Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Optimal sensor selection strategies in the presence of wireless communication links ⋆

Alessandro Nordio * **Alberto Tarable** * **Fabrizio Dabbene** *

* *CNR-IEIIT, Torino, ITALY*
(e-mail: {alessandro.nordio, alberto.tarable, fabrizio.dabbene}@ieiit.cnr.it)

Keywords: Wireless sensor networks, sensor selection, mutual information, convex relaxation

Abstract: The problem of selecting in an optimal way κ sensors from a given set of K sensors providing noisy measurements of some physical variable has received a growing interest in the literature. The problem has been shown to be combinatorial, and several computable relaxations have been presented. In this paper, we consider a particularly interesting variant of the sensor selection problem. Motivated by the increase in the application of wireless sensor networks, i.e. networks of sensors which take *remote* measurements of the quantity of interest and then communicate their values through a (noisy) wireless communication link, we propose a scheme for optimally selecting the wireless sensors taking into account also the available channel state information. The optimality conditions are formally derived in an information-theoretic context, and specific semi-definite programming relaxations leading to computable techniques for large values of κ and K are presented. Also, we derive specific results for the cases of high and low signal-to-noise ratios. Numerical simulations show how knowledge of the channel state information may lead to an increase of the achievable mutual information, and determine a different choice of sensors.

1. INTRODUCTION

In recent years, wireless sensor networks (WSNs) and wireless sensor-and-actuator networks (WSANs) have found application in many fields, like, for example, traffic control, weather forecast, pollution control, etc. In WSNs/WSANs, a number of sensors measure a given physical variable, and transmit their measurements to a gateway, which processes the received information, possibly in order to apply a consequent response to the measured physical system. Often, the number of deployed sensors is larger than needed for processing, so that, at fixed time intervals, the gateway performs sensor selection, with the purpose of enabling a subset of sensors to transmit, while keeping the others in sleep mode. Such a strategy allows saving power and prolonging the lifetime of sensor batteries.

The problem of optimally selecting a subset of sensors from a set of possible choices has received an increasing interest in the literature. Clearly, optimal sensor selection turns out to be a combinatorial problem, so that several papers propose suboptimal approximations or relaxations to make the problem feasible even for a moderate-to-large number of sensors. Moreover, different optimization criteria can be adopted in defining optimal sensor selection. For example, Joshi and Boyd (2009) consider as a parameter the volume of the confidence ellipsoid and adopts a convex relaxation technique to reduce the combinatorial problem to a convex one, which can be solved in polynomial time. Instead, in order to minimize the same parameter a greedy selection algorithm is proposed in Shamaiah et al. (2012), with a guaranteed lower bound on its performance. In Mo et al. (2011), the problem is extended to the case where the sensors measure the output of a dynamical system, and a different criterion is proposed, related to the Kalman-filter error covariance matrix.

Although implicitly assuming that the sensors send their observations to a fusion center, none of these works considers the unavoidable influence of the wireless channel characteristic, which can largely affect the quality of the received signal. Motivated by this observation, in this paper we consider a WSN with K sensors transmitting with multiple antennas to a common receiver (i.e., the network gateway or the system controller). We suppose that there is no interference between signals received from different sensors, i.e., they transmit on orthogonal channels. Among the total of K received signals, the receiver has to choose κ, $\kappa < K$, according to some criterion. To formally define the optimality conditions, we take an information-theoretic approach and adopt as objective function of the wireless sensor selection problem the *mutual information* between the measured variable and the set of selected signals.

In this work we formulate the sensor selection problem in terms of an optimization problem for both high and low signal-to-noise ratios (SNR) on the wireless links. We also show that such an optimization problem can be relaxed into a suboptimal semi-definite program (SDP) (Joshi and Boyd, 2009) or approximated through a greedy algorithm.

The paper is structured as follows. In Section 2, we give a brief review of the main information-theoretic concepts used in the paper. Section 3 contains the considered scenario. In Section 4, we formulate the problem of optimal sensor selection for our scenario. In Section 5, we describe two suboptimal solutions based on relaxation and greedy search. In Section 6, we consider the limit regimes for high and low received signal-to-noise ratio (SNR). In Section 7 we show numerical results with the comparison of different algorithms for sensor selection. Finally, in Section 8, we draw some conclusions.

⋆ 28 March 2014.
This article has been submitted to the 19th IFAC World Congress, Cape Town, South Africa, August 24-29, 2014.

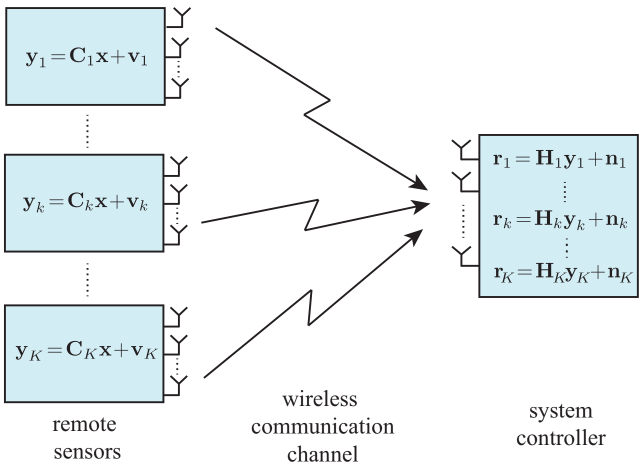

Fig. 1. Wireless sensor network

2. REVIEW OF INFORMATION-THEORY CONCEPTS

In this section, we briefly review the concepts of information theory that will be used throughout the paper.

Let the vector[1] $\mathbf{x}$ of size n be a continuous multivariate random variable (rv) with probability density function (pdf) $f_{\mathbf{x}}(\mathbf{z})$ defined over the region $\mathcal{X} \subseteq \mathbb{R}^n$. The *differential entropy* of $\mathbf{x}$, $h(\mathbf{x})$, is defined as

$$h(\mathbf{x}) \doteq -\int_{\mathcal{X}} f_{\mathbf{x}}(\mathbf{z}) \log f_{\mathbf{x}}(\mathbf{z})\, \mathrm{d}\mathbf{z} \tag{1}$$

provided that the integral exists. The measure units are *bits* or *nats* if the logarithm base is 2 or e, respectively. Roughly speaking, the differential entropy measures the average uncertainty of the realizations of $\mathbf{x}$. If $\mathbf{x}$ is a complex Gaussian rv with covariance $\boldsymbol{\Sigma}$, then $h(\mathbf{x}) = \log |\pi \mathrm{e} \boldsymbol{\Sigma}|$.

