Sensor Selection for Estimation with Correlated Measurement Noise

Sijia Liu, *Student Member, IEEE,* Sundeep Prabhakar Chepuri, *Member, IEEE,* Makan Fardad, *Member, IEEE,* Engin Masazade, *Member, IEEE,* Geert Leus, *Fellow, IEEE,* and Pramod K. Varshney, *Fellow, IEEE*

Abstract

In this paper, we consider the problem of sensor selection for parameter estimation with correlated measurement noise. We seek optimal sensor activations by formulating an optimization problem, in which the estimation error, given by the trace of the inverse of the Bayesian Fisher information matrix, is minimized subject to energy constraints. Fisher information has been widely used as an effective sensor selection criterion. However, existing information-based sensor selection methods are limited to the case of uncorrelated noise or weakly correlated noise due to the use of approximate metrics. By contrast, here we derive the closed form of the Fisher information matrix with respect to sensor selection variables that is valid for any arbitrary noise correlation regime, and develop both a convex relaxation approach and a greedy algorithm to find near-optimal solutions. We further extend our framework of sensor selection to solve the problem of sensor scheduling, where a greedy algorithm is proposed to determine non-myopic (multi-time step ahead) sensor schedules. Lastly, numerical results are provided

S. Liu, M. Fardad and P. K. Varshney are with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, 13244 USA e-mail: {sliu17, makan, varshney}@syr.edu.

S. P. Chepuri and G. Leus are with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, The Netherlands. Email: {s.p.chepuri, g.j.t.leus}@tudelft.nl.

E. Masazade is with the Department of Electrical and Electronics Engineering, Yeditepe University, Istanbul, 34755, Turkey e-mail: engin.masazade@yeditepe.edu.tr.

The work of S. Liu and P. K. Varshney was supported by the U.S. Air Force Office of Scientific Research (AFOSR) under grants FA9550-10-1-0458. The work of M. Fardad was supported by the National Science Foundation (NSF) under awards EAGER ECCS-1545270 and CNS-1329885. The work of S. P. Chepuri and G. Leus is supported by NWO-STW under the VICI program (10382). The work of E. Masazade was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under grant 113E220.

to illustrate the effectiveness of our approach, and to reveal the effect of noise correlation on estimation performance.

Index Terms

Sensor selection, sensor scheduling, parameter estimation, correlated noise, convex relaxation.

I. Introduction

Wireless sensor networks consisting of a large number of spatially distributed sensors have been widely used for environmental monitoring, source localization, and target tracking [1]–[3]. Among the aforementioned applications, sensors observe an unknown parameter or state of interest and transmit their measurements to a fusion center, which then determines the global estimate. However, due to the constraints on the communication bandwidth and sensor battery life, it may not be desirable to have all the sensors report their measurements at all time instants. Therefore, the problem of sensor selection/scheduling arises, which aims to strike a balance between estimation accuracy and sensor activations over space and/or time. The importance of sensor selection has been discussed extensively in the context of various applications, such as target tracking [4], bit allocation [5], field monitoring [6], [7], optimal control [8], power allocation [9], [10], optimal experiment design [11], and leader selection in consensus networks [12].

In this paper, we focus on the problem of sensor selection/scheduling for parameter estimation similar to [12]–[15], but with a key difference in that the measurement noise is correlated in the problem formulation. In [13], the sensor selection problem was elegantly formulated under linear measurement models, and solved via convex optimization. In [14], the problem of sensor selection was generalized to nonlinear measurement models by using the Cramér-Rao bound as the sensor selection criterion. In [12], a particular class of sensor selection problems were transformed into the problem of leader selection in dynamical networks. In [15], the problem of non-myopic scheduling that determined sensor activations over multiple future time steps was addressed for nonlinear filtering with quantized measurements.

In the existing literature [12]–[15], the study of sensor selection/scheduling problems hinges on the assumption of *uncorrelated* measurement noise, which implies that sensor observations are conditionally independent given the underlying parameter. Due to conditional independence, each measurement contributes to Fisher information (equivalently, inverse of the Cramér-Rao bound on the error covariance matrix) in an additive manner [16]. Accordingly, Fisher information becomes a *linear* function with respect to the sensor selection variables (which characterize the subset of sensors we select), and thus the resulting selection problem can be efficiently handled via convex optimization [13], [14]. However, the

sensed data is often corrupted by correlated noise due to the nature of the monitored physical environment [17]. Therefore, development of sensor selection schemes for *correlated* measurements is a critical task.

Recently, it has been shown in [18]–[21] that the presence of correlated noise makes optimal sensor selection/scheduling problems more challenging, since Fisher information is no longer a linear function with respect to the selection variables. In [18]–[20], the problem of sensor selection with correlated noise was formulated so as to minimize an approximate expression of the estimation error subject to an energy constraint or to minimize the energy consumption subject to an approximate estimation constraint. In [21], a reformulation of the multi-step Kalman filter was introduced to schedule sensors for linear dynamical systems with correlated noise.

Different from [18]–[21], here we derive the closed form expression of the estimation error with respect to sensor selection variables under correlated measurement noise, which is valid for any arbitrary noise correlation matrix. This expression is optimized via a convex relaxation method to determine the optimal sensor selection scheme. We also propose a greedy algorithm to solve the corresponding sensor selection problem, where we show that when an inactive sensor is made active, the increase in Fisher information yields an information gain in terms of a rank-one matrix. The proposed sensor selection framework yields a more accurate sensor selection scheme than those presented in [18]–[20], because the schemes of [18]–[20] consider an approximate formulation where the noise covariance matrix is assumed to be independent of the sensor selection variables. We further demonstrate that the prior formulations for sensor selection are valid only when measurement noises are *weakly* correlated. In this scenario, maximization of the trace of the Fisher information matrix used in [20] is equivalent to the problem of maximizing a convex quadratic function over a bounded polyhedron. The resutling problem structure enables the use of optimization methods with reduced computational complexity.

Compared to [21], we adopt the recursive Fisher information to measure the estimation performance of sensor scheduling. However, for non-myopic (multi-time step ahead) schedules, the Fisher information matrices at consecutive time steps are coupled with each other. Due to coupling, expressing the Fisher information matrices in a closed form is intractable. Therefore, we propose a greedy algorithm to seek non-myopic sensor schedules subject to cumulative and individual energy constraints. Numerical results show that our approach yields a better estimation performance than that of [21] for state tracking.

In a preliminary version of this paper [22], we studied the problem of sensor selection using the same framework as in [18]–[20]. Compared to [22], we have the following new contributions in this paper.

- We propose a more general but tractable sensor selection framework that is valid for an arbitrary noise correlation matrix, and present a suite of efficient optimization algorithms.
- We reveal drawbacks of the existing formulations in [18]–[20] for sensor selection, and demonstrate

their validity in only the weak noise correlation regime.

- We extend the proposed sensor selection approach to address the problem of non-myopic sensor scheduling, where the length of the time horizon and energy constraints on individual sensors are taken into account.

The rest of the paper is organized as follows. In Section II, we formulate the problem of sensor selection with correlated noise. In Section III, we present a convex relaxation approach and a greedy algorithm to solve the problem of sensor selection with an arbitrary noise correlation matrix. In Section IV, we present sensor selection approach with weakly correlated noise. In Section V, we extend our framework to solve the problem of non-myopic sensor scheduling. In Section VI, we provide numerical results to illustrate the effectiveness of our proposed methods. In Section VII, we summarize our work and discuss future research directions.

II. Problem Formulation

We wish to estimate a random vector $\mathbf{x} \in \mathbb{R}^n$ with a Gaussian prior probability density function (PDF) $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Observations of $\mathbf{x}$ from m sensors are corrupted by correlated measurement noise. To strike a balance between estimation accuracy and sensor activations, we formulate the problem of sensor selection, where the estimation error is minimized subject to a constraint on the total number of sensor activations.

Consider a linear system

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^m$ is the measurement vector whose mth entry corresponds to a scalar observation from the mth sensor, $\mathbf{H} \in \mathbb{R}^{m \times n}$ is the observation matrix, and $\mathbf{v} \in \mathbb{R}^m$ is the measurement noise vector that follows a Gaussian distribution with zero mean and an *invertible* covariance matrix $\mathbf{R}$. We assume that $\mathbf{x}$ and $\mathbf{v}$ are mutually independent random variables, and the noise covariance matrix is positive definite and thus invertible. We note that the noise covariance matrix is not restricted to being diagonal, so that the measurement noise could be *correlated* among the sensors. We also note that in practice, the first two moments of $\mathbf{x}$ can be learnt from a parametric covariance model, such as a power exponential model together with a training dataset of the parameter [23].

The task of sensor selection is to determine the best subset of sensors to activate in order to minimize the estimation error, subject to a constraint on the number of activations. We introduce a sensor selection vector to represent the activation scheme

$$\mathbf{w} = [w_1, w_2, \ldots, w_m]^T, \quad w_i \in \{0, 1\}, \tag{2}$$

where w_i indicates whether or not the ith sensor is selected. For example, if the ith sensor reports a measurement then $w_i = 1$, otherwise $w_i = 0$. In other words, the *active* sensor measurements can be compactly expressed as

$$\mathbf{y}_w = \mathbf{\Phi}_w \mathbf{y} = \mathbf{\Phi}_w \mathbf{H}\mathbf{x} + \mathbf{\Phi}_w \mathbf{v}, \tag{3}$$

where $\mathbf{y}_w \in \mathbb{R}^{\|\mathbf{w}\|_1}$ is the vector of measurements of selected sensors, $\|\mathbf{w}\|_1$ is the ℓ_1-norm of $\mathbf{w}$ which yields the total number of sensor activations, $\mathbf{\Phi}_w \in \{0,1\}^{\|\mathbf{w}\|_1 \times m}$ is a submatrix of $\mathrm{diag}(\mathbf{w})$ after all rows corresponding to the unselected sensors have been removed, and $\mathrm{diag}(\mathbf{w})$ is a diagonal matrix whose diagonal entries are given by $\mathbf{w}$. We note that $\mathbf{\Phi}_w$ and $\mathbf{w}$ are linked as below

$$\mathbf{\Phi}_w \mathbf{\Phi}_w^T = \mathbf{I}_w \quad \text{and} \quad \mathbf{\Phi}_w^T \mathbf{\Phi}_w = \mathrm{diag}(\mathbf{w}), \tag{4}$$

where $\mathbf{I}_w$ denotes an identity matrix with dimension $\|\mathbf{w}\|_1$.

A. Minimum mean-squared estimation error

We employ the minimum mean square error (MMSE) estimator to estimate the unknown parameter under the Bayesian setup. It is worth mentioning that the use of the Bayesian estimation framework ensures the validity of parameter estimation for an underdetermined system, in which the number of selected sensors is less than the dimension of the parameter to be estimated, namely, $\|\mathbf{w}\|_1 < n$.

Given the Gaussian linear measurement model (1), the prior distribution of the unknown parameter $\mathbf{x}$ and the active sensor measurements (3), the error covariance matrix of the MMSE estimate of $\mathbf{x}$ is given by [24, Theorem 12.1]

$$\mathbf{P}_w = \left(\mathbf{\Sigma}^{-1} + \mathbf{H}^T \mathbf{\Phi}_w^T \mathbf{R}_w^{-1} \mathbf{\Phi}_w \mathbf{H}\right)^{-1}, \tag{5}$$

where the matrix $\mathbf{\Phi}_w \mathbf{H}$ comprises rows of $\mathbf{H}$ for the active sensors, and $\mathbf{R}_w$ denotes the submatrix of $\mathbf{R}$ after all rows and columns corresponding to the inactive sensors have been removed, i.e.,

$$\mathbf{R}_w = \mathbf{\Phi}_w \mathbf{R} \mathbf{\Phi}_w^T. \tag{6}$$

It is clear from (5) that due to the presence of the prior knowledge about $\mathbf{\Sigma}$, the MSE matrix $\mathbf{P}_w$ is always well defined, even if the matrix $\mathbf{H}^T \mathbf{\Phi}_w^T \mathbf{R}_w^{-1} \mathbf{\Phi}_w \mathbf{H}$ is not invertible for an underdetermined system with $\|\mathbf{w}\|_1 \leq n$.

