This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2017.2710051, IEEE Control Systems Letters

Supermodular Batch State Estimation in Optimal Sensor Scheduling

Prince Singh [a] Sze Zheng Yong [b] Emilio Frazzoli [c]

***Abstract*—This work addresses the problem of activating, at each time-step in a finite time horizon problem, a subset of available sensors to generate a "high quality" estimate of the state of a discrete-time linear system operating under limited resources. We propose a sensor schedule that minimizes the mean square estimation error of the batch state vector of the system—the *batch state estimation* (BSE) problem. Due to the presence of limited resources, we address the cardinality-constrained BSE problem, which is inherently combinatorial and computationally intractable when working with large-scale systems. This NP-hard complexity is overcome by employing a greedy algorithm, which returns a near-optimal sensor schedule with performance guarantees when minimizing a *supermodular* objective over *matroids*. To this end, we prove (despite the existence of counter-examples in literature) that our objective function is supermodular when the batch *prior information* matrix is a strictly-diagonally-dominant M-matrix (with a constraint on its inverse and conditions on the measurement model). Hence, we obtain a near-optimal solution to the BSE problem via a greedy algorithm. Additionally, we provide its time complexity.**

***Index Terms*—Batch state estimation; Optimal scheduling; Approximation algorithms; Sensor networks; Large-scale systems.**

I. Introduction

Sensor scheduling algorithms provide a formal way for allocating at each time-step, e.g., in a finite time horizon problem, a fraction of the total number of available sensors in order to observe the state of a dynamical system. Thus, *sensor scheduling* problems are inherently combinatorial and become computationally intractable when dealing with large-scale systems. The need for such algorithms arises for challenging information gathering applications that are required to operate under resource constraints, e.g., limited communication bandwidth and battery life, which limits the number of sensors that can be activated [1]. Thus, one seeks to find an *optimal sensor schedule* that specifies a given number of sensors to be selected at each measurement time-step by optimizing a given performance metric.

The current work addresses the finite time horizon sensor scheduling problem by selecting, for each time-step, the set of sensors that maximizes the quality of the state estimate of a discrete-time linear time-variant system. One measure of "best quality" (in a probabilistic sense) of state estimates are those resulting from sequential minimal variance estimation (i.e., Kalman Filtering), which sums the minimum mean square estimation errors of the individual time-steps over a finite time horizon. However, the estimation error from sequential methods are higher than from batch minimum variance estimation, which minimizes the mean square estimation error of the batch state vector of the system across measurement times [2, Chap. 6], i.e., the trace of batch estimation error covariance. We term this the *batch state estimation* (BSE) problem or equivalently the *multiple time-step optimal sensor scheduling problem*. Hence, the sensor schedule resulting from the BSE problem would be of higher quality than the schedule arising from sequential methods.

The BSE problem is combinatorial (NP-hard complexity) and quickly becomes computationally demanding for large-scale systems; more precisely, the number of sensor subsets that need to be considered grows factorially with the total number of sensors available for use. Therefore, we seek to exploit structural properties of the considered cost function to make use of efficient polynomial-time algorithms that yield a near-optimal sensor schedule with performance guarantees.

Literature Review. The sensor scheduling problem has been addressed by algorithms developed based on: randomization of schedules [3], convex relaxations [4] and tree pruning [5]; however, these often perform poorly. We are interested in utilizing the *submodularity* property of the considered cost function to identify suboptimal sensor schedules that can be approximated in polynomial time via *greedy* algorithms that have performance guarantees [6].

In the context of *sequential* Kalman filtering (KF), several researchers have considered the monotone non-increasing and supermodular property for logdet of the error covariance. [7] considered the single time-step case, [8] the case with a finite observation interval, while [9] showed that this metric is, in general, neither submodular nor supermodular over multiple time-steps. Moreover, counterexamples have been presented against the trace of the estimation error covariance being a supermodular and monotone non-increasing set function in [9] for sensor scheduling and [10] for sensor selection problems. However, [11] recently showed that the aforementioned structural properties of the considered cost function hold under some loose conditions on the system's dynamics in the single time-step KF framework.

Meanwhile, for the *batch* state estimation problem, the logdet of the batch error covariance has been shown to be supermodular in [12]. However, the supermodularity of trace of the batch estimation error covariance has, to our best knowledge, not been considered.

Contributions. The primary contribution of this work lies in extending the results for the single time-step KF in [11] to the

[a] Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA (e-mail: prince1@mit.edu).

[b] School for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ, USA (e-mail: szyong@asu.edu).

[c] Institute for Dynamic Systems and Control, Swiss Federal Institute of Technology (ETH), Zürich, Switzerland (e-mail: efrazzoli@ethz.ch).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2017.2710051, IEEE Control Systems Letters

multiple time-step BSE problem by proving that the trace of the batch estimation error covariance is a supermodular and monotone non-increasing set function if: 1) the batch *prior information* matrix (comprising of the process model) is a strictly-diagonally-dominant M-matrix (with a condition on its inverse), and 2) the measurement matrices have a monomial representation, i.e., each sensor measures a single state of the system. This allows for the BSE problem to be addressed via a greedy heuristic whose resulting solution we prove to be within a factor of $\frac{1}{2}$ from the optimal solution, which constitutes our second contribution. Lastly, we prove that this greedy algorithm has polynomial time complexity that is cubic in the time horizon.[1]

II. Problem Formulation

This section formalizes the BSE problem. In particular, section II-A builds the system and measurement models that we will work with. Then, in Section II-B we state our cardinality-constrained sensor scheduling problem.