Now, let the vector $\mathbf{y}$ of size m be a continuous rv with pdf defined over the region $\mathcal{Y} \subseteq \mathbb{R}^m$. Then *conditional entropy* of $\mathbf{y}$ given $\mathbf{x}$, $h(\mathbf{y}|\mathbf{x})$, is defined as

$$h(\mathbf{y}|\mathbf{x}) \doteq -\int_{\mathcal{X}}\int_{\mathcal{Y}} f_{\mathbf{x},\mathbf{y}}(\mathbf{z}_1, \mathbf{z}_2) \log f_{\mathbf{x},\mathbf{y}}(\mathbf{z}_2|\mathbf{z}_1)\, \mathrm{d}\mathbf{z}_1\, \mathrm{d}\mathbf{z}_2 \tag{2}$$

where $f_{\mathbf{x},\mathbf{y}}(\mathbf{z}_1, \mathbf{z}_2)$ is the joint pdf of $\mathbf{x}$ and $\mathbf{y}$ and $f_{\mathbf{x},\mathbf{y}}(\mathbf{z}_2|\mathbf{z}_1) = \frac{f_{\mathbf{x},\mathbf{y}}(\mathbf{z}_1,\mathbf{z}_2)}{f_{\mathbf{x}}(\mathbf{z}_1)}$ is the conditional pdf of $\mathbf{y}$ given $\mathbf{x}$. Conceptually, the conditional entropy of $\mathbf{y}$ given $\mathbf{x}$ has the meaning of the average uncertainty of $\mathbf{y}$ given that the value of $\mathbf{x}$ is known.

In communication theory, it is often important to estimate the quantity of information flowing through a channel. Let the channel input and output be described by the rv's $\mathbf{x}$ and $\mathbf{y}$, respectively. Then, we define the *mutual information* of such channel as $\mathcal{I}(\mathbf{x}, \mathbf{y})$, which is given by

$$\mathcal{I}(\mathbf{x}; \mathbf{y}) \doteq h(\mathbf{x}) - h(\mathbf{x}|\mathbf{y}) \tag{3}$$

which can be interpreted as the average amount of information about $\mathbf{x}$ (the unknown channel input) that can be extracted from the observation $\mathbf{y}$. The value of $\mathcal{I}(\mathbf{x}; \mathbf{y})$ is always non-negative and is 0 if and only if $\mathbf{x}$ and $\mathbf{y}$ are independent rv's.

[1] Column vectors and matrices are denoted by bold lowercase and bold uppercase letters, respectively. The matrix transpose operator is denoted by $(\cdot)^\mathsf{T}$ while the conjugate transpose operator is denoted by $(\cdot)^\mathsf{H}$. $\mathbf{I}$ indicates the identity matrix and the determinant of a matrix $\mathbf{A}$ is denoted by $|\mathbf{A}|$.

In the most typical example of continuous channel, the output $\mathbf{y}$ is a version of the input $\mathbf{x}$ corrupted with additive noise $\mathbf{n}$, i.e.,

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{4}$$

where $\mathbf{n}$ is complex Gaussian with zero mean and covariance $\boldsymbol{\Sigma}_{\mathbf{n}}$ and independent of $\mathbf{x}$. If $\mathbf{x}$ is complex Gaussian with covariance $\boldsymbol{\Sigma}_{\mathbf{x}}$, then the mutual information $\mathcal{I}(\mathbf{x}; \mathbf{y})$ is given by

$$\begin{aligned}\mathcal{I}(\mathbf{x}; \mathbf{y}) &= \log |\pi \mathrm{e} (\boldsymbol{\Sigma}_{\mathbf{x}} + \boldsymbol{\Sigma}_{\mathbf{n}})| - \log |\pi \mathrm{e} \boldsymbol{\Sigma}_{\mathbf{n}}| \\ &= \log \frac{|\boldsymbol{\Sigma}_{\mathbf{x}} + \boldsymbol{\Sigma}_{\mathbf{n}}|}{|\boldsymbol{\Sigma}_{\mathbf{n}}|}\end{aligned} \tag{5}$$

3. SYSTEM MODEL

Following an approach similar to the one discussed in Joshi and Boyd (2009), we consider the problem of estimating an unknown vector $\mathbf{x} \in \mathbb{R}^{n_x}$ from the measurements performed by K sensors, each one providing n_y linear measurements, corrupted by additive noise. That is, the output of these sensors is given by

$$\mathbf{y}_k = \mathbf{C}_k \mathbf{x} + \mathbf{v}_k, \quad k = 1, \ldots, K$$

where $\mathbf{y}_k$ is a vector of size n_y and $\mathbf{C}_k$ is the k-th observation matrix of size $n_y \times n_x$. The terms $\mathbf{v}_k$, $k = 1, \ldots, K$, represent the measurement noise and are modeled as i.i.d. Gaussian random vectors with zero mean and covariance matrices $\boldsymbol{\Sigma}_{\mathbf{v}_k}$. However, differently from previous works, we consider here the more realistic situation where the sensors act *remotely*, and their measures are to be sent to a common receiver (system controller) through a wireless channel. In particular we assume that all sensors are equipped with a wireless transceiver with n_y antennas and that at the k-th sensor the i-th entry of $\mathbf{y}_k$ can be transmitted by the i-th antenna, $i = 1, \ldots, n_y$.

Transmissions from sensors to the receiver take place on K orthogonal multiple-input multiple-output (MIMO) channels so that they do not interfere with each other. Also, the wireless receiver is equipped with n_r antennas so that the signal $\mathbf{y}_k$, transmitted by the k-th sensor is received as

$$\mathbf{r}_k = \mathbf{H}_k \mathbf{y}_k + \mathbf{n}_k = \mathbf{H}_k \mathbf{C}_k \mathbf{x} + \mathbf{H}_k \mathbf{v}_k + \mathbf{n}_k \tag{6}$$

where $\mathbf{n}_k$ represent additive noise, assumed independent on $\mathbf{v}_k$, and modeled as a complex Gaussian random vector with zero mean and covariance matrix $\sigma^2\mathbf{I}$. The $n_r \times n_y$ matrix $\mathbf{H}_k$ represents the k-th MIMO channel. Specifically the (i, j)-th entry of $\mathbf{H}_k$ is the complex amplitude gain of the link connecting the i-th transmitting antenna to the j-th receiving antenna.

