Submodularity and Greedy Algorithms in Sensor Scheduling for Linear Dynamical Systems [1]

Syed Talha Jawaid [a], Stephen L. Smith [a]

[a] *Department of Electrical and Computer Engineering, University of Waterloo, Waterloo ON, N2L 3G1 Canada*

Abstract

This paper focuses on sensor scheduling for state estimation, which consists of a network of noisy sensors and a discrete-time linear system with process noise. As an energy constraint, only a subset of sensors can take a measurement at each time step. These measurements are fused into a common state estimate using a Kalman filter and the goal is to schedule the sensors to minimize the estimation error at a terminal time. A simple approach is to greedily choose sensors at each time step to minimize the estimation error at the next time step. Recent work has shown that this greedy algorithm outperforms other well known approaches. Results have been established to show that the estimation error is a submodular function of the sensor schedule; submodular functions have a diminishing returns property that ensures the greedy algorithm yields near optimal performance.

As a negative result, we show that most commonly-used estimation error metrics are not, in general, submodular functions. This disproves an established result. We then provide sufficient conditions on the system for which the estimation error is a submodular function of the sensor schedule, and thus the greedy algorithm yields performance guarantees.

1 Introduction

Sensor scheduling problems arise in applications involving the long-term estimation of a physical process through a set of static sensors. Examples include monitoring CO_2 concentrations [22] and monitoring water levels in multiple tanks [21] using a wireless sensor network. In such long-term deployments, energy consumption is a priority and by turning on only a small subset of sensors at each time-step, the battery life of the network can be extended.

Sensor scheduling can be described as follows. We are given a discrete-time linear dynamical system $x_{t+1} = Ax_t + w_t$, where x_t is the n-dimensional state vector and w_t is zero mean Gaussian process noise with known covariance. The goal is to estimate the state x_t through the use of m sensors. Each sensor takes a noisy scalar measurement of the state x_t, where the noise is zero mean Gaussian. A known covariance matrix gives the correlation between sensor noises. At each time step in this sensor scheduling problem we can activate at most $k < m$ sensors to take a measurement. Sensors that are inactive at a particular time step can then sleep, extending their battery life. After each time step t, the k measurements are fused into a single state estimate $\hat{x}_{t|t}$ with covariance $\Sigma_{t|t}$ using a centralized Kalman filter.

Given an initial covariance Σ_0 a terminal time T, and a sensor schedule σ, the final covariance $\Sigma_{T|T}$ is uniquely defined. The quality of a sensor schedule σ is determined as a function of this final covariance: for example the trace, the largest eigenvalue, or the determinant. Each objective can be thought of as a function that takes as input (Σ_0, σ) and outputs a real number. Our goal is to compute the schedule σ that minimizes/maximizes this objective function.

Related work: The sensor scheduling problem (or equivalently, the actuator scheduling problem) dates back at least as far as the early 1970s [2]. Early work cast the problem in an optimal control framework and computed optimal schedules via dynamic programming. The drawback with this approach is that the computation grows exponentially with the terminal time T. In the last decade there has been a resurgence in work on sensor scheduling. A catalyst for this was the study of Kalman filtering with intermittent observations [19]. Based on this work, the authors in [6] provided a method for stochastically selecting measurements via random walk on an appropriately optimized Markov chain.

A variety of approaches have since been proposed for sensor scheduling, including approaches based on con-

Email addresses: stjawaid@uwaterloo.ca (Syed Talha Jawaid), stephen.smith@uwaterloo.ca (Stephen L. Smith).

[1] This paper was not presented at any IFAC meeting. Corresponding author S. L. Smith. Tel. +1-519-888-4567 ext 31215

vex optimization [12], quadratic programming [13], and tree pruning [20]. The work in [13] presents a general framework that allows one to include more complex network and energy constraints. Recent work has looked at periodic schedules [15], on the connection between the sensor selection problem and compressed sensing [3], on provably complete algorithms [10], and on consensus-based algorithms that remove the need for a centralized filter [24].

