

# Sensor Selection via Compressed Sensing

Avishy Carmi and Pini Gurfil

*Faculty of Aerospace Engineering  
Technion - Israel Institute of Technology, Haifa 32000, Israel*

---

**Abstract:** Sensor selection is an NP-hard problem involving the selection of  $S$  out of  $N$  sensors such that optimal filtering performance is attained. We present a novel approach for sensor selection that utilizes a heuristic measure quantifying the incoherence of the vector space spanned by the sensors with respect to the system's principal directions. This approach facilitates the formulation of a convex relaxation problem that can be efficiently modeled and solved using compressed sensing (CS) algorithms. We subsequently develop a new CS algorithm based on subgradient projections. The new CS algorithm for sensor selection is shown to outperform existing methods in a number of applications.

*Keywords:* Sensor selection, Sensor networks, Compressed sensing, Estimability, Subgradient projection methods

---

## 1. INTRODUCTION

Computational advances in the last few decades have dramatically boosted the information processing capabilities of complex systems containing very large numbers of measuring devices, such as sensor networks and multi-agent systems. 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 sensor scheduling and/or a sensor 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. Numerous sensor selection and sensor scheduling strategies have been proposed over the past few decades. The seminal work in [Meier et al.(1967)] recast the sensor scheduling problem as a nonlinear deterministic control problem that turns out to be solvable via a tree-search. Following this, some greedy methods were suggested for coping with the complexity of the tree-search in [Kagami and Ishikawa(2004), Gupta et al.(2004)].

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 by [Joshi and Boyd(2009)]. The most prominent advantage of this approach over the other methods is its applicability as there are many well-established convex optimization techniques.

Solving an 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 [Candes et al.(2006)]. Compressed sensing 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 [Candes et al.(2006)] has shown that under certain conditions, a highly accurate, and even exact solution can be obtained for the original NP-hard problem by solving an  $l_1$  convex relaxation. This result has triggered a massive quest for new efficient convex optimization recipes. Some of the acclaimed compressed sensing algorithms include the Bayesian compressed sensing [Ji et al.(2008)], the gradient projection [Figueiredo et al.(2007)], the gradient pursuit [Blumensath and Davies(2008)], the orthogonal matching pursuit [Pati et al.(1993)], and the least angle regression [Efron et al.(2004)], to name only a few.

In this work, we introduce a novel strategy for solving the sensor selection 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 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.

Another major contribution of this work is the development of a subgradient projection-based method for solving the compressed sensing problem. The remarkably simple approach is adequate for high-dimensional settings, as it exclusively based on matrix and vector multiplications.

We show, by exploring a number of illustrating examples, that the new method outperforms other state-of-the-art compressed sensing techniques when applied to the sensor selection problem.

## 2. PROBLEM OVERVIEW

The sensor selection 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 extremely large, it is necessary to have some sort of a decision scheme for pruning out observations while retaining reasonable performance of the underlying filtering algorithm. Hence, the selection scheme essentially regulates the computational workload while minimizing the attainable estimation error.

The system model is given by the discrete-time linear stochastic process

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

$$y_k(i) = h_k(i)^T x_k + n_k, \quad i = 1, \dots, N \quad (1b)$$

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 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  $P_0 = E[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T]$ , 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  $i$ th sensor,  $y_k(i)$ , is assumed to be contaminated by a zero-mean white Gaussian noise,  $n_k$ , with covariance  $R_k$ . In addition, it is assumed that the observation noise is statistically independent of both the initial state and the process noise at any given time instance,  $E[x_0 n_j^T] = 0$ ,  $E[w_k n_j^T] = 0$ . Using these definitions, we can now formulate the sensor selection problem as follows: Find a subset of  $S$  observations,  $\{y_k(j_i)\}_{i=1}^S$ ,  $j_i \in [1, N]$ , 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 the estimation error covariance in the unbiased case). Hence, our optimization problem may be mathematically formulated as