The signal received by all sensors can then be arranged in a vector, $\mathbf{r} = [{\mathbf{r}_1}^\mathsf{T}, \ldots, {\mathbf{r}_K}^\mathsf{T}]^\mathsf{T}$, of size Kn_r given by

$$\mathbf{r} = \begin{bmatrix} \mathbf{H}_1\mathbf{C}_1 \\ \vdots \\ \mathbf{H}_K\mathbf{C}_K \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{H}_1\mathbf{v}_1 \\ \vdots \\ \mathbf{H}_K\mathbf{v}_K \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_K \end{bmatrix} \tag{7}$$

or, more compactly, as

$$\begin{aligned}\mathbf{r} &= \mathbf{H}\mathbf{C}\mathbf{x} + \mathbf{H}\mathbf{v} + \mathbf{n} \\ &\doteq \boldsymbol{\Phi}\mathbf{x} + \boldsymbol{\eta}\end{aligned} \tag{8}$$

where $\mathbf{\Phi} \doteq \mathbf{HC}$, $\mathbf{H}$ is a $Kn_r \times Kn_y$ matrix given by

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & & \\ & \ddots & \\ & & \mathbf{H}_K \end{bmatrix},$$

$\mathbf{C}$ is a $Kn_y \times Kn_x$ matrix given by $\mathbf{C} \doteq [{\mathbf{C}_1}^\mathsf{T}, \ldots, {\mathbf{C}_K}^\mathsf{T}]^\mathsf{T}$, $\mathbf{v} \doteq [{\mathbf{v}_1}^\mathsf{T}, \ldots, {\mathbf{v}_K}^\mathsf{T}]^\mathsf{T}$ and $\mathbf{n} \doteq [{\mathbf{n}_1}^\mathsf{T}, \ldots, {\mathbf{n}_K}^\mathsf{T}]^\mathsf{T}$. The random vector $\boldsymbol{\eta} = \mathbf{Hv} + \mathbf{n}$ represents the equivalent noise at the receiver (i.e., it accounts for the thermal noise $\mathbf{n}$ and for the measurement noise $\mathbf{v}$) and has covariance matrix

$$\mathbf{\Sigma}_{\boldsymbol{\eta}} = \mathbf{H}\mathbf{\Sigma}_\mathbf{v}\mathbf{H}^\mathsf{H} + \sigma^2\mathbf{I} \tag{9}$$

where $\mathbf{\Sigma}_\mathbf{v} = \mathrm{diag}(\mathbf{\Sigma}_{\mathbf{v}_1}, \ldots, \mathbf{\Sigma}_{\mathbf{v}_K})$ is the covariance matrix of $\mathbf{v}$. The considered setup is summarized in Figure 1.

4. SENSOR SELECTION

The sensor selection problem consists of selecting a suitable subset of sensors of cardinality $\kappa < K$, maximizing an appropriate performance metric. In practice, the system controller accepts to estimate the quantity $\mathbf{x}$ by using the signals received by κ sensors only, out of the K available sensors, at the advantage of a lower complexity. Clearly, better estimates can be obtained if the amount of information transmitted by the selected sensor is higher. Therefore, in the selection problem it is natural to consider the maximization of the mutual information between the selected signals and the measured vector $\mathbf{x}$.

Sensor selection can be mathematically described by left-multiplying the vector $\mathbf{r}$ by the $\kappa n_r \times Kn_r$ *selection matrix* $\mathbf{S}$. Such matrix is made of $\kappa \times K$ blocks of size $n_r \times n_r$ each. The (i,j)-th block of $\mathbf{S}$, $\mathbf{S}_{ij}$, is defined as

$$\mathbf{S}_{ij} \doteq \begin{cases} \mathbf{I} & \text{if sensor } i \text{ is the } j\text{-th selected sensor} \\ \mathbf{0} & \text{otherwise} \end{cases}$$

for $i = 1, \ldots, K$ and $j = 1, \ldots, \kappa$. Also $\mathbf{S}$ is such that for each $j = 1, \ldots, \kappa$ there exists one and only one index i such that $\mathbf{S}_{ij} = \mathbf{I}$ (i.e., exactly κ sensors are selected) and for each $i = 1, \ldots, K$ there is no more than one index j such that $\mathbf{S}_{ij} = \mathbf{I}$ (i.e., sensors are selected no more than once).

Optimal sensor selection can thus be formulated as the problem of choosing the selection matrix $\widehat{\mathbf{S}}$ that maximizes the mutual information $\mathcal{I}(\mathbf{Sr}; \mathbf{x})$, i.e.,

$$\widehat{\mathbf{S}} = \arg\max_{\mathbf{S}} \ \mathcal{I}(\mathbf{Sr}, \mathbf{x}). \tag{10}$$

The maximum mutual information obtained by processing signals received from κ sensors is thus given by $\mathcal{I}(\widehat{\mathbf{S}}\mathbf{r}, \mathbf{x})$. Note that with this formulation, selection is made based on the knowledge of the sensor/channel characteristics only (i.e. the matrices $\mathbf{C}_i, \mathbf{H}_i$ and the noise covariances). This selection is made *prior* to transmitting the actual measurements, thus allowing to switch off the non-selected sensors. As a consequence, sensor selection not only results in a complexity reduction at the controller, but also in a possibly large energy saving in transmission.

4.1 Problem reformulation

By using the definition given in (3) and by assuming that $\mathbf{x}$ is Gaussian distributed with zero mean and covariance matrix $\mathbf{\Sigma}_\mathbf{x}$, the mutual information $\mathcal{I}(\mathbf{Sr}, \mathbf{x})$ can be rewritten in terms of differential entropies as

$$\mathcal{I}(\mathbf{Sr}, \mathbf{x}) = h(\mathbf{Sr}) - h(\mathbf{Sr}|\mathbf{x})$$

We then observe that $\mathbf{Sr}|\mathbf{x}$ is a multivariate complex Gaussian random variable with mean $\mathbf{S\Phi x}$ and covariance $\mathbf{S\Sigma}_{\boldsymbol{\eta}}\mathbf{S}^\mathsf{H}$, while $\mathbf{Sr}$ has zero mean and covariance $\mathbf{S}(\mathbf{\Phi\Sigma_x\Phi}^\mathsf{H} + \mathbf{\Sigma}_{\boldsymbol{\eta}})\mathbf{S}^\mathsf{H}$.