It is known from [16] that the MSE matrix $\mathbf{P}_w$ is the inverse of the Bayesian Fisher information matrix $\mathbf{J}_w$ under the linear Gaussian measurement model with a Gaussian prior distribution. We thus obtain

$$\begin{aligned} \mathbf{J}_w &= \mathbf{P}_w^{-1} \\ &= \mathbf{\Sigma}^{-1} + \mathbf{H}^T \mathbf{\Phi}_w^T \mathbf{R}_w^{-1} \mathbf{\Phi}_w \mathbf{H}, \end{aligned} \tag{7}$$

where the second term is related to the sensor selection scheme. In this paper, for clarity of presentation, we choose to work with $\mathbf{J}_w$ rather than $\mathbf{P}_w$.

It is clear from (6) and (7) that the dependence of $\mathbf{J}_w$ on $\mathbf{w}$ is through $\mathbf{\Phi}_w$. This dependency does not lend itself to easy optimization of scalar-valued functions of $\mathbf{J}_w$ with respect to $\mathbf{w}$. In what follows, we will rewrite $\mathbf{J}_w$ as an explicit function of the selection vector $\mathbf{w}$.

B. Fisher information $\mathbf{J}_w$ as an explicit function of $\mathbf{w}$

The key idea of expressing (7) as an explicit function of $\mathbf{w}$ is to replace $\mathbf{\Phi}_w$ with $\mathbf{w}$ based on their relationship given by (4). Consider a decomposition of the noise covariance matrix [25]

$$\mathbf{R} = a\mathbf{I} + \mathbf{S}, \tag{8}$$

where a positive scalar a is chosen such that the matrix $\mathbf{S}$ is positive definite, and $\mathbf{I}$ is the identity matrix. We remark that the decomposition given in (8) is readily obtained through an eigenvalue decomposition of the positive definite matrix $\mathbf{R}$, and it helps us in deriving the closed form of the Fisher information matrix with respect to $\mathbf{w}$.

Substituting (8) into (6), we obtain

$$\mathbf{R}_w = \mathbf{\Phi}_w(a\mathbf{I} + \mathbf{S})\mathbf{\Phi}_w^T = a\mathbf{I}_w + \mathbf{\Phi}_w\mathbf{S}\mathbf{\Phi}_w^T, \tag{9}$$

where the last equality holds due to (4).

Using (9), we can rewrite a part of the second term on the right hand side of (7) as

$$\begin{aligned}\mathbf{\Phi}_w^T\mathbf{R}_w^{-1}\mathbf{\Phi}_w &= \mathbf{\Phi}_w^T(a\mathbf{I}_w + \mathbf{\Phi}_w\mathbf{S}\mathbf{\Phi}_w^T)^{-1}\mathbf{\Phi}_w \\ &\overset{(1)}{=} \mathbf{S}^{-1} - \mathbf{S}^{-1}(\mathbf{S}^{-1} + a^{-1}\mathbf{\Phi}_w^T\mathbf{\Phi}_w)^{-1}\mathbf{S}^{-1} \\ &\overset{(2)}{=} \mathbf{S}^{-1} - \mathbf{S}^{-1}(\mathbf{S}^{-1} + a^{-1}\operatorname{diag}(\mathbf{w}))^{-1}\mathbf{S}^{-1},\end{aligned} \tag{10}$$

where step (1) is obtained from the matrix inversion lemma[1], and step (2) holds due to (4).

Substituting (10) into (7), the Fisher information matrix can be expressed as

$$\begin{aligned}\mathbf{J}_w =& \mathbf{\Sigma}^{-1} + \mathbf{H}^T\mathbf{S}^{-1}\mathbf{H} \\ &- \mathbf{H}^T\mathbf{S}^{-1}(\mathbf{S}^{-1} + a^{-1}\operatorname{diag}(\mathbf{w}))^{-1}\mathbf{S}^{-1}\mathbf{H}.\end{aligned} \tag{11}$$

It is clear from (11) that the decomposition of $\mathbf{R}$ in (8), together with equations (9)-(10), allows us to make explicit and isolate the dependence of $\mathbf{J}_w$ on $\mathbf{w}$. We also remark that the positive scalar a and

[1]For appropriate matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$, the matrix inversion lemma states that $(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$, which yields $\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{D} = \mathbf{A} - \mathbf{A}(\mathbf{A}+\mathbf{BCD})^{-1}\mathbf{A}$.

positive definite matrix $\mathbf{S}$ can be arbitrarily chosen, and have no effect on the performance of the sensor selection algorithms that will be proposed later on.

C. Formulation of the optimal sensor selection problem

We now state the main optimization problem considered in this work as

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \operatorname{tr}(\mathbf{J}_w^{-1}) \\ \text{subject to} \quad & \mathbf{1}^T\mathbf{w} \leq s, \\ & \mathbf{w} \in \{0,1\}^m, \end{aligned} \tag{P0}$$

where $\mathbf{J}_w \in \mathbb{R}^n$ is given by (11), and $s \leq m$ is a prescribed energy budget given by the maximum number of sensors to be activated. We recall that n is the dimension of the parameter to be estimated, and m is the number of sensors.

We note that (P0) is a nonconvex optimization problem due to the presence of Boolean selection variables. Moreover, if we drop the source statistics $\mathbf{\Sigma}$ from the MSE matrix (5) and impose the assumption $s \geq n$, the proposed formulation (P0) is then applicable for sensor selection in a non-Bayesian framework, where the unknown parameter is estimated through the best linear unbiased estimator [24].

In what follows, we discuss two special cases for the formulations of the sensor selection problem under two different structures of the noise covariance matrix $\mathbf{R}$: a) $\mathbf{R}$ is diagonal, and b) $\mathbf{R}$ has small off-diagonal entries.

D. Formulation for two special cases

When measurement noises are uncorrelated, the noise covariance matrix $\mathbf{R}$ becomes diagonal. From (6) and (7), the Fisher information matrix in the objective function of (P0) simplifies to

$$\begin{aligned} \mathbf{J}_w =& \mathbf{\Sigma}^{-1} + \mathbf{H}^T\mathbf{\Phi}_w^T\mathbf{\Phi}_w\mathbf{R}^{-1}\mathbf{\Phi}_w^T\mathbf{\Phi}_w\mathbf{H} \\ =& \mathbf{\Sigma}^{-1} + \mathbf{H}^T \operatorname{diag}(\mathbf{w})\mathbf{R}^{-1} \operatorname{diag}(\mathbf{w})\mathbf{H} \\ =& \mathbf{\Sigma}^{-1} + \sum_{i=1}^{m} w_i R_{ii}^{-1}\mathbf{h}_i\mathbf{h}_i^T, \end{aligned} \tag{12}$$

where $\mathbf{h}_i^T$ denotes the ith row of $\mathbf{H}$, R_{ii} denotes the ith diagonal entry of $\mathbf{R}$, and the last equality holds due to the fact that

$$w_i^2 = w_i, \ \ i = 1, 2, \ldots, m. \tag{13}$$

It is clear from (12) that each sensor contributes to Fisher information in an additive manner. As demonstrated in [13] and [14], the linearity of the inverse mean squared error (Fisher information) with respect to $\mathbf{w}$ enables the use of convex optimization to solve the problem of sensor selection.

When measurement noises are weakly correlated (namely, $\mathbf{R}$ has small off-diagonal entries), it will be shown in Sec. IV that the Fisher information matrix can be approximately expressed as

$$\hat{\mathbf{J}}_w := \mathbf{\Sigma}^{-1} + \mathbf{H}^T(\mathbf{w}\mathbf{w}^T \circ \mathbf{R}^{-1})\mathbf{H}, \tag{14}$$

where $\circ$ stands for the Hadamard (elementwise) product. The problem of sensor selection with weakly correlated noise becomes

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \operatorname{tr}\left(\mathbf{\Sigma}^{-1} + \mathbf{H}^T(\mathbf{w}\mathbf{w}^T \circ \mathbf{R}^{-1})\mathbf{H}\right)^{-1} \\ \text{subject to} \quad & \mathbf{1}^T\mathbf{w} \leq s, \\ & \mathbf{w} \in \{0,1\}^m. \end{aligned} \tag{P1}$$

Compared to the generalized formulation (P0), the objective function of (P1) is convex with respect to the rank-one matrix $\mathbf{w}\mathbf{w}^T$. Such structure introduces computational benefits while solving (P1). We emphasize that (P1) has been formulated in [18]–[20] for sensor selection with correlated noise, however, using this formulation, without acknowledging that it is only valid when the correlation is weak, can lead to incorrect sensor selection results. We elaborate on the problem of sensor selection with weakly correlated noise in Sec. IV.

III. General Case: Proposed Optimization Methods for Sensor Selection

In this section, we present two methods to solve (P0): the first is based on convex relaxation techniques, and the second is based on a greedy algorithm. First, we show that after relaxing the Boolean constraints the selection problem can be cast as a standard semidefinite program (SDP). Given the solution of the relaxed (P0) we then use the randomization method to generate a near-optimal selection scheme. Next, we show that given a subset of sensors, activating a new sensor always improves the estimation performance. Motivated by this, we present a greedy algorithm that scales gracefully with the problem size to obtain locally optimal solutions of (P0).

A. Convex relaxation

Substituting the expression of Fisher information (11) into problem (P0), we obtain

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \operatorname{tr}\left(\mathbf{C} - \mathbf{B}^T\left(\mathbf{S}^{-1} + a^{-1}\operatorname{diag}(\mathbf{w})\right)^{-1}\mathbf{B}\right)^{-1} \\ \text{subject to} \quad & \mathbf{1}^T\mathbf{w} \leq s, \\ & \mathbf{w} \in \{0,1\}^m, \end{aligned} \tag{15}$$

where for notational simplicity we have defined $\mathbf{C} := \mathbf{\Sigma}^{-1} + \mathbf{H}^T\mathbf{S}^{-1}\mathbf{H}$ and $\mathbf{B} := \mathbf{S}^{-1}\mathbf{H}$.

Problem (15) can be equivalently transformed to [26]

$$\begin{aligned} \underset{\mathbf{w},\mathbf{Z}}{\text{minimize}} \quad & \operatorname{tr}(\mathbf{Z}) \\ \text{subject to} \quad & \mathbf{C} - \mathbf{B}^T \left(\mathbf{S}^{-1} + a^{-1}\operatorname{diag}(\mathbf{w})\right)^{-1} \mathbf{B} \succeq \mathbf{Z}^{-1}, \\ & \mathbf{1}^T\mathbf{w} \leq s, \\ & \mathbf{w} \in \{0,1\}^m, \end{aligned} \tag{16}$$

where $\mathbf{Z} \in \mathbb{S}^n$ is an auxiliary variable, $\mathbb{S}^n$ represents the set of $n \times n$ symmetric matrices, and the notation $\mathbf{X} \succeq \mathbf{Y}$ (or $\mathbf{X} \preceq \mathbf{Y}$) indicates that the matrix $\mathbf{X} - \mathbf{Y}$ (or $\mathbf{Y} - \mathbf{X}$) is positive semidefinite. The first inequality constraint in (16) is obtained from

$$\left(\mathbf{C} - \mathbf{B}^T \left(\mathbf{S}^{-1} + a^{-1}\operatorname{diag}(\mathbf{w})\right)^{-1} \mathbf{B}\right)^{-1} \preceq \mathbf{Z},$$

which implicitly adds the additional constraint $\mathbf{Z} \succeq 0$, since the left hand side of the above inequality is the inverse of the Fisher information matrix.