Greedy algorithms and submodularity: The focus of this paper is on recent work that has shown advantages of using simple greedy algorithms for computing sensor schedules. At each time step t, the greedy algorithm chooses k sensors to minimize the estimate error at time $t+1$. The procedure begins at time step 1, and is repeated until all T time steps of the schedule have been specified. Being greedy, the approach is computationally efficient and simple to implement. Moreover, it has been shown [17,18] that the k sensors chosen at a single time step yield a reduction in estimation error that is within a constant factor of the optimal reduction. This property is due to the *submodularity* [14,4] of the covariance objective function over a single time step. In addition, empirical results in [17] have shown that the greedy algorithm often outperforms more computationally intensive alternatives based on convex optimization [12] over multiple time steps. In [9], a simulation study is performed for estimating a 31-state dynamical system that models a temperature diffusion process using sensor networks containing 20 to 40 sensors. The study compared three algorithms: the greedy algorithm, an optimal schedule based on exhaustive search, and a receding horizon approximate schedule. They found empirically that the runtime of the greedy algorithm was orders of magnitude better than the other two approaches, and the estimation performance exceeded that of the receding horizon approach. The authors also provided theoretical performance guarantees of the greedy algorithm.

In this paper we explore in more depth the connection between sensor scheduling, submodularity, and greedy algorithms. We characterize conditions under which the greedy algorithm gives provable performance guarantees by studying the submodularity of sensor scheduling objective functions.

Contributions: The contributions of this paper are two-fold. First, we provide negative results to show that most sensor schedule objective functions are not, in general, submodular nor monotone functions. This result holds for objectives including the trace of the covariance, the maximum eigenvalue and the log of the determinant, and it disproves the guarantees established in [9]. Second, we provide a set of (restrictive) conditions on the system under which the log of the determinant objective function is submodular, and thus the greedy algorithm has guaranteed performance. An early version of this paper appeared at the conference [11]. This paper expands on the preliminary version in several respects, including complete proofs of all results, details on the submodular counterexamples, and a new interpretation of sensor scheduling in terms of a submodular function over a matroid constraint in Section 6.1.

2 Preliminaries

In this section we review some essential concepts in submodular set functions from [14,4] and submodular sequence functions from [5,1].

2.1 Independence Systems and Matroids

Many combinatorial optimization problems can be formulated as maximizing or minimizing an objective function $f : \mathcal{F} \to \mathbb{R}$ over a set system $(E, \mathcal{F})$. The set $\mathcal{F} \subseteq 2^E$ contains all "allowable" subsets of the base set E. An *independence system* is a set system that is closed under subsets: if $A \in \mathcal{F}$ then $B \subseteq A \implies B \in \mathcal{F}$.

Definition 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 there is an $x \in X \backslash Y$ such that $Y \cup \{x\} \in \mathcal{F}$.*

The *uniform matroid* is defined by the collection of all subsets of E with 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 E is partitioned into n disjoint sets, $\{E_i\}_{i=1}^n$. Given $k \in \mathbb{Z}_+^n$, the partition matroid is defined by the collection $\mathcal{F} := \{A \subseteq E : |A \cap E_i| \leq k_i, \forall i = 1 \dots n\}$.

2.2 Set Functions

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

Definition 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 3 (Submodularity) *The function f is submodular if for all $A \subseteq B \subseteq E$ and for all $x \in E \setminus B$ we have $f(A \cup \{x\}) - f(A) \geq f(B \cup \{x\}) - f(B)$.*

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, $\Delta_f(x|A) \geq \Delta_f(x|B)$ for all $A \subseteq B \subseteq E$.

2.3 Sequence Functions

For our purposes, a *sequence* $A = (a_1, \ldots, a_k)$, of length $k \in \mathbb{Z}_{\geq 0}$ consists of k elements from a base set of elements E, i.e., $a_i \in E$. Two sequences $A = (a_1, \ldots, a_k)$ and $B = (b_1, \ldots, b_\ell)$ defined over the same base set can be *concatenated* into a larger sequence: $A \parallel B = (a_1, \ldots, a_k, b_1, \ldots, b_\ell)$. A *subsequence* of A is a sequence derived from A by deleting some elements but not changing the order of the remaining elements and is denoted $B \subseteq A$. A *sequence function* f defined over a base set E maps from sequences over E to real numbers. The value of a sequence function depends on the order of the elements in the sequence.

Definition 4 (Sequence Monotonicity) *The sequence function f is monotone non-decreasing if for all subsequences A of a sequence B, i.e., $A \subseteq B$, $f(A) \leq f(B)$.*

For a sequence function f, we define the marginal reward of concatenating a sequence C to a sequence A as $\Delta_f(C|A) := f(A \parallel C) - f(A)$. The subscript f will be omitted unless there is ambiguity.