From (5) it follows that

$$\begin{aligned}
\mathcal{I}(\mathbf{Sr}, \mathbf{x}) &= \log\left|\pi e(\mathbf{S\Phi\Sigma_x\Phi}^\mathsf{H}\mathbf{S}^\mathsf{H} + \mathbf{S\Sigma}_{\boldsymbol{\eta}}\mathbf{S}^\mathsf{H})\right| \\
&\quad - \log\left|\pi e(\mathbf{S\Sigma}_{\boldsymbol{\eta}}\mathbf{S}^\mathsf{H})\right| \\
&= \log\left(\left|\mathbf{S\Phi\Sigma_x\Phi}^\mathsf{H}\mathbf{S}^\mathsf{H} + \mathbf{S\Sigma}_{\boldsymbol{\eta}}\mathbf{S}^\mathsf{H}\right|\left|\mathbf{S\Sigma}_{\boldsymbol{\eta}}\mathbf{S}^\mathsf{H}\right|^{-1}\right) \\
&= \log\left(\left|\mathbf{S\Phi\Sigma_x\Phi}^\mathsf{H}\mathbf{S}^\mathsf{H} + \mathbf{S\Sigma}_{\boldsymbol{\eta}}\mathbf{S}^\mathsf{H}\right|\left|(\mathbf{S\Sigma}_{\boldsymbol{\eta}}\mathbf{S}^\mathsf{H})^{-1}\right|\right) \\
&= \log\left|\mathbf{S\Phi\Sigma_x\Phi}^\mathsf{H}\mathbf{S}^\mathsf{H}(\mathbf{S\Sigma}_{\boldsymbol{\eta}}\mathbf{S}^\mathsf{H})^{-1} + \mathbf{I}\right|
\end{aligned} \tag{11}$$

In order to proceed further we observe that the covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\eta}}$ has a block-diagonal structure with blocks of size $n_r \times n_r$. Such a property is due to the fact that both $\mathbf{H}$ and $\mathbf{\Sigma}_\mathbf{v}$ are block-diagonal matrices. In particular $\mathbf{\Sigma}_\mathbf{v}$ is block-diagonal due to the independence of the measurement noises $\mathbf{v}_k$. Therefore we can write $(\mathbf{S\Sigma}_{\boldsymbol{\eta}}\mathbf{S}^\mathsf{H})^{-1} = \mathbf{S\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{S}^\mathsf{H}$, whence it follows that the mutual information (11) can be rewritten as

$$\begin{aligned}
\mathcal{I}(\mathbf{Sr}, \mathbf{x}) &= \log\left|\mathbf{S\Phi\Sigma_x\Phi}^\mathsf{H}\mathbf{S}^\mathsf{H}\mathbf{S\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{S}^\mathsf{H} + \mathbf{I}\right| \\
&\overset{(a)}{=} \log\left|\mathbf{\Sigma_x\Phi}^\mathsf{H}\mathbf{S}^\mathsf{H}\mathbf{S\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{S}^\mathsf{H}\mathbf{S\Phi} + \mathbf{I}\right| \\
&= \log\left|\mathbf{\Sigma_x\Phi}^\mathsf{H}\mathbf{Z\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{Z\Phi} + \mathbf{I}\right|
\end{aligned} \tag{12}$$

where (a) is due to the fact that for any two matrices $\mathbf{A}$ and $\mathbf{B}$, the equality $|\mathbf{I}+\mathbf{AB}| = |\mathbf{I}+\mathbf{BA}|$ holds (Sylvester's Theorem). Moreover $\mathbf{Z} = \mathbf{S}^\mathsf{H}\mathbf{S}$ is a $Kn_r \times Kn_r$ block-diagonal matrix with blocks $\mathbf{Z}_1, \ldots, \mathbf{Z}_K$. Specifically the i-th block of $\mathbf{Z}$, $\mathbf{Z}_i$, can be written as $\mathbf{Z}_i = z_i\mathbf{I}$ where $z_i \in \{0,1\}$, $i = 1, \ldots, K$ are Boolean variables.

Hence, sensor selection can be summarized by the *boolean selection vector* $\mathbf{z} \in \{0,1\}^K$ defined as

$$\mathbf{z} \doteq [z_1 \cdots z_K]^\mathsf{T} \tag{13}$$

were $z_i = 1$ means that the i-th sensor has been selected. Note that the matrix $\mathbf{Z}$ is idempotent (i.e., $\mathbf{Z}^2 = \mathbf{Z}$) and that, because of the block-diagonal structure of $\mathbf{\Sigma}_{\boldsymbol{\eta}}$, $\mathbf{Z\Sigma}_{\boldsymbol{\eta}} = \mathbf{\Sigma}_{\boldsymbol{\eta}}\mathbf{Z}$. It follows that the determinant in the last line of (12) can be rewritten as

$$\begin{aligned}
\left|\mathbf{\Sigma_x\Phi}^\mathsf{H}\mathbf{Z\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{Z\Phi} + \mathbf{I}\right| &= \left|\mathbf{\Sigma_x\Phi}^\mathsf{H}\mathbf{ZZ\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{\Phi} + \mathbf{I}\right| \\
&= \left|\mathbf{\Sigma_x\Phi}^\mathsf{H}\mathbf{Z\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{\Phi} + \mathbf{I}\right|
\end{aligned} \tag{14}$$

Finally, we observe that the matrix $\mathbf{Z\Sigma}_{\boldsymbol{\eta}}^{-1}$ is block-diagonal and its i-th block is given by $z_i\mathbf{\Sigma}_{\boldsymbol{\eta}_i}^{-1}$ where $\mathbf{\Sigma}_{\boldsymbol{\eta}_i}$ is the i-th diagonal block of $\mathbf{\Sigma}_{\boldsymbol{\eta}}$ and is given by

$$\mathbf{\Sigma}_{\boldsymbol{\eta}_i} = \mathbf{H}_i\mathbf{\Sigma}_{\mathbf{v}_i}{\mathbf{H}_i}^\mathsf{H} + \sigma^2\mathbf{I}$$

thus

$$\begin{aligned}
\mathbf{\Phi}^\mathsf{H}\mathbf{Z\Sigma}_{\boldsymbol{\eta}}^{-1}\mathbf{\Phi} &= \sum_{i=1}^{K} z_i{\mathbf{C}_i}^\mathsf{H}{\mathbf{H}_i}^\mathsf{H}\left(\mathbf{H}_i\mathbf{\Sigma}_{\mathbf{v}_i}{\mathbf{H}_i}^\mathsf{H} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{H}_i\mathbf{C}_i \\
&= \sum_{i=1}^{K} z_i\mathbf{A}_i
\end{aligned}$$

where

$$\mathbf{A}_i = {\mathbf{C}_i}^\mathsf{H}{\mathbf{H}_i}^\mathsf{H}\left(\mathbf{H}_i\mathbf{\Sigma}_{\mathbf{v}_i}{\mathbf{H}_i}^\mathsf{H} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{H}_i\mathbf{C}_i. \tag{15}$$