A. Network Dynamics

We assume a sensor network that consists of m sensors—with $G \triangleq \{1, \dots, m\}$ denoting the set of indices to identify a sensor—each of which outputs a scalar measurement and operates in discrete-time to measure the state of a network's dynamics that evolves in discrete-time according to the following discrete-time linear time-variant representation:

$$x_{t+1} = A_t x_t + u_t + w_t, \quad x_0 \sim \mathcal{N}(\hat{x}_0, P_0), \tag{1}$$

where $x_t \in \mathbb{R}^n$ is the system's state at time t, $A_t \in \mathbb{R}^{n\times n}$ prescribes the system dynamics, $u_t \in \mathbb{R}^n$ is a given control input and $w_t \sim \mathcal{N}(0, W_t)$ is the process noise of appropriate dimension with covariance matrix $W_t \succeq 0$. The initial state x_0 of the system follows a Gaussian distribution with given mean $\hat{x}_0$ and given covariance $P_0 \succ 0$.

We now set up the *sensor scheduling model* by requiring that at each t, at most r_t of the m sensors ($r_t \leq m, \forall t$) be activated to observe the state of system (1). To this end, we have the following measurement model:

$$y_t = \mathbb{S}_t C x_t + \mathbb{S}_t v_t, \tag{2}$$

where $C \in \mathbb{R}^{m\times n}$ is the measurement matrix and $v_t \sim \mathcal{N}(0, V_t)$ is the measurement noise of appropriate dimension with covariance matrix $V_t \succ 0$ and is statistically independent of the process noise. Further, $\mathbb{S}_t \in \{0,1\}^{r_t \times m}$ is the binary *sensor selection matrix* that picks out the r_t rows of C, which corresponds to the r_t sensors that need to be activated at time-step t. Now, for the BSE problem with a finite time horizon $t \in (1, \dots, T)$, we first define:

$$\mathcal{S}_t \triangleq \{j \in G : [\mathbb{S}_t]_{ij} = 1 \text{ for some } i \in (1, \dots, r_t)\} \tag{3}$$

[1]*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 $A^\top$. 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 $\mathrm{Cov}(x) = \mathbb{E}\big[(x - \mathbb{E}[x])(x - \mathbb{E}[x])^\top\big]$.

to be the set of indices of the sensors activated at time-step t. Also, we define $\mathcal{S}_{1:T} \triangleq (\mathcal{S}_1, \dots, \mathcal{S}_T)$. To facilitate our proofs, it will be convenient to work with the batch *information* matrix form for the batch state vector $x_{1:T} = [x_1^\top, \dots, x_T^\top]^\top$ with batch *posterior error covariance* matrix $P_{1:T}(\mathcal{S}_{1:T}) \triangleq \mathrm{Cov}(x_{1:T})$, whose closed form is given, e.g., in [12, eq. (12)]. Thus, we have the following definition.

Definition 1 (Batch Information Matrix)**.** *The inverse of the batch* posterior error covariance *matrix* $P_{1:T}(\mathcal{S}_{1:T})$ *is called the batch* posterior information *matrix:* $\Omega_{1:T}(\mathcal{S}_{1:T}) \triangleq P_{1:T}(\mathcal{S}_{1:T})^{-1}$ *and is positive definite by construction. Moreover, the inverse of the batch* prior error covariance *matrix* $\bar{P}_{1:T}$ *is called the batch* prior information *matrix:* $\bar{\Omega}_{1:T} \triangleq \bar{P}_{1:T}^{-1}$ *and represents the "process" contribution.*

In view of Definition 1, the batch *information* matrix can be put in following *additive* structure form

$$\Omega_{1:T}(\mathcal{S}_{1:T}) = \bar{\Omega}_{1:T} + \Delta_{1:T}(\mathcal{S}_{1:T}), \tag{4}$$

where $\Delta_{1:T}(\mathcal{S}_{1:T})$ represents the contribution due to the "measurements". Equation (4) is derived in Appendix A.

The goal of this work is to come up with a sensor scheduling algorithm that selects the set of sensors $\mathcal{S}_{1:T}$ so as to "minimize" the square of the batch state estimation error. We formalize this notion in the following section.

B. Multiple Time-Step Optimal Sensor Scheduling Problem

The minimum variance linear estimator $\hat{x}_{1:T} \triangleq \mathbb{E}[x_{1:T}]$ is obtained by minimizing the square of the estimation error of the batch state vector: $\mathbb{E}[\|x_{1:T} - \hat{x}_{1:T}\|_2^2 | y_{1:T}]$, and this objective corresponds to *trace of the batch posterior estimation error covariance* $P_{1:T}(\mathcal{S}_{1:T})$. Our goal is to identify the sensor set $\mathcal{S}_{1:T}$, by selecting, at each time-step t, a subset r_t (out of m sensors) that minimizes the aforementioned cost. Let us formalize this by introducing the following set function:

$$g(\mathcal{S}_{1:T}) \triangleq \ \mathrm{trace}(P_{1:T}(\mathcal{S}_{1:T})) - \mathrm{trace}(\bar{P}_{1:T}), \tag{5}$$

which captures the change in the square of the batch state estimation error given a sensor set $\mathcal{S}_{1:T}$, where $g(\mathcal{S}_{1:T}) \leq 0$ since the trace of the posterior error covariance is no larger than that of the prior error covariance $\bar{P}_{1:T}$. The minimization of $g(\mathcal{S}_{1:T})$ is equivalent to minimizing trace$(P_{1:T}(\mathcal{S}_{1:T}))$ since the second term in (5) is a constant and is used to *normalize* the cost function (see Definition 4).

The solution of the following problem results in the sensor set $\mathcal{S}_{1:T}$ that minimizes $g(\mathcal{S}_{1:T})$.

