On the Submodularity of Sensor Scheduling for Estimation of Linear Dynamical Systems

Syed Talha Jawaid Stephen L. Smith

***Abstract*— In this paper, we study the problem of using an energy constrained sensor network to estimate the state of a linear dynamical system. The state estimate is computed using a Kalman filter and the goal is to choose a subset of sensors at each time step so as to minimize the *a posteriori* error covariance. Recent work has indicated that the simple greedy algorithm, which chooses the sensor at each time step that maximizes the error covariance reduction, outperforms many other known scheduling algorithms. In addition, it has been suggested that the cost function mapping a sensor sequence to an error covariance cost is submodular; this would imply that the greedy algorithm provides a near optimal sensor schedule.**

As a negative result, we show that the sensor schedule cost is not, in general, a submodular function. This contradicts an established result. We argue that given a linear dynamical system, it is computationally intractable to determine if it will yield a submodular cost. Thus, we provide sufficient and easily checkable conditions under which the dynamical system yields a submodular cost, and thus performance guarantees for the greedy schedule.

I. Introduction

Consider the problem of estimating the time evolution of a physical process: for example, the variation of temperature in a lake or ocean, or the spread of an algae bloom on the surface of a body of water. A common way to monitor such a process is to deploy a network of static sensors. An example is the Argo array [1], which consists of thousands of nodes deployed throughout the global oceans. To obtain the best estimate of the physical process, one option is to have every sensor on and collecting data at all times. However, for long-term deployments, where energy-consumption is a priority, this may not be viable. A proposed alternative is to turn on only a small subset of the sensors at each time step (where time has been discretized). The estimation problem then becomes one of sensor scheduling. Recent work has suggested that a simple greedy approach for selecting sensors to turn on is competitive with the best known algorithms [2]. In this paper we investigate this in detail.

The sensor scheduling problem has received considerable attention in recent years. In [3], the authors look at the problem of using a Kalman filter in the case where measurements are available at each time step with a certain probability. It is shown that the error covariance will be bounded only if the probability of receiving a measurement is above a certain critical value. In [4], the authors provide a method for stochastically selecting measurements for a Kalman filter, based on an intelligently constructed probability distribution, to minimize the expected steady state error covariance.

In [5], a convex relaxation based approach is given to pick measurements for parameter estimation. However, in [2] this approach is empirically shown to be perform worse than a greedy algorithm when optimizing the maximum *a posteriori*. In [2], the authors also show that a greedy algorithm gives a constant factor approximation, since the objective is submodular. In [6] this approach is extended to randomly dropped measurements.

A general framework for sensor scheduling in state estimation is presented in [7]. A number of cost functions can be optimized in this framework, including the final covariance, the average covariance, and the cost of a finite horizon LQG regulator. Network constraints can also be included. The problem is framed as a relaxed quadratic program. A greedy approach is also given, but the error bound is not necessarily constant for unstable systems. In [8], optimal and near-optimal sensor scheduling algorithms are given that rely on tree pruning techniques.

The sensor scheduling problem arises in several applications. In [9], the problem of monitoring CO_2 using a wireless sensor network is demonstrated. The authors use a convex relaxation to approximate the optimal *a posteriori* covariance for a Kalman filter. The sensor scheduling problem is combined with the controller problem in [10] to demonstrate an application for controlling the water level in multiple tanks.

The contributions of this paper are two fold. First, we provide a negative result, showing that the sensor schedule cost is not, in general, a submodular nor a monotone function. This result holds for most objectives including the trace of the covariance and the log of the determinant of the covariance, and it disproves an established result in [11]. Second, we provide sufficient and easily checkable conditions (although, very restrictive) under which the dynamical system yields a submodular cost, and thus performance guarantees for the greedy schedule.

Organization: In Section III we define the problem. In Section IV we provide the main negative results on submodularity and in Section V we provide sufficient conditions when the objective is the log of the determinant of the *a posteriori* covariance matrix.

II. Preliminaries

In this section we review some essential concepts in submodular set and sequence functions.

A. Independence Systems

A large class of combinatorial optimization problems can be formulated as the maximization or minimization of an

This research is partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo ON, N2L 3G1 Canada (stjawaid@uwaterloo.ca; stephen.smith@uwaterloo.ca)

objective function $f : \mathcal{F} \to \mathbb{R}$ over a set system $(E, \mathcal{F})$. The set E is the base set of all elements and $\mathcal{F} \subseteq 2^E$. An **independence system** is a set system that is closed under subsets (i.e., if $A \in \mathcal{F}$ then $B \subseteq A \implies B \in \mathcal{F}$). A useful class of independence systems are called matroids.

Definition II.1 (Matroid)**.** An independence system $(E, \mathcal{F})$ is a matroid if it satisfies the additional property that if $X, Y \in \mathcal{F}$ such that $|X| > |Y|$, then $\exists x \in X \backslash Y$ with $Y \cup \{x\} \in \mathcal{F}$.