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

2.4 Maximization of Submodular Functions

Greedy algorithms have surprisingly good performance in maximizing submodular functions. The *set greedy algorithm* begins with the empty set and repeatedly adds the element $x \in E$ to S that maximizes the marginal return $\Delta(x|S) = f(S \cup \{x\}) - f(S)$. When maximizing a submodular function over a matroid $(E, \mathcal{F})$, we add the element x that maximizes $\Delta(x|S)$ subject to the constraint that $S \cup \{x\} \in \mathcal{F}$. The *sequence greedy algorithm* begins with an empty sequence S, and repeatedly appends the element $x \in E$ that maximizes the marginal return $\Delta(x|S) = f(S \parallel x) - f(S)$.

An *α-approximate greedy algorithm* with $\alpha \leq 1$ finds at each iteration an element x to add/append to S such that $\Delta(x|S) \geq \alpha \max_{\bar{x}} \Delta(\bar{x}|S)$. If $\alpha = 1$, then we recover the set/sequence greedy algorithms. Approximate algorithms are important when it is difficult to determine the true maximizer of the marginal return at each iteration.

We say that an algorithm gives a k-approximation, for some $k < 1$ if it is guaranteed to output a solution with a value of at least k times the optimal.

Lemma 6 (Maximizing a Set Function [14,4]) *Consider the problem of maximizing a submodular, non-negative and monotone function over a matroid. For the uniform matroid, the α-approximate greedy algorithm gives a $1 - \frac{1}{e^\alpha}$ approximation. For a general matroid, the greedy algorithm gives a $\frac{1}{2}$ approximation.*

The following lemma looks at the sequence equivalent to a uniform matroid.

Lemma 7 (Maximizing a Sequence Function [1]) *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.*

3 The Sensor Scheduling Problem

Consider a sensor network consisting of m sensors $M := \{1, \ldots, m\}$, operating in discrete time. Our goal is to estimate the state of a linear time invariant (LTI) system. At each time step t, we can turn on at most k sensors. The measurements taken at each time step are fused into a single estimate using a centralized Kalman filter.

We assume that each sensor takes a scalar measurement. The k sensors selected at time step t can be encoded in the binary selection matrix $S_t \in \{0,1\}^{k \times m}$. Each row of S_t contains a single element equal to 1, and each column contains at most one element equal to 1. The combined LTI system/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 independent zero mean Gaussian vectors with covariance matrices $W, V \in \mathbb{R}^{n \times n}$ such that $W, V \succeq 0$. The matrix $S_t C$ is a $k \times m$ matrix containing k rows of C corresponding to the k sensors selected at time step t.

The state x_t is estimated via the Kalman filter, which produces a state estimate $\hat{x}_{t|t}$ of x_t given the measurements up to and including time t along with the covariance of this estimate $\Sigma_{t|t}$. Given $\Sigma_{t-1|t-1}$, the covariance update consists of two steps: the time update

$$\begin{aligned} \Sigma_{t|t-1} &= A\Sigma_{t-1|t-1}A^\top + W \\ &:= \rho_{\text{time}}(\Sigma_{t-1|t-1}), \end{aligned}$$

and the measurement update,

$$\begin{aligned} \Sigma_{t|t} &= \left(\Sigma_{t|t-1}^{-1} + C^\top S_t^\top (S_t V S_t^\top)^{-1} S_t C\right)^{-1} \\ &:= \rho_{\text{meas}}(\Sigma_{t|t-1}, S_t). \end{aligned}$$

The time and measurement update can be combined to give

$$\Sigma_{t|t} = \left((A\Sigma_{t-1|t-1}A^\top + W)^{-1} + D_t\right)^{-1}, \text{ where}$$
$$D_t = C^\top S_t^\top (S_t V S_t^\top)^{-1} S_t C. \tag{2}$$

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 minimizes/maximizes a function $F : \mathbb{R}^{n\times n} \to \mathbb{R}$ of the final error covariance $\Sigma_{T|T}$ at the terminal time T.

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

$$\Omega_t^\sigma(\Sigma_0) := \rho_{\text{meas}}\Big(\rho_{\text{time}}\big(\Omega_{t-1}^\sigma(\Sigma_0)\big), \sigma_t\Big)$$
$$\Omega_0^\sigma(\Sigma_0) = \Sigma_0, \tag{3}$$

where, with a slight abuse of notation, we have used σ_t, which is a subset of sensors k interchangeably with its corresponding binary selection matrix S_t. We omit the argument Σ_0 from Ω_t where there is no 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.

