Sensor Scheduling via Compressed Sensing

Avishy Carmi
Technion - Israel Institute of Technology, Haifa 32000, Israel
Ariel University Center, Ariel 40700, Israel

Abstract – *We present a novel approach for sensor scheduling which is, in general, a NP-hard problem involving the selection of S out of N sensors such that an optimal filtering performance is attained. Our approach utilizes a heuristic measure that quantifies the incoherence of the vector space defined by the sensors with respect to the system principal directions. This in turn facilitates the formulation of a convex relaxation that can be efficiently solved using a myriad of compressed sensing algorithms.*

Keywords: Sensor selection, Sensor networks, Compressed sensing, Estimability

1 Introduction

Many computational advances in the last few decades have dramatically boosted the information processing capabilities of complex systems involving prohibitively large number of measuring devices (e.g., sensor networks, multi-agent systems [1–3]). In many such systems the sensory information is gathered and fused in a preferably optimal fashion so as to reduce the overall sensitivity to faulty data and to enhance the robustness to component malfunctions. The decision making capabilities attributed to this class of systems greatly depend on the computational resources allocated for processing the gathered sensory data. Occasionally, the processing unit has to deal with excessively large number of observations acquired by the various sensors while at the same time to prune out any redundancies that might have unintentionally been incorporated in (e.g., identical observations which deceivingly give the impression that the amount of information has increased). This in turn may form a bottleneck that significantly damages the overall processing performance of the system.

The bottleneck problem mentioned above is typically alleviated by employing a so-called sensor scheduling/selection scheme. In this regard, the scheduling scheme regulates the computational load by utilizing a relatively small number of sensors as it seeks to minimize the unavoidable information loss. Another important issue that arise in complex sensor constellations is the total power consumption. Usually, in order for the system to sustain its working capabilities over some period of time it is necessary to activate only a limited number of sensors at any given instance.

Numerous sensor scheduling strategies have been proposed over the past few decades. The seminal work in [4] recast the sensor scheduling problem as a non-linear deterministic control problem that turns out to be solvable via a tree-search. Following this, some greedy methods are suggested for coping with the complexity of the tree-search in [5–7]. The approach adopted in [8] allows the sensors to switch randomly according to an optimal probability distribution such that the best expected steady-state performance is attained.

The sensor selection problem is, in general, NP-hard (i.e., there are exactly $\binom{S}{N}$ possibilities of choosing S distinct sensors out of N available ones). This essentially implies that an optimal solution cannot be efficiently computed, in particular when the number of sensors becomes excessively large. Recently, a convex relaxation of the original NP-hard problem has been suggested in [9]. The most prominent advantage of this approach over the other methods is accredited to its applicability as there are a myriad of well-established convex optimization techniques.

Solving a NP-hard subset search problem by means of a convex l_1 relaxation is one of the fundamental concepts in the new emerging theory of compressed sensing [10,11]. Compressed sensing essentially refers to the recovery of a sparse or, more precisely, a compressed representation of a signal (which typically involves a limited number of highly incoherent projections) from a relatively small number of observations, typically less than the signal dimension. The work in [10] has shown that under certain conditions pertaining to the problem characteristics a highly accurate and even exact solution can be obtained for the original NP-hard problem by solving a l_1 convex relaxation. This result has triggered a massive quest for efficient convex optimization recipes. To this end there exist plenty of acclaimed compressed sensing algorithms such as the Bayesian compressed sensing [12], the gradient projection [13], the gradient pursuit [14], and the least angle regression [15], to name only a few.

In this work we introduce a novel strategy for solving the sensor scheduling problem using *any* compressed sensing algorithm. Our approach utilizes a heuristic incoherence measure that is based on the notion of estimability which is the stochastic analogy of observability. This in turn facilitates the formulation of a compressed sensing problem aimed at minimizing the incoherence of the vector space defined by the sensors with respect to the system principal directions.

2 Problem Overview

The sensor scheduling problem is encountered in a wide class of applications in which the (hidden) process of interest is observed by multiple measuring devices. As the number of sensors becomes prohibitively large it is necessary to have some sort of a decision scheme for pruning out observations while retaining a reasonable performance of the underlying filtering algorithm. Hence, the scheduling scheme essentially regulates the computational workload and typically seeks to minimize the attainable estimation error.