We further introduce another auxiliary variable $\mathbf{V} \in \mathbb{S}^n$ such that the first matrix inequality of (16) is expressed as

$$\mathbf{C} - \mathbf{V} \succeq \mathbf{Z}^{-1}, \tag{17}$$

and

$$\mathbf{V} \succeq \mathbf{B}^T \left(\mathbf{S}^{-1} + a^{-1}\operatorname{diag}(\mathbf{w})\right)^{-1} \mathbf{B}. \tag{18}$$

Note that the minimization of $\operatorname{tr}(\mathbf{Z})$ with inequalities (17) and (18) would force the variable $\mathbf{V}$ to achieve its lower bound. In other words, problem (16) is equivalent to the problem in which the inequality constraint in (16) is replaced by the two inequalities (17) and (18). Finally, employing the Schur complement, the inequalities (17) and (18) can be rewritten as the following linear matrix inequalities (LMIs)

$$\begin{bmatrix} \mathbf{C} - \mathbf{V} & \mathbf{I} \\ \mathbf{I} & \mathbf{Z} \end{bmatrix} \succeq 0, \quad \begin{bmatrix} \mathbf{V} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{S}^{-1} + a^{-1}\operatorname{diag}(\mathbf{w}) \end{bmatrix} \succeq 0. \tag{19}$$

Substituting (19) into (16), the sensor selection problem becomes

$$\begin{aligned} \underset{\mathbf{w},\mathbf{Z},\mathbf{V}}{\text{minimize}} \quad & \operatorname{tr}(\mathbf{Z}) \\ \text{subject to} \quad & \text{LMIs in (19)}, \\ & \mathbf{1}^T\mathbf{w} \leq s, \\ & \mathbf{w} \in \{0,1\}^m. \end{aligned} \tag{20}$$

Problem (20) has the form of an SDP except for the last Boolean constraints. As shown in [13], one possibility is to relax each Boolean variable to its convex hull to obtain $\mathbf{w} \in [0,1]^m$. In this case, we

can choose s active sensors given by the first s largest entries of the solution of the relaxed problem, or employ a randomized rounding algorithm [14, Algorithm 3] to generate a Boolean selection vector.

Rather than directly relaxing Boolean selection variables to continuous variables, we can use semidefinite relaxation (SDR) [27] — referred to problems in which the relaxation of a rank constraint leads to an SDP — to better overcome the difficulties posed by the nonconvex constraints of (20). The Boolean constraint (13) on the entries of $\mathbf{w}$ can be enforced by

$$\mathrm{diag}(\mathbf{w}\mathbf{w}^T) = \mathbf{w}, \tag{21}$$

where, with an abuse of notation, $\mathrm{diag}(\cdot)$ returns in vector form the diagonal entries of its matrix argument. By introducing an auxiliary variable $\mathbf{W}$ together with the rank-one constraint

$$\mathbf{W} = \mathbf{w}\mathbf{w}^T, \tag{22}$$

the energy and Boolean constraints in (20) can be expressed as

$$\mathrm{tr}(\mathbf{W}) \leq s, \quad \mathrm{diag}(\mathbf{W}) = \mathbf{w}. \tag{23}$$

After relaxing the (nonconvex) rank-one constraint (22) to $\mathbf{W} \succeq \mathbf{w}\mathbf{w}^T$, we reach the SDP

$$\begin{aligned} \underset{\mathbf{w},\mathbf{W},\mathbf{Z},\mathbf{V}}{\text{minimize}} \quad & \mathrm{tr}\,(\mathbf{Z}) \\ \text{subject to} \quad & \text{LMIs in (19)}, \\ & \mathrm{tr}(\mathbf{W}) \leq s, \\ & \mathrm{diag}(\mathbf{W}) = \mathbf{w}, \\ & \begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^T & 1 \end{bmatrix} \succeq 0, \end{aligned} \tag{24}$$

where the last inequality is derived through the application of a Schur complement to $\mathbf{W} \succeq \mathbf{w}\mathbf{w}^T$.

We can use an interior-point algorithm to solve the SDP (24). In practice, if the dimension of the unknown parameter vector is much less than the number of sensors, the computational complexity of SDP is roughly given by $O(m^{4.5})$ [28]. Once the SDP (24) is solved, we employ a randomization method to generate a near-optimal sensor selection scheme, where the effectiveness of the randomization method has been shown in our extensive numerical experiments. We refer the readers to [27] for more details on the motivation and benefits of randomization used in SDR. The aforementioned procedure is summarized in Algorithm 1, which includes the randomization procedure described in Algorithm 2.

B. Greedy algorithm

We begin by showing in Proposition 1 that even in the presence of correlated measurement noise, the Fisher information increases if an inactive sensor is made active.

Algorithm 1 SDR with randomization for sensor selection

Require: prior information $\mathbf{\Sigma}$, $\mathbf{R} = a\mathbf{I} + \mathbf{S}$ as in (8), observation matrix $\mathbf{H}$ and energy budget s

1: solve the SDP (24) and obtain solution $(\mathbf{w}, \mathbf{W})$

2: call Algorithm 2 for Boolean solution.

Algorithm 2 Randomization method [27]

Require: solution pair $(\mathbf{w}, \mathbf{W})$ from the SDP (24)

1: **for** $l = 1, 2, \ldots, N$ **do**

2: pick a random number $\boldsymbol{\xi}^{(l)} \sim \mathcal{N}(\mathbf{w}, \mathbf{W} - \mathbf{w}\mathbf{w}^T)$

3: map $\boldsymbol{\xi}^{(l)}$ to a sub-optimal sensor selection scheme $\mathbf{w}^{(l)}$

$$w_j^{(l)} = \begin{cases} 1 & \xi_j^{(l)} \geq [\boldsymbol{\xi}^{(l)}]_s \\ 0 & \text{otherwise,} \end{cases} \quad j = 1, 2, \ldots, m,$$

where $w_j^{(l)}$ is the jth element of $\mathbf{w}^{(l)}$, and $[\boldsymbol{\xi}^{(l)}]_s$ denotes the sth largest entry of $\boldsymbol{\xi}^{(l)}$

4: **end for**

5: choose a vector in $\{\mathbf{w}^{(l)}\}_{l=1}^N$ which yields the smallest objective value of (15).

Proposition 1: If $\mathbf{w}$ and $\tilde{\mathbf{w}}$ represent two sensor selection vectors, where $w_i = \tilde{w}_i$ for $i \in \{1, 2, \ldots, m\} \setminus \{j\}$, $w_j = 0$ and $\tilde{w}_j = 1$, then the resulting Fisher information matrix satisfies $\mathbf{J}_{\tilde{w}} \succeq \mathbf{J}_w$. More precisely,

$$\mathbf{J}_{\tilde{w}} - \mathbf{J}_w = c_j \boldsymbol{\alpha}_j \boldsymbol{\alpha}_j^T, \tag{25}$$

and

$$\mathrm{tr}(\mathbf{J}_w^{-1}) - \mathrm{tr}(\mathbf{J}_{\tilde{w}}^{-1}) = \frac{c_j \boldsymbol{\alpha}_j^T \mathbf{J}_w^{-2} \boldsymbol{\alpha}_j}{1 + c_j \boldsymbol{\alpha}_j \mathbf{J}_w^{-1} \boldsymbol{\alpha}_j} \geq 0, \tag{26}$$

where c_j is a positive scalar given by

$$c_j = \begin{cases} R_{jj}^{-1} & \mathbf{w} = \mathbf{0} \\ (R_{jj} - \mathbf{r}_j^T \mathbf{R}_w^{-1} \mathbf{r}_j)^{-1} & \text{otherwise,} \end{cases} \tag{27}$$

and

$$\boldsymbol{\alpha}_j = \begin{cases} \mathbf{h}_j & \mathbf{w} = \mathbf{0} \\ \mathbf{H}^T \mathbf{\Phi}_w^T \mathbf{R}_w^{-1} \mathbf{r}_j - \mathbf{h}_j & \text{otherwise.} \end{cases} \tag{28}$$

In (27)-(28), R_{jj} is the jth diagonal entry of $\mathbf{R}$, $\mathbf{r}_j$ represents the covariance vector between the measurement noise of the jth sensor and that of the active sensors in $\mathbf{w}$, $\mathbf{h}_j^T$ is the jth row of $\mathbf{H}$, $\mathbf{\Phi}_w$ and $\mathbf{R}_w$ are given by (3) and (6), respectively.

Proof: See Appendix A. ■

It is clear from (25) that when an inactive sensor is made active, the increase in Fisher information leads to an information gain in terms of the rank-one matrix given by (25). Such a phenomenon was also discovered in the calculation of sensor utility for adaptive signal estimation [29] and leader selection in stochastically forced consensus networks [12]. Since activating a new sensor does not degrade the estimation performance, the inequality (energy) constraint in (P0) can be reformulated as an equality constraint.

In a greedy algorithm, we iteratively select a new sensor which gives the largest performance improvement until the energy constraint is satisfied with equality. The greedy algorithm is attractive due to its simplicity, and has been employed in a variety of applications [12], [29], [30]. In particular, a greedy algorithm was proposed in [30] for sensor selection under the assumption of uncorrelated measurement noise. We generalize the framework of [30] by taking into account noise correlation. Clearly, in each iteration of the greedy algorithm, the newly activated sensor is the one that maximizes the performance improvement characterized by $\mathrm{tr}(\mathbf{J}_w^{-1}) - \mathrm{tr}(\mathbf{J}_{\tilde{w}}^{-1})$ in (26). We summarize the greedy algorithm in Algorithm 3.

Algorithm 3 Greedy algorithm for sensor selection

Require: $\mathbf{w} = \mathbf{0}$, $\mathcal{I} = \{1, 2, \ldots, m\}$ and $\mathbf{J}_w = \mathbf{\Sigma}^{-1}$

1: **for** $l = 1, 2, \ldots, s$ **do**

2: given $\mathbf{w}$, enumerate all the inactive sensors in $\mathcal{I}$ to determine $j \in \mathcal{I}$ such that $\mathrm{tr}(\mathbf{J}_w^{-1}) - \mathrm{tr}(\mathbf{J}_{\tilde{w}}^{-1})$ in (26) is maximized

3: update $\mathbf{w}$ by setting $w_j = 1$, and update $\mathbf{J}_w$ by adding $c_j \boldsymbol{\alpha}_j \boldsymbol{\alpha}_j^T$ in (25)

4: remove j from $\mathcal{I}$.

5: **end for**

In Step 2 of Algorithm 3, we search $O(m)$ sensors to achieve the largest performance improvement. In (26), the computation of $\mathbf{J}_w^{-1}$ incurs a complexity of $O(n^{2.373})$ [31]. Since Algorithm 3 terminates after s iterations, its overall complexity is given by $O(sm + sn^{2.373})$, where at each iteration, the calculation of $\mathbf{J}_w^{-1}$ is independent of the search for the new active sensor. If the dimension of $\mathbf{x}$ is much less than the number of sensors, the complexity of Algorithm 3 reduces to $O(sm)$. Our extensive numerical experiments show that the greedy algorithm is able to yield good locally optimal sensor selection schemes.

IV. Special Case: Sensor Selection with Weak Noise Correlation

In this section, we show that the existing sensor selection model in [18]–[20] is invalid for an arbitrary noise covariance matrix. We establish that in contrast to the approach proposed in this paper, the existing model in [18]–[20] is only valid when measurement noises are weakly correlated. In this scenario, the proposed sensor selection problem given by (P0) would simplify to (P1). Moreover, if the trace of the Fisher information matrix (also known as information gain defined in [20]) is adopted as the performance measure for sensor selection, we show that the resulting optimization problem can be cast as a special problem of maximizing a convex quadratic function over a bounded polyhedron.

A. Drawbacks of existing formulation

In [18]–[20], several variations of sensor selection problems with correlated noise have been studied, based on whether the quantity to be estimated is a random parameter or a random process, and whether the cost function is energy or estimation error. The common feature in [18]–[20] is that the information matrix was approximated by (14); we repeat equation (14) here for convenience

$$\hat{\mathbf{J}}_w = \mathbf{\Sigma}^{-1} + \mathbf{H}^T(\mathbf{w}\mathbf{w}^T \circ \mathbf{R}^{-1})\mathbf{H}. \tag{29}$$

Compared to our formulation (7), the noise covariance matrix appearing in (29) is independent of the sensor selection variables. In fact, $\hat{\mathbf{J}}_w$ can be thought of as Fisher information under the measurement model

$$\mathbf{y} = \mathbf{\Phi}_w\mathbf{H}\mathbf{x} + \mathbf{v}, \tag{30}$$

where $\mathbf{\Phi}_w$ was defined in (3). Different from (3), the noise from the unselected sensors is spread across the selected sensors. As a result, the measurement model (30) yields $y_i = v_i$ if the ith sensor is inactive. This contradicts the fact that an inactive sensor should keep silent and thus have no effect on the estimation task.