Given an initial covariance Σ_0, and a covariance metric F, we seek a schedule σ that optimizes $F(\Sigma_{T|T})$, where σ contains k sensors per time step. The final covariance $\Sigma_{T|T}$ is computed via (3) as $\Sigma_{T|T} = \Omega_T^\sigma(\Sigma_0)$. To simplify notation, we write $F\big(\Omega_T^\sigma(\Sigma_0)\big)$ as simply $F(\sigma)$, with the understanding that F is evaluated on the covariance obtained via the schedule σ.

Remark 8 (Set or Sequence function) *Over a single time step, the objective $F(\cdot)$ is simply a set function mapping a set of k sensors to a value. Over multiple time steps we need to compute a sequence $\sigma = (\sigma_1, \ldots, \sigma_T)$, where each σ_t is a subset of M containing k sensors. Thus, $F(\Omega_t^\sigma)$ maps a sequence of length $t \leq T$ to a value. In Section 6.1 we discuss an alternative interpretation as a set function over a matroid.*

We are interested in determining conditions under which sensor scheduling can be posed as the maximization of a submodular sequence, and thus when an incremental greedy algorithm can be used to build provably efficient sensor schedules.

4 Submodularity of the Estimation Error

In this section we show that for most of the commonly used metrics, the sensor selection problem is not, in general, submodular.

4.1 Counterexamples for Submodularity

Given a covariance matrix $\Sigma_{t|t}$ we can define the following covariance metrics:

- $F_1 = -\operatorname{trace}(\Sigma_{t|t})$
- $F_2 = \log\det(\Sigma_{t|t}^{-1})$
- $F_3 = -\max\operatorname{eig}(\Sigma_{t|t})$
- $F_4 = \operatorname{trace}(\Sigma_{t+1|t}^{\emptyset} - \Sigma_{t+1|t})$, where $\Sigma_{t+1|t}^{\emptyset}$ is the covariance obtained from t successive time updates without any measurements.

We phrase each optimization as a maximization problem in order to fit with the submodularity literature reviewed in Section 2.4.

Remark 9 (Performance metrics) *If we fix number $p < 1$, then the eigenvalues of the covariance matrix are proportional to the lengths of the axes of the ellipsoid that contains the estimation error $x - \hat{x}$ with probability p [12]. The* volume *of the confidence ellipsoid is directly proportional to the* log *of the determinant of the covariance matrix, which is captured by F_2. The* mean squared error *is given by the sum of the eigenvalues, which is given by the trace 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 [23].*

In [9], the authors show (see Theorems 2 and 3) that the function F_4 is submodular for a single time step as well as sequence submodular 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. The root of the error appears to be as follows. The Riccatti equation (2) is known to be monotone [20] in the sense that if $\Sigma_{t-1|t-1} \preceq \overline{\Sigma}_{t-1|t-1}$, then $\Sigma_{t|t} \preceq \overline{\Sigma}_{t|t}$ (i.e., in the positive semidefinite sense). In [9] this property is incorrectly used to infer monotonicity and submodularity with respect to the set of sensors chosen.

Example 10 (Single Time Step) *Consider the system*

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

Take $X_1 = \{2, 3\} \subset \{2, 3, 4\} = X_2$. Then,

- $\Delta_{F_1}(\{1\}|X_1) = 0.2736$ *but* $\Delta_{F_1}(\{1\}|X_2) = 0.2769$.
- $\Delta_{F_3}(\{1\}|X_1) = 0.1313$ *but* $\Delta_{F_3}(\{1\}|X_2) = 0.1927$.
- $\Delta_{F_4}(\{1\}|X_1) = 0.0684$ *but* $\Delta_{F_4}(\{1\}|X_2) = 0.0692$.

These results show that none of $F_1, F_2,$ or F_4 are submodular, disproving Theorem 2 of [9].

Moreover, if we take $X_1 = \{1, 3\} \subset \{1, 3, 4\} = X_2$, we see that $\Delta(\{2\}|X_1) \geq \Delta(\{2\}|X_2)$ for each of the three functions, showing that the functions 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 sequence submodular over multiple time steps.