Consider the following discrete-time linear stochastic system

$$x_k = A_{k-1}x_{k-1} + B_{k-1}u_{k-1} + G_{k-1}w_{k-1} \tag{1a}$$

$$y_k(i) = h_k(i)^T x_k + n_k, \quad i = 1, \ldots, N \tag{1b}$$

where the deterministic matrices $A_k \in \mathbb{R}^{n\times n}, B_k \in \mathbb{R}^{n\times l}, G_k \in \mathbb{R}^{n\times m}$ and $h_k(i) \in \mathbb{R}^n$. The unobserved signal x_k is an $\mathbb{R}^n$-valued process for which the initial state is normally distributed with mean and covariance $\hat{x}_0 = E[x_0]$ and $P_0 = E\left[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T\right]$, respectively. This process is driven by both a deterministic input $u_k \in \mathbb{R}^l$ and a zero-mean $\mathbb{R}^m$-valued white Gaussian sequence $w_k \sim \mathcal{N}(0, Q)$. The observed scalar process associated with the ith sensor, $y_k(i)$, is assumed to be contaminated by a zero-mean white Gaussian noise, n_k, with covariance R_k. In addition to that, it is assumed that the observation noise is statistically independent of both the initial state and the process noise at any given time instance, $E[x_0 n_j^T] = 0, \quad E[w_k n_j^T] = 0$. Using these definitions we can now formulate the sensor selection problem as follows. We seek a subset of S observations, $\{y_k(j_i)\}_{i=1}^S, \quad j_i \in [1, N]$, that would optimize the performance of a filtering algorithm applied for the estimation of x_k. A widespread measure of the filtering performance is the 2nd moment of the estimation error (which coincides with estimation error covariance in the unbiased case). Hence, our optimization problem may be mathematically formulated as

$$\begin{aligned}&\min_{\{j_i\}_{i=1}^S,\, j_i\in[1,N]} \operatorname{Tr}\left(E\left[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T\right]\right) \\ &= \min_{\{j_i\}_{i=1}^S,\, j_i\in[1,N]} E\Big[E\big[(x_k - \hat{x}_k)^T(x_k - \hat{x}_k) \mid \\ &\qquad \{y_1(j_i)\}_{i=1}^S, \ldots, \{y_k(j_i)\}_{i=1}^S\big]\Big]\end{aligned} \tag{2}$$

where $\hat{x}_k$, Tr and $E[\cdot|\cdot]$ denote the estimator of x_k, the matrix trace operator and the conditional expectation, respectively.

The problem (2) is known to be NP-hard, essentially involving $\binom{S}{N} = S!/[N!(N-S)!]$ possibilities of choosing the sought-after subset $\{y_k(j_i)\}_{i=1}^S$. This obviously implies that an optimal solution might be infeasible in particular when N becomes excessively large (e.g., such as in sensor network). Fortunately, a suboptimal solution to (2) may be efficiently obtained utilizing various relaxation technqiues. Such an approach has been recently introduced in [9] where a convex optimization scheme is employed for solving a sensor selection problem in the framework of static parameter estimation. The idea discussed in [9] involves the explicit expression of the estimation error covariance of a simple least squares (LS) scheme applied for the observation subset $\{y_k(j_i)\}_{i=1}^S, \quad j_i \in [1, N]$, that is

$$P_k = \left(\sum_{i=1}^S h_k(j_i) R_k^{-1} h_k(j_i)^T\right)^{-1} \tag{3}$$

Having this, [9] proceeds by minimizing the determinant of P_k [1] rather than its trace which in turn facilitates the formulation of a relaxed (convex) problem of the form

$$\max_{\substack{\sum_{i=1}^N \beta_i = 1 \\ 0\le\beta_i\le 1}} \log\det\left(\sum_{i=1}^N \beta_i h_k(i) R_k^{-1} h_k(i)^T\right) \tag{4}$$

where β_i can be regarded as some measure of importance associated with the ith sensor. The S chosen sensors are then taken as those having the highest importance score.

The abovementioned approach is not really meant to work for the dynamical system (1) unless its transition matrix A_k is diagonal. This simple observation stems from the fact that in the presence of coupled dynamics (which is manifested by a non-diagonal A_k) the observability requirement of the pair (A_k, H_k) with $H_k = [h_k(j_1), \ldots, h_k(j_S)]$, which guarantees that x_k can be fully estimated based on the sensory information $\{y_k(j_i)\}_{i=1}^S$, does not really coincide with the condition

[1] This is equivalent to maximizing the determinant of $\sum_{i=1}^S h_k(j_i) R_k^{-1} h_k(j_i)^T$.

of having a nonsingular P_k in (3). In other words, the fulfillment of the observability condition

$$\text{Rank}\begin{bmatrix} H_k^T \\ H_k^T A_k \\ \dots \\ H_k^T A_k^{n-1} \end{bmatrix} = n \tag{5}$$

does not imply $\text{Rank}(H_k) = n$ unless A_k is diagonal.

