

Greedy sensor selection: leveraging submodularity

Manohar Shamaiah Siddhartha Banerjee Haris Vikalo
The University of Texas at Austin
Electrical and Computer Engineering, 2501 Speedway, Austin, TX 78712-0240

I. ABSTRACT

We consider the problem of sensor selection in resource constrained sensor networks. The fusion center selects a subset of k sensors from an available pool of m sensors according to the maximum a posteriori or the maximum likelihood rule. We cast the sensor selection problem as the maximization of a submodular function over uniform matroids, and demonstrate that a greedy sensor selection algorithm achieves performance within $(1 - \frac{1}{e})$ of the optimal solution. The greedy algorithm has a complexity of $O(n^3mk)$, where n is the dimension of the measurement space. The complexity of the algorithm is further reduced to $O(n^2mk)$ by exploiting certain structural features of the problem. An application to the sensor selection in linear dynamical systems where the fusion center employs Kalman filtering for state estimation is considered. Simulation results demonstrate the superior performance of the greedy sensor selection algorithm over competing techniques based on convex relaxation.

Index Terms: Submodular functions, Kalman filter

II. INTRODUCTION

Sensor networks have attracted much attention in recent years [1], [2], [3]. A typical feature of such networks is the presence of one or more fusion centers, which aggregate the information from different sensors. Due to various practical considerations, the sensors are generally resource constrained, and hence their communication with the fusion center is limited. To this end, the fusion center schedules only a subset of the available sensors for transmission in each time slot. Different performance metrics lead to various sensor selection problem formulations. Regardless of the formulation, however, sensor selection is essentially combinatorial in nature. Hence finding the optimal solution is generally computationally intensive, leading to a number of heuristics and approximate algorithms.

In this paper, we consider the scenario where the fusion center desires the maximum likelihood (ML) or maximum a posteriori (MAP) estimate of the unknown vector. Only a subset of the sensors which acquire the measurement is allowed to transmit its measurement to the fusion center. Due to the high complexity of obtaining the exact solution, Joshi and Boyd [1] propose a heuristic approach to the problem based on convex relaxation. However, the algorithm proposed in [1] has no guarantees on the performance; moreover, its complexity is cubic in the total number of available sensors, which is prohibitive for large networks. In our work, we show that certain structural properties of the sensor selection problem allow recasting it as a maximization of submodular functions over uniform matroids. This allows us to leverage the results of

Nemhauser and Wolsey [4], who show that for problems with such a structure, a greedy algorithm results in an approximate solution with a performance within a $(1 - \frac{1}{e})$ of the optimal. In addition to the performance guarantees, the greedy algorithm is significantly more computationally efficient than the convex relaxation based approach. In particular, the complexity of the greedy algorithm is $O(n^3mk)$, i.e, it scales linearly with the total number of sensors. We further exploit the structure in the problem to obtain a simplified greedy algorithm with a complexity of $O(n^2mk)$.

A related sensor selection problem arises in state estimation of linear dynamical systems [5], [6]. There, the fusion center employs Kalman filter for state estimation, and the sensor selection problem is concerned with optimizing a metric related to the error covariance matrix in the Kalman recursions. Recently, convex optimization based heuristics were proposed to solve this problem [5]. As an alternative, we employ the aforementioned greedy algorithm and demonstrate that it significantly outperforms the convex optimization based solution. Moreover, by exploiting the connection between the proposed algorithm and sequential processing implementation of Kalman filter [7], we achieve significant complexity reduction. Interestingly, the greedy algorithm has an intuitive interpretation as a procedure which at each step chooses the set of observations in the direction closest to the maximum error.

The paper is organized as follows. The system model and the sensor selection problem formulation are given in Section III. Section IV provides a brief overview of submodular functions and uniform matroids and then presents our analytical and simulation results. Sensor selection for estimation in linear dynamical system using Kalman filtering is presented in Section V, followed by the simulation results. Finally, we conclude the paper in Section VI.

III. SYSTEM MODEL AND SENSOR SELECTION PROBLEM

Consider the problem of estimating a vector $x \in R^n$ from k linear measurements, corrupted by additive noise. The k measurements are to be selected from the set of m measurements, each acquired at one of the sensors in a network. The i^{th} measurement is given by

$$y_i = a'_i x + n_i \quad (1)$$

where a_i denotes the i^{th} measurement vector and n_i represents the zero-mean Gaussian noise with variance σ^2 . Following [1], we assume that the fusion center knows the measurement vectors a_i .