Problem 1 (Cardinality-Constrained Multiple Time-Step Optimal Sensor Scheduling)**.** *For each* $t \in (1, \dots, T)$ *in the batch state estimation problem: given a sensor budget,* $r_t\ (\leq m) \in \mathbb{N}_{>0}$ *on* $|\mathcal{S}_t|$*, i.e., the number of sensors that can be used at time-step t, find the optimal sensor schedule:*

$$\begin{aligned} \underset{\mathcal{S}_t \subseteq G,\ \forall t \in (1,\dots,T)}{\text{minimize}} \quad & g(\mathcal{S}_{1:T}) \\ \text{subject to} \quad & |\mathcal{S}_t| \leq r_t,\ \forall t \in (1, \dots, T). \end{aligned}$$

It is apparent that Problem 1 is combinatorial whose exact solution is computationally intractable when dealing with

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2017.2710051, IEEE Control Systems Letters

large-scale systems. Thus, for an amenable polynomial-time approximate solution (with performance guarantees) by employing a *greedy algorithm*, we resort to proving that $g(\mathcal{S}_{1:T})$ is a supermodular and monotone non-increasing set function for a given choice of sensor set $\mathcal{S}_{1:T}$. To proceed, we impose the following "loose" assumption on the choice of the sensor matrix C defined in (2).

Assumption 1 (Monomial Measurement Matrix and Diagonal Measurement Covariance Matrix)**.** *We assume that*

1) *The measurement matrix C in* (2) *has at most one non-zero entry in each row, and*
2) *The measurement covariance V_t is diagonal.*

Remark 1. *(Observability and Monomial Measurement Matrix) Assumptions 1.1 and 1.2 are commonly made in the sensor scheduling literature and are justified in many practical tasks ranging from robotics [11] to energy [8]. Moreover, if the process model* (1) *and measurement model* (2) *are* time-invariant and *the system is* observable*, then Assumption 1.1 is no longer necessary, since a fully populated measurement matrix C in* (2) *can be conveniently transformed into the* observable canonical *form; thus, yielding a monomial matrix representation (similar to [13, Eq. (18)]).*

III. Supermodularity in Multiple Time-Step Optimal Sensor Scheduling

Before proving our main result, we first review some notions on submodularity of set functions.

A. Submodular Functions

Submodularity is a structural property of *set functions*. For the finite time horizon problem, define the ground set $G_{1:T}$ to be the cartesian product $\{1,\dots,T\} \times G$, then, a set function is defined as $g : 2^{G_{1:T}} \to \mathbb{R}$. Also, we adopt the following notation for the multiple time-step case: for two subsets $A_t, B_t \subseteq G$ at each t, we have that

- $A_{1:T} \subseteq B_{1:T}$ denotes that $A_t \subseteq B_t,\ \forall t \in (1,\dots,T)$.

Definition 2 (Marginal Gain)**.** *For a finite set $G_{1:T}$ and a set function $g : 2^{G_{1:T}} \to \mathbb{R}$, the marginal gain of g at a subset $\mathcal{S}_{1:T} \subseteq G_{1:T}$ with respect to an element $e \in G_{1:T} \setminus \mathcal{S}_{1:T}$ is:*

$$\mathrm{MG}_g(e|\mathcal{S}_{1:T}) \triangleq g(\mathcal{S}_{1:T} \cup \{e\}) - g(\mathcal{S}_{1:T}).$$

Definition 3 (Submodularity and Supermodularity)**.** *Given a finite set $G_{1:T}$, a set function $g : 2^{G_{1:T}} \to \mathbb{R}$ is* submodular *if for all subsets $\mathcal{S}_{1:T} \subseteq \mathcal{S}'_{1:T} \subseteq G_{1:T}$ and an element $e \in G_{1:T} \backslash \mathcal{S}'_{1:T}$, we have*

$$\mathrm{MG}_g(e|\mathcal{S}_{1:T}) \geq \mathrm{MG}_g(e|\mathcal{S}'_{1:T}). \tag{6}$$

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

Equation (6) captures the "diminishing returns" property, i.e., one achieves a higher cost in adding a new element to a smaller set than adding the new element to a larger set.

Definition 4 (Normalized and Monotone)**.** *Given a finite set $G_{1:T}$, a set function $g : 2^{G_{1:T}} \to \mathbb{R}$ is normalized if $g(\emptyset) = 0$, i.e., empty set carries no value, and the set function is monotone non-increasing if for all subsets $\mathcal{S}_{1:T}, \mathcal{S}'_{1:T} \subseteq G_{1:T}$:*

$$\mathcal{S}_{1:T} \subseteq \mathcal{S}'_{1:T} \Rightarrow g(\mathcal{S}_{1:T}) \geq g(\mathcal{S}'_{1:T}),$$

and monotone non-decreasing if

$$\mathcal{S}_{1:T} \subseteq \mathcal{S}'_{1:T} \Rightarrow g(\mathcal{S}_{1:T}) \leq g(\mathcal{S}'_{1:T}).$$

Theorem 1 (Proposition 1.1 in [14])**.** *Given a finite set $G_{1:T}$, a set function $g : 2^{G_{1:T}} \to \mathbb{R}$ is submodular $\iff$ the derived set functions $g_e : 2^{G_{1:T} \backslash \{e\}} \to \mathbb{R}$,*

$$g_e(\mathcal{S}_{1:T}) \triangleq \mathrm{MG}_g(e|\mathcal{S}_{1:T}) \triangleq g(\mathcal{S}_{1:T} \cup \{e\}) - g(\mathcal{S}_{1:T})$$

are monotone non-increasing for all subsets $\mathcal{S}_{1:T} \subseteq G_{1:T}$ and elements $e \in G_{1:T} \backslash \mathcal{S}_{1:T}$.

Next, we prove that the set function $g(\mathcal{S}_{1:T})$ defined in (5) is monotone non-increasing and supermodular in the choice of the sensor set $\mathcal{S}_{1:T}$.