The *uniform matroid* is defined by the collection of all sets of size less then or equal to $m \in \mathbb{Z}_+$, i.e., $\mathcal{F} := \{A \subseteq E : |A| \leq m\}$. Another example is the *partition matroid*. The base set is composed of n disjoint sets, $\{E_i\}_{i=1}^n$. Given $k \in \mathbb{Z}_+^n$, the matroid is defined by the collection $\mathcal{F} := \{A \subseteq E : |A \cap E_i| \leq k_i, \forall i = 1 \dots n\}$.

B. Sequences

For our purposes, a **sequence** $A = (a_1, \dots, a_k)$, $k \in \mathbb{Z}_{\geq 0}$ consists of elements from a base set of elements E, i.e., $a_i \in E$. The **length** of a sequence, $|A|$, is the number of elements in the sequence: if $A = (a_1, \dots, a_k)$, then $|A| = k$. Two sequences $A = (a_1, \dots, a_k)$ and $B = (b_1, \dots, b_\ell)$ defined over the same base set can be **concatenated** into a larger sequence: $A \parallel B = (a_1, \dots, a_k, b_1, \dots, b_\ell)$. A **subsequence** of A is a sequence derived from it by deleting some elements but not changing the order of the remaining elements, e.g., $B = (a_3, a_5)$ is a subsequence of A, and is denoted $B \subseteq A$.

C. Set Functions

Let E be a finite set. A set function, f, defined over E assigns a value to every subset of E, i.e., $f : 2^E \to \mathbb{R}$.

Definition II.2 (Normalized and Monotone)**.** The function f is normalized if $f(\emptyset) = 0$. The function, f, is monotone non-decreasing if for all $A \subseteq B \subseteq E$, $f(A) \leq f(B)$.

Definition II.3 (Submodularity)**.** The function f is submodular if $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$, for all $S, T \subseteq E$.

Submodular functions satisfy the property of *diminishing marginal returns*. That is, the contribution of any element x to the total value of a set decreases as the set gets bigger. More formally, let $\Delta_f(B|A) := f(A \cup B) - f(A)$. Then,

$$\Delta_f(x|A) \geq \Delta_f(x|B), \quad \forall A \subseteq B \subseteq E.$$

D. Sequence Functions

A sequence function defined over a base set E is one that takes in a sequence over E, and outputs a number in $\mathbb{R}$. Note that this is different from a set function in that the order of the elements in the sequence matters, i.e., $f\big((a, b)\big) \neq f\big((b, a)\big)$.

The definitions in the previous section can also be applied to sequence functions. Let E be a finite set and f a function defined on sequences derived from E.

Definition II.4 (Monotonicity)**.** The function, f, is monotone non-decreasing if for all subsequences A of a sequence B, i.e., $A \subseteq B$, $f(A) \leq f(B)$. Similarly, it is monotone non-increasing if $f(A) \geq f(B)$. If neither of these conditions hold, then the function is non-monotone.

The concept of submodularity can also be applied to sequence functions. Denote the marginal value of appending a sequence C to a sequence A as $\Delta_f(C|A) := f(A \parallel C) - f(A)$, where $A \parallel C$ is the concatenation of A and C.

Definition II.5 (Submodularity)**.** The function f is submodular if $\Delta_f(C|A) \geq \Delta(C|B)$, for all $A \subseteq B$.

The subscript f will be omitted unless there is ambiguity.

E. Submodular Function Optimization

An effective method for maximizing a submodular function (set or sequence), is the greedy algorithm. This algorithm begins with an empty set S and repeatedly adds the element x of E to S that maximizes the marginal return $\Delta(x|S)$.

Lemma II.6 (Maximizing Over a Matroid, [12], [13])**.** *Consider the problem of maximizing a submodular, non-negative and monotone non-decreasing function over a matroid. For the uniform matroid, the greedy algorithm gives a* $1 - \frac{1}{e}$ *approximation. For a general matroid, the greedy algorithm gives a* $\frac{1}{2}$ *approximation.*

At times, we may not be able to find the most valuable item from the base set, but rather one that is close to the optimal. This motivates the idea of α-approximate greedy.

Definition II.7 (α-approximate Greedy)**.** Consider an objective function f defined over an independence system $(E, \mathcal{F})$. For $\alpha \in (0, 1]$, an α-approximate greedy algorithm greedily constructs an approximation to the maximum value basis by selecting at each iteration i an element g_i such that

$$\Delta(g_i|G_{i-1}) \geq \alpha \max_{\substack{e \in E \backslash G_{i-1} \\ G_{i-1} \cup \{e\} \in \mathcal{I}}} \Delta(e|G_{i-1}),$$

where $G_i = \bigcup_{j=0}^{i} g_j$ and $g_0 = \emptyset$.

For the case of a sequence submodular functions, the following lemma quantifies the greedy algorithms performance over the sequence equivalent to a uniform matroid.