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

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

The problem (2) is known to be NP-hard, involving  $\binom{S}{N} = S!/[N!(N-S)!]$  combinations of choosing the sought-after subset  $\{y_k(j_i)\}_{i=1}^S$ . This implies that finding an optimal solution might be impractical, in particular when  $N$  becomes excessively large (e.g., such as the case in sensor networks). Fortunately, a suboptimal solution to (2) may be efficiently obtained by utilizing various relaxation techniques. Such an approach has been recently introduced by [Joshi and Boyd(2009)], where a convex optimization

scheme is employed for solving a sensor selection problem in the framework of static parameter estimation. The idea discussed in [Joshi and Boyd(2009)] involves the explicit expression of the estimation error covariance of a simple least squares (LS) scheme applied for the observation subset  $\{y_k(j_i)\}_{i=1}^S$ ,  $j_i \in [1, N]$ , that is

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

Having this, [Joshi and Boyd(2009)] proceeds by minimizing the determinant of  $P_k^{-1}$  rather than its trace, which gives rise to the formulation of a relaxed (convex) program of the form

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

where  $\beta_i$  can be regarded as some measure of importance associated with the  $i$ th sensor. The  $S$  chosen sensors are then taken as those having the highest importance score.

The abovementioned approach is not planned 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  $H_k = [h_k(j_1), \dots, h_k(j_S)]$ , which guarantees that  $x_k$  can be fully estimated based on the sensory information  $\{y_k(j_i)\}_{i=1}^S$ , does not coincide with the condition of having a nonsingular  $P_k$  in (3). In other words, the fulfillment of the observability condition

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

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

The current work purposes a computationally-efficient scheme for obtaining a suboptimal solution of the sensor selection problem for the generalized linear system (1). Our method circumvents the aforementioned limitation of [Joshi and Boyd(2009)] by adopting a heuristic measure that 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 approach.

## 3. ESTIMABILITY AND INCOHERENCE

The notion of estimability is introduced in [Baram and Kailath(1988)]. The idea underlines this concept can be thought of as a stochastic analogy of observability. Formally, let  $\pi_k \doteq E\{[x_k - E(x_k)][x_k - E(x_k)]^T\}$  and  $P_k \doteq E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]$ , where  $\hat{x}_k$  is an estimator of  $x_k$ . The system (1) is said to be estimable if

$$P_k \prec \pi_k, \quad \forall k \geq n \quad (6)$$

i.e., if and only if

$$g^T P_k g < g^T \pi_k g, \quad \forall k \geq n, \quad \forall g \in \mathbb{R}^n \setminus \{0\} \quad (7)$$

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

which, as proven in [Baram and Kailath(1988)], is equivalent to the condition

$$E [y_j x_k^T] g \neq 0, j \leq k, \forall k \geq n, \forall g \in \mathbb{R}^n \setminus \{0\} \quad (8)$$

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

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

$$h_j^T \left( \prod_{i=j}^{k-1} A_i^{-1} \right) g \neq 0, j < k \text{ and } \forall k \geq n, \forall g \in \mathbb{R}^n \setminus \{0\} \quad (9)$$

if the transition matrices,  $A_j, \dots, A_{k-1}$ , and

$$\begin{aligned} \Xi_k &\doteq \pi_k - G_{k-1} Q_{k-1} G_{k-1}^T \\ &\dots - \left( \prod_{i=j+1}^{k-1} A_i^T \right)^T G_j Q_j G_j^T \left( \prod_{i=j+1}^{k-1} A_i^T \right) \end{aligned} \quad (10)$$

are invertible.

The proof is omitted for brevity.