In this work we purpose a computationally efficient scheme for obtaining a possibly suboptimal solution of the sensor selection problem for the generalized linear system (1). Our method circumvents the aforementioned limitation of [9] by adopting a heuristic measure which intermediately affects the performance of the underlying filtering algorithm. This measure, which is termed here *total coherence/incoherence*, is strongly related to the notion of observability or, more precisely, estimability. In what follows, we discuss the latter concept which forms the basis of our heuristic approach.

3 Estimability and Incoherence

The notion of estimability is introduced in [16]. The idea underlines this concept can be thought of as a stochastic analogy of observability [2]. Formally, it layouts the conditions under which the posterior error covariance of an unbiased estimator is strictly smaller than the prior covariance matrix. Defining $\pi_k = E[x_k x_k^T]$ and $P_k = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]$, where $\hat{x}_k$ is an estimator of x_k, the system (1) is said to be estimable if

$$P_k \prec \pi_k, \quad \forall k \geq n \tag{6}$$

i.e., if and only if for any $g \in \mathbb{R}^n, \quad g \neq 0$

$$g^T P_k g < g^T \pi_k g, \quad \forall k \geq n \tag{7}$$

which, as proved in [16], is equivalent to the condition

$$E\left[y_j x_k^T\right] g \neq 0, \quad \text{for some } j \leq k \text{ and } \forall k \geq n \tag{8}$$

An additional rather explicit formulation of (8) is provided in the following proposition.

Proposition 1. *The estimability condition* (8) *is equivalent to*

$$h_j^T A_j^{-1} \cdots A_{k-1}^{-1} g \neq 0, \quad \text{for some } j < k \text{ and } \forall k \geq n \tag{9}$$

for any $g \in \mathbb{R}^n, \quad g \neq 0$*, assuming the transition matrices,* $A_j, \dots, A_{k-1}$*, and* Ξ_k *(see Appendix for its exact definition) are invertible.*

For the sake of consecutiveness the proof is deferred to the appendix. An interesting implication of this proposition is given below.

[2]The relationship between estimability and observability is briefly discussed in [17].

Corollary 1. *For time-invariant systems the condition* (9) *reduces to the Popov-Belevitch-Hautus (PBH) [18] eigenvector test.*

Proof. The corollary follows straightforwardly from Proposition 1 upon recalling that the eigenvectors of $A^{-m} = (A^{-1})^m$ are identical to those of A. □

3.1 Incoherence Heuristic

Let $< a, b >:= a^T b$ be an inner product over $\mathbb{R}^n$. We say that a is incoherent with respect to, or orthogonal to, b whenever $< a, b >= 0$. Following this definition it can be readily recognized that both Proposition 1 and the PBH eigenvector test roughly state that the sensing vector h_k should not be incoherent with respect to any of the system's principal modes, which is characterized by the eigenvectors of A_k, for guaranteeing a proper reconstruction of the state from a finite set of observations. These approaches aim at answering a yes/no type question and do not really say anything about the attainable reconstruction errors in cases where several sensing possibilities, $h_k(i), \quad i = 1, \dots, N$, exist. Nevertheless, a common intuition here might suggest that the more coherent the sensing vector is with respect to the system's principal directions, the lower is the attainable reconstruction error. In order to make use of this thumb rule we define the following measure of incoherence of the ith sensor at time k

$$-\sum_{j=1}^{n} |< h_k(i), \text{Re}(g_k(j)) >| \tag{10}$$

where $g_k(j), \quad j = 1, \dots, n$ denote the eigenvectors of A_k, and $\text{Re}(g)$ stands for the real part of g. The measure (10) can be readily generalized to the multiple sensor case by

$$\Phi_k = -\sum_{l=1}^{n} \sum_{i=1}^{S} |\beta_k(j_i) < h_k(j_i), \text{Re}(g_k(l)) >| \tag{11}$$

which essentially quantifies the incoherence over the entire vector space defined by the selected sensing vectors $\{h_k(j_i)\}_{i=1}^S, \quad j_i \in [1, N]$. The parameters $\beta_k(j_i)$ in (11) are weighting scalars that regulate the contribution of the corresponding sensors to the *total incoherence* Φ_k.

Following the above arguments we may seek to minimize the total incoherence instead of the covariance measure in our original problem (2). In virtue of the definition (11) a possibly suboptimal solution to (2) can be efficiently computed as described next.

3.2 Sparse Formulation

It turns out that replacing the original objective in (2) with the total incoherence measure (11) to yield

$$\min_{\{j_i\}_{i=1}^S,\, j_i \in [1,N]} \Phi_k \tag{12}$$

facilitates the formulation of a relaxed problem that can be easily solved using a myriad of optimization techniques. Notice, however, that the solution of (12), and obviously of any relaxation thereof, is most likely to be suboptimal in the sense of the original objective in (2).