The Fisher information in (29) can also be interpreted as [18, Sec. 3]

$$\begin{aligned}\hat{\mathbf{J}}_w &= \mathbf{\Sigma}^{-1} + \sum_{i,j\in\mathcal{S}} \bar{R}_{ij}\mathbf{h}_i\mathbf{h}_j^T, \\ &= \mathbf{\Sigma}^{-1} + \mathbf{H}^T\mathbf{\Phi}_w^T(\mathbf{\Phi}_w\mathbf{R}^{-1}\mathbf{\Phi}_w^T)\mathbf{\Phi}_w\mathbf{H},\end{aligned} \tag{31}$$

where $\mathcal{S}$ is the set of selected sensors, and $\bar{R}_{ij}$ denotes the (i,j)th entry of $\mathbf{R}^{-1}$. In (31), $\mathbf{R}^{-1}$ is computed first and then truncated according to the sensor selection scheme. This is an incorrect way of modeling the noise covariance matrix for active sensors, since the matrix $\mathbf{R}$ should be truncated first and then inverted as demonstrated in (7).

Both of the interpretations (30) and (31) indicate that the existing formulation in [18]–[20] is inaccurate for modeling the problem of sensor selection with correlated noise. A natural question that arises from the preceding discussion is whether there exist a condition that ensures the validity of the Fisher information matrix (29) as presented in [18]–[20]? We will show in the next section that the formulation reported in [18]–[20] becomes valid only when sensor selection is restricted to the weak noise correlation regime.

B. Validity of existing formulation: weak correlation

We consider the scenario of weakly correlated noise, in which the noise covariance matrix $\mathbf{R}$ has small off-diagonal entries, namely, noises are weakly correlated across the sensors. For ease of representation, we express the noise covariance matrix as

$$\mathbf{R} = \mathbf{\Lambda} + \epsilon \mathbf{\Upsilon}, \tag{32}$$

where $\mathbf{\Lambda}$ is a diagonal matrix which consists of the diagonal entries of $\mathbf{R}$, $\epsilon\mathbf{\Upsilon}$ is a symmetric matrix whose diagonal entries are zero and off-diagonal entries correspond to those of $\mathbf{R}$, the parameter ϵ is introduced to govern the strength of noise correlation across the sensors, and $\mathbf{\Lambda}$ and $\mathbf{\Upsilon}$ are independent of ϵ. Clearly, the covariance of weakly correlated noises can be described by (32) for some small value of ϵ since $\mathbf{\Upsilon}$ is ϵ-independent. As $\epsilon \to 0$, the off-diagonal entries of $\mathbf{R}$ are forced to go to zero.

Proposition 2 below shows that the correct expression (7) of Fisher information is equal to the expression (29), as presented in [18]–[20], up to first order in ϵ as $\epsilon \to 0$.

Proposition 2: If measurement noises are weakly correlated and $\mathbf{R} = \mathbf{\Lambda} + \epsilon\mathbf{\Upsilon}$, then the Fisher information matrix (7) can be expressed as

$$\mathbf{J}_w = \hat{\mathbf{J}}_w + O(\epsilon^2) \quad \text{as } \epsilon \to 0,$$

where $\hat{\mathbf{J}}_w$ is given by (29).

Proof: See Appendix B. ■

It is clear from Proposition 2 that (P1) is valid only when the noise correlation is weak. Proceeding with the same logic as in the previous section for the introduction of constraints (22)-(23), we relax (P1) to the SDP

$$\begin{aligned} \underset{\mathbf{w},\mathbf{W},\mathbf{Z}}{\text{minimize}} \quad & \operatorname{tr}(\mathbf{Z}) \\ \text{subject to} \quad & \begin{bmatrix} \mathbf{\Sigma}^{-1} + \mathbf{H}^T(\mathbf{W} \circ \mathbf{R}^{-1})\mathbf{H} & \mathbf{I} \\ \mathbf{I} & \mathbf{Z} \end{bmatrix} \succeq 0, \\ & \operatorname{tr}(\mathbf{W}) \le s, \;\; \operatorname{diag}(\mathbf{W}) = \mathbf{w}, \\ & \begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^T & 1 \end{bmatrix} \succeq 0, \end{aligned} \tag{33}$$

where $\mathbf{Z} \in \mathbb{S}^n$ is an auxiliary optimization variable. Given the solution pair $(\mathbf{w}, \mathbf{W})$ of problem (33), we can use the randomization method in Algorithm 2 to construct a near-optimal sensor selection scheme. The computational complexity of solving problem (33) is close to that of solving the SDP (24). However, as will be evident later, the sensor selection problem with weakly correlated noise can be further simplified if the trace of the Fisher information matrix is used as the performance measure. In this scenario, the obtained problem structure enables the use of more computationally inexpensive algorithms, e.g., bilinear programing, to solve the sensor selection problem.

C. Sensor selection by maximizing trace of Fisher information

Instead of minimizing the estimation error, the trace of Fisher information (so-called T-optimality [32]) also has been used as a performance metric in problems of sensor selection [20], [33], [34]. According to [35, Lemma 1], the trace of Fisher information constitutes a lower bound to the trace of error covariance matrix given by $\mathbf{J}_w^{-1}$ in (7). That is,

$$\mathrm{tr}(\mathbf{J}_w^{-1}) \geq \frac{n^2}{\mathrm{tr}(\mathbf{J}_w)}. \tag{34}$$

Motivated by (34) and the generalized information gain used in [20], we propose to minimize the lower bound of the objective function in (P1), which leads to the problem

$$\begin{aligned} \underset{\mathbf{w}}{\text{maximize}} \quad & \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1} + \mathbf{H}^T(\mathbf{w}\mathbf{w}^T \circ \mathbf{R}^{-1})\mathbf{H}\right) \\ \text{subject to} \quad & \mathbf{1}^T\mathbf{w} \leq s, \\ & \mathbf{w} \in \{0,1\}^m. \end{aligned} \tag{P2}$$

It is worth mentioning that the sensor selection scheme obtained from (P2) may not be optimal in the MMSE sense. However, the trace operator is linear and introduces computational benefits in optimization. Reference [20] has shown that (P2) is not convex even if Boolean selection variables are relaxed. However, there is no theoretical justification and analysis provided in [20] on the problem structure. In what follows, we demonstrate that the Boolean constraint in (P2) can be replaced by its convex hull $\mathbf{w} \in [0,1]^m$ without loss of performance, to obtain an equivalent optimization problem.

Proposition 3: (P2) is equivalent to

$$\begin{aligned} \underset{\mathbf{w}}{\text{maximize}} \quad & \mathbf{w}^T\boldsymbol{\Omega}\mathbf{w} \\ \text{subject to} \quad & \mathbf{1}^T\mathbf{w} \leq s, \\ & \mathbf{w} \in [0,1]^m, \end{aligned} \tag{35}$$

where $\boldsymbol{\Omega}$ is a positive semidefinite matrix given by $\mathbf{A}(\mathbf{R}^{-1} \otimes \mathbf{I}_n)\mathbf{A}^T$, $\otimes$ denotes the Kronecker product, $\mathbf{A} \in \mathbb{R}^{m \times mn}$ is a block-diagonal matrix whose diagonal blocks are given by $\{\mathbf{h}_i^T\}_{i=1}^m$, and $\mathbf{h}_i^T$ denotes the ith row of the measurement matrix $\mathbf{H}$.

Proof: See Appendix C. ■

It is clear from Proposition 3 that (P2) eventually approaches the problem of maximizing a convex quadratic function over a bounded polyhedron. It is known [36] that finding a globally optimal solution of (35) is NP-hard. Therefore, we resort to local optimization methods, such as bilinear programming and SDR, to solve problem (35). To be specific, bilinear programming is a special case of alternating convex optimization, where at each iteration we solve two linear programs. Since bilinear programming is based on linear programming, it scales gracefully with problem size but with a possibility of only finding local optima. If we rewrite the constraints of problem (35) as quadratic forms in $\mathbf{w}$, (P2) can be further transformed into a nonconvex *homogeneous* quadratically constrained quadratic program (QCQP), which refers to a QCQP without involving linear terms of optimization variables. In this scenario, SDR can be applied to solve the problem. Compared to the application of SDR in (33), the homogeneous QCQP leads to an SDP with a smaller problem size. We refer the readers to [22, Sec. V] and [20, Sec. V] for more details on the application of bilinear programming and SDR.

V. Non-myopic Sensor Scheduling

In this section, we extend the sensor selection framework with correlated noise to the problem of non-myopic sensor scheduling, which determines sensor activations for multiple future time steps. Since the Fisher information matrices at consecutive time steps are coupled with each other, expressing them in a closed form with respect to the sensor selection variables becomes intractable. Therefore, we employ a greedy algorithm to seek locally optimal solutions of the non-myopic sensor scheduling problem.

Consider a discrete-time dynamical system

$$\mathbf{x}_{t+1} = \mathbf{F}_t \mathbf{x}_t + \mathbf{u}_t \tag{36}$$

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t, \tag{37}$$

where $\mathbf{x}_t \in \mathbb{R}^n$ is the target state at time t, $\mathbf{y}_t \in \mathbb{R}^m$ is the measurement vector whose ith entry corresponds to a scalar observation from the ith sensor at time t, $\mathbf{F}_t$ is the state transition matrix from time t to time $t+1$, and $\mathbf{H}_t$ denotes the observation matrix at time t. The inputs $\mathbf{u}_t$ and $\mathbf{v}_t$ are white, Gaussian, zero-mean random vectors with covariance matrices $\mathbf{Q}$ and $\mathbf{R}$, respectively. We note that the covariance matrix $\mathbf{R}$ may not be diagonal, since the noises experienced by different sensors could be spatially correlated. We also remark that although the dynamical system (36)-(37) is assumed to be linear, it will be evident later that the proposed sensor scheduling framework is also applicable to non-linear dynamical systems.

The PDF of the initial state $\mathbf{x}_0$ at time step t_0 is assumed to be Gaussian with mean $\hat{\mathbf{x}}_0$ and covariance matrix $\hat{\mathbf{P}}_0$, where $\hat{\mathbf{x}}_0$ and $\hat{\mathbf{P}}_0$ are estimates of the initial state and error covariance from the previous measurements obtained using filtering algorithms, such as a particle filter or a Kalman filter [37], [38]. At time step t_0, we aim to find the optimal sensor schedule over the next τ time steps $t_0+1, t_0+2, \ldots, t_0+\tau$. Hereafter, for notational simplicity, we assume $t_0 = 0$. The sensor schedule can be represented by a vector of binary variables

$$\mathbf{w} = [\mathbf{w}_1^T, \mathbf{w}_2^T, \ldots, \mathbf{w}_\tau^T]^T \in \{0,1\}^{\tau m}, \tag{38}$$

where $\mathbf{w}_t = [w_{t,1}, w_{t,2}, \ldots, w_{t,m}]^T$ characterizes the sensor schedule at time $1 \leq t \leq \tau$. In what follows, we assume that $\tau > 1$. If $\tau = 1$, the non-myopic sensor scheduling problem reduces to the sensor selection problem for one snapshot or the so-called myopic scheduling problem. This case has been studied in the previous sections.

In the context of state tracking [16], [39], the Fisher information matrix has the following recursive form

$$\mathbf{J}_t = (\mathbf{Q} + \mathbf{F}_{t-1}\mathbf{J}_{t-1}^{-1}\mathbf{F}_{t-1}^T)^{-1} + \mathbf{G}_t \tag{39}$$

$$\mathbf{G}_t = \mathbf{H}_t^T \mathbf{\Phi}_{w_t}^T (\mathbf{\Phi}_{w_t} \mathbf{R} \mathbf{\Phi}_{w_t}^T)^{-1} \mathbf{\Phi}_{w_t} \mathbf{H}_t, \tag{40}$$

for $t = 1, 2, \ldots, \tau$, where $\mathbf{J}_t$ denotes the Fisher information at time t, $\mathbf{G}_t$ denotes the part of Fisher information matrix which incorporates the updated measurement, and $\mathbf{\Phi}_{w_t}$ is a submatrix of $\mathrm{diag}(\mathbf{w}_t)$ where all the rows corresponding to the unselected sensors are removed. It is clear from (10) that the term involving $\mathbf{\Phi}_{w_t}$ in (40) can be further expressed as an explicit form with respect to $\mathbf{w}_t$.

