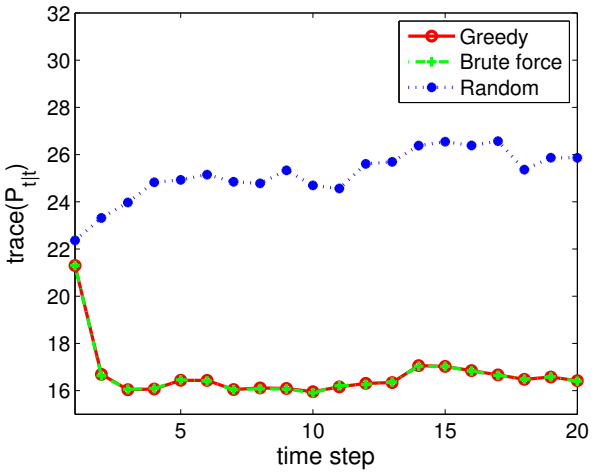

Fig. 3: Multi robot localization: MSE comparison for different sensor schedules. The results are averaged over 100 runs.

$\Omega_{0|0}$ is an M-matrix, implies that $\Omega_{t|t-1}$ remains an M-matrix for all $t \geq 0$ (a simple proof is given in the supplemental material [9]). Therefore, according to the empirical evidence of Section IV-A, we conclude that supermodularity holds and the greedy algorithm applied to our GPS-measurement-selection problem obtains near-optimal results.

For our tests we set $n = 10$ and $r = 4$. Moreover, we assume that time $t = 0$ the robots are uniformly distributed in a square environment and they move according to the random walk (17), and that the initial covariance is $P_{0|0} = I_{2n\times 2n}$.

Fig. 3 compares the MSE resulting from the GPS-measurement selection obtained by three approaches applied to the sensor scheduling Problem 1:

1) the greedy Algorithm 1 (solid red line);
2) a brute-force search algorithm, which gives the optimal solution of Problem 1 (dashed green line);
3) a random selection algorithm that selects r out of n robots uniformly at random (dotted blue line).

Fig. 3 shows the MSE for a 20-time-step simulation. The sensor scheduling problem is solved at each time step. The MSE is averaged over 100 Monte Carlo runs. We note that the performance of the greedy algorithm is practically indistinguishable from the optimal selection, and remarkably better than a random selection, further confirming the practicality of the findings in this paper.

V. Conclusion and Future Work

We study a single-time-step sensor scheduling problem in which one needs to select r out of N available sensors to minimize the MSE of an optimal estimator. Our goal is to study specific properties of the MSE, namely *supermodularity*, which enables efficient near-optimal selection of the sensor set via greedy algorithms. While it is known that the MSE is not supermodular (nor submodular) in general, we show that if the prior information matrix is a strictly-diagonally-dominant M-matrix with strictly ultrametric inverse, then the MSE (trace of the covariance) is indeed supermodular. We elucidate on our theoretical results with

numerical simulations on random linear systems and we provide a concrete application to a multi robot localization problem. Empirical evidence shows that the only requirement that is actually needed for supermodularity is that the information matrix is an M-matrix. Future work includes relaxing current assumptions on the prior information matrix to meet the empirical evidence.

REFERENCES

[1] Andreas Krause, Ram Rajagopal, Anupam Gupta, and Carlos Guestrin. Simultaneous placement and scheduling of sensors. In *Proceedings of the 2009 International Conference on Information Processing in Sensor Networks*, pages 181–192. IEEE Computer Society, 2009.

[2] Vijay Gupta, Timothy H Chung, Babak Hassibi, and Richard M Murray. On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage. *Automatica*, 42(2):251–260, 2006.

[3] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming*, 14(1):265–294, 1978.

[4] Manohar Shamaiah, Siddhartha Banerjee, and Haris Vikalo. Greedy sensor selection: Leveraging submodularity. In *49th IEEE conference on decision and control (CDC)*, pages 2572–2577. IEEE, 2010.

[5] Siddharth Joshi and Stephen Boyd. Sensor selection via convex optimization. *IEEE Transactions on Signal Processing*, 57(2):451–462, 2009.

[6] Syed Talha Jawaid and Stephen L Smith. Submodularity and greedy algorithms in sensor scheduling for linear dynamical systems. *Automatica*, 61:282–288, 2015.

[7] Vasileios Tzoumas, Ali Jadbabaie, and George J Pappas. Sensor placement for optimal kalman filtering: Fundamental limits, submodularity, and algorithms. *arXiv preprint arXiv:1509.08146*, 2015.

[8] Haotian Zhang, Raid Ayoub, and Shreyas Sundaram. Sensor selection for optimal filtering of linear dynamical systems: Complexity and approximation. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 5002–5007. IEEE, 2015.

[9] 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 *supplemental material, available online at:* `https://dl.dropboxusercontent.com/u/4309261/shared/Singh16acc-supplemental.pdf`, 2016.

[10] László Lovász. Submodular functions and convexity. In *Mathematical Programming The State of the Art*, pages 235–257. Springer, 1983.

[11] Uriel Feige. A threshold of ln n for approximating set cover. *Journal of the ACM (JACM)*, 45(4):634–652, 1998.

[12] Anthony Atkinson, Alexander Donev, and Randall Tobias. *Optimum experimental designs, with SAS*, volume 34. Oxford University Press, 2007.

[13] Karl Löwner. Über monotone matrixfunktionen. *Mathematische Zeitschrift*, 38(1):177–216, 1934.

[14] Tyler H Summers, Fabrizio L Cortesi, and John Lygeros. On submodularity and controllability in complex dynamical networks. *IEEE Transactions on Control of Network Systems*, 3(1):91–101, 2016.

[15] Pattrawut Chansangiam. Operator monotone functions: Characterizations and integral representations. *arXiv preprint arXiv:1305.2471*, 2013.

[16] Claude Dellacherie, Servet Martinez, and Jaime San Martin. *Inverse M-matrices and ultrametric matrices*, volume 2118. Springer, 2014.

[17] Servet Martínez, Gérard Michon, and Jaime San Martín. Inverse of strictly ultrametric matrices are of stieltjes type. *SIAM Journal on Matrix Analysis and Applications*, 15(1):98–106, 1994.

[18] Antonio Galves, Servet Martinez, and Pierre Picco. Fluctuations in derrida's random energy and generalized random energy models. *Journal of statistical physics*, 54(1-2):515–529, 1989.

[19] Prabir Barooah and Joao P Hespanha. Graph effective resistance and distributed control: Spectral properties and applications. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 3479–3485. IEEE, 2006.

[20] Carlos Guestrin, Andreas Krause, and Ajit Paul Singh. Near-optimal sensor placements in gaussian processes. In *Proceedings of the 22nd international conference on Machine learning*, pages 265–272. ACM, 2005.

[21] Kristian Kersting, Christian Plagemann, Patrick Pfaff, and Wolfram Burgard. Most likely heteroscedastic gaussian process regression. In *Proceedings of the 24th international conference on Machine learning*, pages 393–400. ACM, 2007.

[22] Hadi Jamali-Rad, Andrea Simonetto, Xiaoli Ma, and Geert Leus. Distributed sparsity-aware sensor selection. *IEEE Transactions on Signal Processing*, 63(22):5951–5964, 2015.