Sensor Selection for Optimal Filtering of Linear Dynamical Systems: Complexity and Approximation

Haotian Zhang, Raid Ayoub and Shreyas Sundaram

***Abstract*— We consider the problem of selecting an optimal set of sensors to estimate the states of linear dynamical systems. Specifically, the goal is to choose (at design-time) a subset of sensors (satisfying certain budget constraints) from a given set in order to minimize the steady state error covariance produced by a Kalman filter. In this paper, we show that this sensor selection problem is NP-hard, even under the additional assumption that the system is stable. We then provide bounds on the worst-case performance of sensor selection algorithms based on the system dynamics, and show that certain typical objective functions are not submodular or supermodular in general. While this makes it difficult to evaluate the performance of greedy algorithms for sensor selection, we show via simulations that a certain greedy algorithm performs well in practice. We also propose a variant of the greedy algorithm which is based on the Lyapunov equation and show that the corresponding (relaxed) cost function is modular.**

I. Introduction

A key problem in control system design is to select a suitable set of sensors or actuators (either at design-time or at run-time) in order to achieve certain performance objectives [1]. For the objective of estimating the state of a given linear Gauss-Markov system, there has been a growing literature in the past decade that studies how to dynamically select sensors at run-time to minimize the error covariance of the corresponding Kalman filter. This is known as the *sensor scheduling problem*, due to the fact that a different set of sensors can be chosen at each time-step (e.g., see [2], [3]).

The design-time sensor selection problem (where the sensors are not allowed to change over time) has been studied in various forms, including cases where the objective is to guarantee a certain structural property of the system [4] or to optimize energy or information theoretic metrics [5], [6]. In [7], the authors considered the problem of estimating a *static* random variable and showed that the corresponding problem can be modeled as a Boolean convex problem. They then proposed various heuristics for sensor selection by using convex relaxation techniques. Sensor selection for parameter estimation was also studied in [8] where the uncertainty is due to deterministic and bounded disturbances. However, the results in [7], [8] do not directly translate to state estimation for dynamical systems and no performance guarantees are provided on the provided procedure.

In this paper, we consider the design-time sensor selection problem for optimal filtering of linear dynamical systems. It is often claimed in the literature that such optimization problems are intractable [7], [8], [9]; however, to the best of our knowledge, there is still no explicit characterization of the complexity of this problem for given (numerically specified) systems. Thus, our first contribution is to show that it is NP-hard to find the optimal solution of cost-constrained sensor selection problems, even under the assumption that the system is stable. We then provide insights into what factors of the system affect the performance of sensor selection algorithms by using the concept of the *sensor information matrix* [9]. We show that the worst-case performance can be bounded by a parameter that depends only on the system matrix, and that the performance of a sensor selection algorithm cannot be arbitrarily bad if the system matrix is well conditioned, even under very large noise.

Since in general it may be intractable to find the optimal selection of sensors, a reasonable tradeoff is to design appropriate approximation algorithms. For the (run-time) sensor scheduling problem, it has been shown that certain types of cost function are submodular [3] and thus certain greedy algorithms can be applied to obtain guaranteed performance. However, we show that the cost function of the (design-time) sensor selection problem does not necessarily have this property, precluding the direct application of classical results from the theory of combinatorial optimization. Nevertheless, we show via simulations that a certain greedy algorithm performs well in practice. We also consider the problem of optimizing an upper bound of the original cost function based on the Lyapunov equation and propose a variant of the greedy algorithm. We show that the relaxed cost function is modular and that the running time of the corresponding greedy algorithm scales more slowly with the number of states in the system as compared to the original greedy algorithm, at the cost of a decrease in performance.

Notation For a square matrix M, let M^T, $\text{trace}(M)$, $\det(M)$, $\{\lambda_i(M)\}$ and $\{\sigma_i(M)\}$ be its transpose, trace, determinant, set of eigenvalues and set of singular values, respectively. The eigenvalues $\{\lambda_i(M)\}$ of M are ordered with nonincreasing real parts, i.e., $\Re\lambda_1(M) \geq \cdots \geq \Re\lambda_n(M)$; the same order applies to the set of singular values $\{\sigma_i(M)\}$. A positive semi-definite matrix M is denoted by $M \succeq 0$. The identity matrix with dimension n is denoted by $I_{n\times n}$. For a vector v, let $\text{diag}(v)$ be the diagonal matrix with diagonal entries being the elements of v.

This material is based upon work supported by a grant from Intel Corporation.

Haotian Zhang is with the Department of Electrical and Computer Engineering at the University of Waterloo. Email: h223zhan@uwaterloo.ca.

Raid Ayoub is with the Strategic CAD Labs, Intel Corporation. Email: raid.ayoub@intel.com.

Shreyas Sundaram is with the School of Electrical and Computer Engineering at Purdue University. Email: sundara2@purdue.edu.

II. Problem Formulation

Consider the discrete-time linear system

$$x_{k+1} = Ax_k + w_k, \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the system state, $w_k \in \mathbb{R}^n$ is a zero-mean white Gaussian noise process with $E\left[w_k w_k^T\right] = W$ for all $k \in \mathbb{N}$, and $A \in \mathbb{R}^{n\times n}$ is the system dynamics matrix. We assume throughout that the pair $(A, W^{\frac{1}{2}})$ is stabilizable.