Remark 11 (A Test for Submodularity?) *Given an LTI system and an initial covariance, a natural question is whether we can determine if the associated sensor scheduling objective function is submodular. Unfortunately, for set functions there is no known polynomial-time algorithm for testing submodularity [16]. Thus, unless some special structure of the Kalman update can be exploited, it is unlikely that there exists an efficient method to test for set or sequence submodularity of a particular problem. This motivates the following section in which we provides checkable conditions for sequence submodularity.*

5 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.

5.1 Submodularity Over a Single Time Step

The single time-step problem is one of choosing a subset k of the m measurements. Let $\sigma_t \subset M$ be the k sensors chosen at time step t. Then, from equation (2), the objective function is

$$F(\sigma_t) = \log\det(\rho_{\text{time}}(\Omega_{t-1}^{\sigma})^{-1} + D_t).$$

In [17] it is shown that if V is diagonal, then this objective is monotone non-decreasing and submodular in the selected measurements σ_t. As a result, the authors deduce that using a greedy algorithm gives a $(1 - \frac{1}{e})$-approximation [17, Lemma 1]. However, as stated this result is technically not correct, as the objective function is not normalized (i.e., the value of empty set is not necessarily zero). This yields an additional error term that depends on the value of the empty set: GREEDY $\geq (1 - \frac{1}{e})\text{OPT} + F(\emptyset)\epsilon$, where $\epsilon > 0$ depends on k. This can be fixed by redefining the objective function as

$$F'(\sigma_t) := \log\det\left(\rho_{\text{time}}(\Omega_{t-1}^{\sigma})\right)F(\sigma_t), \tag{5}$$

where $\log\det(\rho_{\text{time}}(\Omega_t^{\sigma}))$ is a normalization factor so that $F(\emptyset) = 0$.

Lemma 12 *For a single time step, the function $F'(\sigma_t)$ is monotone non-decreasing, normalized and submodular function over the base set M.*

By this lemma and Lemma 6, picking the best set of k sensors at a particular time step can be approximated using a greedy algorithm to within $(1-\frac{1}{e})$ of the optimal.

5.2 Sequence Submodularity Over Multiple Time Steps

We now assume that at each time step the optimal set of k sensors can be chosen in an attempt to study the properties of the objective function over multiple time steps. Given a sensor schedule σ of length $t \leq T$, we define the sequence function to be

$$\begin{aligned} F(\sigma) = F(\Omega_t^{\sigma}(\Sigma_0)) &= \log\det(\Omega_0^{\sigma}) - \log\det(\Omega_t^{\sigma}) \\ &= \log\left(\det(\Sigma_0)\det(\Omega_t^{\sigma})^{-1}\right). \end{aligned} \tag{6}$$

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 objective is not sequence submodular. For example, even for a diagonal and stable A, if W is nonzero, then the objective will not be sequence submodular.

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

$$\Omega_t^{\sigma} = \left((A^{-\top})^t\Sigma_0^{-1}A^{-t} + \sum_{i=1}^{t}(A^{-\top})^{(t-i)}D_iA^{-(t-i)}\right)^{-1}, \tag{7}$$

where $D_t = C^{\top}S_t^{\top}(S_tVS_t^{\top})^{-1}S_tC$ is as defined in equation (2). We are now ready to give conditions on the function for sequence submodularity. The proof is contained in the appendix.

Theorem 13 *For the function (6), 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 $A^{\top}D_iA \preceq D_i$ for all possible measurement matrices, then the function is monotone non-decreasing and sequence submodular.*

If the conditions of Theorem 13 hold then by Lemma 7, greedily selecting the measurement matrix at each time step gives a $(1-\frac{1}{e})$-approximation.

Remark 14 (The restrictiveness of Theorem 13) *In addition to the requirement of no process noise, the two conditions in Theorem 13 are that $A\Sigma_0 A^\top \preceq \Sigma_0$ and $A^\top D_i A \preceq D_i$ for all possible measurement matrices. The Lyapunov stability theorem tells us that the eigenvalues of A have magnitude strictly less than 1 if and only if there exists $P \succ 0$ such that $A^\top P A \prec P$ [7, 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 $A^\top D_i A \preceq D_i$ conditions in Theorem 13 require checking $\binom{m}{k}$ matrices. We can ask whether we can design the sensor network (i.e., the matrix C) such that $A^\top D_i A \preceq D_i$ for all possible measurement matrices. While it is possible to construct C in certain cases, we do not yet know of conditions on A for the existence of a satisfying C, or a method for computing such a C.