A. Maximum a posteriori criterion

Suppose the prior density of x is $\mathcal{N}(0, \Sigma_x)$. The maximum a posteriori probability (MAP) estimate of x is given by [8]

$$\hat{x}_{MAP} = \left(\sigma^{-2} \sum_{i \in S} a_i a_i' + \Sigma_x^{-1} \right)^{-1} \sum_{i \in S} y_i a_i, \quad (2)$$

where S denotes the set of selected measurements. The estimation error covariance matrix is given by

$$\Sigma_{map}(S) = \left(\sigma^{-2} \sum_{i \in S} a_i a_i' + \Sigma_x^{-1} \right)^{-1}. \quad (3)$$

An often used scalar measure of the quality of estimation is based on the volume of η confidence ellipsoid or its mean radius [8]. Both of these are functions of $\log \det(\Sigma_{map}(S))$. The sensor selection can then be posed as the problem of selecting a subset of $k \geq n$ sensors from the set of m sensors, such that the log volume (or the mean radius) of the confidence ellipsoid is minimized. This can be expressed as the following optimization problem,

$$\begin{aligned} \max & \quad \log \det \left(\sigma^{-2} \sum_{i \in S} a_i a_i' + \Sigma_x^{-1} \right), \\ \text{subject to} & \quad |S| = k. \end{aligned} \quad (4)$$

where $|S|$ denotes the cardinality of the set S .

By introducing binary variables $z_i \in \{0, 1\}$ to indicate the membership of each sensor in the selected subset, (4) can be written as

$$\begin{aligned} \max & \quad \log \det \left(\sigma^{-2} \sum_{i=1}^m z_i a_i a_i' + \Sigma_x^{-1} \right), \\ \text{subject to} & \quad z_i \in \{0, 1\}, \quad \sum_{i=1}^m z_i = k. \end{aligned} \quad (5)$$

The convex relaxation of the above optimization problem is given by

$$\begin{aligned} \max & \quad \log \det \left(\sigma^{-2} \sum_{i=1}^m z_i a_i a_i' + \Sigma_x^{-1} \right), \\ \text{subject to} & \quad 0 \leq z_i \leq 1, \quad i = 1, 2, \dots, m \\ & \quad \sum_{i=1}^m z_i = k \end{aligned} \quad (6)$$

In [1], (6) was solved by reformulating it as a semi-definite program (SDP). The solution to the SDP may take fractional values, in which case some kind of sorting and rounding need to be employed in order to obtain the desired solution. The complexity of the SDP algorithm scales as $O(m^3)$.

B. Maximum-Likelihood criterion

The maximum-likelihood (ML) criterion leads to the estimator of the form

$$\hat{x}_{ML} = \left(\sigma^{-2} \sum_{i \in S} a_i a_i' \right)^{-1} \sum_{i \in S} y_i a_i, \quad (7)$$

with the corresponding error covariance given by

$$\Sigma_{ml}(S) = \sigma^2 \left(\sum_{i \in S} a_i a_i' \right)^{-1} \quad (8)$$

The corresponding sensor selection problem is formulated as the following optimization,

$$\begin{aligned} \max & \quad \log \det \left(\sum_{i=1}^m z_i a_i a_i' \right), \\ \text{subject to} & \quad z_i \in \{0, 1\} \\ & \quad \sum_{i=1}^m z_i = k. \end{aligned} \quad (9)$$

Its convex relaxation leads to the following optimization,

$$\begin{aligned} \max & \quad \log \det \left(\sum_{i \in S} z_i a_i a_i' \right), \\ \text{subject to} & \quad 0 \leq z_i \leq 1, \quad i = 1, 2, \dots, m \\ & \quad \sum_{i=1}^m z_i = k. \end{aligned} \quad (10)$$

Note that the ML estimation is as same as the MAP estimation in the limit of $\Sigma_x = vI, v \rightarrow \infty$ (compare formulations (5) and (9)). Hence we approximate the ML estimation as the following optimization problem,

$$\begin{aligned} \max & \quad \log \det \left(\sum_{i=1}^m z_i a_i a_i' + \varepsilon I \right), \\ \text{subject to} & \quad z_i \in \{0, 1\}, \quad \sum_{i=1}^m z_i = k, \end{aligned} \quad (11)$$

where $\varepsilon > 0$ is chosen to be a very small constant. The reason for this modification is that a non-zero ε , albeit very small, ensures applicability of the submodular approach to solving (11), which we describe next.