There is a set $\mathcal{Q}$ consisting of q sensors. Each sensor $i \in \mathcal{Q}$ provides a scalar measurement of the form

$$y_k^i = C_i x_k + v_k^i, \tag{2}$$

where $C_i \in \mathbb{R}^{1\times n}$ is the state measurement matrix for that sensor, and $v_k^i \in \mathbb{R}$ is a zero-mean white Gaussian noise process. For convenience, we use $C \in \mathbb{R}^{q\times n}$ to denote the matrix where the i-th row is C_i, and $v_k \in \mathbb{R}^q$ to denote the vector where the i-th component is v_k^i. We denote $E[v_k v_k^T] = V$ and take $E[v_k w_j^T] = 0$ for all $j, k \in \mathbb{N}$.

Each sensor $i \in \mathcal{Q}$ has an associated *cost* $r_i \in \mathbb{R}_{\geq 0}$, representing, for example, monetary costs of purchasing and installing that sensor. Define the cost vector $r \triangleq [r_1 \cdots r_q]^T$. We also assume there is a *sensor budget* $\beta \in \mathbb{R}_{\geq 0}$, representing the total cost that can be spent on sensors from $\mathcal{Q}$.

For any given subset of sensors, the Kalman filter provides the optimal estimate of the state using the measurements from those sensors (in the sense of minimizing mean square estimation error, under the stated assumptions on the noise processes). Let $z \in \{0,1\}^q$ be the indicator vector of the chosen sensors, i.e., $z_i = 1$ if and only if sensor $i \in \mathcal{Q}$ is chosen, and define the selection matrix $Z \triangleq \operatorname{diag}(z)$. Define $\tilde{C} \triangleq ZC$, $\tilde{V} \triangleq ZVZ^T$, and let $\Sigma_{k+1|k}(z)$ be the filtered error covariance (at time-step k) of the Kalman filter when the set of sensors indicated by the vector z are chosen. Then the error covariance $\Sigma_{k+1|k}(z)$ satisfies the Riccati mapping [10]:

$$\Sigma_{k+1|k}(z) = A^T \Sigma_{k|k-1}(z) A + W - A^T \Sigma_{k|k-1}(z)\tilde{C}^T(\tilde{C}\Sigma_{k|k-1}(z)\tilde{C}^T + \tilde{V})^{-1}\tilde{C}\Sigma_{k|k-1}(z)A. \tag{3}$$

Note that the inverse in the Riccati mapping is interpreted as a pseudo-inverse if the argument is not invertible.

If the pair $(A, \tilde{C})$ is detectable (and given the stabilizability of $(A, W^{\frac{1}{2}})$), the error covariance $\Sigma_{k+1|k}(z)$ will converge to a unique limit $\Sigma(z)$, which satisfies the following *discrete algebraic Riccati equation (DARE)*:

$$\Sigma(z) = A^T\Sigma(z)A + W - A^T\Sigma(z)\tilde{C}^T(\tilde{C}\Sigma(z)\tilde{C}^T + \tilde{V})^{-1}\tilde{C}\Sigma(z)A. \tag{4}$$

Using the matrix inversion lemma [11], the DARE (4) can also be written as

$$\Sigma(z) = W + A^T(\Sigma^{-1}(z) + \underbrace{\tilde{C}^T\tilde{V}^{-1}\tilde{C}}_{R(z)})^{-1}A, \tag{5}$$

where the matrix $R(z)$ is the so-called sensor information matrix corresponding to the indicator vector z.[1]

Definition 1 (Feasible Sensor Selection): The sensor selection $z \in \{0,1\}^q$ is said to be **feasible** if $\Sigma_{k+1|k}(z)$ converges to a finite limit (denoted by $\Sigma(z)$) as $k \to \infty$, and this limit does not depend on $\Sigma_{0|0}(z)$. □

We now propose the following *Kalman filtering sensor selection (KFSS) problem.*

Problem 1 (KFSS Problem): Given a system matrix $A \in \mathbb{R}^{n\times n}$, a measurement matrix $C \in \mathbb{R}^{q\times n}$, a positive semidefinite system noise covariance matrix $W \in \mathbb{R}^{n\times n}$, a positive semidefinite sensor noise covariance matrix $V \in \mathbb{R}^{q\times q}$, a cost vector $r \in \mathbb{R}^q_{\geq 0}$, and a budget $\beta \in \mathbb{R}_{\geq 0}$, the KFSS problem is to solve the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & \operatorname{trace}(\Sigma(z)) \\ \text{subject to} \quad & r^T z \leq \beta \\ & z \in \{0,1\}^q \end{aligned} \tag{6}$$

where $\Sigma(z)$ is given by equation (4) and z is the decision variable. When z is not feasible, define $\operatorname{trace}(\Sigma(z)) = \infty$.

III. Complexity of the KFSS Problem

In this section, we study the complexity of the KFSS problem and show that the problem is NP-hard. We will use the following well-known result on Kalman filtering [10].

Lemma 1: When the pair $(A, W^{\frac{1}{2}})$ is stabilizable, the indicator vector z is feasible if and only if the pair $(A, \tilde{C})$ is detectable.

To show the complexity of the KFSS problem, we will relate it to the problems described below.