5.3 Greedy Approximation

Assuming that the conditions of Theorem 13 are met, sequentially selecting the set of k sensors at each time step will give a $(1-\frac{1}{e})$-approximation (≈ 0.6321). However, there are $\binom{m}{k} = O(m^k)$ possible sets of size k. An alternative 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 5.1. 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 15 *Consider a sensor scheduling problem satisfying the assumptions of Theorem 13. If the objective function is given by (6), then the greedy algorithm yields a schedule with value at least $1-e^{-(1-1/e)} \approx 0.47$ of the optimal in $O(Tkmn^2 + Tn^3)$ time.*

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 12, 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 (5) 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 $\alpha = (1-\frac{1}{e})$, $\log\det(\rho_{\text{time}}(\Omega_{t-1}^\sigma))\det(\Sigma_t^g)^{-1} \geq \alpha \log\det(\rho_{\text{time}}(\Omega_{t-1}^\sigma))\det(\Sigma_t^o)^{-1}$ which implies

$$\log\det(\Omega_{t-1}^\sigma)\det(\Sigma_t^g)^{-1} \geq \alpha\log\det(\Omega_{t-1}^\sigma)\det(\Sigma_t^o)^{-1} + (\alpha-1)\log\det(A)^2,$$

since $W=0$ so $\rho_{\text{time}}(\Omega_{t-1}^\sigma) = A\Omega_{t-1}^\sigma A^\top$.

Let $\sigma^g = (\sigma_1^g, \ldots, \sigma_T^g)$ be the sequence of measurements by greedily selecting the k sensors at each time step and $\sigma^o = (\sigma_1^o, \ldots, \sigma_T^o)$ be the optimal schedule. Therefore, $\Delta(\sigma_t^g|\sigma_{[1,t-1]}^g) \geq \alpha \max_{\sigma_i} \Delta(\sigma_i|\sigma_{[1,t-1]}^g) + \epsilon$, where $\epsilon = (\alpha-1)\log\det(A)^2$. Note that if $\epsilon = 0$, then we can apply Lemma 7 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 following the proof of Lemma 7 (given in [1]):

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

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

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

The only thing that remains is to bound the second "error" term of $T(1-\alpha^{-1})(1-e^{-\alpha})\log\det(A)^2 \approx (-0.2727)T\log\det(A)^2$ in this expression: To do this, note that since $\Sigma_0 \succ 0$,

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

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 so as to avoid repeated calculation of inverses and determinants [17]. With this method, the *a posteriori* covariance matrix and its objective value can be computed in $O(n^2mk)$ per time step. The time update requires two matrix multiplications which

naively require $O(n^3)$ time. Therefore, the total runtime is $O(Tkmn^2 + Tn^3)$. □

6 Future Directions and Conclusions

Finally, we present an alternative interpretation of sensor scheduling as a set function over a matroid, which presents an interesting avenue for future work, along with some conclusions.

6.1 Sensor Scheduling as a Set Function over a Matroid

We have focused on the sequential greedy algorithm in which k sensors are chosen for time step 1, followed by k for time step 2, and so on until the scheduled is completed. The corresponding notion of submodularity is sequence submodularity. An alternative greedy algorithm can be posed based on set submodularity. The idea is to begin with a schedule $\sigma = (\sigma_1, \ldots, \sigma_T)$ where each $\sigma_t = \emptyset$ and for kT iterations do the following:

(1) for each time slot t in 1 to T with $|\sigma_t| < k$ and for each sensor s in M not already in σ_t
(2) add s to σ_t and call the new schedule σ'
(3) calculate the increase $F(\Omega_T^{\sigma'}) - F(\Omega_T^{\sigma})$
(4) update σ to schedule with maximum increase

This is the greedy algorithm over a partition matroid defined as follows: Create T distinct copies of the sensor set M, denoted $M_1, \ldots, M_T$, and define the base set as $E = \cup_{t=1}^{T} M_t$, which contains Tm elements. Define the independent sets to be $\mathcal{F} = \{A \subseteq E : |A \cap M_t| \leq k,\ \forall t \in \{1, \ldots, T\}\}$. Then, the goal is to find the set in $\mathcal{F}$ that maximizes a function of the final covariance $\Sigma_{T|T}$. If the objective is submodular and monotone increasing over the matroid, then by Lemma 6 the above algorithm would give a $1/2$ approximation to the optimal schedule.