IV. GREEDY SENSOR SELECTION

A. Submodular function maximization over uniform matroid.

In this section we review definitions and results related to submodular functions and matroids [9].

Definition 1 [Submodularity]: Let S be a finite set and 2^S denote power set. A set function $f : 2^S \rightarrow R$ is said to be submodular iff

$$\forall A, B \subseteq S, \quad f(A \cup B) + f(A \cap B) \leq f(A) + f(B).$$

For finite set S this is equivalent to $\forall A \subseteq B \subseteq S, \forall j \in S \setminus B$,

$$f(A + j) - f(A) \geq f(B + j) - f(B) \quad (12)$$

i.e., the function f satisfies the diminishing increments property. The submodular function f is monotone if $f(A) \leq f(B), \forall A \subseteq B$.

Definition 2 [Matroid]: A matroid is a pair $M = (S; \mathcal{I})$ where $\mathcal{I} \subseteq 2^S$ and

- 1) $\forall B \in \mathcal{I}, A \subset B \Rightarrow A \in \mathcal{I}$
- 2) $\forall A, B \in \mathcal{I}, |A| < |B| \Rightarrow \exists x \in B \setminus A, A + x \in \mathcal{I}$

A matroid is an abstraction of combinatorial objects, generating notion of linear independence in vector spaces. If m denotes the cardinality of the set S (i.e., $|S| = m$), then for any integer $k \leq m$, we can define $\mathcal{I} = \{J | J \subseteq S, |J| \leq k\}$; $M = (S; \mathcal{I})$ is called a uniform matroid $M_{m,k}$.

B. The greedy sensor selection algorithm

The optimization problem (4) is of the form $\max\{f(S) \mid |S| \leq k\}$. If $f(S)$ is a monotone submodular function then (4) corresponds to maximization of a submodular function over a uniform matroid constraint [9], [4]. For this problem, it was shown in [4] that a greedy algorithm results in a solution with the objective value within $(1 - \frac{1}{e})$ of the optimal value¹. At each step the greedy algorithm chooses the measurement from the available measurements which maximizes the objective when included with previously chosen measurements. In the following lemma, we show that the objective function of the optimization (4) is submodular and hence the greedy algorithm (formalized as Algorithm 1) results in the guaranteed $(1 - \frac{1}{e})$ optimal solution.

Lemma 1: $f(S) = \log \det(\sigma^{-2} \sum_{i \in S} a_i a_i' + \Sigma_x^{-1})$ is a monotone submodular function, and hence the solution to (4) via Algorithm 1 is $(1 - \frac{1}{e})$ optimal. Moreover, the complexity of Algorithm 1 is $O(n^3 mk)$.

Proof: See Appendix.

In general, $k = O(n)$, i.e., the number of sensors to be selected is of the same order as the dimension of the unknown vector, in which case the complexity of Algorithm 1 is essentially $O(n^4 m)$. If the number of the available measurements m is very large compared to n (i.e., $m > n$), the complexity of Algorithm 1 is significantly lower than the complexity of the convex optimization algorithm (the later is $O(m^3)$). We further reduce the complexity of Algorithm 1 by noting that the step 2 in Algorithm 1 requires computation of the determinant of a rank 1 matrix. This can be simplified as follows. Let $M_s = \sigma^{-2} \sum_{i \in K_s} a_i a_i' + \Sigma_x^{-1}$, then

$$\max_{j \in S^s} \log \det(M_s + a_j a_j') = \max_{j \in S^s} (a_j' M_s^{-1} a_j). \quad (13)$$

Hence, we need to propagate (or store) only M_s^{-1} , which is obtained using the following recursion

$$M_s^{-1} = M_{s-1}^{-1} - \frac{M_{s-1}^{-1} a_{k_{s-1}} a_{k_{s-1}}' M_{s-1}^{-1}}{1 + a_{k_{s-1}}' M_{s-1}^{-1} a_{k_{s-1}}}, \quad (14)$$

where $M_0^{-1} = \Sigma_x$ and $a_{k_{s-1}}$ is the optimal vector chosen by the greedy algorithm at step $s-1$. This modification is summarized as Algorithm 2.