Problem 2: Given a matrix $A \in \mathbb{R}^{n\times n}$, the problem of finding a diagonal matrix $M \in \mathbb{R}^{n\times n}$ with the fewest nonzero elements such that the pair (A, M) is controllable (resp. stabilizable, detectable) is referred to as the minimum controllability (resp. minimum stabilizability, minimum detectability) problem.

We now characterize the complexity of the KFSS problem.

Theorem 1: The KFSS problem is NP-hard. □

Proof: We first give a reduction from the minimum detectability problem to the KFSS problem. Given $A \in \mathbb{R}^{n\times n}$ for the minimum detectability problem and some $p \in \{1,\ldots,n\}$, the instance for the corresponding KFSS problem with parameter p is the system matrix A, the set $\mathcal{Q}$ of n sensors with the measurement matrix $C = I_{n\times n}$, the system noise covariance matrix $W = I_{n\times n}$, the sensor noise covariance matrix $V = I_{n\times n}$, the cost vector $r = [1\cdots 1]^T$ and the budget $\beta = p$. Suppose there is an algorithm $\mathcal{A}$ that determines the minimum value of $\operatorname{trace}(\Sigma(z))$ over all z satisfying $r^T z \leq p$ (recall that $\operatorname{trace}(\Sigma(z)) = \infty$ for any selection z that is not feasible). If the output of algorithm $\mathcal{A}$ is not ∞, then the solution of the specified KFSS problem with parameter p is feasible; by Lemma 1, we know that the solution to the minimum detectability problem (i.e., the minimum number of nonzero entries of the diagonal matrix

[1] Note that the sensor information matrix is different from the *Fisher information matrix*, which is the inverse of the error covariance matrix [10].

$M \in \mathbb{R}^{n\times n}$ such that (A, M) is detectable) is at most p. In order to solve the minimum detectability problem, we need to call algorithm $\mathcal{A}$ at most n times (i.e., increase p from 1 to n). Thus, if the minimum detectability problem is NP-hard, then the KFSS problem is also NP-hard.

The NP-hardness of the minimum detectability problem follows from the proof of NP-hardness of the minimum controllability problem in [12]. Specifically, given $n_1, n_2 \in \mathbb{Z}_{\geq 1}$ and a collection $\mathcal{C}$ of n_1 nonempty subsets of $\{1, \cdots, n_2\}$, let $A(\mathcal{C}) = U^{-1}\operatorname{diag}(1, \cdots, n_1 + n_2 + 1)U$, where U is some invertible matrix related to $\mathcal{C}$.[2] In [12], the author proved that $\mathcal{C}$ has a hitting set with cardinality s if and only if there exists a diagonal matrix B with no more than s nonzero entries such that $(A(\mathcal{C}), B)$ is controllable. Since the hitting set problem is NP-hard, the minimum controllability problem is also NP-hard. Note that the set of eigenvalues of $A(\mathcal{C})$ is $\{1, \cdots, n_1 + n_2 + 1\}$, which are all unstable. Thus, to find a matrix B such that $(A(\mathcal{C}), B)$ is stabilizable is equivalent to finding a matrix B such that $(A(\mathcal{C}), B)$ is controllable, which implies that the minimum stabilizability problem is NP-hard. By the duality of stabilizability and detectability, the minimum detectability problem is also NP-hard, completing the proof. ■

Note that the above result shows that it is NP-hard to find a feasible solution for the KFSS problem, even when all sensors have identical costs. The following result shows that the KFSS problem is still NP-hard if the matrix A is stable (so that *all* sensor selections are feasible), but when the sensor costs can be arbitrary.

Theorem 2: The KFSS problem is NP-hard even under the additional assumption that the matrix A is stable. □

Proof: We show the result by giving the reduction from the optimization form of the 0-1 knapsack problem to the KFSS problem. First note that if the matrices A, C, W, V are all diagonal, then the resulting steady state error covariance Σ is also diagonal. Specifically, when $A = aI_{n\times n}$ with $0 < a < 1$ being some constant, $C = I_{n\times n}$, $V = 0$, and $W = \operatorname{diag}([w_1 \cdots w_n])$, we know that Σ is diagonal with $\Sigma_{ii} = w_i$ if $z_i = 1$ and $\Sigma_{ii} = \frac{w_i}{1-a^2}$ if $z_i = 0$. Thus, the reduction of estimation error by adding sensor i is $\Sigma_i(z_i = 0) - \Sigma_i(z_i = 1) = \frac{a^2}{1-a^2}w_i, \forall i$. Given the number of items n, the set of values $\{\alpha_i\}$, the set of weights $\{\beta_i\}$ and the weight budget B for the 0-1 knapsack problem, the corresponding instance for the KFSS problem is the *stable* system matrix $A = \frac{1}{2}I_{n\times n}$ (i.e., take the constant $a = \frac{1}{2}$), the set $\mathcal{Q}$ of n sensors with the measurement matrix $C = I_{n\times n}$, the system noise covariance matrix $W = \operatorname{diag}([w_1 \cdots w_n])$ with $w_i = \frac{1-a^2}{a^2}\alpha_i = 3\alpha_i$, the sensor noise covariance matrix $V = 0$, the cost vector $r = [\beta_1 \cdots \beta_n]^T$ and the budget $\beta = B$. Then we can see that an indicator vector z for the 0-1 knapsack problem is optimal if and only if it is optimal for the corresponding KFSS problem. Since the optimization form of the 0-1 knapsack problem is NP-hard, the KFSS problem is NP-hard even under the additional assumption that the matrix A is stable. ■

[2]Note that U is constructed based on the incidence matrix of $\mathcal{C}$; we omit the construction details and refer to the proof of Theorem 1.1 in [12].