We begin by rewriting (12) as a recovery problem for which the weighting parameters form a sparse vector $\beta_k = [\beta_k(1), \ldots, \beta_k(N)]^T$ with $\beta_k(j_i) \neq 0$ for a set of S indices $\{j_i\}_{i=1}^S$ where $S << N$ (i.e., there are exactly S elements in the support of β_k where S is much smaller than N, the total number of available sensors). Using the convention $\| \beta_k \|_0$ for the cardinality of the support of β_k, (12) can now be expressed as

$$\min_{\|\beta_k\|_0 = S} \left[-\sum_{l=1}^{n} \sum_{i=1}^{N} | < \beta_k(i) h_k(i), \mathrm{Re}(g_k(l)) > | \right] \quad (13)$$

Before proceeding any further we would like to clarify a certain issue which, from our standpoint, is crucial to the understanding of our argument. The main goal of this work, as it was already emphasized in the introductory part, is to uniquely formulate the sensor selection problem in the framework of compressed sensing for which there are numerous solution techniques. This objective underlies our forthcoming derivations in which, as the reader will witness, we sacrifice optimality in the sense of the original objective in (2). Having stated this, we continue by replacing the objective in (13) with an upper bound. Firstly, we observe that

$$\begin{aligned} -\sum_{l=1}^{n} \sum_{i=1}^{N} | < \beta_k(i) h_k(i), \mathrm{Re}(g_k(l)) > | \leq \\ -\left| \sum_{l=1}^{n} \sum_{i=1}^{N} < \beta_k(i) h_k(i), \mathrm{Re}(g_k(l)) > \right| = \\ -| < H_k \beta_k, V_k > | \end{aligned} \quad (14)$$

where

$$H_k = \underbrace{\begin{bmatrix} h_k(1), \ldots, h_k(N) \\ \vdots \\ h_k(1), \ldots, h_k(N) \\ \vdots \\ h_k(1), \ldots, h_k(N) \end{bmatrix}}_{(n^2) \times N}, \quad V_k = \underbrace{\begin{bmatrix} \mathrm{Re}(g_k(1)) \\ \vdots \\ \vdots \\ \mathrm{Re}(g_k(n)) \end{bmatrix}}_{(n^2) \times 1} \quad (15)$$

Furthermore, the law of cosines implies

$$\begin{aligned} -| < H_k \beta_k, V_k > | \leq \\ \frac{1}{2} \left[\| V_k - H_k \beta_k \|_2^2 - \| V_k \|_2^2 - \| H_k \beta_k \|_2^2 \right] \leq \\ \frac{1}{2} \left[\| V_k - H_k \beta_k \|_2^2 - \| V_k \|_2^2 \right] \end{aligned} \quad (16)$$

Finally, combining (14) – (16) we obtain the following problem

$$\min_{\|\beta_k\|_0 = S} \| V_k - H_k \beta_k \|_2^2 \quad (17)$$

Unexceptionally, *for prohibitively large N*, an optimizer of (17) cannot be computed efficiently. Nonetheless, an appropriate convex relaxation of (17) can be readily formulated by replacing the 0-norm with the l_1-norm. This problem, which is typically referred to as compressed sensing (CS), was extensively investigated in the recent literature. In what follows, we briefly discuss the ideas underlying CS with an emphasis on their usage in the context of our problem.

4 Compressed Sensing

In the past decade, the l_1 norm was suggested as a sparseness-promoting term in the seminal work introducing the acclaimed LASSO operator [19] and the basis pursuit [20]. Recasting the sparse recovery problem using the l_1 norm yields a convex relaxation of the original NP-hard problem (17), which can be efficiently solved via a myriad of well-established optimization techniques. Commonly, there are two equivalent convex formulations that follow from (17): The quadratically-constrained linear program, which takes the form

$$\min_{\beta_k} \| \beta_k \|_1 \text{ s.t. } \| V_k - H_k \beta_k \|_2^2 \leq \epsilon \quad (18)$$

and the quadratic program

$$\min_{\beta_k} \| V_k - H_k \beta_k \|_2^2 \text{ s.t. } \| \beta_k \|_1 \leq \epsilon' \quad (19)$$

It can be shown that for proper values of the tuning parameters ϵ and ϵ' the solution of both these problems coincide.

Recently, [10, 11] have shown that an accurate solution of (17) can almost always be obtained by solving the convex relaxation (18) assuming that the sensing matrix $H_k \in \mathbb{R}^{(n^2) \times N}$ obeys the so-called restricted isometry property (RIP). The RIP roughly implies that the columns of a given matrix nearly behave like an orthonormal basis. This desired property is possessed by several random constructions, which guarantee the uniqueness of the sparse solution. In particular, an exact recovery is highly probable when using such matrices provided that a relation of the type

$$S = \mathcal{O}(n^2 / \log(N/n^2)) \quad (20)$$

holds. For an extensive overview of several RIP constructions and their role in CS, the reader is referred to [10, 11].