Lemma II.8 (Maximizing a Sequence [14])**.** *Consider a normalized monotone non-decreasing submodular sequence function* f *defined over the base set of elements* E*. The problem of selecting the maximum value sequence of size* T *can be approximated to within* $1 - \frac{1}{e^\alpha}$ *using an* α*-approximate greedy algorithm.*

III. The Sensor Scheduling Problem

Consider a sensor network consisting of m sensors $M := \{1, \dots, m\}$, operating in discrete time. Our goal is to estimate the state of a linear time invariant (LTI) system. To do this, we can turn on at most k sensors at each time step t. We assume that the measurements taken by each sensor at each time step are fused into a single estimate. This is commonly performed at a fusion center.

To formalize this problem, the combination of the LTI system and sensor network can be written as

$$\begin{aligned} x_{t+1} &= Ax_t + w_t \\ y_{t+1} &= S_t C x_t + S_t v_t, \end{aligned} \tag{1}$$

where $t \in \mathbb{Z}_{\geq 0}$, $x_t \in \mathbb{R}^n$, $y_t \in \mathbb{R}^k$, $A \in \mathbb{R}^{n \times n}$, and $C \in \mathbb{R}^{m \times n}$. The process noise w_t and measurement noise v_t

are zero mean and independent Gaussian noise vectors with covariance matrices $W, V \in \mathbb{R}^{n\times n}$ such that $W, V \succeq 0$.

Each row of the C matrix corresponds to a single sensor, and thus we assume that each sensor takes scalar measurements. The k sensors selected at time step t are encoded in the binary selection matrix $S_t \in \{0,1\}^{k\times m}$. Each row of S_t contains only a single element that is equal to 1, and every column contains at most one element that is equal to 1. Thus, S_tC is a $k \times m$ matrix containing k rows of C corresponding to the k sensors selected at time step t.

To estimate the state x_t of the dynamical system, we use the Kalman filter. To this end, let us define two functions $\rho^T : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ (we use T for time and $\top$ for transpose) and $\rho^M_X : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$, where X is a set of sensors from which we can construct S_t, as

$$\begin{aligned}\rho^T(\Sigma_{t-1|t-1}) &:= A\Sigma_{t-1|t-1}A^\top + W\\ \rho^M_X(\Sigma_{t|t-1}) &:= \Sigma_{t|t-1} - \Sigma_{t|t-1}C^\top S_t^\top (S_tC\Sigma_{t|t-1}C^\top S_t^\top \\ &\qquad + S_tVS_t^\top)^{-1}S_tC\Sigma_{t|t-1}\\ &= \left(\Sigma^{-1}_{t|t-1} + C^\top S_t^\top (S_tVS_t^\top)^{-1}S_tC\right)^{-1}.\end{aligned}$$

The covariance update for the Kalman filter consists of two steps: the time update, $\Sigma_{t|t-1} = \rho^T(\Sigma_{t-1|t-1})$, results in the *a priori* estimate, and the measurement update, $\Sigma_{t|t} = \rho^M_X(\Sigma_{t|t-1})$, gives the *a posteriori* estimate. It is known that the Kalman filter gives the best mean squared error of the state estimate among all linear estimators.

Given a time horizon T, our problem is to compute a sensor schedule $\sigma = (\sigma_1, \ldots, \sigma_T)$, where $\sigma_i \subset M$, $|\sigma_i| = k$ for each $i \in 1, \ldots, T$. We seek a schedule σ that maximizes a function of the information matrix $\Sigma^{-1}_{T|T}$. This is equivalent to minimizing a function of the error covariance $\Sigma_{T|T}$.

The initial *a posteriori* covariance estimate is a positive definite matrix $\Sigma_0 \succ 0$. Given a sensor schedule σ and a time $t \le T$, we can define a function $\Omega^\sigma_t : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ such that $\Omega^\sigma_t(\Sigma_0) = \Sigma_{t|t}$. The function is defined recursively as follows:

$$\begin{aligned}\Omega^\sigma_t(\Sigma_0) &:= \rho^M_{\sigma_t}(\rho^T(\Omega^\sigma_{t-1}(\Sigma_0)))\\ &= \left((A\Omega^\sigma_{t-1}(\Sigma_0)A^\top + W)^{-1} + D_t\right)^{-1},\\ \Omega^\sigma_0(\Sigma_0) &= \Sigma_0,\\ D_t &:= C^\top S_t^\top (S_tVS_t^\top)^{-1}S_tC,\end{aligned} \tag{2}$$

where S_t is the binary selection matrix given by the k sensors in σ_t and D_t is defined for convenience. We will omit the initial covariance parameter for Ω_t from now on unless there is ambiguity. Note that in this definition the first time update comes before the first measurement update, so in effect, an initial measurement is skipped. This can be avoided by using a "dummy" initial covariance of $A^{-1}(\Sigma_{0|-1} - W)A^{-\top}$ where $\Sigma_{0|-1}$ is the initial *a priori* estimate.

Formally, we seek to optimize $F(\Omega^\sigma_T(\Sigma_0))$, for some covariance metric F, under the constraint that at most k sensors can be chosen at each time step. For the rest of this section, $F(\Omega^\sigma_t(\Sigma_0))$ and $F(\sigma)$ will be used interchangeably to refer to the value of the covariance after running the sequence σ.