The complexity of Algorithm 2 is $O(n^2 mk)$. First, note that the complexity of computing the quadratic form 13 is $O(n^2)$ and there are $O(m)$ such computations. Next, using 14, we can compute M_s^{-1} in $O(n^2)$ computations (as opposed to $O(n^3)$ for direct matrix inversion). To see this, denote $r_{s-1} = M_{s-1}^{-1} a_{k_{s-1}}$, which implies that $M_{s-1}^{-1} a_{k_{s-1}} a_{k_{s-1}}' M_{s-1}^{-1} = r_{s-1} r_{s-1}'$, a product of rank 1 matrices. The complexity of computing 14 is thus $O(n^2)$, and therefore of selecting each sensor is $O(n^2 m)$. Finally, there are k such iterations. Hence the overall complexity of Algorithm 2 is $O(n^2 mk)$.

C. Simulation results

In this section, we compare the performance of Algorithm 2 and the convex optimization based algorithm employed for

¹This has been recently generalized to arbitrary matroid constraint[9]

Algorithm 1 Basic greedy algorithm

1. *Initialization:*
 $s = 1, K_s = \{\}, S^s = S$
 2. *Determine greedily the next measurement:*
 $k_s = \arg \max_{j \in S^s} \log \det(\sigma^{-2} \sum_{i \in K_s} a_i a_i' + a_j a_j' + \Sigma_x^{-1})$
 3. *Update the measurement set:*
 $S^{s+1} = S^s \setminus k_s, K^{s+1} = K^s \cup k_s, s \leftarrow s + 1$, go to step 2 if $s \leq k$.
-

Algorithm 2 Simplified greedy algorithm

1. *Initialization:*
 $s = 1, K_s = \{\}, S^s = S, M_s = \Sigma_x^{-1}$
 2. *Determine greedily the next measurement:*
 $k_s = \arg \max_{j \in S^s} (a_j' M_s^{-1} a_j)$
 3. *Update the measurement set and M_s :*
 $S^{s+1} = S^s \setminus k_s, K^{s+1} = K^s \cup k_s, s \leftarrow s + 1, M_s^{-1} = M_{s-1}^{-1} - \frac{M_{s-1}^{-1} a_{k_{s-1}} a_{k_{s-1}}' M_{s-1}^{-1}}{1 + a_{k_{s-1}}' M_{s-1}^{-1} a_{k_{s-1}}}$, go to step 2 if $s \leq k$.
-

solving the sensor selection problem (9).

The elements of a_i are generated as iid Gaussian random variables $\mathcal{N}(0, \frac{1}{n})$. To obtain a set of k sensors scheduled for transmission, we solve the modified ML objective (11) with $\varepsilon = 10^{-3}$. The performance of the greedily selected sensors is compared using the actual objective (9) which they induce.

Figure 1 shows objective obtained by the two algorithms for different random realizations of the a_i . Here we set $m = 150$, $n = 20$, and $k = 20$. The objective value of the Algorithm 2 is almost always better than that of the convex optimization algorithm. Figure 2 shows the performance comparison of the average objective versus the dimension of x (n) for a fixed number of sensors $k = 30$. The objective value is averaged over 100 realizations of the measurement set. Figure 2 shows that, as n approaches k , Algorithm 2 outperforms the convex optimization approach. In Figure 3, we plot the average objective as a function of the number of sensors k for $n = 20$. It can be inferred that the Algorithm 2 achieves given value of the objective function with fewer sensors than the convex relaxation approach. However, as the number of sensors selected increases, both the approaches result in the same performance.

V. APPLICATION: SENSOR SELECTION USING KALMAN FILTER AT THE FUSION CENTER

Consider a sensor network in which fusion center selects k sensors for transmitting the measurements of a linear dynamical system. The fusion center employs Kalman filter to track the state of the system. The system model is given by

$$x_{t+1} = A_t x_t + w_t, \quad y_t = S_t H_t x_t + v_t,$$

where $x_t \in R^n$ is the state vector, $y_t \in R^k$ is the measurement vector, w_t and v_t are the Gaussian noises with covariance Q_t and R_t respectively. $A_t \in R^{n \times n}$ is the state transition matrix and $H_t \in R^{m \times n}$ is the set of all sensors available (m) at time t . $S_t \in R^{k \times m}$ is the binary sensor selection matrix at time t whose nonzero entries extract the selected sensor