### 3.1 Incoherence Heuristic

Let  $\langle a, b \rangle := a^T b$  be an inner product over  $\mathbb{R}^n$ . We say that  $a$  is incoherent with respect to, or orthogonal to,  $b$  whenever  $\langle a, b \rangle = 0$ . 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,  $h_k(i)$ ,  $i = 1, \dots, N$ , 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  $i$ th sensor at time  $k$

$$- \sum_{j=1}^n |\langle h_k(i), \text{Re}(g_k(j)) \rangle| \quad (11)$$

where  $g_k(j)$ ,  $j = 1, \dots, n$  denote the eigenvectors of  $A_k$ , and  $\text{Re}(g)$  stands for the real part of  $g$ . The measure (11) can be readily generalized to the multiple sensor case by

$$\Phi_k = - \sum_{l=1}^n \sum_{i=1}^S |\beta_k(j_i) \langle h_k(j_i), \text{Re}(g_k(l)) \rangle| \quad (12)$$

which essentially quantifies the incoherence over the entire vector space defined by the selected sensing vectors  $\{h_k(j_i)\}_{i=1}^S$ ,  $j_i \in [1, N]$ . The parameters  $\beta_k(j_i)$  in (12) are weighting scalars that regulate the contribution of the corresponding sensors to the *total incoherence*  $\Phi_k$ .

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 (12), a

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 (12) to yield

$$\min_{\{\beta_i\}_{i=1}^S, j_i \in [1, N]} \Phi_k \quad (13)$$

facilitates the formulation of a relaxed problem that can be easily solved using a myriad of optimization techniques. Notice, however, that the solution of (13), 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 (13) as a recovery problem for which the weighting parameters form a sparse vector  $\beta_k = [\beta_k(1), \dots, \beta_k(N)]^T$  with  $\beta_k(j_i) \neq 0$  for a set of  $S$  indices  $\{j_i\}_{i=1}^S$  where  $S \ll N$  (i.e., there are exactly  $S$  elements in the support of  $\beta_k$  where  $S$  is much smaller than  $N$ , the total number of available sensors). Using the notation convention  $\|\beta_k\|_0$  to denote the cardinality of the support of  $\beta_k$ , (13) can now be expressed as

$$\min_{\|\beta_k\|_0=S} \left[ - \sum_{l=1}^n \sum_{i=1}^N |\langle \beta_k(i) h_k(i), \text{Re}(g_k(l)) \rangle| \right] \quad (14)$$

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 we sacrifice optimality in the sense of the original objective in (2). Having stated this, we continue by replacing the objective in (14) with an upper bound. Firstly, we observe that

$$\begin{aligned} & - \sum_{l=1}^n \sum_{i=1}^N |\langle \beta_k(i) h_k(i), \text{Re}(g_k(l)) \rangle| \leq \\ & \quad - \left| \sum_{l=1}^n \sum_{i=1}^N \langle \beta_k(i) h_k(i), \text{Re}(g_k(l)) \rangle \right| = \\ & \quad - |\langle H_k \beta_k, V_k \rangle| \end{aligned} \quad (15)$$

where

$$H_k = \underbrace{\begin{bmatrix} h_k(1), \dots, h_k(N) \\ \vdots \\ h_k(1), \dots, h_k(N) \\ \vdots \\ h_k(1), \dots, h_k(N) \end{bmatrix}}_{(n^2) \times N}, \quad V_k = \underbrace{\begin{bmatrix} \text{Re}(g_k(1)) \\ \vdots \\ \text{Re}(g_k(n)) \end{bmatrix}}_{(n^2) \times 1} \quad (16)$$

Furthermore, the law of cosines implies

$$\begin{aligned} & - |\langle H_k \beta_k, V_k \rangle| \leq \\ & \quad \frac{1}{2} [\|V_k - H_k \beta_k\|_2^2 - \|V_k\|_2^2 - \|H_k \beta_k\|_2^2] \leq \\ & \quad \frac{1}{2} [\|V_k - H_k \beta_k\|_2^2 - \|V_k\|_2^2] \end{aligned} \quad (17)$$

Finally, combining (15) – (17) we obtain the following problem

$$\min_{\|\beta_k\|_0=S} \|V_k - H_k \beta_k\|_2^2 \quad (18)$$

Unexceptionally, for prohibitively large  $N$ , an optimizer of (18) cannot be computed efficiently. Nonetheless, an appropriate convex relaxation of (18) 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.