Remark 1: *In case of non-linear measurement models, the term $\mathbf{G}_t$ in the Fisher information matrix becomes*

$$\mathbf{G}_t = \mathbf{E}_{\mathbf{x}_t}[(\nabla_{\mathbf{x}_t^T}\mathbf{h})^T \mathbf{\Phi}_{w_t}^T (\mathbf{\Phi}_{w_t} \mathbf{R} \mathbf{\Phi}_{w_t}^T)^{-1} \mathbf{\Phi}_{w_t} (\nabla_{\mathbf{x}_t^T}\mathbf{h})],$$

where $\mathbf{h}(\cdot)$ is a nonlinear measurement function, and $\nabla_{\mathbf{x}_t^T}\mathbf{h}$ is the Jacobian matrix of $\mathbf{h}$ with respect to $\mathbf{x}_t$. In this equation, the expectation with respect to $\mathbf{x}_t$ is commonly calculated with the help of the prediction state $\hat{\mathbf{x}}_t := \mathbf{F}_{t-1}\mathbf{F}_{t-2}\cdots\mathbf{F}_0\hat{\mathbf{x}}_0$ [38], [40]. To be concrete, we approximate the PDF of $\mathbf{x}_t$ with $p(\mathbf{x}_t) = \delta(\mathbf{x}_t - \hat{\mathbf{x}}_t)$, where $\delta(\cdot)$ is a δ-function. The matrix $\mathbf{G}_t$ is then given by

$$\mathbf{G}_t = \hat{\mathbf{H}}_t^T \mathbf{\Phi}_{w_t}^T (\mathbf{\Phi}_{w_t} \mathbf{R} \mathbf{\Phi}_{w_t}^T)^{-1} \mathbf{\Phi}_{w_t} \hat{\mathbf{H}}_t, \tag{41}$$

where $\hat{\mathbf{H}}_t := \nabla_{\mathbf{x}_t^T}\mathbf{h}(\hat{\mathbf{x}}_t)$.

We note that the Fisher information matrices at consecutive time steps are coupled with each other due to the recursive structure in (39). Therefore, $\mathbf{J}_t$ is a function of all selection variables $\{\mathbf{w}_k\}_{k=1}^t$. The

recursive structure makes the closed form of Fisher information intractable. This is in sharp contrast with the problem of myopic sensor selection, where expressing the Fisher information matrix in a closed form is possible.

We now pose the non-myopic sensor scheduling problem

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \ & \frac{1}{\tau}\sum_{t=1}^{\tau}\text{tr}(\mathbf{J}_t^{-1}) \\ \text{subject to} \ & \mathbf{1}^T\mathbf{w} \leq s, && \text{(42a)} \\ & \textstyle\sum_{t=1}^{\tau} w_{t,i} \leq s_i, \ i=1,2,\ldots,m, && \text{(42b)} \\ & \mathbf{w} \in \{0,1\}^{m\tau}, \end{aligned}$$

where $\mathbf{J}_t$ is determined by (39)-(40), the *cumulative* energy constraint (42a) restricts the total number of activations for all sensors over the entire time horizon, and the *individual* energy constraint (42b) implies that the ith sensor can report at most s_i measurements over τ time steps.

To solve problem (42) in a numerically efficient manner, we employ a greedy algorithm that iteratively activates one sensor at a time until the energy constraints are satisfied with equality. The proposed greedy algorithm can be viewed as a generalization of Algorithm 3 by incorporating the length of the time horizon and individual energy constraints.

We elaborate on the greedy algorithm. In the initial step, we assume $\mathbf{w}=\mathbf{0}$ and split the set of indices of $\mathbf{w}$ into m subsets $\{\mathcal{I}_i\}_{i=1}^m$, where we use the entries of the set $\mathcal{I}_i$ to keep track of all the time instants at which the ith sensor is inactive. The set $\mathcal{I}_i$ is initially given by $\{i, i+m, \ldots, i+(\tau-1)m\}$ for $i=1,2,\ldots,m$. There exists a one-to-one correspondence between an index $j \in \mathcal{I}_i$ and a time instant $t \in \{1,2,\ldots,\tau\}$ at which the ith sensor can be scheduled, where $j=i+(t-1)m$. At every iteration of the greedy optimization algorithm, we update $\mathcal{I}_i$ for $i=1,2,\ldots,m$ such that it only contains indices of zero entries of $\mathbf{w}$. The quantity $\tau-|\mathcal{I}_i|$ gives the number of times that the ith sensor has been used, where $|\cdot|$ denotes the cardinality of a set. The condition $\tau-|\mathcal{I}_i| \geq s_i$ indicates a violation of the individual energy constraint. Note that the union $\{\mathcal{I}_1 \cup \mathcal{I}_2 \cup \ldots \cup \mathcal{I}_m\}$ gives all the remaining time instants at which the sensors can be activated. We enumerate all the indices in the union to determine the index j^* such that the objective function of (42) is minimized as $w_{j^*}=1$. We summarize the greedy algorithm for non-myopic sensor scheduling in Algorithm 4.

The computational complexity of Algorithm 4 is dominated by Step 3. Specifically, we evaluate the objective function of (42) using $O(\tau m)$ operations. And the computation of the Fisher information matrix requires a complexity of $O(\tau m^{2.373})$, where $O(\tau)$ accounts for the number of recursions, and $O(m^{2.373})$ is the complexity of matrix inversion in (41) [31]. We emphasize that different from Proposition 1,

Algorithm 4 Greedy algorithm for sensor scheduling

Require: $\mathbf{w} = \mathbf{0}$ and $\mathcal{I}_i = \{i, i+m, \ldots, i+(\tau-1)m\}$ for $i = 1, 2, \ldots, m$

1: **for** $l = 1, 2, \ldots, \min\{s, \sum_{i=1}^{m} s_i\}$ **do**

2: if $\tau - |\mathcal{I}_i| \geq s_i$, then replace $\mathcal{I}_i$ with an empty set for $i = 1, 2, \ldots, m$,

3: enumerate indices of $\mathbf{w}$ in $\{\mathcal{I}_1 \cup \mathcal{I}_2 \cup \ldots \cup \mathcal{I}_m\}$ to select j^* such that the objective function of (42) is minimized when $w_j = 1$,

4: remove j from $\mathcal{I}_{i^*}$, where i^* is given by the remainder of $\frac{j}{m}$ for $i^* \neq m$, and $i^* = m$ if the remainder is 0.

5: **end for**

expressing the closed form of the performance improvement in a greedy manner becomes intractable, since the Fisher information matrices are coupled with each other over the time horizon. Therefore, the computation cost of Algorithm 4 is given by $O(\tau^2 m^{3.373})$ per iteration.

For additional perspective, we compare the computational complexity of Algorithm 4 with the method in [21], where a reweighted ℓ_1 based quadratic programming (QP) was used to obtain locally optimal sensor schedules under linear (or linearized) dynamical systems with correlated noise. It was shown in [21] that the computational complexity of QP was ideally given by $O(m^{2.5}\tau^5)$ for every reweighting ℓ_1 iteration. We note that the computational complexity of the greedy algorithm increases slightly in terms of the network size by a factor $m^{0.873}$, while it decreases significantly in terms of the length of the time horizon by a factor τ^3.

VI. Numerical Results

In this section, we demonstrate the effectiveness of the proposed approach for sensor selection/scheduling with correlated measurement noise. In our numerical examples, we assume that the sensors are randomly deployed in a square region, where each of them provides the measurement of an unknown parameter or state. For parameter estimation, we use the linear MMSE estimator [24, Sec. 12] to estimate the unknown parameter. For state tracking, we use the extended Kalman filter [24, Sec. 13] to track the target state.

Sensor selection for parameter estimation: We consider a network with $m \in \{20, 50\}$ sensors to estimate the vector of parameters $\mathbf{x} \in \mathbb{R}^n$ with $n = 2$, where sensors are randomly deployed over a 50×50 lattice. The prior PDF of $\mathbf{x}$ is given by $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = [10, 10]^T$ and $\boldsymbol{\Sigma} = \mathbf{I}$. For simplicity, the row vectors of the measurement matrix $\mathbf{H}$ are chosen randomly, and independently,

from the distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}/\sqrt{n})$ [13]. The covariance matrix of the measurement noise is set by an exponential model [41]

$$R_{ij} = \mathrm{cov}(v_i, v_j) = \sigma_v^2 \, e^{-\rho \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|_2}, \tag{43}$$

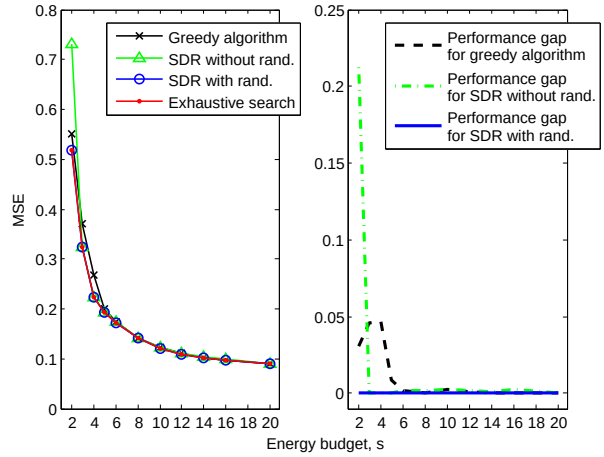
for $i, j = 1, 2, \ldots, m$, where $\sigma_v^2 = 1$, $\boldsymbol{\beta}_i \in \mathbb{R}^2$ is the location of the ith sensor in the 2D plane, $\|\cdot\|_2$ denotes the Euclidean norm, and ρ is the correlation parameter which governs the strength of spatial correlation, namely, a larger (or smaller) ρ corresponds to a weaker (or stronger) correlation.

We choose $N = 100$ while performing the randomization method. Also, we employ an exhaustive search that enumerates all possible sensor selection schemes to obtain the globally optimal solution of (P0). The estimation performance is measured through the empirical MSE, which is averaged over 1000 numerical trials.

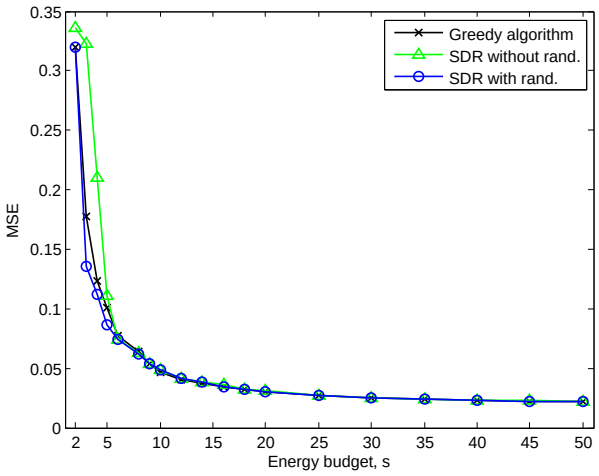
In Fig. 1, we present the MSE as a function of the energy budget by solving (P0) with correlation parameter $\rho = 0.1$. In Fig. 1-(a) for the tractability of exhaustive search, we consider a small network with $m = 20$ sensors. We compare the performance of the proposed greedy algorithm and SDR with randomization to that of SDR without randomization and exhaustive search. In particular, the right plots of Fig. 1-(a) show the performance gaps for the obtained locally optimal solutions compared to the globally optimal solutions resulting from an exhaustive search. We observe that the SDR method with randomization outperforms the greedy algorithm and yields optimal solutions. The randomization method also significantly improves the performance of SDR in sensor selection. This is not surprising, and our numerical observations agree with the literature [27], [42] that demonstrate the power and utility of randomization in SDR.

In Fig. 1-(b), we present the MSE as a function of the energy budget for a relatively large network ($m = 50$). Similar to the results of Fig. 1-(a), the SDR method with randomization yields the lowest estimation error. We also observe that the MSE ceases to decrease significantly when $s \geq 20$. This indicates that a subset of sensors suffices to provide satisfactory estimation performance, since the presence of correlation among sensors introduces information redundancy and makes observations less diverse.