In the rest of this paper, we focus on the case where the matrix A is stable in order to obtain further insights.

IV. Upper Bounds on the Performance of Sensor Selection Algorithms

In this section, we study worst-case bounds on the performance of sensor selection algorithms to solve the KFSS problem. Specifically, we consider the ratio $r(\Sigma) \triangleq \frac{\operatorname{trace}(\Sigma_{\text{worst}})}{\operatorname{trace}(\Sigma_{\text{opt}})}$, where Σ_{opt} and Σ_{worst} are the solutions of the DARE corresponding to the optimal selection of sensors and the worst-case selection, respectively. Note that for any sensor selection algorithm, the performance of that algorithm is within $r(\Sigma)$ times the optimal performance. In other words, the quantity $r(\Sigma)$ characterizes the 'spectrum' of the performance of all feasible selections. We will be using the following results.

Lemma 2 ([13]): For matrices $M, N \in \mathbb{R}^{n\times n}$ with M symmetric and $N \succeq 0$, we have $\lambda_n(M)\operatorname{trace}(N) \leq \operatorname{trace}(MN) \leq \lambda_1(M)\operatorname{trace}(N)$.

Lemma 3 ([14]): A square matrix A is Schur stable if and only if there exists a nonsingular matrix P such that $\sigma_1(P^{-1}AP) < 1$. Moreover, the condition number $\kappa = \frac{\sigma_1(P)}{\sigma_n(P)}$ of P is upper bounded by some constant γ_A which only depends on the matrix A.

Lemma 4 ([15]): For $\Sigma \succeq 0$ satisfying the DARE (5) with $W \succ 0$, we have $\Sigma \succeq A^T(W^{-1} + R(z))^{-1}A + W$.

We now give an upper bound for $r(\Sigma)$. To incorporate the nature of the sensor set $\mathcal{Q}$, we will use the sensor information matrix $R(z)$ from (5) which encapsulates both the sensor matrix $\tilde{C}$ and the corresponding noise covariance $\tilde{V}$ corresponding to the indicator vector z.

Theorem 3: For the given cost vector r and budget β, let $\mathcal{R} = \{R(z)\}$ be the set of all sensor information matrices such that the constraint $r^T z \leq \beta$ is satisfied. Denote $\lambda_1^{\max} = \max\{\lambda_1(R) | R \in \mathcal{R}\}$. Then for the stable system (1) with $W \succ 0$, the following bound for $r(\Sigma)$ holds:

$$r(\Sigma) \leq \frac{\alpha_A(1 + \lambda_1^{\max})\operatorname{trace}(W)}{n\sigma_n^2(A)\lambda_n(W) + (1 + \lambda_1^{\max})\operatorname{trace}(W)}, \quad (7)$$

where α_A is some constant that only depends on A. □

Proof: Note that since A is stable, all indicator vectors $z \in \{0,1\}^q$ are feasible (in the sense of Definition 1). We first give an upper bound for $\operatorname{trace}(\Sigma_{\text{worst}})$. Consider the case where $z = \mathbf{0}$ (i.e., no sensors are chosen). In this case, the DARE (4) becomes the Lyapunov equation $\Sigma(\mathbf{0}) = A^T\Sigma(\mathbf{0})A + W$. Define $\bar{\Sigma} = P^T\Sigma(\mathbf{0})P$ and $\bar{W} = P^TWP$, where P is nonsingular and satisfies $\sigma_1(P^{-1}AP) < 1$. Note that since the matrix A is stable, by Lemma 3, such a matrix P always exists. Let $D = P^{-1}AP$. Then we get $\bar{\Sigma} = D^T\bar{\Sigma}D + \bar{W}$. By Lemma 2, we know that $\operatorname{trace}(D^T\bar{\Sigma}D) = \operatorname{trace}(DD^T\bar{\Sigma}) \leq \sigma_1^2(D)\operatorname{trace}(\bar{\Sigma})$ and thus $\operatorname{trace}(\bar{\Sigma}) \leq \frac{\operatorname{trace}(\bar{W})}{1-\sigma_1^2(D)}$. Since the matrix PP^T is symmetric (and positive definite) and $W, \Sigma \succeq 0$, by Lemma 2, we know that $\operatorname{trace}(\bar{\Sigma}) = \operatorname{trace}(PP^T\Sigma(\mathbf{0})) \geq \sigma_n^2(P)\operatorname{trace}(\Sigma(\mathbf{0}))$ and $\operatorname{trace}(\bar{W}) = \operatorname{trace}(PP^TW) \leq \sigma_1^2(P)\operatorname{trace}(W)$. Combining the above analysis, we obtain

$$\begin{aligned}\operatorname{trace}(\Sigma_{\text{worst}}) \leq \operatorname{trace}(\Sigma(\mathbf{0})) &\leq \frac{\sigma_1^2(P)}{\sigma_n^2(P)} \frac{\operatorname{trace}(W)}{1-\sigma_1^2(D)} \\ &\leq \frac{\gamma_A^2 \operatorname{trace}(W)}{1-\sigma_1^2(D)} \triangleq \alpha_A \operatorname{trace}(W),\end{aligned}$$

where the second last inequality is due to Lemma 3.