B. Supermodularity in Batch State Estimation

Previous work [9] dealing with the single time-step case provided counterexamples to show that trace of the error covariance is not a supermodular set function in general. Additionally, it has been shown through a simple example in [11] that, for the single time-step case, the sensor set resulting from the maximization of the trace of the information matrix (i.e., the T-optimality criterion [15]) is not equivalent to sensor set resulting from the minimization of the trace of the error covariance. Our goal here is to provide conditions on the system dynamics under which the trace of the batch error covariance is a supermodular set function. We begin with the following proposition, which we prove in Appendix B.

Proposition 1 (Decrease in Batch Estimation Error)**.** *The set function $g(\mathcal{S}_{1:T})$ in* (5) *is normalized and monotonically non-increasing in the choice of the sensor set $\mathcal{S}_{1:T}$.*

We now present definitions of two classes of matrices that the main result below rests upon (see proof in Appendix C).

Definition 5 (Strictly-Diagonally Dominant M-matrix [16])**.** *A matrix $M \in \mathbb{R}^{n \times n}$ is an M-matrix if each: 1) off-diagonal entry is non-positive ($M_{ij} \leq 0$), and 2) eigenvalue has a positive real part. An M-matrix M is said to be strictly diagonally dominant if the following inequalities hold:*

$$|M_{ii}| > \sum_{j \neq i} |M_{ij}|, \quad \forall i = 1,\dots,n. \tag{7}$$

Definition 6 (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) *$U_{ij} \geq \min\{U_{ik}, U_{kj}\},\ \forall (i,k,j) \in \{1,\dots,n\}$*
3) *$U_{ii} > U_{ik},\ \forall (i \neq k) \in \{1,\dots,n\}$*

Theorem 2 (Supermodularity Condition)**.** *Under Assumption 1, the set function $g(\mathcal{S}_{1:T})$ defined in* (5) *is supermodular with respect to the sensor set $\mathcal{S}_{1:T}$ if the batch prior information matrix $\bar{\Omega}_{1:T}$ in Definition 1 is a strictly-diagonally-dominant M-matrix with strictly ultrametric inverse.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2017.2710051, IEEE Control Systems Letters

One can easily check the conditions of Theorem 2 based on Definitions 5 and 6. For instance, the assumptions are satisfied for a relatively common assumption of a diagonal prior covariance matrix (hence, strictly ultrametric) that results in a diagonal prior information matrix (thus, an M-matrix). Moreover, [18, Sec. 8.14] showed that for the blocking experiment design problem, information matrices are Laplacian and thus M-matrices. Note also that the empirical evidence presented in [11] hints that the *strictly-diagonally-dominant* and strict ultrametric inverse requirements on $\bar{\Omega}_{1:T}$ can be relaxed while still guaranteeing supermodularity; hence the performance guarantee in Theorem 3 still applies.

C. Greedy Algorithm

We are now ready to present the *greedy algorithm* that can be used to approximate the solution of Problem 1. As the solution to Problem 1 is NP-hard [19], we propose a polynomial time approximation algorithm (i.e., Algorithm 1) and prove its associated performance guarantees for the present problem. To this end, we first define the following:

Definition 7 (Independence Systems and Matroids [20])**.** *Given a finite set N together with a collection of sets $\mathcal{F} \subseteq 2^N$, the pair $(N, \mathcal{F})$ is:*

- *an independence system, if*
  - *the empty set is independent, i.e., $\emptyset \in \mathcal{F}$, and*
  - *every subset of an independent set N is independent, i.e., for each $Y \subseteq X \subseteq N$, $X \in \mathcal{F} \Rightarrow Y \in \mathcal{F}$.*
- *a matriod, if it is an independence system and for any $X, Y \in \mathcal{F}$ and $|X| > |Y|$, there exists an element $z \in X \backslash Y$ such that $Y \cup \{z\} \in \mathcal{F}$.*

The following *greedy* heuristic, adapted from [6, pg. 5], approximately solves Problem 1 for the pair $(N, \mathcal{F})$ and has been defined for the present setting as

$$N \triangleq G_{1:T}, \mathcal{F} \triangleq \{\mathcal{S}_{1:T} \in 2^N : |\mathcal{S}_t| \leq r_t, \forall t \in (1, \ldots, T)\}, \quad (8)$$

where $G_{1:T}$ has been defined in Section III-A and the set $\mathcal{F}$ denotes the particular sensor set belonging to the power set of N (or equivalently, $G_{1:T}$) that satisfies the prescribed sensor budget requirements for the finite time horizon problem.

Algorithm 1 iteratively builds the sensor set $\mathcal{S}_{1:T} \triangleq (t, \mathcal{S}_t)_{t=1}^T$, with $\mathcal{S}_t$ being the set of sensors activated at time-step t, from the set $N = G_{1:T}$. Note that the batch state estimation problem is solved for each $e \in N^{t-1}$ in Line 5 of Algorithm 1 and we propose the use of the Rauch-Tung-Striebel (RTS) smoother algorithm [2, Section 6.1.1.2], comprising of forward (Kalman Filter) and backward passes, which has lower time complexity than batch matrix inversion.

Next, we provide the performance guarantees and the time complexity of Algorithm 1 (will be proven in Appendix D).

Theorem 3 (Performance Guarantee and Time Complexity)**.** *This theorem is comprised of two parts:*

- Performance Guarantee: *If the set function $g(\mathcal{S}_{1:T}) \leq 0$ in* (5) *is normalized, monotone non-increasing and supermodular in the choice of the sensor set $\mathcal{S}_{1:T}$, then, Algorithm 1 enjoys the following guarantee:*

Algorithm 1: Greedy Algorithm for Problem 1.