Remark III.1 (Set or Sequence function). When optimizing over multiple time steps, the objective is a sequence function since the time steps generate an ordered sequence of measurements. Over a single time step the objective is a simply a set function. •

We are interested in characterizing the conditions under which sensor selection can be posed as a submodular optimization problem. This will allow us to apply known results in submodular set function optimization to obtain bounds for various algorithms such as the greedy algorithm.

IV. Submodularity of the Kalman Update

In this section we show that for most of the commonly used metrics, the sensor selection problem is not, in general, submodular. We then comment on the difficulty of determining whether or not a particular system is submodular.

A. Counterexamples for Submodularity

We consider the following sensor scheduling functions, where the objective is maximization:

- $F_1 = -\operatorname{trace}(\Sigma_{T|T})$
- $F_2 = \log\det(\Sigma^{-1}_{T|T})$
- $F_3 = -\max\operatorname{eig}(\Sigma_{T|T})$
- $F_4 = \operatorname{trace}(\Sigma^\emptyset_{T+1|T} - \Sigma_{T+1|T})$, where $\Sigma^\emptyset_{t|t-1}$ is the final covariance if no sensors are selected.

The following remark gives some intuition about these metrics.

Remark IV.1 (Performance metrics). The eigenvalues of covariance matrix are proportional to the lengths of the axes of the ellipsoid that contains the estimation error $x - \hat{x}$ with a certain probability [5]. The *volume* of the confidence ellipsoid is directly proportional to the determinant of the covariance matrix, which is captured by F_2. The *mean squared error* is given by the sum of the eigenvalues of the covariance matrix, which is equivalent to its trace, as in F_1 and F_4. The *worst-case error covariance* is proportional to the maximum eigenvalue of the covariance, which is objective F_3. A detailed comparison of performance measures can be found in [15].

Since we are investigating the submodularity of sensor scheduling, we phrase each optimization as a maximization problem for which the results in Section II-E would apply. •

In [11], the authors show (see Theorems 2 and 3, [11]) that the function F_4 is submodular for a single time step as well as over multiple time steps (in the latter, the schedule consists of one sensor per time step). We give a counterexample to show that this claim is false.

Example IV.2 (Single Time Step)**.** Consider the system

$$A = \begin{bmatrix}0.5 & 0\\ 0 & 0.5\end{bmatrix}, \quad C = \begin{bmatrix}1.0 & 0.5 & 0.7 & 0.3\\ 0 & 0.5 & 0.3 & 0.7\end{bmatrix}^\top, \quad W = \Sigma_0 = I_{2\times 2}, \quad V = I_{4\times 4}. \tag{3}$$

Take $X_1 = \{2,3\} \subset \{2,3,4\} = X_2$. So, $\Delta_{F_1}(\{1\}|X_1) = 0.2736$ but $\Delta_{F_1}(\{1\}|X_2) = 0.2769$. Also $\Delta_{F_3}(\{1\}|X_1) = 0.1313$ but $\Delta_{F_3}(\{1\}|X_2) = 0.1927$. Also $\Delta_{F_4}(\{1\}|X_1) =$

0.0684 but $\Delta_{F_4}(\{1\}|X_2) = 0.0692$. These results contradict the decreasing marginal benefits property of submodularity, and disprove Theorem 2 of [11]. In fact, if we take $X_1 = \{1,3\} \subset \{1,3,4\} = X_2$, we will see that $\Delta(\{2\}|X_1) \geq \Delta(\{2\}|X_2)$ for each of the three functions. This means that they are also not supermodular. ▲

The previous example considered sensor selection over a single time step and showed that objectives F_1, F_3, and F_4 are all, in general, not submodular. Unsurprisingly, the results are even worse over multiple time steps. Using the sequences $\sigma_1 = (1,4,1,4) \subset (1,4,1,4,4) = \sigma_2$, and looking at the marginal change in value of each function if sensor 4 is chosen in the next time step, we can show that none of the four metrics are submodular.

B. A Test for Submodularity?

An interesting question is, given an LTI system of the form of (1), and an initial covariance, can we determine whether or not the cost function associated with the sensor scheduling problem will be submodular? The system model (1) induces a cost function F. The function maps sequences σ to real numbers. Thus, more generally, we can ask, given a sequence (or set) function, is it computationally tractable to determine whether or not it is submodular?

For set functions, there is no known polynomial time algorithm [16]. Since every set function can be expressed as a sequence function, this result also implies that no tractable algorithm exists for testing sequence submodularity. Without any additional structure on the sequence function, one would have to exhaustively check the submodularity definition for every possible sequence B and subsequence A. Therefore, it appears likely that, unless some particular structure of the Kalman update can be exploited, we can not in general even test whether or not a sensor scheduling problem will be submodular. This motivates the following section in which we provide a set of easily checkable sufficient conditions on the system for which the cost is submodular.

V. Submodular Sensor Scheduling Conditions

In the previous section we saw that the sensor selection problem is not, in general, submodular. Over a single time step, the only objective for which we did not provide a counterexample was F_2. In this section we look for conditions on which the $\log\det$ objective function is submodular.