Next we derive a lower bound for $\operatorname{trace}(\Sigma_{\text{opt}})$. For any selection of sensors z, since $W \succ 0$ and $R(z) \succeq 0$, there exists a matrix U that can diagonalize W^{-1} and $R(z)$ simultaneously, i.e., there exists a matrix U such that $W^{-1} = U^{-1}(U^{-1})^T$ and $R(z) = U^{-1}\Lambda(U^{-1})^T$, where $\Lambda = \operatorname{diag}([\lambda_1(R(z)) \cdots \lambda_n(R(z))])$ [11]. Then we have

$$\begin{aligned}\operatorname{trace}(\Sigma(z)) &\geq \operatorname{trace}(A^T(W^{-1}+R(z))^{-1}A + W) \\ &\geq \sigma_n^2(A)\operatorname{trace}((W^{-1}+R(z))^{-1}) + \operatorname{trace}(W) \\ &= \sigma_n^2(A)\operatorname{trace}(U^T(I+\Lambda)^{-1}U) + \operatorname{trace}(W) \\ &\geq \sigma_n^2(A)\sigma_n^2(U)\sum_i \frac{1}{1+\lambda_i(R(z))} + \operatorname{trace}(W) \\ &\geq \frac{n\sigma_n^2(A)\lambda_n(W) + (1+\lambda_1^{\max})\operatorname{trace}(W)}{1+\lambda_1^{\max}}.\end{aligned}$$

Note that the first inequality is due to Lemma 4 and the second and fourth inequalities are due to Lemma 2. Further note that the derived lower bound for $\operatorname{trace}(\Sigma(z))$ holds for any sensor selection z and thus holds for $\operatorname{trace}(\Sigma_{\text{opt}})$.

The result follows by combining the upper bound for $\operatorname{trace}(\Sigma_{\text{worst}})$ and the lower bound for $\operatorname{trace}(\Sigma_{\text{opt}})$. ■

Using the result in Theorem 3, we can provide the following simpler (and looser) upper bound for $r(\Sigma)$ which highlights the role of the system dynamics matrix A.

Corollary 1: If the given system (1) is stable, there exists a constant α_A which only depends on the matrix A such that $r(\Sigma) \leq \alpha_A$.

Remark 1: Note that since the upper bound for $r(\Sigma)$ only depends on the system matrix A, no sensor selection algorithm will provide arbitrarily bad performance as long as A is well conditioned, regardless of the statistics of the noise processes and the nature of the sensor set $\mathcal{Q}$. Further note that if A is normal (i.e., $A^TA = AA^T$), then we can always choose P such that $\gamma_A = 1$ and thus $r(\Sigma) \leq \frac{1}{1-\lambda_1^2(A)}$; in other words, the ratio $r(\Sigma)$ only depends on the degree of stability of the system. In particular, if $A = 0$, then the state x_{k+1} in (1) is uncorrelated with x_k, and thus measurements of the current state are not useful in predicting the next state. This is corroborated by the fact that $r(\Sigma) = 1$ in this case. □

V. Greedy Algorithms

In this section, we explore potential approximation algorithms to solve the KFSS problem. We focus on the case where $r = [1 \cdots 1]^T$ and $\beta = p$ for some $p \in \mathbb{N}$ (i.e., our goal is to choose p sensors out of the total q sensors to optimize the performance of the Kalman filter). In [3], the authors showed that the cost function associated with the single-step sensor scheduling problem is submodular and thus the greedy algorithm provides a near optimal solution. Thus, in this section, we study a DARE based greedy algorithm for sensor selection, given as Algorithm 1, which iteratively picks sensors that provide the largest incremental decrease in the steady state error covariance.

Algorithm 1 DARE based Greedy Algorithm

Input: Matrices A, C, W, V and constant p
Output: A set $\mathcal{S}$ of chosen sensors
$k \leftarrow 0$, $\mathcal{S} \leftarrow \emptyset$
for $k \leq p$ **do**
  for $i \in \mathcal{Q} \cap \bar{\mathcal{S}}$ **do**
    Solve $\operatorname{trace}(\Sigma_{\mathcal{S}}^i) = \operatorname{trace}(\Sigma(\mathcal{S} \cup \{i\}))$
  end for
  Choose j with $\operatorname{trace}(\Sigma_{\mathcal{S}}^j) = \min_i \operatorname{trace}(\Sigma_{\mathcal{S}}^i)$
  $\mathcal{S} \leftarrow \mathcal{S} \cup \{j\}$, $\mathcal{Q} \leftarrow \mathcal{Q} \setminus \{j\}$, $k \leftarrow k+1$
end for

Note that it is easy to show that for the case where the matrices A, C, W, V are all diagonal, Algorithm 1 outputs the optimal solution. Outside of these trivial cases, there are few tools available to give performance guarantees on greedy algorithms. One such tool is the concept of submodularity, which has been used in the analysis of greedy algorithms for the sensor scheduling problem, as mentioned above. Here, we briefly review this concept (see [16] for a comprehensive discussion) and show that the trace of the steady state error covariance (and other related metrics) does not satisfy this property in general.