In Fig. 2, we solve the problem of sensor selection with weak noise correlation ($\rho = 0.5$), and present the MSE as a function of the energy budget $s \in \{2, 3, \ldots, 50\}$. We compare the performance of three optimization approaches: SDR with randomization for solving (P1), bilinear programming (BP) for solving (P2), and SDR with randomization for solving (P2). We recall that (P1) is to minimize the trace of the error covariance matrix and (P2) is to maximize the trace of Fisher information. As we can see, approaches that maximize the trace of Fisher information yield worse estimation performance than those that minimize the estimation error. This is because (P2) ignores the contribution of prior information $\boldsymbol{\Sigma}$



(a)



(b)

Fig. 1: MSE versus energy budget with correlation parameter $\rho = 0.1$.

in sensor selection. We also note that although BP (a linear programming based approach) has the lowest computational complexity, it leads to the worst optimization performance.

In Fig. 3, we present the MSE as a function of the correlation parameter ρ, where $m = 50$ and $s \in \{7, 13\}$. We consider sensor selection schemes by using SDR with randomization to solve problems (P0) and (P1), respectively. For comparison, we also present the estimation performance when all the sensors are selected. As demonstrated in Fig. 3, we consider two correlation regimes: weak correlation and strong correlation. We observe that in the weak correlation regime, solutions of both (P0) and (P1) yield the same estimation performance. In the strong correlation regime, solutions of (P1) could lead to worse estimation performance for sensor selection. We also observe that the sensitivity to the strategy

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2016.2550005, IEEE Transactions on Signal Processing

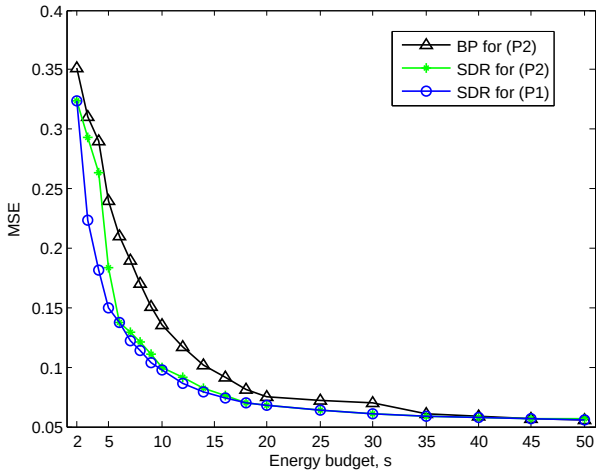


Fig. 2: MSE versus energy budget for sensor selection with weak noise correlation $\rho = 0.5$.

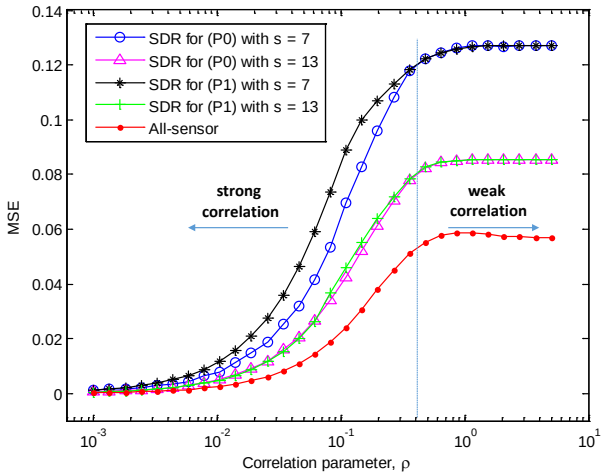


Fig. 3: MSE versus the strength of correlation for $s \in \{7, 13\}$.

of sensor selection reduces if the strength of correlation becomes extremely large, e.g., $\rho \leq 0.05$. More interestingly, the estimation performance is improved as the correlation becomes stronger. This is because for strongly correlated noise, noise cancellation could be achieved by subtracting one observation from the other [43]. Further if we fix the value of ρ, the estimation error decreases when the energy budget increases, and the performance gap between solutions of (P0) and (P1) reduces.

Sensor scheduling for state tracking

In this example, we track a target with $m = 30$ sensors over 30 time steps. We assume that the target state is a 4×1 vector $\mathbf{x}_t = [x_{t,1}, x_{t,2}, x_{t,3}, x_{t,4}]^T$, where $(x_{t,1}, x_{t,2})$ and $(x_{t,3}, x_{t,4})$ denote the target

location and velocity at time step t. The state equation (36) follows a white noise acceleration model [38]

$$\mathbf{F}_t = \begin{bmatrix} 1 & 0 & \Delta & 0 \\ 0 & 1 & 0 & \Delta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \ \mathbf{Q} = q \begin{bmatrix} \frac{\Delta^3}{3} & 0 & \frac{\Delta^2}{2} & 0 \\ 0 & \frac{\Delta^3}{3} & 0 & \frac{\Delta^2}{2} \\ \frac{\Delta^2}{2} & 0 & \Delta & 0 \\ 0 & \frac{\Delta^2}{2} & 0 & \Delta \end{bmatrix},$$

where Δ and q denote the sampling interval and the process noise parameter, respectively. In our simulations, we set $\Delta = 1$ and $q = 0.01$. The prior PDF of the initial state is assumed to be Gaussian with mean $\hat{\mathbf{x}}_0 = [1, 1, 0.5, 0.5]^T$ and covariance $\hat{\mathbf{\Sigma}}_0 = \mathrm{diag}(1, 1, 0.1, 0.1)$. The measurement equation follows a power attenuation model [44],

$$h_i(\mathbf{x}_t) = \sqrt{\frac{P_0}{1 + (x_{t,1} - \beta_{i,1})^2 + (x_{t,2} - \beta_{i,2})^2}} \tag{44}$$

for $i = 1, 2, \ldots, m$, where $P_0 = 10^4$ is the signal power of the source, and the pair $(\beta_{i,1}, \beta_{i,2})$ is the position of the ith sensor. The covariance matrix of the measurement noise is given by (43) with $\rho = 0.035$.

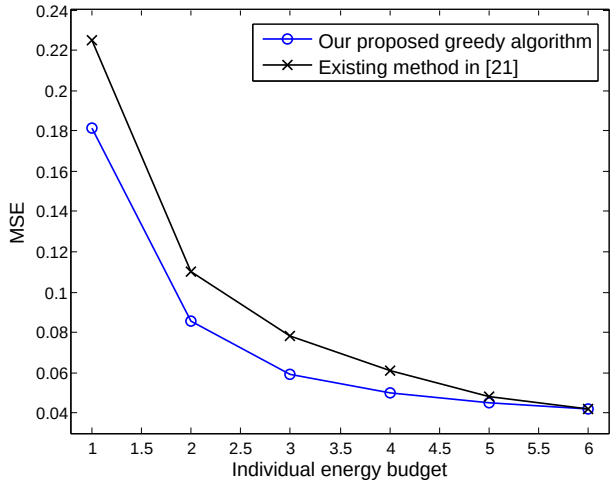


Fig. 4: MSE versus individual energy budget in target tracking.

In the sensor scheduling problem (42), we assume $s = \sum_{i=1}^{m} s_i$ and $s_1 = s_2 = \cdots = s_m$. In order to implement the proposed greedy algorithm and the existing method in [21], the nonlinear measurement function (44) is linearized at the prediction state $\hat{\mathbf{x}}_t = \mathbf{F}_{t-1}\mathbf{F}_{t-2}\cdots\mathbf{F}_0\hat{\mathbf{x}}_0$ as suggested in Remark 1. We determine sensor schedules for every $\tau = 6$ future time steps, and then update the estimate of the target state based on the selected measurements via an extended Kalman filter [45]. The estimation performance

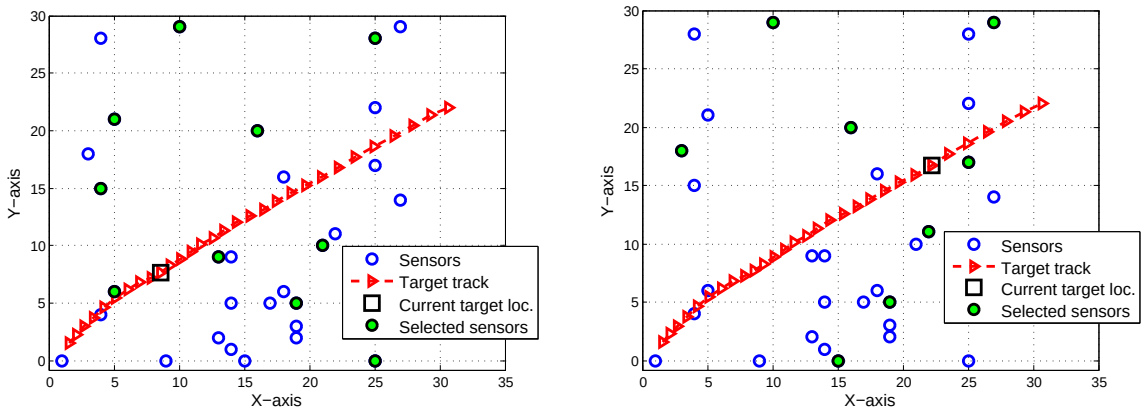


Fig. 5: Sensor schedules when $s_i = 2$: (a) $t = 10$, (b) $t = 24$.

is measured through the empirical MSE, which is obtained by averaging the estimation error over 30 time steps and 1000 simulation trials.

In Fig. 4, we present the MSE as a function of the individual energy budget. We compare the performance of our proposed greedy algorithm with that of the sensor scheduling method in [21]. We remark that the method in [21] relies on a reformulation of linearized dynamical systems and an ℓ_1 relaxation in optimization. In contrast, the proposed greedy algorithm is independent of the dynamical system models and convex relaxations. We observe that the greedy algorithm outperforms the method in [21]. This result together with the previous results in Fig. 1 and 2 have implied that the greedy algorithm could yield satisfactory estimation performance. Sensor schedules at time steps $t = 10$ and 24 are shown in Fig. 5. We observe that some sensors closest to the target are selected due to their high signal power. However, from the entire network point of view, the active sensors tend to be spatially distributed rather than aggregating in a small neighborhood around the target. This is because observations from neighboring sensors are strongly correlated in space and may lead to information redundancy in target tracking.

VII. Conclusion

In this paper, we studied the problem of sensor selection/scheduling with correlated measurement noise. We proposed a general but tractable framework to design optimal sensor activations. We pointed out some drawbacks of the existing frameworks for sensor selection with correlated noise, and showed that the existing formulation is valid only for the special case of weak noise correlation. Further, we extended our framework to the problem of non-myopic sensor scheduling, where a greedy algorithm was developed

to design non-myopic sensor schedules. Numerical results were provided to illustrate the effectiveness of our approach and the impact of noise correlation on the performance of sensor selection.

In future work, we will study applications of sensor selection with correlated noise, such as localization in multipath environments, sensor collaboration in distributed estimation, and clock synchronization in wireless sensor networks. It would also be of interest to seek theoretical guarantees for the performance of the greedy algorithm. Furthermore, in order to reduce the computational burden at the fusion center, developing a decentralized architecture where the optimization procedure can be carried out in a distributed way and by the sensors themselves is another direction of future research.

APPENDIX A
PROOF OF PROPOSITION 1

Given the sensor selection scheme $\tilde{\mathbf{w}}$, it is clear from (7) that Fisher information can be written as

$$\mathbf{J}_{\tilde{w}} = \mathbf{\Sigma}^{-1} + [\mathbf{H}_w^T, \mathbf{h}_j]\mathbf{R}_v^{-1} \begin{bmatrix} \mathbf{H}_w \\ \mathbf{h}_j^T \end{bmatrix}, \ \mathbf{R}_{\tilde{w}} := \begin{bmatrix} \mathbf{R}_w & \mathbf{r}_j \\ \mathbf{r}_j^T & r_{jj} \end{bmatrix} \tag{45}$$

where $\mathbf{H}_w := \mathbf{\Phi}_w \mathbf{H}$.