By substituting (15), (14) in (12), the mutual information can be expressed as

$$\mathcal{I}(\mathbf{Sr}, \mathbf{x}) = \log \left| \sum_{i=1}^{K} z_i \mathbf{\Sigma}_{\mathbf{x}} \mathbf{A}_i + \mathbf{I} \right| \tag{16}$$

The previous developments allow us to state the next proposition, which provides a formulation of the optimal sensor selection problem in terms of the selection vector $\mathbf{z}$.

Proposition 1. (Optimal wireless sensor selection). *The selection vector $\hat{\mathbf{z}}$ that maximizes the mutual information is given by the solution of the following optimization problem*

$$\begin{aligned} \widehat{\mathbf{z}} = \arg \max_{\|\mathbf{z}\|_0 = \kappa} \log & \left| \sum_{i=1}^{K} z_i \mathbf{\Sigma}_{\mathbf{x}} \mathbf{A}_i + \mathbf{I} \right| \\ \text{subject to} \quad & z_i \in \{0, 1\}, \ \ i = 1, \ldots, K \end{aligned} \tag{17}$$

where $\|\mathbf{z}\|_0$ denotes the ℓ_0 norm, corresponding to the number of nonzero elements in $\mathbf{z}$.

5. NUMERICAL SOLUTION

It should be noted that solving the optimal wireless sensor selection problem formulated in Proposition 1 amounts at solving a hard non-convex optimization problem, due to its intrinsic combinatorial nature. In fact, it requires in principle to evaluate the performance index for each of the $\binom{K}{\kappa}$ possible choices of sensor measurements. Specific branch-and-bound techniques, in the spirit of Lawler and Wood (1966); Welch (1982), can be devised for the numerical solution of this problem, but this approach is clearly not practical unless κ and K are relatively small. In the other cases, different approximations or relaxations are possible, as discussed in the next subsections.

5.1 Convex relaxation

It is immediate to observe that the optimization problem (17) can be relaxed into a semi-definite program (SDP) by relaxing the requirement that the selection vector $\mathbf{z}$ is to be binary. Indeed, the cost function $\log \left| \sum_{i=1}^{K} z_i \mathbf{\Sigma}_{\mathbf{x}} \mathbf{A}_i + \mathbf{I} \right|$ is a concave function for $\mathbf{z} \in \mathbb{R}^K$. This approach, which is the analogous of that introduced in Joshi and Boyd (2009), leads to the following concave relaxation

$$\begin{aligned} \widehat{\mathbf{z}}^{\mathrm{CR}} = \arg \max_{\|\mathbf{z}\|_1 = \kappa} \log & \left| \sum_{i=1}^{K} z_i \mathbf{\Sigma}_{\mathbf{x}} \mathbf{A}_i + \mathbf{I} \right| \\ \text{subject to} \quad & z_i \in [0, 1], \quad i = 1, \ldots, K \end{aligned} \tag{18}$$

Note that in the above maximization problem, the ℓ_0-norm is replaced by the ℓ_1-norm. This is a standard technique applied in optimization to derive convex relaxations to combinatorial problems, and has been applied, e.g., in the context of compressive sensing in Donoho (2006). The complexity of this SDP algorithm scales as $\mathcal{O}(K^3)$, and hence it can be applied also to rather large networks. Note that the solution of this SDP relaxation will in general take fractional values, and some kind of sorting and rounding is necessary to obtain the desired solution. The simplest approach consists in selecting the κ elements of $\widehat{\mathbf{z}}^{\mathrm{CR}}$ with the largest values. A more sophisticated approach, which in general leads to better solutions, consists in applying an iterative procedure, where each iteration solves a *weighted* ℓ_1-problem. This procedure has been suggested in Candés et al. (2008) to enhance the sparsity of the solution of ℓ_1-minimization problems.

5.2 Greedy algorithm

A second approach to derive a computable approximate solution to the wireless sensor selection problem discussed in Proposition 1 consists in applying a greedy procedure, that is by choosing the sensors one at a time, until κ sensors are finally selected. At the ν-th step ($1 \leq \nu \leq \kappa$), the selected sensor is the one maximizing the objective function when combined with the previously chosen $\nu - 1$. The procedure is described in Algorithm 1. As it can be seen, in this case the objective function can be rewritten as the following set-function (where a set-function is a function whose input is a given set $\mathcal{K}$)

$$I(\mathcal{K}) \doteq \log \left| \sum_{i \in \mathcal{K}} \mathbf{\Sigma}_{\mathbf{x}} \mathbf{A}_i + \mathbf{I} \right| \tag{19}$$

Algorithm 1 GREEDY SENSOR SELECTION

(1) Initialization: $\nu = 1, \mathcal{K}^{(1)} = \emptyset, \overline{\mathcal{K}}^{(1)} = \{1, \ldots, K\}$
(2) Select greedily the next sensor:

$$\widehat{k}_\nu = \arg \max_{j \in \overline{\mathcal{K}}^{(\nu)}} I(\mathcal{K}^{(\nu)} \cup j) \tag{20}$$

(3) Update the measurement set:

$$\mathcal{K}^{(\nu+1)} = \mathcal{K}^{(\nu)} \cup \widehat{k}_\nu, \quad \overline{\mathcal{K}}^{(\nu+1)} = \overline{\mathcal{K}}^{(\nu)} \setminus \widehat{k}_\nu \tag{21}$$

(4) Set $\nu \leftarrow \nu + 1$, IF $\nu \leq \kappa$ GOTO 2.