1 **Input:** N, $\mathcal{F}$, $g(\mathcal{S}_{1:t})$;
2 **Output:** Greedy solution $(\mathcal{S}_{1:T})$;
3 $\mathcal{S}_{1:T}^0 \leftarrow \emptyset$, $N^0 \leftarrow N$, iteration $i \leftarrow 1$;
4 **while** $N^{i-1} \neq \emptyset$ **or** $\mathcal{S}_{1:T}^{i-1} \cup \{j\} \notin \mathcal{F}$ $(\forall j \in N^{i-1})$ **do**
5 Select $e(i) \in \arg\min_{e \in N^{i-1}} \mathrm{MG}_g(e|\mathcal{S}_{1:T}^{i-1})$ with ties settled arbitrarily;
6 **if** $|\mathcal{S}_{1:T}^{i-1} \cup \{e(i)\}| \notin \mathcal{F}$ **then**
7 $N^{i-1} \leftarrow N^{i-1} \backslash \{e(i)\}$;
8 continue (go to line 4) ;
9 **else**
10 $\mathcal{S}_{1:T}^{i} \leftarrow \mathcal{S}_{1:T}^{i-1} \cup \{e(i)\}$;
11 $N^{i} \leftarrow N^{i-1} \backslash \{e(i)\}$;
12 **end**
13 $i \leftarrow i + 1$;
14 **end**

$$g(\mathcal{S}_{1:T}^G) \leq \tfrac{1}{2} g(\mathcal{S}_{1:T}^*) \Leftrightarrow \mathit{trace}(P_{1:T}(\mathcal{S}_{1:T}^G)) \leq \tfrac{1}{2}(\mathit{trace}(P_{1:T}(\mathcal{S}_{1:T}^*)) + \mathit{trace}(\bar{P}_{1:T}))'$$

where $\mathcal{S}_{1:T}^*$ is the optimal solution to Problem 1 and $\mathcal{S}_{1:T}^G$ is the greedy solution resulting from Algorithm 1.

- Time complexity: *Algorithm 1 has a time complexity of* $\mathcal{O}\big(mn^{2.37}T^2 \sum_{t=1}^T r_t\big) = \mathcal{O}\big(mn^{2.37}T^3 \max_{t=\{1,\ldots,T\}} r_t\big)$.

The above performance guarantee is the worst-case bound and the greedy algorithm often performs much better in practice. The time complexity of Algorithm 1 is also lower than the state-of-the-art multiple time-step scheduling algorithm based on semidefinite relaxation in [21] that was reported by [12] to be $\mathcal{O}\big((\max_{t \in (1,\ldots,T)} r_t^2) T^2 (nT)^{2.5}\big)$. Additionally, no performance bounds were given for the convex relaxation-based sensor scheduling algorithm in [21].

IV. Conclusion and Future Work

For the batch state estimation (BSE) problem, we derived sufficient conditions on the prior information matrix so that the cost function we seek to minimize—trace of the batch error covariance—is a supermodular and monotone non-increasing set function; thus, we obtain a near-optimal solution to the BSE problem via a greedy heuristic, whose performance guarantees and time complexity can be rigorously proven. A natural extension of this work is to devise scheduling algorithms for nonlinear systems in the face of measurement noise and outliers to facilitate on-board decision making for complex robotics applications, e.g., Simultaneous Localization And Mapping (SLAM). Additional future work includes the simulation of large-scale systems.

References

[1] A. Krause, R. Rajagopal, A. Gupta, and C. Guestrin. Simultaneous placement and scheduling of sensors. In *International Conference on Information Processing in Sensor Networks*, pages 181–192, 2009.

[2] J. L. Crassidis and J. L. Junkins. *Optimal estimation of dynamic systems*. CRC press, 2011.

[3] V. Gupta, T. Chung, B. Hassibi, and R. Murray. On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage. *Automatica*, 42(2):251–260, 2006.

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

[5] M. P. Vitus, W. Zhang, A. Abate, J. Hu, and C. J. Tomlin. On efficient sensor scheduling for linear dynamical systems. *Automatica*, 2012.

[6] M. L. Fisher, G. L. Nemhauser, and L. A. Wolsey. An analysis of approximations for maximizing submodular set functions-II. In *Polyhedral combinatorics*, pages 73–87. Springer, 1978.

[7] M. Shamaiah, S. Banerjee, and H. Vikalo. Greedy sensor selection: Leveraging submodularity. In *IEEE CDC*, pages 2572–2577, 2010.

[8] V. Tzoumas, A. Jadbabaie, and G. J. Pappas. Sensor placement for optimal Kalman filtering: Fundamental limits, submodularity, and algorithms. *arXiv preprint arXiv:1509.08146*, 2015.

[9] S. Jawaid and S. Smith. Submodularity and greedy algorithms in sensor scheduling for linear dynamical systems. *Automatica*, 2015.

[10] H. Zhang, R. Ayoub, and S. Sundaram. Sensor selection for optimal filtering of linear dynamical systems: Complexity and approximation. In *IEEE Conference on Decision and Control*, pages 5002–5007, 2015.

[11] 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 *IEEE American Control Conference, suppl. material: http://tinyurl.com/h4p2hwm*, 2017.

[12] V. Tzoumas, A. Jadbabaie, and G. J. Pappas. Near-optimal sensor scheduling for batch state estimation: Complexity, algorithms, and limits. In *IEEE Conference on Decision and Control (CDC)*, pages 2695–2702, 2016.

[13] D. Luenberger. Canonical forms for linear multivariable systems. *IEEE Transactions on Automatic Control*, 12(3):290–293, 1967.

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

[15] A. Atkinson, A. Donev, and R. Tobias. *Optimum experimental designs, with SAS*, volume 34. Oxford Univ. Press, 2007.

[16] C. Dellacherie, S. Martinez, and J. San Martin. *Inverse M-matrices and ultrametric matrices*, volume 2118. Springer, 2014.