We have determined some properties of this algorithm, and plan to continue our investigation in future work. First, we have found simple counterexamples for submodularity over this matroid by i) setting $W = V = \Sigma_0 = I$; ii) randomly generating A and C as 3×4 and 6×4 matrices with each entry uniform between zero and one; and iii) using a schedule of length $T = 2$ and $k = 3$ sensors per time step. Second, the algorithm requires significantly more computation than the sequential greedy since i) at each iteration, each of the mT elements in the base set must be considered (in contrast to just m elements per iteration of the sequential greedy); and ii) the objective at each iteration is a function of $\Sigma_{T|T}$, and thus T time and measurement updates must be performed each time a new sensor is added. However, the algorithm is polynomial time, and the conditions for submodularity to hold over the matroid appear to be less strict than those for sequence submodularity.

6.2 Conclusions

In this paper, we studied the problem of using an energy constrained sensor network to estimate the state of an uncertain process. We showed that contrary to recent work, the sensor schedule objective is not, in general, a submodular function. We then provided a set of sufficient and easily checkable conditions under which the sensor schedule objective is sequence submodular. If the objective is sequence submodular, the greedy algorithm performs within approximately ≈ 0.47 of the optimal.

A Proofs of Main Results

We begin by establishing the following lemma.

Lemma 16 *Given a positive semidefinite matrix P, then for all integers j, l such that $0 \leq j < l$, the following hold: (1) If $XPX^\top \preceq P$ then $X^l P (X^\top)^l \preceq X^j P (X^\top)^j$; (2) If $XPX^\top \succeq P$ then $X^l P (X^\top)^l \succeq X^j P (X^\top)^j$.*

PROOF. The two statements are symmetric and so we will just prove the first; the proof of the second is the same except with the inequalities reversed. We first show by induction that $X^i P (X^\top)^i \preceq X^{i-1} P (X^\top)^{i-1}$. The statement holds for $i = 1$. Assume it holds for $i = k$. Following the same argument as in [8, Obs. 7.1.6], we know that given that $A, B \succeq 0$ and $C \in \mathbb{C}^{n \times m}$, we have $A \succeq B \implies C^* A C \succeq C^* B C$ As a result,

$$X^k P (X^\top)^k \preceq X^{k-1} P (X^\top)^{k-1} \implies$$
$$X X^k P (X^\top)^k (X^\top) \preceq X X^{k-1} P (X^\top)^{k-1} X^\top.$$

Therefore, the claim holds for $i = k + 1$. Applying this recursively creates a sequence of orderings and the statement to be proven follows. □

PROOF. [Proof of Theorem 13] The value of empty set is easy to see, $F(\Omega_0^{\emptyset}(\Sigma_0)) = \log\big(\det(\Sigma_0)\det(\Sigma_0^{-1})\big) = 0$.

Take $A \subset B$ where B is a sequence of measurements. Let $B = (1, \ldots, b)$, with the corresponding measurements $\{D_i\}_{i \in B}$, and $\{A(i)\}_{i=1}^{a}$ are the indices in B that are part of A. Note that by definition of a 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. For monotonicity, the requirement is $F(A) \leq F(B)$:

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

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

$$\log\det(\Omega^{Ax}_{a+1})^{-1} - \log\det(\Omega^{A}_{a})^{-1} \geq \log\det(\Omega^{Bx}_{b+1})^{-1} - \log\det(\Omega^{B}_{b})^{-1}$$

which is satisfied if and only if

$$\log\big(\det(\Omega^{Ax}_{a+1})^{-1}\det(\Omega^{A}_{a})\big) \geq \log\big(\det(\Omega^{Bx}_{b+1})^{-1}\det(\Omega^{B}_{b})\big).$$

Now, applying the covariance update (3), we have

$$\Omega^{Ax}_{a+1}(\Sigma_0) = \Omega^{x}_{1}(\Omega^{A}_{a}) = \big((A\Omega^{A}_{a}A^\top)^{-1} + D_x\big)^{-1},$$
$$\Omega^{Bx}_{b+1}(\Sigma_0) = \Omega^{x}_{1}(\Omega^{B}_{b}) = \big((A\Omega^{B}_{b}A^\top)^{-1} + D_x\big)^{-1}.$$