4.1 Sensor Scheduling via CS

In the preceding section we have formulated two equivalent convex formulations (18) and (19) of the NP-hard problem (17). Either of these relaxations can be efficiently solved using various CS methods such as the Bayesian CS (BCS) [12], the recently introduced gradient projection [13] and gradient pursuit [14], the least angle regression [15], and the CSKF [21], to name only a few. The obtained sparse solution β_k can be then used to select an optimal sensor constellation. Thus, recalling the role of the weights $\beta_k(i)$ in the total incoherence measure in (13), we select those sensors for which the associated weights are the most significant in terms of magnitude. In detail, we sort $|\beta_k(i)|, \quad i = 1, \dots, N$ in a descending order

$$|\beta_k(j_1)| \geq |\beta_k(j_2)| \geq \cdots \geq |\beta_k(j_N)| \tag{21}$$

and select S out of the top ranked sensors, $\{y_k(j_i)\}_{i=1}^{S}$. Notice that this approach is somewhat similar to the one adopted in [9].

4.2 Random Sensor Networks

As it is demonstrated in the ensuing, the CS-based sensor scheduling scheme suggested inhere works remarkably well even when the underlying sensing matrix does not obey the RIP. Nevertheless, this is a desired property that may dramatically improve the performance in terms of accuracy and speed. To this end, the commonly used constructions that do posses this property are random by nature [3]. The random matrices described in Section 3.4 in [11] are known to satisfy the RIP with overwhelming probability assuming a relation of the type (20) holds. Nonetheless, such sensing matrices are typically not encountered in virtually many engineering applications. In this regard, an exception is to be made whenever a random component is introduced into the sensing device. Such an approach was recently adopted in the so-called single-pixel camera [22]. This camera which heavily relies on CS theory implements a randomly-controlled array of mirrors for recovering a compressed image based on a relatively small number of observations.

The idea of incorporating random sensing mechanisms within conventional applications was triggered by an informal discussion in [11]. There, the author has distinctively pointed out that this concept might lead to a whole new philosophy in data acquisition. His unique view of the subject is somewhat hinted in the term *compressive sampling* which appears in the title in [11].

Inspired by this concept, we describe here a sensor constellation comprising of randomly controlled measuring devices. Consider, for instance, a network of N sensors, $\{h_k(i)\}_{i=1}^{N}$, for which each individual element in $h_k(i) \in \mathbb{R}^n$ is randomly picked over a discrete binary space. In particular, assume that there are only two equally probable values that a single element in $h_k(i)$ may take. Then, in this case, it can be shown that the overall sensing matrix defined in (15) satisfies the RIP up to a certain level (this construction is similar to the binary measurements matrix described in Section 3.4 in [11]).

[3] A few rather deterministic approaches for constructing RIP matrices have been recently discussed in the literature.

5 Numerical Study

In this section the performance of the CS-based sensor selection scheme is assessed in various synthetic scenarios. The linear system considered here is constructed as follows. The transition matrix A_k is composed out of the orthonormal eigenvectors of a symmetric random matrix UU^T where the entries of U are uniformly distributed over $[-1, 1]$ (i.e., the eigenvalues of A_k lie on the unit circle). The process x_k itself is driven by a zero-mean Gaussian white sequence with covariance $Q = 0.1^2 I_{n\times n}$. In most scenarios the total number of available sensors is $N = 200$ out of which either 5 or 10 are selected. In the first few examples, the jth element of the sensing vector $h_k(i)$ is set as $sin\left(2\pi/N(i + j/n)\right)$. This vector is then normalized to yield $\| \ h_k(i) \ \|_2 = 1$ for suppressing large deviations in signal to noise ratio over the entire set of sensors. The main purpose motivating such a composition is to obtain a possibly rank deficient sensing matrix $[h_k(1), \dots, h_k(N)]^T$ which will obviously make the problem much more challenging (Note that if instead of this, the entries of $h_k(i)$ would have been taken as independent identically distributed samples, then on the overall we could not expect to spot a prominent difference in estimation performance for different sensor subsets). The observations themselves are contaminated by a zero-mean white noise with covariance 0.05^2.

In all scenarios we have used the standard Kalman filter (KF) to obtain the estimates $\hat{x}_k$ based exclusively on the selected sensors' observations. Our CS-based selection scheme employs either the BCS of [12] or the CSKF of [21] for solving the relaxed problem (19). The sensors are then selected based on β_k as described in Section 4.1. Notice that whenever the underlying system is time-invariant, it is sufficient to execute the selection scheme only once, preferably at the beginning of the filtering procedure. In other cases, the selection scheme should be invoked whenever a change occurs in either the transition or sensing matrices.