If $\mathbf{w} \neq \mathbf{0}$, the inverse of $\mathbf{R}_{\tilde{w}}$ in (45) is given by

$$\mathbf{R}_{\tilde{w}}^{-1} = c_j \begin{bmatrix} {c_j}^{-1}\mathbf{R}_w^{-1} + \mathbf{R}_w^{-1}\mathbf{r}_j\mathbf{r}_j^T\mathbf{R}_w^{-1} & -\mathbf{R}_w^{-1}\mathbf{r}_j \\ -\mathbf{r}_j^T\mathbf{R}_w^{-1} & 1 \end{bmatrix} \tag{46}$$

where $c_j := 1/(r_{jj} - \mathbf{r}_j^T\mathbf{R}_w^{-1}\mathbf{r}_j)$, and $c_j > 0$ following from the Schur complement of $\mathbf{R}_{\tilde{w}}$. Substituting (46) into (45), we obtain

$$\mathbf{J}_{\tilde{w}} = \mathbf{J}_w + c_j \boldsymbol{\alpha}_j \boldsymbol{\alpha}_j^T, \tag{47}$$

where $\mathbf{J}_w = \mathbf{\Sigma}^{-1} + \mathbf{H}_w^T\mathbf{R}_w^{-1}\mathbf{H}_w$ as indicated by (7), and $\alpha_j := \mathbf{H}_w^T\mathbf{R}_w^{-1}\mathbf{r}_j - \mathbf{h}_j$.

If $\mathbf{w} = \mathbf{0}$, namely, $\mathbf{J}_w = \mathbf{\Sigma}^{-1}$, we can immediately obtain from (45) that

$$\mathbf{J}_{\tilde{w}} = \mathbf{J}_w + \frac{1}{r_{jj}}\mathbf{h}_j^T\mathbf{h}_j. \tag{48}$$

Equations (47) and (48) imply that $\mathbf{J}_{\tilde{w}} - \mathbf{J}_w \succeq \mathbf{0}$ since $c_j > 0$.

We apply the matrix inversion lemma to (47). This yields

$$\mathbf{J}_{\tilde{w}}^{-1} = [\mathbf{J}_w + c_j\boldsymbol{\alpha}_j\boldsymbol{\alpha}_j^T]^{-1} = \mathbf{J}_w^{-1} - \frac{c_j\mathbf{J}_w^{-1}\boldsymbol{\alpha}_j\boldsymbol{\alpha}_j^T\mathbf{J}_w^{-1}}{1 + c_j\boldsymbol{\alpha}_j\mathbf{J}_w^{-1}\boldsymbol{\alpha}_j}.$$

The improvement in estimation error is then given by

$$\mathrm{tr}(\mathbf{J}_w^{-1}) - \mathrm{tr}(\mathbf{J}_{\tilde{w}}^{-1}) = \frac{c_j\boldsymbol{\alpha}_j^T\mathbf{J}_w^{-2}\boldsymbol{\alpha}_j}{1 + c_j\boldsymbol{\alpha}_j\mathbf{J}_w^{-1}\boldsymbol{\alpha}_j}.$$

■

APPENDIX B
PROOF OF PROPOSITION 2

Our goal is to simplify the Fisher information matrix given by (7) under the assumption of weak noise correlation. According to (32), we obtain

$$\begin{aligned}
\mathbf{R}_w^{-1} =&(\mathbf{\Phi}_w\mathbf{R}\mathbf{\Phi}_w^T)^{-1}\\
=&(\mathbf{\Phi}_w\mathbf{\Lambda}\mathbf{\Phi}_w^T+\epsilon\mathbf{\Phi}_w\mathbf{\Upsilon}\mathbf{\Phi}_w^T)^{-1}\\
\overset{(1)}{=}&(\mathbf{I}+\epsilon\mathbf{\Phi}_w\mathbf{\Lambda}^{-1}\mathbf{\Phi}_w^T\mathbf{\Phi}_w\mathbf{\Upsilon}\mathbf{\Phi}_w^T)^{-1}\mathbf{\Phi}_w\mathbf{\Lambda}^{-1}\mathbf{\Phi}_w^T\\
\overset{(2)}{=}&(\mathbf{I}-\epsilon\mathbf{\Phi}_w\mathbf{\Lambda}^{-1}\mathbf{\Phi}_w^T\mathbf{\Phi}_w\mathbf{\Upsilon}\mathbf{\Phi}_w^T)\mathbf{\Phi}_w\mathbf{\Lambda}^{-1}\mathbf{\Phi}_w^T\\
&+O(\epsilon^2)\quad(\text{as }\epsilon\to 0)\\
\overset{(3)}{=}&\mathbf{\Phi}_w\mathbf{\Lambda}^{-1}\mathbf{\Phi}_w^T-\epsilon\mathbf{\Phi}_w\mathbf{\Lambda}^{-1}\mathbf{D}_w\mathbf{\Upsilon}\mathbf{D}_w\mathbf{\Lambda}^{-1}\mathbf{\Phi}_w^T\\
&+O(\epsilon^2)\quad(\text{as }\epsilon\to 0),
\end{aligned}\tag{49}$$

where $\mathbf{D}_w := \mathrm{diag}(\mathbf{w})$. In (49), step (1) holds since we use the facts that $\mathbf{\Lambda}$ is a diagonal matrix and $(\mathbf{\Phi}_w\mathbf{\Lambda}\mathbf{\Phi}_w^T)^{-1}=\mathbf{\Phi}_w\mathbf{\Lambda}^{-1}\mathbf{\Phi}_w^T$; step (2) is obtained from the Taylor series expansion $(\mathbf{I}+\epsilon\mathbf{X})^{-1}=\sum_{i=0}^{\infty}(-\epsilon\mathbf{X})^i$ as $\epsilon\to 0$ (namely, the spectrum of $\epsilon\mathbf{X}$ is contained inside the open unit disk); step (3) is true since $\mathbf{\Phi}_w^T\mathbf{\Phi}_w=\mathbf{D}_w$ as in (4).

Substituting (49) into (7), we obtain

$$\begin{aligned}
\mathbf{J}_w =&\mathbf{\Sigma}^{-1}+\mathbf{H}^T\mathbf{\Phi}_w^T\mathbf{\Phi}_w\mathbf{\Lambda}^{-1}\mathbf{\Phi}_w^T\mathbf{\Phi}_w\mathbf{H}\\
&-\epsilon\mathbf{H}^T\mathbf{\Phi}_w^T\mathbf{\Phi}_w\mathbf{\Lambda}^{-1}\mathbf{D}_w\mathbf{\Upsilon}\mathbf{D}_w\mathbf{\Lambda}^{-1}\mathbf{\Phi}_w^T\mathbf{\Phi}_w\mathbf{H}\\
&+O(\epsilon^2)\quad(\text{as }\epsilon\to 0)\\
\overset{(1)}{=}&\mathbf{\Sigma}^{-1}+\mathbf{H}^T(\mathbf{D}_w\mathbf{\Lambda}^{-1}\mathbf{D}_w-\epsilon\mathbf{D}_w\mathbf{\Lambda}^{-1}\mathbf{\Upsilon}\mathbf{\Lambda}^{-1}\mathbf{D}_w)\mathbf{H}\\
&+O(\epsilon^2)\quad(\text{as }\epsilon\to 0)\\
=&\mathbf{\Sigma}^{-1}+\mathbf{H}^T\mathbf{D}_w(\mathbf{\Lambda}^{-1}-\epsilon\mathbf{\Lambda}^{-1}\mathbf{\Upsilon}\mathbf{\Lambda}^{-1})\mathbf{D}_w\mathbf{H}\\
&+O(\epsilon^2)\quad(\text{as }\epsilon\to 0)\\
\overset{(2)}{=}&\mathbf{\Sigma}^{-1}+\mathbf{H}^T\mathbf{D}_w\mathbf{R}^{-1}\mathbf{D}_w\mathbf{H}+O(\epsilon^2)\quad(\text{as }\epsilon\to 0)\\
\overset{(3)}{=}&\mathbf{\Sigma}^{-1}+\mathbf{H}^T(\mathbf{w}\mathbf{w}^T\circ\mathbf{R}^{-1})\mathbf{H}+O(\epsilon^2)\quad(\text{as }\epsilon\to 0),
\end{aligned}$$

where step (1) is achieved by using the fact that $\mathbf{D}_w\mathbf{\Lambda}^{-1}=\mathbf{\Lambda}^{-1}\mathbf{D}_w=\mathbf{D}_w\mathbf{\Lambda}^{-1}\mathbf{D}_w$, step (2) holds due to $\mathbf{R}^{-1}=\mathbf{\Lambda}^{-1}-\epsilon\mathbf{\Lambda}^{-1}\mathbf{\Upsilon}\mathbf{\Lambda}^{-1}+O(\epsilon^2)$, and step (3) is true since $\mathbf{D}_w$ is diagonal and has only binary elements. ■

APPENDIX C
PROOF OF PROPOSITION 3

We begin by simplifying the objective function in (P2),

$$\begin{aligned}\phi(\mathbf{w}) &:= \mathrm{tr}(\mathbf{\Sigma}^{-1}) + \mathrm{tr}\left((\mathbf{w}\mathbf{w}^T \circ \mathbf{R}^{-1})(\mathbf{H}^T\mathbf{H})\right) \\ &= \mathrm{tr}(\mathbf{\Sigma}^{-1}) + \sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j \bar{R}_{ij}\mathbf{h}_i^T\mathbf{h}_j \\ &= \mathrm{tr}(\mathbf{\Sigma}^{-1}) + \mathbf{w}^T\mathbf{\Omega}\mathbf{w}, \end{aligned} \tag{50}$$

where $\bar{R}_{ij}$ is the (i,j)th entry of $\mathbf{R}^{-1}$, and $\bar{R}_{ij}\mathbf{h}_i^T\mathbf{h}_j$ corresponds to the (i,j)th entry of $\mathbf{\Omega}$ which yields the succinct form

$$\mathbf{\Omega} = \mathbf{A}(\mathbf{R}^{-1} \otimes \mathbf{I}_n)\mathbf{A}^T. \tag{51}$$

In (51), $\otimes$ denotes the Kronecker product, $\mathbf{A} \in \mathbb{R}^{m \times mn}$ is a block-diagonal matrix whose diagonal blocks are given by $\{\mathbf{h}_i^T\}_{i=1}^m$, and $\mathbf{\Omega} \succeq 0$ due to $\mathbf{R}^{-1} \otimes \mathbf{I}_n \succeq 0$.

According to (50), (P2) can be rewritten as

$$\begin{aligned} \underset{\mathbf{w}}{\text{maximize}} \quad & \mathbf{w}^T\mathbf{\Omega}\mathbf{w} \\ \text{subject to} \quad & \mathbf{1}^T\mathbf{w} \leq s, \\ & \mathbf{w} \in \{0,1\}^m. \end{aligned} \tag{52}$$

Next, we prove that problem (35) is equivalent to problem (52). We recall that the former is a relaxation of the latter, where the former entails the maximization of a convex quadratic function over a bounded polyhedron $\mathcal{P} := \{\mathbf{w} | \mathbf{1}^T\mathbf{w} \leq s, \mathbf{w} \in [0,1]^m\}$. It has been shown in [46] that optimal solutions of such a problem occur at vertices of the polyhedron $\mathcal{P}$, which are zero-one vectors. This indicates that solutions of problem (35) are feasible for problem (52). Therefore, solutions of (35) are solutions of (52), and vice versa. ■

REFERENCES

[1] L. Oliveira and J. Rodrigues, "Wireless sensor networks: a survey on environmental monitoring," *Journal of Communications*, vol. 6, no. 2, 2011.

[2] Y. Zou and K. Chakrabarty, "Sensor deployment and target localization in distributed sensor networks," *ACM Transactions on Embedded Computing Systems*, vol. 3, no. 1, pp. 61–91, Feb. 2004.

[3] T. He, P. Vicaire, T. Yan, L. Luo, L. Gu, G. Zhou, S. Stoleru, Q. Cao, J. A. Stankovic, and T. Abdelzaher, "Achieving real-time target tracking using wireless sensor networks," in *Proceedings of IEEE Real Time Technology and Applications Symposium*, 2006, pp. 37–48.

[4] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 61–72, Mar. 2002.

[5] E. Masazade, R. Niu, and P. K. Varshney, "Dynamic bit allocation for object tracking in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5048–5063, Oct. 2012.