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

[18] G. M. Constantine. *Combinational theory and statistical design*. John Wiley & Sons, Inc., 1986.

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

[20] D. J. A. Welsh. Matroid theory. *London Math. Soc. Monogr.*, 1976.

[21] V. Roy, A. Simonetto, and G. Leus. Spatio-temporal sensor management for environmental field estimation. *Sig Proc*, 128:369–381, 2016.

[22] T. Summers, F. Cortesi, and J. Lygeros. On submodularity and controllability in complex dynamical networks. *IEEE Transactions on Control of Network Systems*, 3(1):91–101, 2016.

APPENDIX

A. Derivation of the Batch Information Matrix (4)

Let $w_{1:T} = [w_1^\top, \dots, w_T^\top]^\top$. Then, the batch form of (1) can be re-written as

$$u_{1:T} = \bar{H}_{1:T}\ x_{1:T} - w_{1:T}, \tag{9}$$

where the k-th row of $\bar{H}_{1:T} \in \mathbb{R}^{nT \times nT}$ comprises of the tuples $(-A_k, I_{n\times n})$ with $-A_{k+1}$ being placed below $I_{n\times n}$ of the k-th row. Now, let $\mathbb{S}_{1:T}$ denote a matrix of diagonal form whose diagonal entries are comprised of the sensor selection matrices: $\mathbb{S}_1, \dots, \mathbb{S}_T$. Similarly, $C_{1:T}$ denotes a block diagonal matrix with the measurement matrix C on each of its T diagonal entries and define $v_{1:T} = [v_1^\top, \dots, v_T^\top]^\top$. Then, the batch form of (2) can be written as

$$y_{1:T} = H_{1:T}\ x_{1:T} + \mathbb{S}_{1:T}\ v_{1:T}, \tag{10}$$

where $\mathbb{S}_{1:T}C_{1:T} \triangleq H_{1:T} \in \mathbb{R}^{\sum_{t=1}^{T} r_t \times nT}$ is a matrix of diagonal form whose diagonal elements comprise of the matrices: $\mathbb{S}_1 C, \dots, \mathbb{S}_T C$. Finally, combining (9) and (10),

$$\begin{bmatrix} u_{1:T} \\ y_{1:T} \end{bmatrix} = \begin{bmatrix} \bar{H}_{1:T} \\ H_{1:T} \end{bmatrix} x_{1:T} + \begin{bmatrix} -w_{1:T} \\ \mathbb{S}_{1:T}\ v_{1:T} \end{bmatrix}, \tag{11}$$

then, the "optimal" (i.e., in the sense of minimum variance) least squares estimate of (11) based upon the two information subsets $(u_{1:T}, y_{1:T})$ is given by application of the *Gauss-Markov* theorem [2, Chap. 2],

$$\hat{x}_{1:T} = \left(\bar{H}_{1:T}^\top W_{1:T}^{-1} \bar{H}_{1:T} + H_{1:T}^\top \left(\mathbb{S}_{1:T} V_{1:T} \mathbb{S}_{1:T}^\top\right)^{-1} H_{1:T} \right)^{-1} \left[\bar{H}_{1:T}^\top W_{1:T}^{-1} u_{1:T} + H_{1:T}^\top \mathbb{S}_{1:T} V_{1:T}^{-1}\ y_{1:T} \right]$$

and it is well known [2, pg. 21] that the batch *information* matrix is given by the term inside the inverse of solution above. Now, noting that $\left(\mathbb{S}_{1:T} V_{1:T} \mathbb{S}_{1:T}^\top\right)^{-1} = \mathbb{S}_{1:T} V_{1:T}^{-1} \mathbb{S}_{1:T}^\top$ because $\mathbb{S}_{1:T}, V_{1:T}$ are block diagonal matrices with $\mathbb{S}_{1:T}$ being a binary matrix and using the definition $H_{1:T} \triangleq \mathbb{S}_{1:T} C_{1:T}$ in the above expression, we have

$$\begin{aligned} \Omega_{1:T}(\mathbb{S}_{1:T}) &= \bar{H}_{1:T}^\top W_{1:T}^{-1} \bar{H}_{1:T} + C_{1:T}^\top J V_{1:T}^{-1} J C_{1:T} \\ &\triangleq \bar{\Omega}_{1:T} + \Delta_{1:T}(\mathbb{S}_{1:T}), \end{aligned} \tag{12}$$

where $J \triangleq \mathbb{S}_{1:T}^\top \mathbb{S}_{1:T}$, which is the product of monomial matrices and hence has a diagonal form. Further, $W_{1:T}, V_{1:T}$ represent block diagonal matrices, of appropriate dimensions, and are comprised of the noise covariance matrices of the batch process noise $w_{1:T}$ and the batch measurement noise $v_{1:T}$, respectively. Moreover, the "measurement" contribution term, defined as $\Delta_{1:T}(\mathbb{S}_{1:T})$, can be further simplified:

$$\Delta_{1:T}(\mathbb{S}_{1:T}) \triangleq C_{1:T}^\top J V_{1:T}^{-1} J C_{1:T} = C_{1:T}^\top J V_{1:T}^{-1} C_{1:T}, \tag{13}$$

where the first equality follows from $J, V_{1:T}$ being block diagonal matrices (thus, $J V_{1:T}^{-1} = V_{1:T}^{-1} J$) with the property $J^2 = J$ due to the definition of $\mathbb{S}_{1:T}$.