In what follows we have compared the attainable root mean square estimation error (RMSE) [4] of the KF equipped with our CS-based method with that of a KF employing either the convex optimization approach of [9] or a subset of distinct randomly selected sensors. In all examples, excluding the first one, the comparisons are based on 300 Monte Carlo (MC) runs.

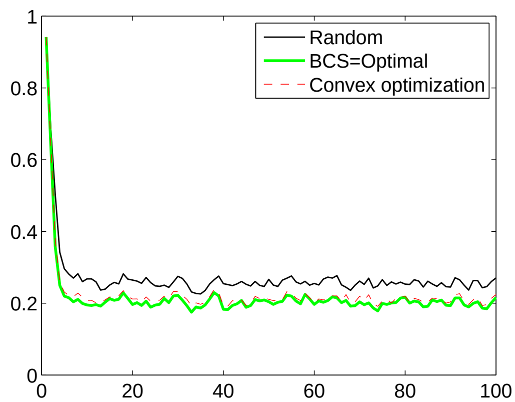
5.1 Small Scale Example

In this example we consider a time-invariant system with a state dimension of $n = 4$ and a total of $N = 30$ sensors out of which only 4 are selected. The main purpose of this small scale scenario is to assess the optimality of the CS-based selection scheme in terms of the objective in (2). Here we have exactly 27405 ways of choosing a subset of 4 distinct sensors out of the

[4] That is $\sqrt{E\left[(x_k - \hat{x}_k)^T (x_k - \hat{x}_k)\right]}$

available 30. The optimal subset can be therefore obtained using an excessive search over the entire range of possibilities.

The results of this example are summarized in Fig. 1. This figure shows the RMSE computed based on 50 MC runs when using the CS-based selection scheme (employing the BCS [12]), the convex optimization approach of [9], and a subset of randomly selected sensors (which were picked at the beginning of each run). In this specific example, the CS approach manages to identify the optimal subset. The performance of the convex optimization approach is nearly optimal whereas the random selection scheme yields the worse RMSE.



(a) 30 sensors, 4 selected

Figure 1. The RMSE obtained when using the various sensor scheduling schemes.

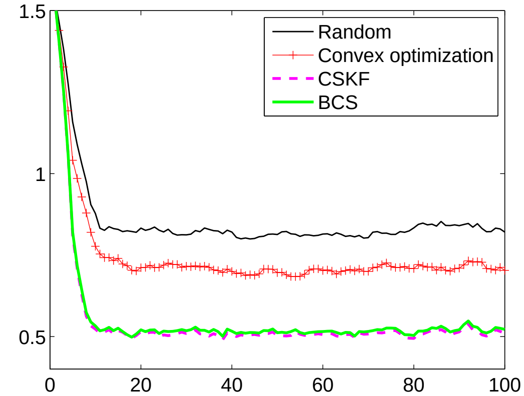
5.2 Time-Invariant Example

In this scenario we consider a time-invariant system with a state dimension $n = 10$ and a total of $N = 200$ sensors out of which either 5 or 10 are selected. Here, the predetermined matrices, A and $h(i)$, remain unchanged in all runs. As before, we compare the performance of our CS-based method (employing either the BCS or the CSKF algorithms) with both the convex optimization and the random selection approaches.

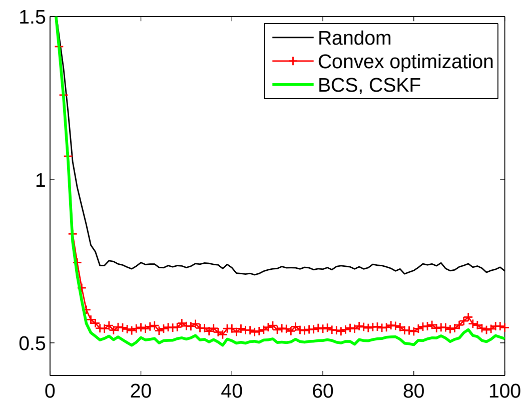
The results of this example are summarized in Fig. 2. This figure clearly demonstrates the superiority of the CS approach over the other two methods. The slight improvement of using the CSKF algorithm instead of the BCS in solving (19) can be noticed in Fig. 2a.