Definition 2: Let E be a finite set and define the set function $f : 2^E \rightarrow \mathbb{R}$. The set function f is normalized if $f(\mathbf{0}) = 0$, and is monotone if for every $X \subseteq Y \subseteq E$, $f(X) \leq f(Y)$. □

Definition 3 (Submodularity): The set function f is submodular if for all $X, Y \subseteq E$, $f(X) + f(Y) \geq f(X \cap Y) + f(X \cup Y)$, is supermodular if $-f$ is submodular, and is modular if it is both submodular and supermodular. □

Theorem 4 ([16]): If the cost function f to be maximized is normalized, monotone and submodular, then the performance of the greedy algorithm is within a factor of $1 - \frac{1}{e}$ of the optimal. □

A. Lack of Submodularity of the Cost Function

We will consider the problem of *maximizing* three different performance metrics: $F_1(z) = -\operatorname{trace}(\Sigma(z)), F_2(z) = \log\det(\Sigma^{-1}(z))$ and $F_3(z) = -\max \lambda_i(\Sigma(z))$, where F_2 captures the volume of the confidence ellipsoid and F_3 captures the worst-case error covariance.[3] Note that maximizing F_1 is equivalent to minimizing $-F_1$ as in the KFSS problem. In [3], the authors showed that the metric F_2 is submodular for the single-step sensor scheduling problem while F_1 and F_3 are neither submodular nor supermodular. One question of interest is whether any of these metrics is submodular

[3]Note that in order to apply the result in Theorem 4, these metrics need to be normalized; we omit the normalization factor as it does not influence the results in this section.

or supermodular for the KFSS problem (where $\Sigma(z)$ is the solution of the DARE) when the matrix A is stable. However, the following counterexample shows that these metrics are neither supermodular nor submodular in general.

Example 1: Consider an instance of the KFSS problem with $A = \begin{bmatrix} 0.3 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}, C = \begin{bmatrix} 1 & 0.5 & 0.7 & 0 \\ 0 & 0.5 & 0.3 & 0.7 \end{bmatrix}^T, W = I_{2\times 2}, V = I_{4\times 4}$. Note that $\mathrm{eig}(A) = \{0.1298, 0.7702\}$ and thus A is stable. For metric F_i and two sets X, Y, let the change of utility by adding Y to X be $\Delta_{F_i}(Y|X)$, i.e., $\Delta_{F_i}(Y|X) = F_i(X \cup Y) - F_i(X)$. One can check that $\Delta_{F_i}(\{1\}|\{2,3\}) < \Delta_{F_i}(\{1\}|\{2,3,4\})$ and $\Delta_{F_i}(\{1\}|\{2\}) > \Delta_{F_i}(\{1\}|\{2,3\})$, $i \in \{1,2,3\}$, which contradict the submodularity and supermodularity of the corresponding metrics, respectively. □

The above negative result implies that one may not be able to use classical results from combinatorial optimization to analyze Algorithm 1; despite this, our simulations in Section VI show that this algorithm performs well in practice.

B. Approximated KFSS Problem

Note that due to the nonlinear nature of the DARE, it is in general difficult to obtain its analytical solution. Thus, in this subsection, we will consider an approximation for the cost function in the KFSS problem and explore its structural properties. Specifically, the solution of the DARE can be approximated by the solutions of two Lyapunov equations as follows.

Lemma 5 ([17]): Let P_w and $P_v(z)$ satisfy the Lyapunov equations $P_w = A^T P_w A + W$ and $P_v(z) = A^T P_v(z) A + R(z)$, respectively. For $\Sigma(z) \succeq 0$ satisfying the DARE (4), we have $(P_v(z) + P_w^{-1})^{-1} \preceq \Sigma(z) \preceq P_v(z)^{-1} + P_w$.

As a heuristic, we consider the problem of minimizing the upper bound $\mathrm{trace}(P_v(z)^{-1} + P_w)$ on $\mathrm{trace}(\Sigma)$ (rather than $\mathrm{trace}(\Sigma)$ itself). Since the term P_w does not depend on the specific selection of sensors, we further approximate the problem by seeking to maximize $\mathrm{trace}(P_v(z))$.

Problem 3 (Approximated KFSS Problem): Given a system matrix $A \in \mathbb{R}^{n\times n}$, a measurement matrix $C \in \mathbb{R}^{q\times n}$, a positive semidefinite system noise covariance matrix $W \in \mathbb{R}^{n\times n}$, a positive semidefinite sensor noise covariance matrix $V \in \mathbb{R}^{q\times q}$ and the number of sensors to be chosen p, the approximated KFSS problem is to solve the following optimization problem:

$$\begin{aligned} \text{maximize} \quad & \mathrm{trace}(P_v(z)) \\ \text{subject to} \quad & 1^T z \le p \\ & z \in \{0,1\}^q \end{aligned} \tag{8}$$

where $P_v(z)$ is defined in Lemma 5.

Here we propose the following Lyapunov equation based greedy algorithm, given as Algorithm 2. We show that the cost function of the approximated KFSS problem is modular when the covariance matrix V is diagonal and thus Algorithm 2 provides near optimal performance for the approximated KFSS problem in this case. Note that the proof of Theorem 5 is a relatively straightforward extension of the results in [5] on the continuous-time Lyapunov equation to the discrete-time case.