#### 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 [Tibshirani(1996)] and the basis pursuit [Chen et al.(1998)]. Recasting the sparse recovery problem using the  $l_1$  norm yields a convex relaxation of the original NP-hard problem (18), which can be efficiently solved via a myriad of well-established optimization techniques. Commonly, there are two equivalent convex formulations that follow from (18): The quadratically-constrained linear program, which takes the form

$$\min_{\beta_k} \|\beta_k\|_1 \text{ s.t. } \|V_k - H_k\beta_k\|_2^2 \leq \epsilon \quad (19)$$

and the quadratic program

$$\min_{\beta_k} \|V_k - H_k\beta_k\|_2^2 \text{ s.t. } \|\beta_k\|_1 \leq \epsilon' \quad (20)$$

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

Recently, [Candes et al.(2006)] have shown that an accurate solution of (18) can almost always be obtained by solving the convex relaxation (19) 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

$$S = \mathcal{O}(n^2 / \log(N/n^2)) \quad (21)$$

holds. For an extensive overview of several RIP constructions and their role in CS, the reader is referred to [Candes et al.(2006)].

##### 4.1 Sensor Selection via Compressed Sensing

In the preceding section we have formulated two equivalent convex formulations (19) and (20) of the NP-hard problem (18). Either of these relaxations can be efficiently solved using various CS methods such as the Bayesian CS (BCS) [Ji et al.(2008)], the recently introduced sequential subspace optimization (SESOP) [Elad et al.(2007)] and iterative hard thresholding (IHT) [Blumensath et al.(2007)], the gradient projection [Figueiredo et al.(2007)] and gradient pursuit [Blumensath and Davies(2008)], the least angle regression [Efron et al.(2004)], and the orthogonal matching pursuit (OMP) [Pati et al.(1993)], to name only a few. The obtained solution  $\beta_k$  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 (14), we select those sensors for which the associated weights

are the most significant in terms of magnitude. In detail, we sort  $|\beta_k(i)|$ ,  $i = 1, \dots, N$  in a descending order

$$|\beta_k(j_1)| \geq |\beta_k(j_2)| \geq \dots \geq |\beta_k(j_N)| \quad (22)$$

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 [Joshi and Boyd(2009)].

##### 4.2 Subgradient Projections for Compressed Sensing

Subgradient projection (SP) methods refer to a set of iterative convex optimization techniques. At a first glance these methods seem to have a structure similar to the standard descent approaches (e.g., steepest descent). Nevertheless, there are two prominent differences that render them viable for a far wider variety of applications. As distinct from standard descent schemes, SP methods relax the requirement of differentiable objectives as they essentially rely on the computation of the instantaneous subgradient [Censor and Zenios(1997)]. In addition, their step size is independent of the search direction, which typically results in a considerable reduction of the associated computational overhead especially in high dimensional settings.

*The Basic SP-CS Iteration* Let  $g(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$  be a convex function with domain  $\mathbb{R}^n$ . Then, the SP iteration assumes the form [Censor and Zenios(1997)]

$$\beta^{i+1} = \beta^i - \lambda_i \frac{g(\beta^i)}{\|t^i\|_2^2} t^i \quad (23)$$

where  $\lambda_i$  and  $t^i$  denote a positive relaxation parameter and the instantaneous subgradient computed at  $\beta^i$ , respectively. The application of (23) for solving the problems (19) and (20) involves the formulation of an equivalent unconstrained objective [James et al.(2009)]

$$g(\beta^i) = \|\beta^i\|_1 + \tau \|V_k - H_k\beta^i\|_2^2 \quad (24)$$

where  $\tau > 0$  is a tuning parameter. The associated subgradient is therefore given by

$$t^i = \text{sign}(\beta^i) - 2\tau H_k^T (V_k - H_k\beta^i) \quad (25)$$

where  $\text{sign}(\beta^i) \in \mathbb{R}^n$  comprises of the corresponding signs of the elements of  $\beta^i$ .