B. Proof of Proposition 1

Proof. To show that the set function $g(\mathcal{S}_{1:T})$ defined in (5) is normalized, for the finite time horizon case, note that $\bar{\Omega}_{1:T}$ in (4) is a constant matrix and so $\bar{P}_{1:T} \triangleq \bar{\Omega}_{1:T}^{-1}$ is a constant matrix. Thus, it is straightforward to see that $g(\emptyset) = \mathrm{trace}(P_{1:T}(\emptyset)) - \mathrm{trace}(\bar{P}_{1:T}) = \mathrm{trace}(\bar{P}_{1:T}) - \mathrm{trace}(\bar{P}_{1:T}) = 0$; hence, this set function is normalized by Definition 4.

We now prove that $g(\mathcal{S}_{1:T})$ is monotone non-increasing in accordance with Definition 4: $\mathcal{S}_{1:T} \subseteq \mathcal{S}'_{1:T} \Rightarrow g(\mathcal{S}'_{1:T}) \leq g(\mathcal{S}_{1:T})$. From Definition 1 for a sensor set $\bar{\mathcal{S}}$, we have $P_{1:T}(\bar{\mathcal{S}}) \triangleq \Omega_{1:T}(\bar{\mathcal{S}})^{-1} = \left(\bar{\Omega}_{1:T} + \Delta_{1:T}(\bar{\mathcal{S}})\right)^{-1}$. Further, we have $\Delta_{1:T}(\mathcal{S}'_{1:T}) \succeq \Delta_{1:T}(\mathcal{S}_{1:T})$ since more information is gained by using more sensors and by the positive definiteness of the prior information matrix $\Rightarrow \bar{\Omega}_{1:T} + \Delta_{1:T}(\mathcal{S}'_{1:T}) \succeq \bar{\Omega}_{1:T} + \Delta_{1:T}(\mathcal{S}_{1:T}) \Rightarrow P_{1:T}(\mathcal{S}'_{1:T}) \preceq P_{1:T}(\mathcal{S}_{1:T}) \Rightarrow P_{1:T}(\mathcal{S}'_{1:T}) - \bar{P}_{1:T} \preceq P_{1:T}(\mathcal{S}_{1:T}) - \bar{P}_{1:T}$ since we have subtracted a positive definite constant matrix $\bar{P}_{1:T}$ on both sides of the inequality and finally, taking trace of both sides, we have our desired result: $g(\mathcal{S}'_{1:T}) \leq g(\mathcal{S}_{1:T})$. ∎

C. Proof of Theorem 2

Proof. We now prove that $g(\mathcal{S}_{1:T})$ is supermodular in this choice of the sensor set. To this end, we equivalently prove that $f(\mathcal{S}_{1:T}) \triangleq -g(\mathcal{S}_{1:T})$ is submodular in this choice of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2017.2710051, IEEE Control Systems Letters

sensor set. Thus, for any subset $\mathcal{S}_{1:T} \subseteq G_{1:T}$ and element $e \in G_{1:T} \backslash \mathcal{S}_{1:T}$, we have the following derived set function:

$$\begin{aligned} & f_e(\mathcal{S}_{1:T}) \triangleq f(\mathcal{S}_{1:T} \cup \{e\}) - f(\mathcal{S}_{1:T}) \\ & = -\mathrm{trace}(P_{1:T}(\mathcal{S}_{1:T} \cup \{e\})) + \mathrm{trace}(P_{1:T}(\mathcal{S}_{1:T})) \\ & = -\mathrm{trace}(\{\Omega_{1:T}(\mathcal{S}_{1:T} \cup \{e\})\}^{-1}) + \mathrm{trace}(\{\Omega_{1:T}(\mathcal{S}_{1:T})\}^{-1}) \\ & = -\mathrm{trace}\left(\{\Omega(\mathcal{S}_{1:T}) + \Delta_e\}^{-1}\right) + \mathrm{trace}(\{\Omega(\mathcal{S}_{1:T})\}^{-1}), \end{aligned}$$

where the second equality follows from $P_{1:T}(\mathcal{S}_{1:T}) \triangleq \Omega_{1:T}(\mathcal{S}_{1:T})^{-1}$ (Definition 1) and in the last line we have dropped the subscript $1{:}T$ for notational simplicity and have used the additivity property of the information matrix from (4). Further, $\Delta_e = \Delta_{1:T}(\mathbb{S}^e_{1:T})$ is as per (13), where only one sensor e at some particular time-step (i.e., a pair $(t,i) \in \{1,\ldots,T\} \times G$) is added for the entire time horizon and $\mathbb{S}^e_{1:T} \in \mathbb{R}^{1 \times mT}$ is a row vector with a '1' in its $((t-1)m+i)$-th entry and '0' elsewhere. Since $\mathbb{S}^e_{1:T}$ is of rank 1 and Δ_e is computed from $\mathbb{S}^e_{1:T}$ and is diagonal by the monomial assumption of $C_{1:T}$ in Assumption 1, $\Delta_e \triangleq c_e c_e^\top$ is a rank one matrix with $c_e \in \mathbb{R}^{nT \times 1}$.

Now, based on Theorem 1, proving submodularity of $f(\mathcal{S}_{1:T})$ is equivalent to proving that the marginal gain $f_e(\mathcal{S}_{1:T})$ is monotonically non-increasing. To this end, we proceed along the lines of the proof of Theorem 5 in [22]: for two sensor sets $\mathcal{S}_{1:T}, \mathcal{S}'_{1:T} \subseteq G_{1:T}$ such that $\mathcal{S}_{1:T} \subseteq \mathcal{S}'_{1:T}$ and a scalar $\gamma \in [0,1]$, define $\Omega(\gamma) \triangleq \Omega(\mathcal{S}_{1:T}) + \gamma\left(\Delta(\mathcal{S}'_{1:T}) - \Delta(\mathcal{S}_{1:T})\right)$ and the following scalar function:

$$\hat{f}_e(\gamma) \triangleq -\mathrm{trace}\left(\{\Omega(\gamma) + \Delta_e\}^{-1}\right) + \mathrm{trace}(\Omega(\gamma)^{-1}).$$