Algorithm 2 Lyapunov Equation based Greedy Algorithm

Input: Matrices A, C, W, V and constant p
Output: A set $\mathcal{S}$ of chosen sensors
$k \leftarrow 0$, $\mathcal{S} \leftarrow \emptyset$
for $k \le p$ **do**
  for $i \in \mathcal{Q} \cap \bar{\mathcal{S}}$ **do**
    Solve $\mathrm{trace}(P_v^i) = \mathrm{trace}(P_v(\mathcal{S} \cup \{i\}))$
  end for
  Choose j with $\mathrm{trace}(P_v^j) = \max_i \mathrm{trace}(P_v^i)$
  $\mathcal{S} \leftarrow \mathcal{S} \cup \{j\}$, $\mathcal{Q} \leftarrow \mathcal{Q} \setminus \{j\}$, $k \leftarrow k+1$
end for

Lemma 6 ([16]): A set function $f : 2^{\mathcal{Q}} \to \mathbb{R}$ is modular if $\forall \mathcal{S} \subset \mathcal{Q}$, $f(\mathcal{S}) = f(\mathbf{0}) + \sum_{s\in\mathcal{S}} f(s)$.

Theorem 5: Let $R_{\mathcal{S}} = C_{\mathcal{S}}^T V_{\mathcal{S}}^{-1} C_{\mathcal{S}}$ be the sensor information matrix associated with the set $\mathcal{S}$ of chosen sensors and let $P_v(\mathcal{S})$ be the solution of $P_v = A^T P_v A + R_{\mathcal{S}}$. Define $\mathrm{trace}(P_v(\mathbf{0})) = 0$. If the matrix A is stable and the covariance matrix V is diagonal and $V \succ 0$, then the function $\mathrm{trace}(P_v(\mathcal{S}))$ is normalized, monotone and modular. □

Proof: For any $\mathcal{S} \subset \mathcal{Q}$, since the covariance matrix V is diagonal and $V \succ 0$, we know that $R_{\mathcal{S}}$ can be decomposed as $R_{\mathcal{S}} = \sum_{s\in\mathcal{S}} \frac{1}{v_s} C_s^T C_s$ where C_s and v_s are the row of C and measurement noise covariance corresponding to sensor s, respectively. Thus, we get

$$\begin{aligned} P_v(\mathcal{S}) &= \sum_{k=0}^{\infty} A^k C_{\mathcal{S}}^T V_{\mathcal{S}}^{-1} C_{\mathcal{S}} (A^T)^k \\ &= \sum_{s\in\mathcal{S}} \sum_{k=0}^{\infty} A^k \frac{1}{v_s} C_s^T C_s (A^T)^k = \sum_{s\in\mathcal{S}} P_v(s). \end{aligned}$$

Note that the first equality is due to the analytic solution of the discrete-time Lyapunov equation [18]. Since the trace operator is a linear function and $\mathrm{trace}(P_v(\mathbf{0})) = 0$, by using Lemma 6, we know that the function $\mathrm{trace}(P_v(\mathcal{S}))$ is normalized, monotone and modular. ■

Corollary 2: Let the set of sensors chosen by Algorithm 2 be $\mathcal{S}$ and the optimal solution of the approximated KFSS problem be $\mathrm{trace}(P_v^{\mathrm{opt}})$. If the covariance matrix V is diagonal and $V \succ 0$, then $\mathrm{trace}(P_v(\mathcal{S})) \ge (1 - \frac{1}{e}) \,\mathrm{trace}(P_v^{\mathrm{opt}})$.

VI. Simulation

In this section, we provide simulation results for the performance of the DARE based greedy algorithm (Algorithm 1) and the Lyapunov equation based greedy algorithm (Algorithm 2) and discuss their complexity.

A. Performance Evaluation

In order to illustrate the performance of algorithms considered in this paper, we use another sensor selection strategy, the semi-definite programming (SDP) approach [19], as a benchmark. The objective of the SDP approach is to minimize the H_2 norm from the noise to the estimation error and sparsity is achieved by adding a penalty function; see

TABLE I

PERFORMANCE COMPARISON OF DIFFERENT ALGORITHMS OVER 25 RANDOMLY GENERATED SYSTEMS. FOR ALGORITHM A, THE TABLE PRESENTS THE AVERAGE, VARIANCE AND WORST-CASE VALUES OF $\frac{\text{trace}(\Sigma_A)}{\text{trace}(\Sigma_{\text{OPT}})}$ OVER THE 25 RUNS.

	Average	Variance	Worst-case
Algorithm 1	2.5	6.1	11.8
Algorithm 2	7	22.8	22.3
SDP	5	20	18.5

TABLE II

COMPARISON OF THE AVERAGE RUNNING TIME (IN SECONDS) OVER 25 RUNS.

	$n = 50$	$n = 100$	$n = 150$	$n = 200$
Algorithm 1	3.7	18.3	83.5	166.6
Algorithm 2	0.2	0.8	2.2	6.2

[19] for more details. We randomly generate 25 systems all having dimension 10 (i.e., $n = 10$). For each system, the goal is to choose 5 sensors out of a total of 20 (i.e., $q = 20$, $p = 5$, and $r = [1 \cdots 1]^T$). The results are summarized in Table I; note that the value is the ratio of the performance of each algorithm over the optimal solution (found by brute-force). From Table I, we can see that in general, Algorithm 1 outperforms the other two algorithms, and Algorithm 2 (which attempts to minimize an upper bound on $\text{trace}(\Sigma)$) generally performs the worst. However, we will see in the next subsection that the benefit of Algorithm 2 is that its run-time scales more slowly with the number of states than those of the other algorithms.