A. Single Time Step

The single time-step problem is one of choosing a subset k of the m measurements. This problem is addressed in [2], and here we provide a small correction to their result. The authors show that $\log\det(\rho^T(\Omega^\sigma_{t-1})^{-1} + M_X)$ is monotone non-decreasing and submodular in the selected measurements[1] assuming that V is a diagonal matrix. As a result, the authors deduce that using a greedy algorithm gives a $(1-\frac{1}{e})$-approximation [2, Lemma 1]. As stated, this result does not hold, as the objective function is not normalized (i.e., the value of empty set is not necessarily zero). giving an additional error term that depends on the initial value: $\text{GREEDY} \geq (1-\frac{1}{e})\text{OPT} + G(\emptyset)\epsilon$, where $\epsilon > 0$ depends on k. The fix, however, is very simple. We can define a new objective function,

$$G_{t+1}(X) := \log\det(\rho^T(\Omega^\sigma_t))\det(\rho^T(\Omega^\sigma_t)^{-1} + M_X), \quad (4)$$

where $\det(\rho^T(\Omega^\sigma_t))$ is a normalization factor so $G_t(\emptyset) = 0$.

Lemma V.1. *For a single time step, the function $G_t(X)$ is monotone non-decreasing and submodular with an initial value of 0.*

By this lemma, picking the best set of sensors at a particular time step can be approximated using a greedy algorithm to within $(1-\frac{1}{e})$ of the optimal [12].

B. Multiple Time Steps

We now assume that at each time step the optimal sensor can be chosen in an attempt to study the properties of the objective function over multiple time steps. We will use the following extension of (4) as the objective function for multiple time step optimization,

$$\begin{aligned} F(\sigma) = F(\Omega^\sigma_T(\Sigma_0)) &= \log\det(\Omega^\sigma_0) - \log\det(\Omega^\sigma_T) \\ &= \log\left(\det(\Sigma_0)\det(\Omega^\sigma_T)^{-1}\right). \end{aligned} \quad (5)$$

We assume that A is non-singular and that there is no process noise, i.e., $w_t = 0$. When these assumptions are not satisfied, it is easy to construct seemingly trivial examples for which the cost is not submodular. For example, even for a diagonal and stable A, if W is nonzero, then the cost will not be submodular.

Using these assumptions, we can solve the recursive covariance update to obtain

$$\Omega^\sigma_t = \left((A^{-\top})^t\Sigma_0^{-1}A^{-t} + \sum_{i=1}^{t}(A^{-\top})^{(t-i)}M_iA^{-(t-i)}\right)^{-1}. \quad (6)$$

The proof of the following Lemma is omitted and can be found in [17].

Lemma V.2. *Given a positive semidefinite matrix P, then for all nonnegative integers j, l such that $0 \leq j < l$, the following hold: 1) If $XPX^\top \preceq P$ then $X^lP(X^\top)^l \preceq X^jP(X^\top)^j$; 2) If $XPX^\top \succeq P$ then $X^lP(X^\top)^l \succeq X^jP(X^\top)^j$.*

Theorem V.3. *For the function* (5)*, the value of the empty set is 0. Also, with the assumptions that A is full rank and $W = 0$, if $A\Sigma_0A^\top \preceq \Sigma_0$ and $A^\top M_iA \preceq M_i$ for all possible measurement matrices, then the function is monotone non-decreasing and submodular.*

Proof. The value of empty set is easy to see, $F(\Omega^\emptyset_0(\Sigma_0)) = \log\det(\Sigma_0)\det(\Sigma_0^{-1}) = 0$.

Take $A \subset B$ where B is a sequence of measurements. Let $B = (1,\ldots,b)$, with the corresponding measurements $\{M_i\}_{i\in B}$, and $\{A(i)\}_{i=1}^a$ are the indices in B that are part of A. Note that by definition of subsequence, $A(i) < A(i+1)$ for all i, i.e., the order in which elements appear in B must be the same as the order in which they appear in A.

[1]Here X is the set of sensors to be chosen at time t and M_X – through an abuse of notation – is the corresponding measurement matrix, similar to the one defined in (2).

For monotonicity, the requirement is $F(A) \leq F(B)$:

$$\log\det(\Sigma_0)\det(\Omega_a^A)^{-1} \leq \log\det(\Sigma_0)\det(\Omega_b^B)^{-1}$$
$$\iff \det(\Omega_a^A) \geq \det(\Omega_b^B)$$

For submodularity, the requirement is $F(A \,\|\, x) - F(A) \geq F(B \,\|\, x) - F(B)$. This can be written as,

$$\log\det(\Omega_{a+1}^{Ax})^{-1} - \log\det(\Omega_a^A)^{-1} \geq \log\det(\Omega_{b+1}^{Bx})^{-1} - \log\det(\Omega_b^B)^{-1}$$

which is satisfied if and only if

$$\log\det(\Omega_{a+1}^{Ax})^{-1}\det(\Omega_a^A) \geq \log\det(\Omega_{b+1}^{Bx})^{-1}\det(\Omega_b^B).$$