The convergence properties of the SP method (23) greatly depends on the selection of the relaxation sequence  $\{\lambda_i\}_{i \geq 0}$ . It turns out that properly regulating these parameters can assure the convergence of (23) towards the optimizer of (24). In what follows we provide some insights pertaining to this issue.

*Theorem 1.* Let  $\{\beta^i\}_{i \geq 0}$  be a sequence produced by a repeated application of the SP iteration (23). Assume also that  $g(\beta^i) > g(\beta^*)$  with  $g(\beta^*)$  being the minimum of  $g(\cdot)$ . For such  $\{\beta^i\}_{i \geq 0}$  there exists a sequence of non-negative relaxation parameters  $\{\lambda_i\}_{i \geq 0}$  with  $(\sup_i \lambda_i) \leq c < 2$  satisfying

$$\|\beta^{i+1} - \beta^*\|_2^2 \leq \|\beta^i - \beta^*\|_2^2, \quad \forall i \geq 0 \quad (26)$$

The proof is omitted for brevity.

#### 5. NUMERICAL STUDY

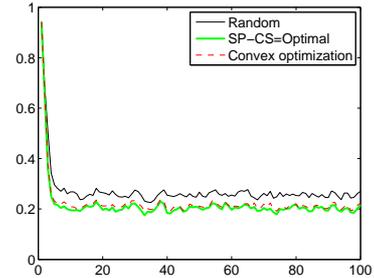
The linear system considered here is constructed as follows. The transition matrix  $A_k$  is composed out of the orthonormal eigenvectors of a symmetric random matrix

$UU^T$  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  $j$ th element of the sensing vector  $h_k(i)$  is set as  $\sin(2\pi/N(i + j/n))$ . 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 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 SP-CS method described previously (using no more than 5 iterations with  $\lambda_i = 10^{-3}$  and  $\tau = 50$ . See (23) – (25)) or the BCS of [Ji et al.(2008)] for solving the relaxed problem (20). The sensors are then selected based on  $\beta_k$  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)<sup>2</sup> of the KF equipped with our CS-based method with that of a KF employing either the convex optimization approach of [Joshi and Boyd(2009)] 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 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 SP-CS), the convex optimization approach of [Joshi and Boyd(2009)], and a subset of randomly selected sensors (which were picked at the beginning of each run). In this specific example, the CS scheme 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.

<sup>2</sup> That is  $\sqrt{E[(x_k - \hat{x}_k)^T(x_k - \hat{x}_k)]}$

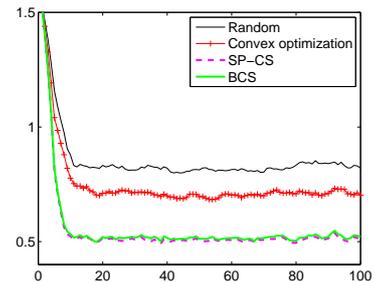


(a) 30 sensors, 4 selected

Fig. 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 SP-CS or the BCS 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.



(a) 5 selected

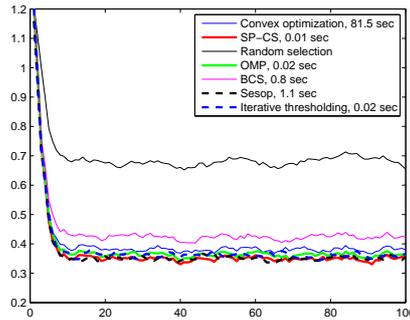
Fig. 2. The RMSE obtained when using the various sensor scheduling schemes.