B. Complexity Analysis

Note that the complexity of solving the DARE and the Lyapunov equation is $O(n^5)$ and $O(n^3)$, respectively, where n is the number of states [20]. If we aim to choose p sensors from a set of q sensors, then the complexity of Algorithm 1 and Algorithm 2 are $O(pqn^5)$ and $O(pqn^3)$, respectively. Thus, the running time of the Lyapunov equation based greedy algorithm scales more slowly than the DARE based greedy algorithm as n grows. For example, see Table II for a simulation which supports this conclusion; the simulation is conducted on a typical 2.4-GHz personal computer and the goal is also to choose 5 sensors out of 20. Note that in Table II, we do not include the running time of the SDP approach, as we found our solver ran out of memory when the number of states n exceeded 50. As argued in [19], when the weight of the sparsity penalty function is *fixed*, the complexity of the SDP approach is $O((n+q)^6)$ (recall that q is the total number of sensors), which is worse than the DARE based greedy algorithm. Moreover, the process of choosing an appropriate weight for the sparsity penalty function (in order to obtain the desired level of sparsity) requires additional computation.

VII. CONCLUSION

In this paper, we studied the KFSS problem for linear dynamical systems. We showed that this problem is NP-hard (even under the assumption that the system is stable) and provided upper bounds for the performance of the worst-case selection of sensors and highlighted the factors that dominate the worst-case performance. We then studied two greedy algorithms for sensor selection. For the DARE based greedy algorithm, we provided a negative result showing that the corresponding cost function is neither supermodular nor submodular; however, simulations indicate that this algorithm performs well in practice. For the Lyapunov equation based greedy algorithm, we showed that this algorithm has guaranteed performance with respect to the relaxed cost function; although this algorithm performs less well than the original algorithm, the run-time scales better with the system size. Since the DARE based greedy algorithm outperforms other algorithms in general, further study on determining provable bounds on this algorithm is of interest.

REFERENCES

[1] M. Van De Wal and B. De Jager, "A review of methods for input/output selection," *Automatica*, vol. 37, no. 4, pp. 487–510, 2001.

[2] 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.

[3] S. T. Jawaid and S. L. Smith, "On the submodularity of sensor scheduling for estimation of linear dynamical systems," in *Proc. of American Control Conference*, 2014, pp. 4139–4144.

[4] S. Pequito, S. Kar, and A. Aguiar, "A structured systems approach for optimal actuator-sensor placement in linear time-invariant systems," in *Proc. of American Control Conference*, 2013, pp. 6108–6113.

[5] T. H. Summers and J. Lygeros, "Optimal sensor and actuator placement in complex dynamical networks," in *Proc. of IFAC World Congress*, 2014, pp. 3784–3789.

[6] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies," *The Journal of Machine Learning Research*, vol. 9, pp. 235–284, 2008.

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

[8] Y. Wang, M. Sznaier, and F. Dabbene, "A convex optimization approach to worst-case optimal sensor selection," in *Proc. of 52nd IEEE Conference on Decision and Control*, 2013, pp. 6353–6358.

[9] M. F. Huber, "Optimal pruning for multi-step sensor scheduling," *IEEE Trans. on Automatic Control*, vol. 57, no. 5, pp. 1338–1343, 2012.

[10] B. Anderson and J. Moore, *Optimal Filtering*. Dover Books, 1979.

[11] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Camb. Univ., 2012.

[12] A. Olshevsky, "Minimal controllability problems," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 3, pp. 249–258, 2014.

[13] F. Zhang and Q. Zhang, "Eigenvalue inequalities for matrix product," *IEEE Transactions on Automatic Control*, vol. 51, no. 9, pp. 1506–1509, 2006.

[14] J. Liu and J. Zhang, "The open question of the relation between square matrix's eigenvalues and its similarity matrix's singular values in linear discrete system," *International Journal of Control, Automation and Systems*, vol. 9, no. 6, pp. 1235–1241, 2011.

[15] N. Komaroff, "Iterative matrix bounds and computational solutions to the discrete algebraic Riccati equation," *IEEE Transactions on Automatic Control*, vol. 39, no. 8, pp. 1676–1678, 1994.

[16] L. Lovász, "Submodular functions and convexity," in *Mathematical Programming The State of the Art*. Springer, 1983, pp. 235–257.

[17] R. S. Bucy and P. D. Joseph, *Filtering for Stochastic Processes with Applications to Guidance*. Americ. Math. Soc., 1987, vol. 326.

[18] J. P. Hespanha, *Linear Systems Theory*. Princeton University Press, 2009.

[19] N. K. Dhingra, M. R. Jovanović, and Z. Q. Luo, "An ADMM algorithm for optimal sensor and actuator selection," in *Proc. of 53rd IEEE Conference on Decision and Control*, 2014, pp. 4039–4044.

[20] P. Benner, J.-R. Li, and T. Penzl, "Numerical solution of large-scale Lyapunov equations, Riccati equations, and linear-quadratic optimal control problems," *Numerical Linear Algebra with Applications*, vol. 15, no. 9, pp. 755–777, 2008.