The use of Algorithm 1 in the context of sensor selection has been proposed in Shamaiah et al. (2012), where its suboptimality properties are discussed in details, based on the concept of submodularity [2]. Submodularity plays for discrete functions the same role than convexity for continuous functions, see e.g. the survey Krause and Golovin (2014), and has been leveraged in various problems in the contexts of optimal sensor placement (Krause et al., 2011) and leader selection (Clark et al., 2014).

Lemma 1. (Suboptimality of the greedy algorithm). *Let $\hat{\mathcal{I}}_{\mathrm{GR}}$ be the mutual information obtained with the solution of Algorithm 1 and let $\hat{\mathcal{I}}$ be the optimal solution of the problem in* (17)*. Then,*

$$\hat{\mathcal{I}}_{\mathrm{GR}} \geq \left(1 - \frac{1}{\mathrm{e}}\right) \hat{\mathcal{I}} \tag{22}$$

Proof. The lemma is immediately proved by observing that $I(\mathcal{K})$ in (19) is submodular and monotone for given $\mathcal{K}$, and applying the reasoning in (Shamaiah et al., 2012, Lemma 1).

6. HIGH- AND LOW-SNR REGIMES

In this section, we consider the sensor selection problem in the case of high and low signal-to-noise ratio (SNR) on the wireless channel. Specifically we first consider the case $\sigma^2 \to 0$ (high-SNR regime) and then the case $\sigma^2 \to +\infty$ (low-SNR regime).

For the high-SNR regime, we will make the further assumption that $n_r \geq n_y$ and that the channel matrices $\mathbf{H}_i$ are full rank.

[2] For a given finite set $\mathcal{K}$, a set-function $f : 2^{\mathcal{K}} \to \mathbb{R}$, where $2^{\mathcal{K}}$ denotes the power set, is said to be *submodular* iff $\forall A \subseteq B \subseteq \mathcal{K}$, and $\forall j \in \mathcal{K} \setminus B$, it holds $f(A + j) - f(A) \geq f(B + j) - f(B)$, that is the function f satisfies the so-called diminishing increments property. Note that the submodular function f is monotone if $f(A) \leq f(B), \forall A \subseteq B$.

Proposition 2. (High-SNR regime). *If $n_r \geq n_y$ and the matrices $\mathbf{H}_i$, $i = 1, \ldots, K$, are all full rank then, for $\sigma^2 \to 0$, the optimal sensor selection problem defined in* (17) *can be rewritten as*

$$\widehat{\mathbf{z}}^{\mathrm{H}} = \underset{\|\mathbf{z}\|_0=\kappa}{\arg\max} \ \log\left|\sum_{i=1}^{K} z_i \mathbf{\Sigma_x} \mathbf{C}_i{}^{\mathsf{H}} \mathbf{\Sigma}_{\mathbf{v}_i}^{-1} \mathbf{C}_i + \mathbf{I}\right|$$
$$\text{subject to}$$
$$z_i \in \{0,1\}, \ \ i = 1, \ldots, K \tag{23}$$

Proof. The term $\left(\mathbf{H}_i \mathbf{\Sigma}_{\mathbf{v}_i} \mathbf{H}_i{}^{\mathsf{H}} + \sigma^2 \mathbf{I}\right)^{-1} \mathbf{H}_i$ in (15) can be rewritten as

$$\begin{aligned}
&\left(\mathbf{H}_i \mathbf{\Sigma}_{\mathbf{v}_i} \mathbf{H}_i{}^{\mathsf{H}} + \sigma^2 \mathbf{I}\right)^{-1} \mathbf{H}_i \\
&\quad = \frac{1}{\sigma^2}\left(\frac{\mathbf{H}_i \mathbf{\Sigma}_{\mathbf{v}_i} \mathbf{H}_i{}^{\mathsf{H}}}{\sigma^2} + \mathbf{I}\right)^{-1} \mathbf{H}_i \\
&\quad = \left(\frac{\mathbf{H}_i \mathbf{\Sigma}_{\mathbf{v}_i} \mathbf{H}_i{}^{\mathsf{H}}}{\sigma^2} + \mathbf{I}\right)^{-1} \frac{\mathbf{H}_i \mathbf{\Sigma}_{\mathbf{v}_i}^{1/2} \mathbf{\Sigma}_{\mathbf{v}_i}^{-1/2}}{\sigma^2} \\
&\quad = \left(\widetilde{\mathbf{H}}_i \widetilde{\mathbf{H}}_i{}^{\mathsf{H}} + \mathbf{I}\right)^{-1} \widetilde{\mathbf{H}}_i \frac{\mathbf{\Sigma}_{\mathbf{v}_i}^{-1/2}}{\sigma}
\end{aligned} \tag{24}$$

where $\widetilde{\mathbf{H}}_i = \mathbf{H}_i \mathbf{\Sigma}_{\mathbf{v}_i}^{1/2} / \sigma$. We then observe that for any matrix $\widetilde{\mathbf{H}}_i$ the following equality holds

$$\left(\widetilde{\mathbf{H}}_i \widetilde{\mathbf{H}}_i{}^{\mathsf{H}} + \mathbf{I}\right)^{-1} \widetilde{\mathbf{H}}_i = \widetilde{\mathbf{H}}_i \left(\widetilde{\mathbf{H}}_i{}^{\mathsf{H}} \widetilde{\mathbf{H}}_i + \mathbf{I}\right)^{-1}$$

hence

$$\begin{aligned}
\left(\mathbf{H}_i \mathbf{\Sigma}_{\mathbf{v}_i} \mathbf{H}_i{}^{\mathsf{H}} + \sigma^2 \mathbf{I}\right)^{-1} \mathbf{H}_i &= \widetilde{\mathbf{H}}_i \left(\widetilde{\mathbf{H}}_i{}^{\mathsf{H}} \widetilde{\mathbf{H}}_i + \mathbf{I}\right)^{-1} \frac{\mathbf{\Sigma}_{\mathbf{v}_i}^{-1/2}}{\sigma} \\
&= \mathbf{H}_i \left(\mathbf{\Sigma}_{\mathbf{v}_i} \mathbf{H}_i{}^{\mathsf{H}} \mathbf{H}_i + \sigma^2 \mathbf{I}\right)^{-1}
\end{aligned} \tag{25}$$

It follows that in the high-SNR regime the term involving the matrix $\mathbf{H}_i$ in (15) can be rewritten as

$$\begin{aligned}
&\lim_{\sigma^2 \to 0} \mathbf{H}_i{}^{\mathsf{H}} \left(\mathbf{H}_i \mathbf{\Sigma}_{\mathbf{v}_i} \mathbf{H}_i{}^{\mathsf{H}} + \sigma^2 \mathbf{I}\right)^{-1} \mathbf{H}_i \\
&\quad = \lim_{\sigma^2 \to 0} \mathbf{H}_i{}^{\mathsf{H}} \mathbf{H}_i \left(\mathbf{\Sigma}_{\mathbf{v}_i} \mathbf{H}_i{}^{\mathsf{H}} \mathbf{H}_i + \sigma^2 \mathbf{I}\right)^{-1} \\
&\quad = \mathbf{\Sigma}_{\mathbf{v}_i}^{-1}
\end{aligned} \tag{26}$$

provided that $\mathbf{H}_i{}^{\mathsf{H}} \mathbf{H}_i$ is invertible. Under such conditions, (17) turns to (23).