[6] H. Zhang, J. Moura, and B. Krogh, "Dynamic field estimation using wireless sensor networks: Tradeoffs between estimation error and communication cost," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2383–2395, June 2009.

[7] S. Liu, A Vempaty, M. Fardad, E. Masazade, and P. K. Varshney, "Energy-aware sensor selection in field reconstruction," *IEEE Signal Processing Letters*, vol. 21, no. 12, pp. 1476–1480, 2014.

[8] S. Liu, M. Fardad, P. K. Varshney, and E. Masazade, "Optimal periodic sensor scheduling in networks of dynamical systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3055–3068, June 2014.

[9] G. Thatte and U. Mitra, "Sensor selection and power allocation for distributed estimation in sensor networks: Beyond the star topology," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2649–2661, July 2008.

[10] S. Liu, S. Kar, M. Fardad, and P. K. Varshney, "Sparsity-aware sensor collaboration for linear coherent estimation," *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2582–2596, May 2015.

[11] K. Chaloner and I. Verdinelli, "Bayesian experimental design: A review," *Statistical Science*, vol. 10, no. 3, pp. 273–304, 1995.

[12] F. Lin, M. Fardad, and M. R. Jovanović, "Algorithms for leader selection in stochastically forced consensus networks," *IEEE Transactions on Automatic Control*, vol. 59, no. 7, pp. 1789–1802, July 2014.

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

[14] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 684–698, Feb. 2015.

[15] E. Masazade, R. Niu, and P. K. Varshney, "An approximate dynamic programming based non-myopic sensor selection method for target tracking," in *Proc. the 46th Annual Conf. Information Sciences and Systems*, March 2012, pp. 1–6.

[16] H. L. Van Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering Tracking*, Wiley-IEEE press, 2007.

[17] A. Jindal and K. Psounis, "Modeling spatially correlated data in sensor networks," *ACM Transactions on Sensor Networks*, vol. 2, pp. 466–499, Nov. 2006.

[18] E. Rigtorp, "Sensor selection with correlated noise," M.S. thesis, KTH Royal Institute of Technology, Aug. 2010.

[19] H. Jamali-Rad, A. Simonetto, G. Leus, and X. Ma, "Sparsity-aware sensor selection for correlated noise," in *Proceedings of 17th International Conference on Information Fusion (FUSION)*, July 2014, pp. 1–7.

[20] X. Shen and P. K. Varshney, "Sensor selection based on generalized information gain for target tracking in large sensor networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 363–375, Jan 2014.

[21] Y. Mo, R. Ambrosino, and B. Sinopoli, "Sensor selection strategies for state estimation in energy constrained wireless sensor networks," *Automatica*, vol. 47, no. 7, pp. 1330–1338, 2011.

[22] S. Liu, E. Masazade, M. Fardad, and P. K. Varshney, "Sensor selection with correlated measurements for target tracking in wireless sensor networks," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 4030–4034.

[23] J. O. Berger, V. De Oliveira, and B. Sanso, "Objective bayesian analysis of spatially correlated data," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1361–1374, Nov. 2001.

[24] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[25] S. P. Chepuri and G. Leus, "Sparse sensing for distributed gaussian detection," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 2394–2398.

[26] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[27] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, May 2010.

[28] A. Nemirovski, "Interior point polynomial time methods in convex programming," 2012 [Online], Available: *http://www2.isye.gatech.edu/ nemirovs/Lect_IPM.pdf*.

[29] A. Bertrand, J. Szurley, P. Ruckebusch, I. Moerman, and M. Moonen, "Efficient calculation of sensor utility and sensor removal in wireless sensor networks for adaptive signal estimation and beamforming," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5857–5869, Nov. 2012.

[30] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: Leveraging submodularity," in *Proc. of 49th IEEE Conference on Decision and Control (CDC)*, Dec. 2010, pp. 2572–2577.

[31] V. V. Williams, "Multiplying matrices faster than Coppersmith-Winograd," in *Proc. the 44th symposium on Theory of Computing*, 2012, pp. 887–898.

[32] F. E. Udwadia, "Methodology for optimum sensor locations for parameter identification in dynamic systems," *Journal of Engineering Mechanics*, vol. 120, no. 2, pp. 368–390, Feb. 1994.

[33] X. Shen, S. Liu, and P. K. Varshney, "Sensor selection for nonlinear systems in large sensor networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 4, pp. 2664–2678, October 2014.

[34] T. H. McLoughlin and M. Campbell, "Solutions to periodic sensor scheduling problems for formation flying missions in deep space," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 2, pp. 1351–1368, April 2011.

[35] J. Fang and H. Li, "Power constrained distributed estimation with correlated sensor data," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3292–3297, Aug. 2009.

[36] P. M. Pardalos, "Global optimization algorithms for linearly constrained indefinite quadratic problems," *Computers & Mathematics with Applications*, vol. 21, no. 67, pp. 87 – 97, 1991.

[37] E. Masazade, R. Niu, and P. K. Varshney, "Dynamic bit allocation for object tracking in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5048–5063, Oct. 2012.

[38] E. Masazade, M. Fardad, and P. K. Varshney, "Sparsity-promoting extended Kalman filtering for target tracking in wireless sensor networks," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 845–848, Dec. 2012.

[39] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior Cramér-Rao bounds for discrete-time nonlinear filtering," *IEEE Transactions on Signal Processin*, vol. 46, no. 5, pp. 1386–1396, may 1998.

[40] S. P. Chepuri and G. Leus, "Sparsity-promoting adaptive sensor selection for non-linear filtering," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 5080–5084.

[41] S. Liu, E. Masazade, M. Fardad, and P. K. Varshney, "Sparsity-aware field estimation via ordinary Kriging," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 3976–3980.

[42] A. d'Aspremont and S. Boyd, "Relaxations and randomized methods for nonconvex QCQPs," Stanford, CA: Stanford Univ., Autumn 2003 [Online], Available: *http://web.stanford.edu/class/ee392o/relaxations.pdf*.

[43] F. Peng and B. Chen, "Decentralized estimation with correlated additive noise: Does dependency always imply redundancy?," in *Proceedings of Asilomar Conference on Signals, Systems and Computers*, Nov 2013, pp. 677–681.

[44] R. Niu and P. K. Varshney, "Target location estimation in sensor networks with quantized data," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4519–4528, Dec. 2006.

[45] B. M. Yu, K. V. Shenoy, and M. Sahani, "Derivation of extended kalman filtering and smoothing equations," 2004.

[46] H. Konno, "Maximization of a convex quadratic function under linear constraints," *Mathematical Programming*, vol. 11, no. 1, pp. 117–127, 1976.

Sijia Liu (S'13) was born in Zhengzhou, Henan, China, on December 19, 1985. He received the B.S. and M.S. degrees in electrical engineering from Xi'an Jiaotong University in 2008 and 2011, respectively. Since 2011, he has been working towards his Ph.D. degree in electrical and computer engineering at Syracuse University. He was a research intern in the Wireless Access Lab at Futurewei Technologies Corporation in Bridgewater, NJ in summer 2015.

His research interests include statistical signal processing, convex/nonconvex optimization, resource management for wireless sensor networks, and machine learning. He was one of the Best Student Paper Finalists at Asilomar Conference on Signals, Systems, and Computers, 2013. He was the winner of the Nunan Research Poster Competition at Syracuse University in 2012.

Sundeep Prabhakar Chepuri (S'11-M'16) was born in India in 1986. He received his M.Sc. degree *(cum laude)* in electrical engineering and Ph.D. degree *(cum laude)* from the Delft University of Technology, The Netherlands, in July 2011 and January 2016, respectively. He has held positions at Robert Bosch, India, during 2007-2009, and Holst Centre/imec-nl, The Netherlands, during 2010-2011. He is currently a postdoctoral scholar with the Circuits and Systems group at the Faculty of Electrical Engineering, Mathematics and Computer Science of the Delft University of Technology, The Netherlands. His general research interest lies in the field of mathematical signal processing, statistical inference, sensor networks, and wireless communications.

Dr. Chepuri received the Best Student Paper Award for his publication at the ICASSP 2015 conference in Australia.

Makan Fardad (M'08) received the B.S. and M.S. degrees in electrical engineering from Sharif University of Technology and Iran University of Science and Technology, respectively. He received the Ph.D. degree in mechanical engineering from the University of California, Santa Barbara in 2006. He was a Postdoctoral Associate at the University of Minnesota before joining the Department of Electrical Engineering and Computer Science at Syracuse University in 2008. His research interests are in modeling, analysis, and optimization of large-scale dynamical networks.

Engin Maşazade (S'03-M'10) received his B.S. degree from the Department of Electronics and Communications Engineering, Istanbul Technical University, Istanbul, Turkey, in 2003. He received his M.S. and Ph.D. degrees from the Electronics Engineering Program, Sabanci University, Istanbul, Turkey, in 2006 and 2010, respectively.

He is currently an assistant professor with the Department of Electrical and Electronics Engineering, Yeditepe University, Istanbul, Turkey. Before joining Yeditepe University he was a postdoctoral research associate with the Department of Electrical Engineering and Computer Science, Syracuse University, NY, USA. His research interests include distributed detection, localization, and tracking for wireless sensor networks, and dynamic resource management in sensor/communication networks. In 2013, he received the TUBITAK (Scientific and Technological Research Council of Turkey) CAREER Award.

Geert Leus (M'01-SM'05-F'12) received the MSc and PhD degree in applied sciences from the Katholieke Universiteit Leuven, Belgium, in June 1996 and May 2000, respectively. Currently, Geert Leus is an "Antoni van Leeuwenhoek" Full Professor at the Faculty of Electrical Engineering, Mathematics and Computer Science of the Delft University of Technology, The Netherlands. His research interests are in the area of signal processing for communications. Geert Leus received a 2002 IEEE Signal Processing Society Young Author Best Paper Award and a 2005 IEEE Signal Processing Society Best Paper Award. He is a Fellow of the IEEE and a Fellow of EURASIP. Geert Leus was the Chair of the IEEE Signal Processing for Communications and Networking Technical Committee, and an Associate Editor for the IEEE Transactions on Signal Processing, the IEEE Transactions on Wireless Communications, the IEEE Signal Processing Letters, and the EURASIP Journal on Advances in Signal Processing. Currently, he is a Member-at-Large to the Board of Governors of the IEEE Signal Processing Society and a member of the IEEE Sensor Array and Multichannel Technical Committee. He finally serves as the Editor in Chief of the EURASIP Journal on Advances in Signal Processing.

Pramod K. Varshney (S'72-M'77-SM'82-F'97) was born in Allahabad, India, on July 1, 1952. He received the B.S. degree in electrical engineering and computer science (with highest honors), and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana-Champaign in 1972, 1974, and 1976 respectively.

Since 1976, he has been with Syracuse University, Syracuse, NY, where he is currently a Distinguished Professor of Electrical Engineering and Computer Science and the Director of CASE: Center for Advanced Systems and Engineering. He is also an Adjunct Professor of Radiology at Upstate Medical University, Syracuse. His current research interests are in distributed sensor networks and data fusion, detection and estimation theory, wireless communications, image processing, radar signal processing, and remote sensing. He is the author of Distributed Detection and Data Fusion (New York: Springer-Verlag, 1997).

Dr. Varshney was a James Scholar, a Bronze Tablet Senior, and a Fellow while at the University of Illinois. He is a member of Tau Beta Pi and is the recipient of the 1981 ASEE Dow Outstanding Young Faculty Award. He was elected to the grade of Fellow of the IEEE in 1997 for his contributions in the area of distributed detection and data fusion. He was the Guest Editor of the Special Issue on Data Fusion of the IEEE PROCEEDINGS January 1997. In 2000, he received the Third Millennium Medal from the IEEE and Chancellor's Citation for exceptional academic achievement at Syracuse University. He is the recipient of the IEEE 2012 Judith A. Resnik Award, an honorary Doctor of Engineering degree from Drexel University in 2014, and ECE Distinguished Alumni Award from UIUC in 2015. He is on the Editorial Boards of the Journal on Advances in Information Fusion and IEEE Signal Processing Magazine. He was the President of International Society of Information Fusion during 2001.