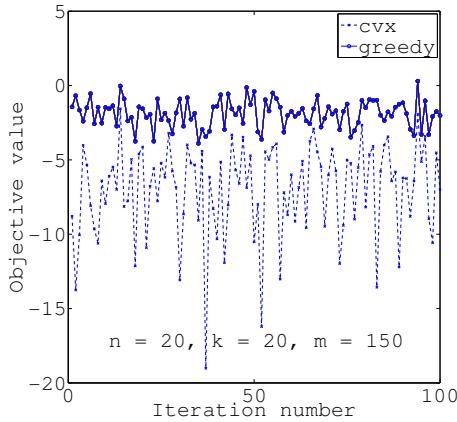


Fig. 1: Instantaneous objective values, $m = 150, n = 20, k = 20$.

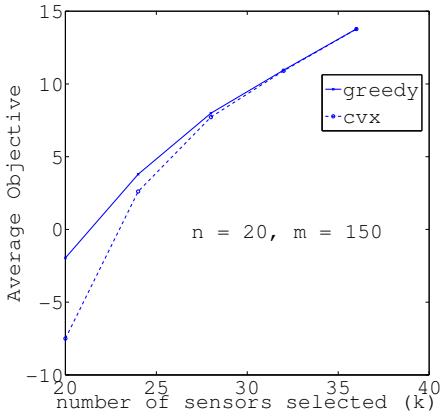


Fig. 2: Objective value vs k , $m = 150, n = 20$.

measurements that are transmitted to the fusion center. Each row of S_t has one and only one nonzero entry. Each column of S_t can have at most one nonzero entry. The goal of the fusion center is to design the sensor selection matrix S_t , so as to minimize the filtered error covariance. Let $P_{t+1|t}$ and $P_{t+1|t+1}$ be the prediction and filtered error covariance respectively at time instant $k+1$. Then

$$P_{t+1|t} = A_t P_{t|t} A_t^T + Q_t, \quad (15)$$

$$P_{t+1|t+1} = (P_{t+1|t}^{-1} + H_{t+1}^T R_{t+1}^{-1} Z_{t+1} H_{t+1})^{-1} \quad (16)$$

where $Z_{t+1} = S_{t+1}^T S_{t+1} \in R^{m \times m}$ is a binary diagonal matrix. Assuming the measurements are independent across sensors (i.e., R_t is diagonal), we have

$$\begin{aligned} P_{t+1|t+1} &= (P_{t+1|t}^{-1} + H_{t+1}^T R_{t+1}^{-\frac{1}{2}} Z_{t+1} R_{t+1}^{-\frac{1}{2}} H_{t+1})^{-1} \\ &= (P_{t+1|t}^{-1} + C_{t+1}^T Z_{t+1} C_{t+1})^{-1} \\ &= (P_{t+1|t}^{-1} + \sum_{i=1}^m z_{t+1}(i) c_{t+1}(i) c_{t+1}'(i))^{-1}, \end{aligned} \quad (17)$$

where $C_{t+1} = R_{t+1}^{-\frac{1}{2}} H_{t+1}$, $c_{t+1}(i)$ is the i^{th} row of C_{t+1} , and $z_{t+1}(i) \in \{0, 1\}$ are the diagonal entries of Z_{t+1} . The

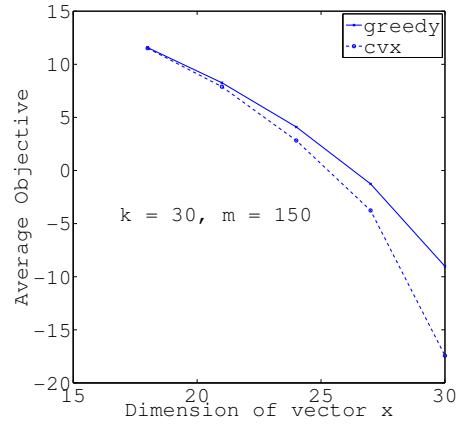


Fig. 3: Objective value vs n , $m = 150, k = 30$.