Now, applying the covariance update formula (2), we have

$$\Omega_{a+1}^{Ax}(\Sigma_0) = \Omega_1^x(\Omega_a^A) = \left((A\Omega_a^A A^\top)^{-1} + M_x\right)^{-1},$$
$$\Omega_{b+1}^{Bx}(\Sigma_0) = \Omega_1^x(\Omega_b^B) = \left((A\Omega_b^B A^\top)^{-1} + M_x\right)^{-1}.$$

Substituting back and multiplying both sides by $\det(A)\det(A^\top)$, the condition for submodularity becomes

$$\det(A\Omega_a^A A^\top)\det((A\Omega_a^A A^\top)^{-1} + M_x) \geq \det(A\Omega_b^B A^\top)\det((A\Omega_b^B A^\top)^{-1} + M_x).$$

Thus, we have submodularity if and only if the following is satisfied $\det(I + (A\Omega_a^A A^\top)M_x) \geq \det(I + (A\Omega_b^B A^\top)M_x)$ Taking $M_x := L^\top L$ (since $M_x \succeq 0$) and applying the matrix determinant lemma[2], the condition becomes $\det(I + (LA)\Omega_a^A(LA)^\top) \geq \det(I + (LA)\Omega_b^B(LA)^\top)$. A sufficient condition for both monotonicity and submodularity to hold then is $\Omega_a^A \succeq \Omega_b^B$. Applying (6), this is equivalent to

$$(A^{-\top})^a\Sigma_0^{-1}A^{-a} + \sum_{i=1}^{a}(A^{-\top})^{(a-i)}M_{A(i)}A^{-(a-i)} \preceq (A^{-\top})^b\Sigma_0^{-1}A^{-b} + \sum_{j=1}^{b}(A^{-\top})^{(b-j)}M_jA^{-(b-j)}. \tag{7}$$

Now, let's look at the individual terms in inequality (7). We have $\Sigma_0 \succ 0$ and $a \leq b$. Assuming that $A\Sigma_0 A^\top \preceq \Sigma_0$ we can apply Lemma V.2 to find that $A^a\Sigma_0(A^\top)^a \succeq A^b\Sigma_0(A^\top)^b$. Taking the inverse of both sides (which requires that A is full rank) we obtain

$$(A^{-\top})^a\Sigma_0^{-1}A^{-a} \preceq (A^{-\top})^b\Sigma_0^{-1}A^{-b}.$$

For every $A(i)$, the term $M_{A(i)}$ will appear on both sides of inequality (7). We can show that $(a-i) \leq (b-j)$. Since $A \subset B$, the first element of A ($i = 1$) can be at most in position $b-a+1$ in B. Similarly, the second element in A can be at most in position $b-a+2$ in B. Therefore, the inequality $j \leq b-a+i$ holds. Assuming that $A^{-\top}M_{A(i)}A^{-1} \succeq M_{A(i)}$, since we know that $M_{A(i)} \succeq 0$, we can apply Lemma V.2 again to obtain

$$(A^{-\top})^{a-i}M_{A(i)}A^{-(a-i)} \preceq (A^{-\top})^{b-j}M_{A(i)}A^{-(b-j)}.$$

[2]Given $A_{n\times n}$, $U_{n\times k}$ and $V_{n\times k}$. Assuming A^{-1} exists, $\det(A + UV^\top) = \det(I + V^\top A^{-1}U)\det(A)$. For the special case of $A = I$, this is just *Sylvester's Theorem of Determinants*.

Note that the condition $A^{-\top}M_{A(i)}A^{-1} \succeq M_{A(i)}$ is equivalent to $A^\top M_{A(i)}A \preceq M_{A(i)}$.

Therefore, under the assumptions made, the inequality (7) holds and the function is monotone non-decreasing and submodular. □

If the conditions of Theorem V.3 hold then by Lemma II.8, greedily selecting the measurement matrix at each time step gives a $(1 - \frac{1}{e})$-approximation.

Remark V.4 (The restrictiveness of Theorem V.3). In addition to the requirement of no process noise, the two conditions in Theorem V.3 are that $A\Sigma_0 A^\top \preceq \Sigma_0$ and $A^\top M_i A \preceq M_i$ for all possible measurement matrices.

The first condition is related to the stability of A. The theorem for Lyaponov stability of discrete time systems states that the eigenvalues of A have magnitude strictly less than 1 if and only if there exists $P \succ 0$ such that $A^\top PA \prec P$ [18, Thm. 8.4]. Thus, if A is exponentially stable, then we can pick an initial covariance to satisfy $A\Sigma_0 A^\top \preceq \Sigma_0$. Conversely, if A is unstable then no such Σ_0 exists.