Now, in order to show that $\hat{f}_e(\gamma)$ is monotonically non-increasing, we rely on the fundamental theorem of calculus and require that the following inequality holds: $f_e(\mathcal{S}'_{1:T}) = \hat{f}_e(1) = \hat{f}_e(0) + \int_0^1 \frac{d}{d\gamma}\hat{f}_e(\gamma)\, d\gamma \leq \hat{f}_e(0) = f_e(\mathcal{S}_{1:T})$, for which a sufficient condition is that $\frac{d}{d\gamma}\hat{f}_e(\gamma) \leq 0$. Performing straightforward calculations similar to that done for Proposition 1 in [11], we obtain the following trace inequality for supermodularity: $\frac{d}{d\gamma}\hat{f}_e(\gamma) = \mathrm{trace}\big[\Psi_e(\gamma) \cdot \bar{\mathcal{F}}\big] \leq 0$ where $\Psi_e(\gamma) \triangleq \{\Omega(\gamma) + \Delta_e\}^{-2} - \Omega(\gamma)^{-2}$ and $\bar{\mathcal{F}} \triangleq \Delta(\mathcal{S}'_{1:T}) - \Delta(\mathcal{S}_{1:T}) \succeq 0$ since more information is gained by using more sensors. By Corollary 1 of [11], a sufficient condition to guarantee the above trace inequality is that $\Psi_e(\gamma) \preceq 0$, which is true iff c_e (such that $\Delta_e \triangleq c_e c_e^\top$) is either zero or is an eigenvector of $\Omega(\gamma)$. As these conditions are very restrictive, Assumption 1 is imposed, which implies that $\bar{\mathcal{F}}$ is a matrix of diagonal form comprising entirely of non-negative elements on its diagonal, which can be verified from (13). And it is straightforward to see that the above trace inequality is satisfied if the diagonal entries of $\Psi_e(\gamma)$ are non-positive. Finally, adopting Theorem 4 of [11] for the batch case, the trace inequality is satisfied if the prior information matrix $\bar{\Omega}$ defined in (4) is a strictly-diagonally-dominant M-matrix with strictly ultrametric inverse. ■

D. Proof of Theorem 3

Proof. [of performance guarantee] We will need the following result on submodular maximization over matroids.

Lemma 1 (Theorem 2.1 of [6])**.** *Given an independence system $(N, \mathcal{F})$ described as the intersection of P matroids, and given $h : 2^N \to \mathbb{R}$ a normalized, monotone non-decreasing, submodular set function; there exists a polynomial time greedy algorithm that returns an approximate solution to:*

$$\max_{S \subseteq N} \{h(S) : S \in \cap_{p=1}^{P} \mathcal{F}_p, (N, \mathcal{F}_p) \text{ matroids } \forall p = (1,\ldots,P)\}, \tag{14}$$

that satisfies $\dfrac{h(S^*) - h(S^G)}{h(S^*) - h(\emptyset)} \leq \dfrac{P}{P+1}$, where S^* is the optimal solution to (14) *and S^G the greedy solution.*

The following facilitates the performance guarantee proof.

Lemma 2. *Problem 1 is a special case of the optimization problem* (14) *with $P = 1$.*

Proof. We need to appropriately re-define the independence system $(N, \mathcal{F})$ and the set function $h(S)$ so that (14) becomes equivalent to Problem 1. Given the set $G_{1:T}$ defined in Section III-A along with the pair $(N, \mathcal{F})$ defined in (8):

- For $S \equiv \mathcal{S}_{1:T} \subseteq N \equiv G_{1:T}$, let $h(S) = -g(\mathcal{S}_{1:T})$ with $g(\mathcal{S}_{1:T})$ defined in (5), which is a normalized, monotone non-decreasing and submodular set function as proven in Proposition 1 and Theorem 2, respectively.
- It is straightforward to verify that the pair $(N, \mathcal{F})$ in (8) is both an independence system and a matroid as per Definition 7. Finally, this independence system can be written using only one matroid, i.e., $P = 1$.

The above two points completes this proof. ■

Proof. [of time complexity] We employ the Rauch-Tung-Striebel (RTS) smoother algorithm [2, Section 6.1.1.2] in order to compute the time complexity of Algorithm 1. The RTS smoother comprises of the forward (Kalman Filter) and backward passes, which are each run for T time-steps. In the instance of having no sensor measurements at a given time-step, the forward pass comprises solely of the prediction step (with no update step). Now, at each time-step, for each pass of the smoother, $g(\mathcal{S}_{1:t})$ defined in (5) needs to be computed and its complexity is governed by generating the error covariance matrix, which relies on computing the multiplication and inverse of $(n \times n)$ matrices (based on the RTS smoother equations) and so is an $\mathcal{O}(n^{2.37})$ operation when using the optimized Coppersmith-Winograd algorithm.

Additionally, in order to find the minimum element that minimizes the marginal gain, the marginal gain is computed for all elements in the set N—defined in (8)—resulting in an operational time complexity of $\mathcal{O}(mT)$. This optimization needs to be done only once and the minimizers can be ranked and stored beforehand so that they can be used when executing line 8 within each iteration of Algorithm 1. Thus, for the finite time horizon problem, the complexity of the RTS smoother is $\mathcal{O}(mn^{2.37}T^2)$. Finally, the while loop is terminated once all the elements in $\mathcal{F}$ have been exhaustively searched (i.e., the prescribed sensor budget has been met); thereby, resulting in the overall time complexity of Algorithm 1 to be $\mathcal{O}\big(mn^{2.37}T^2 \sum_{t=1}^{T} r_t\big) = \mathcal{O}\big(mn^{2.37}T^3 \max_{t=\{1,\ldots,T\}} r_t\big)$. ■