objective is to choose $z_{t+1}(i)$ so as to minimize the error metric $P_{t+1|t+1}$. One such criteria leads to optimization

$$\begin{aligned} \max \quad & \log \det(P_{t+1|t+1}^{-1}) \\ \text{subject to} \quad & z_{t+1}(i) \in \{0, 1\}, \quad \sum_{i=1}^m z_{t+1}(i) = k. \end{aligned} \quad (18)$$

Algorithm 3 Sensor selection for Kalman filter using greedy algorithm

1. Initialization:

$$s = 1, P_{t+1|t+1}^s = P_{t+1|t}$$

2. Determine greedily the next measurement:

$$k_s = \arg \max_{j \in S^s} c_{t+1}'(j) P_{t+1|t+1}^s c_{t+1}(j)$$

3. Covariance update:

$$P_{t+1|t+1}^{s+1} = P_{t+1|t+1}^s - \frac{P_{t+1|t+1}^s c_{t+1}'(k_s) P_{t+1|t+1}^s}{1 + c_{t+1}'(k_s) P_{t+1|t+1}^s c_{t+1}(k_s)}$$

4. Update the measurement set:

$$S^{s+1} = S^s \setminus k_s, \quad s \leftarrow s + 1, \quad \text{go to step 2 if } s \leq k$$

5. Update the prediction step after k greedy steps:

$$\begin{aligned} P_{t+1|t+1}^k &= P_{t+1|t+1}^k, \\ P_{t+2|t+1} &= A_{t+1} P_{t+1|t+1} A_{t+1}^T + Q_{t+1} \end{aligned}$$

The fusion center selects sensors for transmission at each time instant using the above metric and then updates $P_{t+1|t}$ according to (15). The optimization problem (18) corresponds to maximization of a monotone submodular function subject to a uniform matroid constraint (in particular, the MAP formulation (4)) and hence we can apply the greedy algorithm to solve (18). In particular, we augment Algorithm 2 with an additional step (15). This is formalized as Algorithm 3. The steps of Algorithm 3 are reminiscent of the sequential processing in Kalman filter handling multiple measurements [7]. The step 2

in the Algorithm 3 chooses the measurement in the sequential processing such that it is maximally aligned with direction of the maximum error. This algorithm closely resembles the V-Lambda filtering [10]. Note that the intermediate covariance values $P_{t+1|t+1}^{s+1}$ will be used for the sequential processing of the transmitted measurements.

A. Simulation results

We compare the performance of Algorithm 3 with the approach where the sensor selection at each step of Kalman recursion is obtained via convex relaxation of the objective function. Elements of the measurement matrix H_t are generated as i.i.d Gaussian random variables with zero mean and variance $\frac{1}{n}$. For the system dynamics, we assume $A = I_{n \times n}$ and $Q = 9I_{n \times n}$. The observation noise variance of each measurement is uniformly distributed in $[0.5, 2]$, and is known at the fusion center. The resulting root-mean-square-error (RMSE) of the estimators is computed by averaging over 50 Monte Carlo runs, with a time horizon $T = 20$ for each realization. Figure 4 shows the performance comparison as a function of the number of sensors selected for $n = 15$. Algorithm 3 performs significantly better than the convex relaxation based algorithm, especially when the number of selected sensors is close to n . Figure 5, shows the RMSE performance as a function of n for a fixed number of sensors $k = 30$. As n grows, the performance of the convex relaxation based approach deteriorates much faster than that of Algorithm 3.

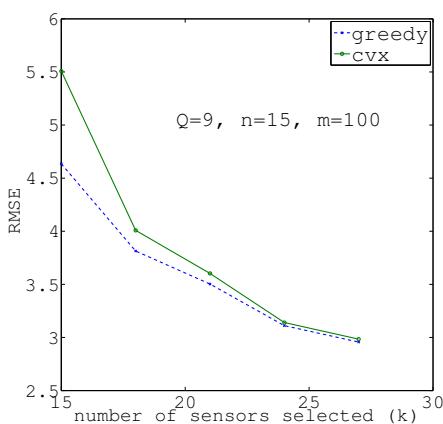


Fig. 4: Average RMSE versus the number of selected sensors ($m = 100, n = 15$).