The $A^\top M_i A \preceq M_i$ conditions in Theorem V.3 require checking $\binom{m}{k}$ matrices. We can ask whether we can design the sensor network (i.e., the matrix C) such that $A^\top M_i A \preceq M_i$ for all possible measurement matrices. Although it is possible to construct C for certain cases, we do not yet know of explicit conditions on A for the existence of a satisfying C, or a general method for computing such a C. •

C. Greedy Approximation

Assuming that the conditions of Theorem V.3 are met, sequentially selecting the best measurement matrix at each time step will give a $(1 - \frac{1}{e})$-approximation (≈ 0.6321). One drawback is that there are $\binom{m}{k} = O(m^k)$ possible measurement matrices at each step. An alternative approach is to greedily select k measurements at each time step. Thus, at given time step we construct a set of k sensors as in Section V-A. This is done greedily across time steps. This approach leads to a faster runtime but a less tight bound, as shown in the following theorem.

Theorem V.5. *Consider the sensor scheduling problem such that $W = 0$, V is diagonal, A is full rank, $A\Sigma_0 A^\top \preceq \Sigma_0$ and $A^\top M_i A \preceq M_i$ for all possible measurement matrices M_i. Using* (5) *as the objective function leads to a $1 - \frac{1}{e^{1-1/e}}$-factor approximation in $O(Tkmn^2 + Tn^3)$ time if the k sensors are chosen greedily at each time step.*

Proof. Sequentially selecting the best measurement matrix at each time step corresponds to solving the optimization problem $\max_{\sigma_t}\Delta(\sigma_t|\sigma_{[1,t-1]}) \equiv \max_{\sigma_t}\log\det(\Sigma_0)\det(\Omega_t^\sigma)^{-1} - \log\det(\Sigma_0)\det(\Omega_{t-1}^\sigma)^{-1} \equiv \max_{\sigma_t}\log\det(\Omega_{t-1}^\sigma)\det(\Omega_t^\sigma)^{-1}$, at each time step, where $\sigma_{[1,t-1]}$ is the sensor sequence from 1 to $t-1$.

By Lemma V.1, since the measurement noise matrix V is diagonal, building the measurement matrix greedily at a particular time step can be solved to within $(1 - \frac{1}{e})$ of the optimal using (4) as the objective function. Let Σ_t^g and Σ_t^o be the resulting covariance after applying the sequence of

measurements $\sigma_{[1,t-1]}$ and then selecting the greedy and optimal measurements respectively at time step t. Therefore, taking $\alpha = (1 - \frac{1}{e})$, $\log\det(\rho^T(\Omega^\sigma_{t-1}))\det(\Sigma^g_t)^{-1} \geq \alpha \log\det(\rho^T(\Omega^\sigma_{t-1}))\det(\Sigma^o_t)^{-1}$ which implies $\log\det(\Omega^\sigma_{t-1})\det(\Sigma^g_t)^{-1} \geq \alpha\log\det(\Omega^\sigma_{t-1})\det(\Sigma^o_t)^{-1} + (\alpha-1)\log\det(A)^2$, since $W = 0$ so $\rho^T(\Omega^\sigma_{t-1}) = A\Omega^\sigma_{t-1}A^\top$.

Let $\sigma^g = (\sigma^g_1, \ldots, \sigma^g_T)$ be the sequence of measurements by greedily selecting the k sensors at each time step and $\sigma^o = (\sigma^o_1, \ldots, \sigma^o_T)$ be the optimal schedule. Therefore, $\Delta(\sigma^g_t|\sigma^g_{[1,t-1]}) \geq \alpha \max_{\sigma_i} \Delta(\sigma_i|\sigma^g_{[1,t-1]}) + \epsilon$,where $\epsilon = (\alpha-1)\log\det(A)^2$. Note that if $\epsilon = 0$, then we can apply Lemma II.8 to deduce that $F(\sigma^g) \geq (1 - \frac{1}{e^\alpha})F(\sigma^o)$. However, since $\epsilon \neq 0$, we cannot apply this directly. Instead, we solve for a bound of the greedy schedule by imitating the proof of Lemma II.8 (given in [14]).

$$\begin{aligned}
\Delta(\sigma^g_t|\sigma^g_{[1,t-1]}) &\geq \alpha \max_{\sigma_i} \Delta(\sigma_i|\sigma^g_{[1,t-1]}) + \epsilon \\
&\geq \alpha \max_{\sigma_i \in \sigma^o} \Delta(\sigma_i|\sigma^g_{[1,t-1]}) + \epsilon \\
&\geq \frac{\alpha}{T}\Delta(\sigma^o|\sigma^g_{[1,t-1]}) + \epsilon \\
&\geq \frac{\alpha}{T}\left(F(\sigma^o) - F(\sigma^g_{[1,t-1]})\right) + \epsilon.
\end{aligned}$$

This implies that $F(\sigma^g_{[1,t]}) \geq \frac{\alpha}{T}F(\sigma^o) + (1-\frac{\alpha}{T})F(\sigma^g_{[1,t-1]}) + \epsilon$. Solving this recurrence relation,

$$\begin{aligned}
&F(\sigma^g) = F(\sigma^g_{[1,T]}) \\
&\geq \left(\frac{\alpha}{T}F(\sigma^o) + \epsilon\right)\sum_{i=0}^{T-1}(1-\frac{\alpha}{T})^i \\
&= F(\sigma^o)\left(1 - (1-\frac{\alpha}{T})^T\right) + \frac{T}{\alpha}\left(1-(1-\frac{\alpha}{T})^T\right)\epsilon \\
&\geq F(\sigma^o)(1-e^{-\alpha}) + \frac{T}{\alpha}(1-e^{-\alpha})\epsilon.
\end{aligned}$$