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

$$\det(A\Omega^{A}_{a}A^\top)\det((A\Omega^{A}_{a}A^\top)^{-1} + D_x) \geq \det(A\Omega^{B}_{b}A^\top)\det((A\Omega^{B}_{b}A^\top)^{-1} + D_x).$$

Thus, we have submodularity if and only if the following is satisfied $\det(I + (A\Omega^{A}_{a}A^\top)D_x) \geq \det(I + (A\Omega^{B}_{b}A^\top)D_x)$ Taking $D_x := L^\top L$ (since $D_x \succeq 0$) 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 (7), this is equivalent to

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

Now, let's look at the individual terms in inequality (A.1). We have $\Sigma_0 \succ 0$ and $a \leq b$. Assuming that $A\Sigma_0A^\top \preceq \Sigma_0$ we can apply Lemma 16 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 $D_{A(i)}$ will appear on both sides of inequality (A.1). 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.

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

Therefore, the inequality $j \leq b - a + i$ holds. Assuming that $A^{-\top}D_{A(i)}A^{-1} \succeq D_{A(i)}$, since we know that $D_{A(i)} \succeq 0$, we can apply Lemma 16 again to obtain

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

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

Therefore, under the assumptions made, the inequality (A.1) holds and the function is monotone non-decreasing and submodular. □

Acknowledgements

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

References

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

[2] M. Athans. On the determination of optimal costly measurement strategies for linear stochastic systems. *Automatica*, 8(4):397–412, 1972.

[3] A. Carmi and P. Gurfil. Sensor selection via compressed sensing. *Automatica*, 49(11):3304 – 3314, 2013.

[4] M. L. Fisher, G. L. Nemhauser, and L. A. Wolsey. An analysis of approximations for maximizing submodular set functions - II. In *Polyhedral Combinatorics*, volume 8 of *Mathematical Programming Studies*, pages 73–87. 1978.

[5] D. Golovin and A. Krause. Adaptive submodularity: theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, 42(1):427–486, 2011.

[6] 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*, 42(2):251–260, 2006.

[7] J. P. Hespanha. *Linear systems theory.* Princeton University Press, 2009.

[8] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, 2010.

[9] M. F. Huber, A. Kuwertz, F. Sawo, and U. D. Hanebeck. Distributed greedy sensor scheduling for model-based reconstruction of space-time continuous physical phenomena. In *Int. Conf. on Information Fusion*, pages 102 –109, 2009.

[10] S. T. Jawaid and S. L. Smith. A complete algorithm for the infinite horizon sensor scheduling problem. In *American Control Conference*, pages 437–442, Portland, OR, June 2014.

[11] S. T. Jawaid and S. L. Smith. On the submodularity of sensor scheduling for estimation of linear dynamical systems. In *American Control Conference*, pages 4139–4144, Portland, OR, June 2014.

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

[13] Y. Mo, R. Ambrosino, and B. Sinopoli. Sensor selection strategies for state estimation in energy constrained wireless sensor networks. *Automatica*, 47(7):1330–1338, 2011.

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

[15] L. Orihuela, A. Barreiro, F. Gómez-Estern, and F. R. Rubio. Periodicity of Kalman-based scheduled filters. *Automatica*, 50(10):2672 – 2676, 2014.

[16] C. Seshadhri and J. Vondrák. Is submodularity testable? *Algorithmica*, 69:1–25, 2014.

[17] M. Shamaiah, S. Banerjee, and H. Vikalo. Greedy sensor selection: Leveraging submodularity. In *IEEE Conf. on Decision and Control*, pages 2572 –2577, 2010.

[18] M. Shamaiah, S. Banerjee, and H. Vikalo. Greedy sensor selection under channel uncertainty. *IEEE Wireless Communications Letters*, 1(4):376–379, 2012.

[19] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry. Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control*, 49(9):1453–63, 2004.

[20] M. P. Vitus, W. Zhang, A. Abate, J. Hu, and C. J. Tomlin. On efficient sensor scheduling for linear dynamical systems. *Automatica*, 48(10):2482–2493, 2012.

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

[22] 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*, pages 501–506, 2008.

[23] 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*, 48(3):2594–2613, 2012.

[24] W. Yang, G. Chen, X. Wang, and L. Shi. Stochastic sensor activation for distributed state estimation over a sensor network. *Automatica*, 50(8):2070 – 2076, 2014.