VI. CONCLUSION

In this paper, we formulated maximum-a-posteriori and maximum-likelihood sensor selection problems as optimizations of submodular functions over uniform matroids. Relying on the existing results for optimization of submodular functions, we proposed a greedy sensor selection algorithm which finds a solution achieving objective within $(1 - \frac{1}{e})$ from the optimal one. Exploiting the structure of the problem, we simplified the original greedy algorithm to significantly reduce its complexity. Moreover, we considered the sensor selection

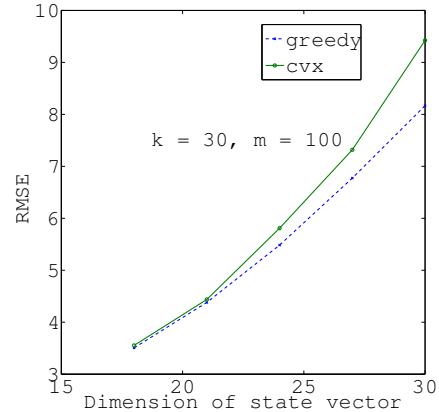


Fig. 5: Average RMSE versus the dimension of the state vector x ($m = 100, k = 30$).

problem in the context of state estimation in linear dynamical systems via Kalman filtering. In each step of the Kalman filter algorithm, the problem was formalized as an optimization of a submodular function over uniform matroids and solved via the greedy algorithm. Simulation results demonstrate that the proposed approach outperforms competing convex relaxation techniques, while needing smaller computational complexity.

REFERENCES

- [1] S. Joshi and S. Boyd, "Sensor Selection via Convex Optimization," *IEEE Trans on Signal Processing*, vol. 57, no. 2, pp 451-462, 2009.
- [2] H. Rowaihy, S. Eswaran, M. Johnson, D. Verma, A. Bar-Noy, T. Brown, and T. L. Porta1, "A Survey of Sensor Selection Schemes in Wireless Sensor Networks," *Proc. SPIE*, 2007, vol. 6562
- [3] A. Krause and C. Guestrin, "Near-optimal Observation Selection using Submodular Functions," *American Association for Artificial Intelligence*, 2007.
- [4] G. L. Nemhauser and L. A. Wolsey, "Best algorithms for approximating the maximum of a submodular set function," *Math. Oper. Research*, 3(3):177-188, 1978.
- [5] J. E. Weimer, B. Sinopoli, and B. H. Krogh, "A Relaxation Approach to Dynamic Sensor Selection in Large-Scale Wireless Networks," *International Conference on Distributed Computing Systems Workshops*, 2008.
- [6] N. Moshtagh, L. Chen, and R. Mehra, "Optimal Measurement Selection For Any-time Kalman Filtering With Processing Constraints," *CDC*, 2009.
- [7] T. Kailath, A. H. Sayed, and B. Hassibi, "Linear Estimation," *Prentice Hall*, 2000.
- [8] Stephen Boyd and Lieven Vandenberghe "Convex Optimization," *Cambridge University Press*, 2004
- [9] G. Calinescu, C. Chekuri, M. Pal, and J. Vondrak, "Maximizing a Monotone Submodular Function subject to a Matroid Constraint," *SICOMP*, 2009
- [10] Y. Oshman, "Optimal sensor selection strategy for discrete-time state estimators," *IEEE Trans on Aerospace and Electronic Systems*, vol. 30, no. 2, pp. 307-314, 1994.
- [11] R. A. Horn and C. R. Johnson "Matrix analysis," *Cambridge University Press*, 1990.

APPENDIX

We first show $f(S) = \log \det\left(\sigma^{-2} \sum_{i \in S} z_i a_i a_i' + \Sigma_x^{-1}\right)$ is a *monotone submodular* function². Without a loss of generality, we assume $\sigma = 1$.

Monotonicity: Let $A = \{a_1, a_2, \dots, a_k\}$, and so

$$F(A) = \log \det\left(\Sigma_x^{-1} + \sum_{i=1}^k a_i a_i'\right). \quad (19)$$

Introduce $B = \{\bar{a}_1, \bar{a}_2, \dots, \bar{a}_l\}$. We need to show that $F(A \cup B) \geq F(A)$, for all sets B . To see this note that

$$\begin{aligned} & F(A \cup B) - F(A) \\ &= \log \frac{\det\left(\Sigma_x^{-1} + \sum_{i=1}^k a_i a_i' + \sum_{i=1}^l \bar{a}_i \bar{a}_i'\right)}{\det\left(\Sigma_x^{-1} + \sum_{i=1}^k a_i a_i'\right)} \\ &= \log \det\left(I + \left(\Sigma_x^{-1} + \sum_{i=1}^k a_i a_i'\right)^{-1} \sum_{i=1}^l \bar{a}_i \bar{a}_i'\right). \end{aligned}$$