5.3 Unobservable Configurations

In this example we consider a time-invariant system that resembles the previously mentioned one in its dimensionality, however with a slight change in the composition of both the sensing vectors $h(i)$ and the transition matrix A. Here, we let $A = UU^T$ where U is uniformly sampled as previously described. In addition to that, we let $h(i)$, $i = 1, \ldots, N$ be either one of



(a) 5 selected



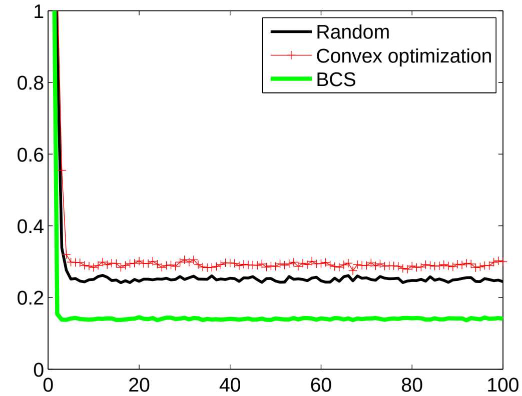
(b) 10 selected

Figure 2. The RMSE obtained when using the various sensor scheduling schemes.

the (orthonormal) eigenvectors of A. Following this, it is quite obvious that the sensor space consists of subsets for which this system becomes strictly unobservable (e.g., when the selected subset lacks at least one of the eigenvectors of A).

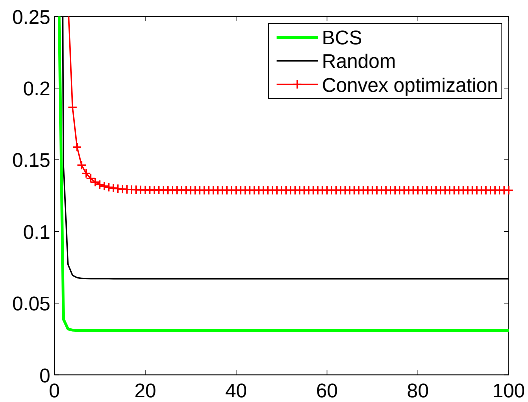
Two scenarios are examined the first of which involves some predetermined matrices, A and $h(i)$, that remain fixed in all runs. The results for this specific case are depicted in Fig. 3. In a companion example we construct a new set of system matrices at the beginning of each run. The results for this case are illustrated in Fig. 4 where the mean trace of the estimation error covariance is depicted.

As before, the CS-based approach outperforms the other methods. From both Figs. 3 and 4, it can be clearly seen that as opposed to the previous examples, in which the convex optimization approach has shown an improvement over the random selection scheme, in this scenario this method fails to work, essentially yielding the worst filtering performance.



(a) 10 selected

Figure 3. The RMSE obtained when using the various sensor scheduling schemes.



(a) 10 selected

Figure 4. The mean trace of the estimation error covariance when using the various sensor scheduling schemes.

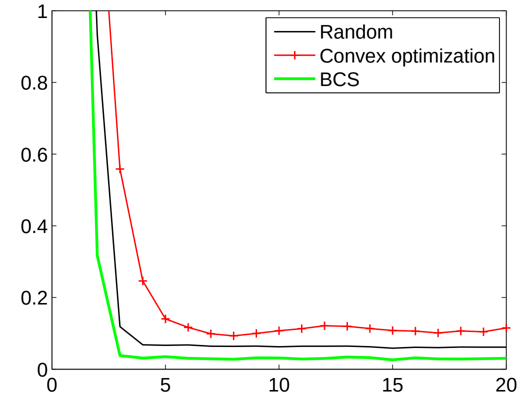
5.4 Time-Varying Example

As the title suggests, in this example we consider a time-varying system. The dimensionality and dynamics of this system are similar to those described in the preceding example with the only exception that here the matrices A_k and $h_k(i)$ are constructed at each and every time step.

The results of this scenario are shown in Fig. 5 where the mean trace of the estimation error covariance is depicted. Unexceptionally, in this example as in the preceding ones, the CS-based approach attains the best performance with respect to the other two methods.

6 Conclusions

A novel compressed sensing-based approach for sensor scheduling is presented. The new method ,which



(a) 10 selected

Figure 5. The mean trace of the estimation error covariance when using the various sensor scheduling schemes.

is essentially based on a heuristic measure, is shown to outperform the convex optimization approach of [9] and the trivial random selection scheme. The computational efficiency of the new method in large scale settings exclusively depends on the compressed sensing algorithm used.