Note that in the high-SNR regime the sensor selection problem does not depend on the channel matrices $\mathbf{H}_i$, as far as they are full rank and $n_r \leq n_y$. This result is also of importance whenever the channel rapidly varies over time, so that it cannot be tracked optimally or, in other words, when the channel coherence time is much lower than the time required to choose the optimal set of sensors.

Proposition 3. (Low-SNR regime). *If $\sigma^2 \to \infty$, the optimal sensor selection problem in* (17) *reduces to*

$$\widehat{\mathbf{z}}^{\mathrm{L}} = \underset{\|\mathbf{z}\|_0=\kappa}{\arg\max} \ \sum_{i=1}^{K} z_i \mathsf{Tr}\left\{\mathbf{\Sigma_x} \mathbf{C}_i{}^{\mathsf{H}} \mathbf{H}_i{}^{\mathsf{H}} \mathbf{H}_i \mathbf{C}_i\right\}$$
$$\text{subject to}$$
$$z_i \in \{0,1\}, \ \ i = 1, \ldots, K \tag{27}$$

Proof. We start from the expression of the mutual information given in (12) and we observe that, for $\sigma^2 \to \infty$, $\mathbf{\Sigma}_{\boldsymbol{\eta}}^{-1} \simeq \sigma^{-2}\mathbf{I}$. Under this condition and from (12) we have:

$$\begin{aligned}
\mathcal{I}(\mathbf{Sr}, \mathbf{x}) &\simeq \log\left|\sigma^{-2} \mathbf{\Sigma_x} \mathbf{\Phi}^{\mathsf{H}} \mathbf{Z} \mathbf{Z} \mathbf{\Phi} + \mathbf{I}\right| \\
&= \log\left|\sigma^{-2} \mathbf{\Sigma_x} \mathbf{\Phi}^{\mathsf{H}} \mathbf{Z} \mathbf{\Phi} + \mathbf{I}\right| \\
&\simeq \sigma^{-2} \mathsf{Tr}\left\{\mathbf{\Sigma_x} \mathbf{\Phi}^{\mathsf{H}} \mathbf{Z} \mathbf{\Phi}\right\} \\
&= \sigma^{-2} \sum_{i=1}^{K} z_i \mathsf{Tr}\left\{\mathbf{\Sigma_x} \mathbf{C}_i{}^{\mathsf{H}} \mathbf{H}_i{}^{\mathsf{H}} \mathbf{H}_i \mathbf{C}_i\right\}
\end{aligned} \tag{28}$$

where we used the fact that for $\epsilon \to 0$, $|\mathbf{I} + \epsilon \mathbf{A}| = 1 + \epsilon \mathsf{Tr}\{\mathbf{A}\} + o(\epsilon^2)$, and $\log(1 + \epsilon \mathsf{Tr}\{\mathbf{A}\}) \simeq \epsilon \mathsf{Tr}\{\mathbf{A}\}$. Under the $\arg\max$ operator the constant terms independent of $\mathbf{z}$ can be removed so that the optimal sensor selection can be written as in (27).

Note that the problem in (27) can easily be solved by a greedy algorithm that takes the κ largest values of

$$p_i = \mathsf{Tr}\left\{\mathbf{\Sigma_x} \mathbf{C}_i{}^{\mathsf{H}} \mathbf{H}_i{}^{\mathsf{H}} \mathbf{H}_i \mathbf{C}_i\right\},$$

$i = 1, \ldots, K$. Also, the solution depends on the channel matrices $\mathbf{H}_i$ (as opposed to the high-SNR case). Moreover, the value of p_i can be interpreted as the total useful power reaching the receiver from the i-th sensor.

7. NUMERICAL RESULTS

In this section, we show some simulation results to assess the performance of the different sensor selection algorithms described in the previous sections.

In our tests, three sensors have to be chosen out of a set of twenty, i.e., $K = 20$ and $\kappa = 3$. The vector $\mathbf{x}$ to be estimated has length $n_x = 4$ and a covariance matrix $\mathbf{\Sigma_x} = \mathbf{I}$. Each sensor performs $n_y = 2$ scalar measurements, while the receiver is equipped with $n_r = 2$ antennas. The measurement noise vectors have covariance matrices $\mathbf{\Sigma}_{\mathbf{v}_i} = \mathbf{I}$, $k = 1, \ldots, K$. The entries of the $\mathbf{C}_k$ matrices, $k = 1, \ldots, K$, are chosen i.i.d. from a zero-mean real Gaussian distribution, and then the $\mathbf{C}_k$'s are normalized to a constant Frobenius norm and kept fixed for the whole simulation. The entries of the $\mathbf{H}_k$ matrices, $k = 1, \ldots, K$, are chosen i.i.d. from a zero-mean unit-variance circular complex Gaussian distribution, as in the case where the wireless channel is affected by Rayleigh fading. We define the channel signal-to-noise ratio as $\mathrm{SNR} = 1/\sigma^2$.

In a first test, whose results are depicted in Fig. 2 we kept fixed the set of $\mathbf{H}_k$'s for the whole simulation. The figure shows the mutual information achieved by the optimal sensor selection of Prop. 1, together with the mutual information achieved by the high-SNR and low-SNR versions of the same optimal selection, as defined in Props. 2 and 3, respectively. In the lower part of Fig. 2, we show the selected sensor set by each algorithm as a function of SNR. As it can be seen, the low-SNR algorithm of Prop. 3 merges with the optimal selection only for SNRs lower than -25 dB, while it is largely suboptimal for medium-to-high SNRs. The high-SNR algorithm of Prop. 2 merges with the optimal selection for $\mathrm{SNR} = 14$ dB, while it loses up to 1.5 bits of mutual information for medium SNR values. The slope changes in the optimal selection curve correspond to changes in the selected sensor set.