### 5.3 Medium Scale Example

A comparison of some of the previously mentioned CS methods is provided in the following medium scale example comprising of 5000 sensors. The time-invariant system here is composed as before with  $n = 6$  states and 5 selected sensors. The results in this case are shown in Fig. 3. From this figure it can be readily recognized that the best performance in terms of both speed and accuracy is attained by the SP-CS method (using no more than 5 iterations).

## 6. CONCLUSIONS

A novel compressed-sensing-based approach for sensor scheduling was presented. The new method, which is essentially based on a heuristic measure, is shown to outperform the convex optimization approach of [Joshi and Boyd(2009)] and the trivial random selection scheme.



(a) 5000 sensors, 5 selected

Fig. 3. The RMSE obtained when using various CS methods for solving the sensor selection problem.

#### REFERENCES

- [Baram and Kailath(1988)] Baram, Y. and Kailath, T. (1988). Estimability of linear systems. *IEEE Transactions on Automatic Control*, 33, 1116 – 1121.
- [Blumensath and Davies(2008)] Blumensath, T. and Davies, M.E. (2008). Gradient pursuits. *IEEE Transactions on Signal Processing*, 56(6), 2370 – 2382.
- [Blumensath et al.(2007)] Blumensath, T., Yaghoobi, M., and Davies, M. (2007). Iterative hard thresholding and  $l_0$  regularisation. *IEEE International Conference on Acoustics, Speech and Signal Processing*.
- [Candes et al.(2006)] Candes, E.J., Romberg, J., and Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52, 489–509.
- [Censor and Zenios(1997)] Censor, Y. and Zenios, S.A. (1997). *Parallel optimization: Theory, algorithms, and applications*. Oxford University Press, New York, NY, USA.
- [Chen et al.(1998)] Chen, S.S., Donoho, D.L., and Saunders, M.A. (1998). Atomic decomposition by basis pursuit. *SIAM Journal of Scientific Computing*, 20(1), 33 – 61.
- [Efron et al.(2004)] Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, 32(2), 407 – 499.
- [Elad et al.(2007)] Elad, M., Matalon, B., and Zibulevsky, M. (2007). Coordinate and Subspace Optimization Methods for Linear Least Squares with Non-Quadratic Regularization. *Applied and Computational Harmonic Analysis*, 23, 346–367.
- [Figueiredo et al.(2007)] Figueiredo, M.A.T., Nowak, R.D., and Wright, S.J. (2007). Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 1, 586 – 597.
- [Gupta et al.(2004)] Gupta, V., Chung, T., Hassibi, B., and Murray, R.M. (2004). Sensor scheduling algorithms requiring limited computation. In the proceedings of the international conference on acoustics, speech and signal processing.
- [James et al.(2009)] James, G.M., Radchenko, P., and Lv, J. (2009). DASSO: connections between the Dantzig Selector and LASSO. *Journal of the Royal Statistical Society*, 71, 127–142.
- [Ji et al.(2008)] Ji, S., Xue, Y., and Carin, L. (2008). Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56, 2346 – 2356.
- [Joshi and Boyd(2009)] Joshi, S. and Boyd, S. (2009). Sensor selection via convex optimization. *IEEE Transactions on Signal Processing*, 57, 451 – 462.
- [Kagami and Ishikawa(2004)] Kagami, S. and Ishikawa, M. (2004). A sensor selection method considering communication delays. 206 – 211. In the proceedings of the IEEE international conference on robotics and automation.
- [Kailath(1980)] Kailath, T. (1980). *Linear Systems*. Prentice Hall, Englewood Cliffs, New Jersey.
- [Meier et al.(1967)] Meier, L., Peschon, J., and Dressler, R. (1967). Optimal control of measurement subsystems. *IEEE Transactions on Automatic Control*, 12, 528 – 536.
- [Pati et al.(1993)] Pati, Y.C., Rezifar, R., and Krishnaprasad, P.S. (1993). Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. 27th Asilomar Conf. on Signals, Systems and Comput.
- [Tibshirani(1996)] Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267–288.