Therefore, we need to show that

$$c = \log \det\left(I + \left(\Sigma_x^{-1} + \sum_{i=1}^k a_i a_i'\right)^{-1} \sum_{i=1}^l \bar{a}_i \bar{a}_i'\right) \geq 0$$

Let $M_1 = (\Sigma_x^{-1} + \sum_{i=1}^k a_i a_i')^{-1}$ and $M_2 = \sum_{i=1}^l \bar{a}_i \bar{a}_i'$. It is easy to see M_2 is positive semi definite. Hence, we can find a matrix M_3 such that $M_2 = M_3 M_3'$. Using this we can write

$$\begin{aligned} c &= \log \det\left(I + M_1 M_3 M_3'\right) \\ &= \log \det\left(I + M_3' M_1 M_3\right), \end{aligned}$$

where in obtaining the last expression we used the Sylvester's determinant theorem [11] ($\det(I + AB) = \det(I + BA)$). Since M_1 is positive definite, the matrix $M_3' M_1 M_3$ is positive semidefinite. All the eigen-values of $I + M_3' M_1 M_3$ are at least unity, and hence we obtain the desired result.

Submodularity: Here we prove that $f(S)$ satisfies (12). Let $A = \{a_1, a_2, \dots, a_k\}$, $B = \{a_1, a_2, \dots, a_l\}$, where $l \geq k$, i.e, $A \subseteq B$. Choose a generic element $a_g \notin B$. We need to show that

$$F(A \cup a_g) - F(A) \geq F(B \cup a_g) - F(B). \quad (20)$$

We can write

$$\begin{aligned} & F(A \cup a_g) - F(A) - F(B \cup a_g) + F(B) \\ &= \log \frac{\det(M_1 + a_g a_g') \det(M_2)}{\det(M_2 + a_g a_g') \det(M_1)}, \end{aligned}$$

where

$$\begin{aligned} M_1 &= \Sigma_x^{-1} + \sum_{i=1}^k a_i a_i', \\ M_2 &= \Sigma_x^{-1} + \sum_{i=1}^l a_i a_i' = M_1 + \sum_{i=k+1}^l a_i a_i' = M_1 + M_3, \\ M_3 &= \sum_{i=k+1}^l a_i a_i' \end{aligned}$$

²It is to be noted that $f(S)$ is well defined for all subsets S including the null set.

So, we need to prove that

$$\frac{\det(M_1 + a_g a_g') \det(M_2)}{\det(M_2 + a_g a_g') \det(M_1)} \geq 1$$

We have

$$\begin{aligned} & \frac{\det(M_1 + a_g a_g') \det(M_2)}{\det(M_2 + a_g a_g') \det(M_1)} \\ &= \frac{\det(M_1 + a_g a_g') \det(M_1 + M_3)}{\det(M_1 + M_3 + a_g a_g') \det(M_1)} \\ &= \frac{\det(I + M_1^{-1} a_g a_g')}{\det(I + (M_1 + M_3)^{-1} a_g a_g')} \\ &= \frac{(1 + a_g' M_1^{-1} a_g)}{(1 + a_g' (M_1 + M_3)^{-1} a_g)} \end{aligned}$$

Hence we need to show that

$$\frac{(1 + a_g' M_1^{-1} a_g)}{(1 + a_g' (M_1 + M_3)^{-1} a_g)} \geq 1$$

i.e,

$$a_g' (M_1^{-1} - (M_1 + M_3)^{-1}) a_g \geq 0$$

Since for invertible positive definite matrices $M > N$ it holds $M^{-1} < N^{-1}$, the matrix $M_1^{-1} - (M_1 + M_3)^{-1}$ is positive semidefinite and hence the above inequality holds. This proves the desired result.

To complete the proof of Lemma 1, we next discuss complexity of Algorithm 1. At step 2 of Algorithm 1 we evaluate the determinant of positive definite matrix. There are $O(m)$ such determinants to be computed. The complexity of finding the determinant of a $n \times n$ matrix is, in general, $O(n^3)$. When choosing k sensors, Algorithm 1 has k greedy steps and hence its complexity is $O(n^3 m k)$.