In a second test, depicted in Fig. 3 we average over 100 Rayleigh-fading channel realizations. The figure shows the mutual information achieved by the optimal sensor selection of

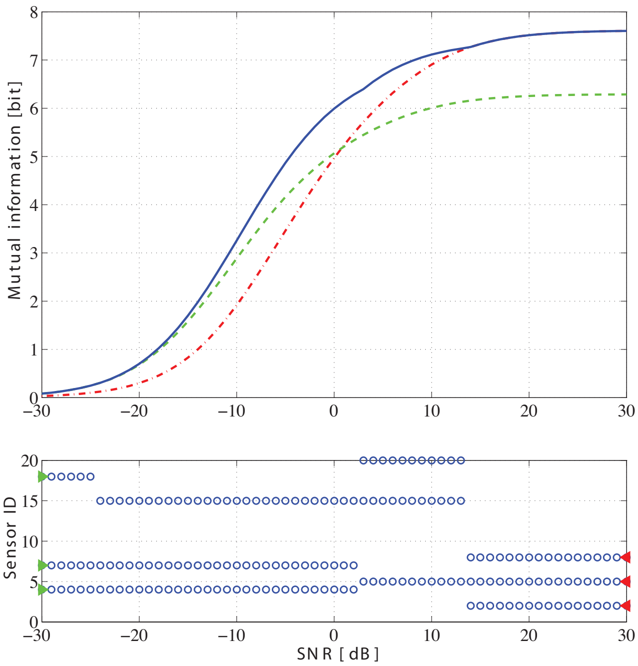

Fig. 2. (Top) Achievable mutual information versus SNR for the optimal wireless sensor selection (solid blue curve) and its counterparts at high SNR (dash-dotted red) and low SNR (dashed green). (Bottom) Optimal wireless sensor selection for different SNR (blue circle markers), and low- and high-SNR counterparts (green and red triangle markers). Fixed channel realization.

Proposition 1, a re-weighted implementation of the convex relaxation described in Section 5.1 and the greedy sensor selection of Algorithm 1. As it can be seen, the two approximated algorithms reach the optimal performance up to $\mathrm{SNR} = 4$ dB, while they become slightly suboptimal for higher SNR, while neither of the two dominates the other. Notice that both suboptimal solutions exhibit bumps, i.e., the achieved mutual information is not everywhere an increasing function of SNR, a fact which is probably related to the choice of the $\mathbf{C}_k$'s. Finally notice that the penalty incurred by the approximated algorithms always remains within 1 bit of mutual information.

8. CONCLUSIONS

In this paper, we studied the problem of optimal sensor selection in the case when the sensed information is to be transmitted through a noisy wireless channel. The optimality conditions have been derived in an information-theoretic framework, and numerical simulations clearly show the importance of taking into account channel state information to perform the best sensor selection. Further studies will consider the problem of remote Kalman filtering in the same wireless context. Finally, the possibility of introducing additional constraints in the choice of the sensors can be considered, as in Wang et al. (2013).

ACKNOWLEDGMENTS

The authors would like to thank Roberto Tempo for the very useful discussions and suggestions, and Mario Sznaier for motivating the research on sensor selection.

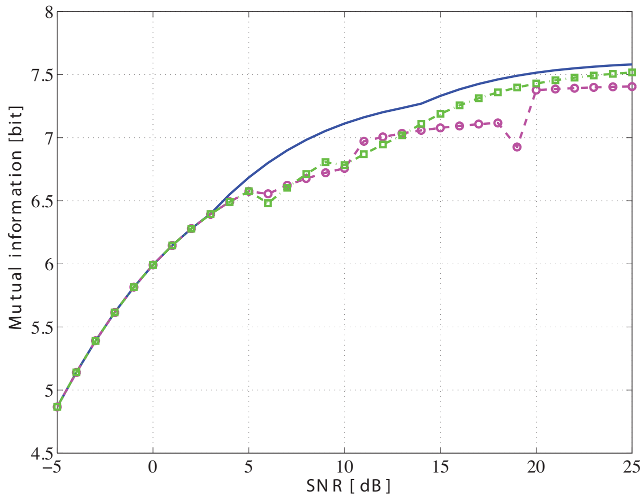

Fig. 3. Comparison among the optimal sensor selection (blue curve), the convex-relaxation algorithm (red) and the greedy sensor selection (green) in terms of achievable mutual information versus SNR. Average over 100 channel realizations.

REFERENCES

E.J. Candés, M.B. Wakin, and S.P. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier Analysis and Applications*, 14:877–905, 2008.

A. Clark, L. Bushnell, and R. Poovendran. A supermodular optimization framework for leader selection under link noise in linear multi-agent systems. *IEEE Transactions on Automatic Control*, 59(2):283–297, 2014.

D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

S. Joshi and S. Boyd. Sensor selection via convex optimization. *IEEE Trans. on Sig. Proc.*, 57:451–462, 2009.

A. Krause and D. Golovin. Submodular function maximization. In *Tractability: Practical Approaches to Hard Problems (to appear)*. Cambridge University Press, February 2014.

A. Krause, R. Rajagopal, A. Gupta, and C. Guestrin. Simultaneous optimization of sensor placements and balanced schedules. *IEEE Transactions on Automatic Control*, 56(10): 2390–2405, October 2011.

E.L. Lawler and D.E. Wood. Branch-and-bound methods: A survey. *Operations Research*, 14:699–719, 1966.

Y. Mo, R. Ambrosino, and B. Sinopoli. Sensor selection strategies for state estimation in energy constrained wireless sensor networks. *Automatica*, 47:1330–1338, 2011.

M. Shamaiah, S. Banerjee, and H. Vikalo. Greedy sensor selection under channel uncertainty. *IEEE Wireless Communications Letters*, 1(4):376–379, 2012.

Y. Wang, M. Sznaier, and F. Dabbene. A convex optimization approach to worst-case optimal sensor selection. In *Proceedings IEEE Conference on Decision and Control*, 2013.

W. Welch. Branch-and-bound search for experimental designs based on D-optimality and other criteria. *Technometrics*, 24 (1):41–48, 1982.