Therefore, the greedy schedule is within a factor of $(1-e^{-\alpha})$ of the optimal but there is an error term of $T(1-\alpha^{-1})(1-e^{-\alpha})\log\det(A)^2 \approx (-0.2727)T\log\det(A)^2$. Note, however, that since $\Sigma_0 \succ 0$,

$$\begin{aligned}
&A\Sigma_0A^\top \preceq \Sigma_0 \implies \det(A\Sigma_0A^\top) \leq \det(\Sigma_0) \\
&\implies \det(A)^2 \leq 1 \implies \log\det(A)^2 \leq 0.
\end{aligned}$$

As a result, $F(\sigma^g) \geq (1-e^{-\alpha})F(\sigma^o)$ and substituting the value of α gives the constant factor bound of ≈ 0.4685.

For the complexity, there are kmT iterations; for each of the T time steps, k measurements need to be selected from a set of m. The optimization for each time step can be performed intelligently to avoid having to repeatedly calculate inverses and determinants. This results in a runtime of $O(n^2mk)$ per time step [2] such that the output is the *a posteriori* covariance matrix. The time update requires two matrix multiplications which naively require $O(n^3)$ time. Therefore, the total runtime is $O(Tkmn^2 + Tn^3)$. □

VI. Conclusions

In this paper, we studied the problem of using an energy constrained sensor network to estimate the state of a linear dynamical system. We showed that contrary to recent work, the sensor schedule cost is not, in general, a submodular function. We then provided a set of sufficient and easily checkable conditions under which the sensor schedule cost is submodular. If the cost is submodular, the greedy algorithm performs within approximately $1/2$ of the optimal.

For future work we would like to determine the exact complexity of testing whether or not a function is submodular. It appears that such a test is computationally intractable, but to the best of our knowledge this has not been formally shown.

References

[1] H. J. Freeland and P. F. Cummins, "Argo: A new tool for environmental monitoring and assessment of the world's oceans, and example from the N.E. Pacific," *Progress in Oceanography*, vol. 64, pp. 31–44, 2005.

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

[3] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE Trans Automatic Ctrl*, vol. 49, no. 9, pp. 1453–63, 2004.

[4] V. Gupta, T. H. Chung, B. Hassibi, and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage," *Automatica*, vol. 42, no. 2, pp. 251–260, 2006.

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

[6] 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.

[7] Y. Mo, R. Ambrosino, and B. Sinopoli, "Sensor selection strategies for state estimation in energy constrained wireless sensor networks," *Automatica*, vol. 47, no. 7, pp. 1330–1338, 2011.

[8] M. P. Vitus, W. Zhang, A. Abate, J. Hu, and C. J. Tomlin, "On efficient sensor scheduling for linear dynamical systems," *Automatica*, vol. 48, no. 10, pp. 2482–2493, 2012.

[9] J. E. Weimer, B. Sinopoli, and B. H. Krogh, "A relaxation approach to dynamic sensor selection in large-scale wireless networks," in *International Conf. on Distributed Computing Systems*, 2008, pp. 501–506.

[10] J. Weimer, J. Araújo, A. Hernandez, and K. Johansson, "Periodic constraint-based control using dynamic wireless sensor scheduling," in *IEEE Conf. on Decision and Control*, 2011, pp. 4789–4796.

[11] M. Huber, A. Kuwertz, F. Sawo, and U. Hanebeck, "Distributed greedy sensor scheduling for model-based reconstruction of space-time continuous physical phenomena," in *International Conf. on Information Fusion*, 2009, pp. 102 –109.

[12] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions - I," *Mathematical Programming*, vol. 14, pp. 265–294, 1978.

[13] M. L. Fisher, G. L. Nemhauser, and L. A. Wolsey, "An analysis of approximations for maximizing submodular set functions - II," in *Polyhedral Combinatorics*, ser. Mathematical Programming Studies, 1978, vol. 8, pp. 73–87.

[14] S. Alaei and A. Malekian, "Maximizing sequence-submodular functions and its application to online advertising," *CoRR*, vol. abs/1009.4153v1, 2010.

[15] C. Yang, L. Kaplan, and E. Blasch, "Performance measures of covariance and information matrices in resource management for target state estimation," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2594–2613, 2012.

[16] C. Seshadhri and J. Vondrák, "Is submodularity testable?" *Algorithmica*, pp. 1–25, 2011.

[17] S. T. Jawaid, "Informative path planning and sensor scheduling for persistent monitoring tasks," Master's thesis, University of Waterloo, 2013, http://hdl.handle.net/10012/7896.

[18] J. P. Hespanha, *Linear systems theory*. Princeton university press, 2009.