A Proof of Proposition 1

Let us assume, without any loss of generality, that $B_k = 0, \quad \forall k$ in (1a). Now, observing (1a) we may write

$$
\begin{aligned}
&x_{k-1} = A_{k-1}^{-1} x_k - A_{k-1}^{-1} G_{k-1} w_{k-1} \\
&\qquad x_{k-2} = A_{k-2}^{-1} x_{k-1} - A_{k-2}^{-1} G_{k-2} w_{k-2} = \\
&A_{k-2}^{-1} A_{k-1}^{-1} x_k - A_{k-2}^{-1} A_{k-1}^{-1} G_{k-1} w_{k-1} - A_{k-2}^{-1} G_{k-2} w_{k-2} \\
&\qquad\qquad \vdots \\
&x_j = A_j^{-1} \cdots A_{k-1}^{-1} x_k - A_j^{-1} \cdots A_{k-1}^{-1} G_{k-1} w_{k-1} - \cdots \\
&\qquad\qquad - A_j^{-1} G_j w_j
\end{aligned} \tag{22}
$$

for any $j < k$. Further substituting the explicit expression of y_j, and of x_j thereof, within the estimability condition (8), yields

$$
\begin{aligned}
&E\left[y_j x_k^T\right] g = E\left[(h_j^T x_j + n_j) x_k^T\right] g = \\
&\qquad E\big[h_j^T A_j^{-1} \cdots A_{k-1}^{-1} x_k x_k^T\big] g - \\
&E\big[h_j^T A_j^{-1} \cdots A_{k-1}^{-1} G_{k-1} w_{k-1} x_k^T\big] g - \cdots \\
&\quad - E\big[h_j^T A_j^{-1} G_j w_j x_k^T\big] g + E\big[n_j x_k^T\big] g \neq 0
\end{aligned} \tag{23}
$$

From (1a) we notice that

$$
x_k = A_{k-1} \cdots A_j x_j + G_{k-1} w_{k-1} + \cdots + A_{k-1} \cdots A_{j+1} G_j w_j \tag{24}
$$

Substituting (24) within some of the terms on the right hand side of (23) yields

$$E\left[y_j x_k^T\right] g = h_j^T A_j^{-1} \cdots A_{k-1}^{-1} \bar{g} \neq 0 \qquad (25)$$

where $\bar{g} = \Xi_k g$ with $\Xi_k := \pi_k - G_{k-1} Q_{k-1} G_{k-1}^T - \cdots - A_{k-1} \cdots A_{j+1} G_j Q_j G_j^T A_{j+1}^T \cdots A_{k-1}^T$, which thereby completes the proof.

References

[1] J. Kahn, R. Katz, and K. Pister, "Next century challenges: Mobile networking for smart dust," In the proceedings of the ACM MobiCom conference, Seattle, WA, 1999.

[2] S. I. Roumeliotis and G. A. Bekey, "Distributed multi-robot localization," *IEEE Transactions on Robotics and Automation*, vol. 18, pp. 781 – 795, 2002.

[3] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, pp. 6 – 23, 1997.

[4] L. Meier, J. Peschon, and R. Dressler, "Optimal control of measurement subsystems," *IEEE Transactions on Automatic Control*, vol. 12, pp. 528 – 536, 1967.

[5] Y. Oshman, "Optimal sensor selection strategy for discrete-time state estimators," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, pp. 307 – 314, 1994.

[6] S. Kagami and M. Ishikawa, "A sensor selection method considering communication delays," In the proceedings of the IEEE international conference on robotics and automation, 2004, pp. 206 – 211.

[7] V. Gupta, T. Chung, B. Hassibi, and R. M. Murray, "Sensor scheduling algorithms requiring limited computation," In the proceedings of the international conference on acoustics, speech and signal processing, 2004.

[8] V. Gupta, T. Chung, B. Hassibi, and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage," *Automatica*, vol. 42, pp. 251 – 260, 2006.

[9] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Transactions on Signal Processing*, vol. 57, pp. 451 – 462, February 2009.

[10] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, 2006.

[11] E. J. Candes, "Compressive sampling," Madrid, Spain, 2006, European Mathematical Society, Proceedings of the International Congress of Mathematicians.

[12] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, pp. 2346 – 2356, June 2008.

[13] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, pp. 586 – 597, December 2007.

[14] T. Blumensath and M. E. Davies, "Gradient pursuits," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2370 – 2382, 2008.

[15] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407 – 499, 2004.

[16] Y. Baram and T. Kailath, "Estimability of linear systems," *IEEE Transactions on Automatic Control*, vol. 33, pp. 1116 – 1121, December 1988.

[17] D. Goshen-Meskin and I. Y. Bar-Itzhack, "On the connection between estimability and observability," *IEEE Transactions on Automatic Control*, vol. 37, pp. 1225 – 1226, August 1992.

[18] T. Kailath, *Linear Systems*, Prentice Hall, Englewood Cliffs, New Jersey, 1980.

[19] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[20] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal of Scientific Computing*, vol. 20, no. 1, pp. 33 – 61, 1998.

[21] A. Carmi, P. Gurfil, and D. Kanevsky, "Methods for sparse signal recovery using kalman filtering with embedded pseudo-measurement norms and quasi-norms," *IEEE Transactions on Signal Processing*, In print.

[22] M. F. Durate, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, 2008.