

Sparsity-Promoting Sensor Selection for Non-Linear Measurement Models

Sundeep Prabhakar Chepuri, *Student Member, IEEE*, and Geert Leus, *Fellow, IEEE*

Abstract—The problem of choosing the best subset of sensors that guarantees a certain estimation performance is referred to as sensor selection. In this paper, we focus on observations that are related to a general non-linear model. The proposed framework is valid as long as the observations are independent, and its likelihood satisfies the regularity conditions. We use several functions of the Cramér–Rao bound (CRB) as a performance measure. We formulate the sensor selection problem as the design of a sparse vector, which in its original form is a nonconvex ℓ_0 -(quasi) norm optimization problem. We present relaxed sensor selection solvers that can be efficiently solved in polynomial time. The proposed solvers result in sparse sensing techniques. We also propose a projected subgradient algorithm that is attractive for large-scale problems. The developed theory is applied to sensor placement for localization.

Index Terms—Convex optimization, Cramér–Rao bound, non-linear models, projected subgradient algorithm, sensor networks, sensor placement, sensor selection, sparse sensing, sparsity.

I. INTRODUCTION

ADVANCES in sensor technology have enabled a large spectrum of applications and services related to safety and security, surveillance, environmental and climate monitoring, to list a few. The sensor nodes are spatially deployed and operate as a network, with each sensor node capable of sensing, processing, and communicating to other nodes or a central processing unit. As a network, their fundamental task is distributed data sampling (i.e., to sense the environment) from which we seek to extract relevant information. The sensors generate a prohibitively large dataset, which is usually gathered at a fusion center. This gathered data has to be optimally processed, rejecting the redundant, identical, or faulty measurements.

Sensor selection is a fundamental design task in sensor networks. The number of sensors are often limited either by economical constraints (hardware costs), or the availability of physical or storage space. In order to reduce the hardware costs,

as well as the resulting communications and processing overhead, one would like to smartly deploy the sensors taking into account the associated performance trade-off. Sensor selection also enables the design of spatio-temporal samplers that guarantee a certain performance measure such as energy-efficiency, information measure, estimation accuracy, or detection probability. For example, the sensor placement problem can be interpreted as a sensor selection problem in which the best subset of the available sensor locations are selected subject to a specific performance constraint. Sensor selection is pertinent to various diverse fields, especially to applications dealing with large-scale networks like network monitoring [2], [3], location-aware services like target localization and tracking [4]–[6], field (e.g., heat, sound, precipitation) estimation [7], [8], and environmental monitoring in general. The fundamental questions of interest are:

- q1. Where to deploy the limited sensors available?
- q2. Do we need to process all the acquired measurements?

To this end, we focus on gathering only the most informative data via *sparse sensing* for a general (non-linear) inverse problem.

A. Related Prior Works

A large volume of literature exists on sensor selection [9, and references therein]. The sensor selection problem is often formulated as an optimization problem based on some well-known performance measures from experimental design [9], [10], [11, p. 384]. The sensor selection problem is expressed as the following optimization problem:

$$\arg \min_{\mathbf{w} \in \{0,1\}^M} f(\mathbf{E}(\mathbf{w})) \quad \text{s.t.} \quad \mathbf{1}_M^T \mathbf{w} = K, \quad (1)$$

where \mathbf{w} is a selection vector of length M , and $f(\mathbf{E}(\mathbf{w}))$ is a scalar cost function related to the error covariance matrix \mathbf{E} . The error covariance matrix is optimized to select the best subset of K sensors out of M available sensors. Different functions $f(\mathbf{E}(\mathbf{w}))$ can be used, and the typical choices for $f(\mathbf{E}(\mathbf{w}))$ are related to:

- 1) *A-optimality*: minimizes the sum of eigenvalues of \mathbf{E} with $f(\mathbf{E}(\mathbf{w})) := \text{tr}\{\mathbf{E}(\mathbf{w})\}$.
- 2) *E-optimality*: minimizes the maximum eigenvalue of \mathbf{E} with $f(\mathbf{E}(\mathbf{w})) := \lambda_{\max}\{\mathbf{E}(\mathbf{w})\}$.
- 3) *D-optimality*: minimizes the determinant of \mathbf{E} with $f(\mathbf{E}(\mathbf{w})) := \ln \det\{\mathbf{E}(\mathbf{w})\}$.

All the above measures are equally reasonable, although neither of them completely characterizes the error covariance. There is no general answer to the question of how does one performance metric compare with the other.

Manuscript received September 19, 2013; revised March 24, 2014 and September 01, 2014; accepted November 19, 2014. Date of publication December 10, 2014; date of current version January 08, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. John McAllister. This work was supported in part by STW under the FASTCOM project (10551) and in part by NWO-STW under the VICI program (10382). A conference precursor of this manuscript has been published in the *Proceedings of the Twenty-First European Signal Processing Conference*, September 2013 [1].

The authors are with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft 2628CD, The Netherlands (e-mail: s.p.chepuri@tudelft.nl; g.j.t.leus@tudelft.nl).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2014.2379662

The problem in (1) is a combinatorial optimization problem involving $\binom{K}{M}$ searches, and it is clearly intractable even for small-scale problems. To simplify this problem, the nonconvex Boolean constraint $\mathbf{w} \in \{0, 1\}^M$ is relaxed to a convex box constraint $\mathbf{w} \in [0, 1]^M$. The relaxed optimization problem has been studied in [9] for additive Gaussian linear models, where the matrix \mathbf{E} can be expressed in closed form. More importantly, the above listed performance measures for additive Gaussian linear models are independent of the unknown parameter. Moreover, in practice, the exact number of sensors K to select might not be known. Nevertheless, this number K can always be tuned to achieve a desired performance.

The above selection problem is applied to sensor placement for power grid monitoring in [12]. Alternative greedy approaches exploiting the submodularity of the objective function [13]–[17] are also proposed to solve the sensor selection problem. Sensor selection for dynamical systems often referred to as sensor polling or scheduling, is studied in [18]–[20]. In [21], the sensor placement problem for linear models is addressed as the design of a sensing matrix that optimizes a measure related to the orthogonality of its rows. All the above literature (in general) deals with measurements that are related to additive Gaussian linear models. Experimental design for non-linear models within the Bayesian and sequential design frameworks is discussed in [22]. In [18], sensor selection for target tracking based on extended Kalman filtering has been proposed, in which the selection is performed by designing a sparse gain matrix. Moreover [18] is limited to an additive Gaussian non-linear model, that is linearized around the (noisy) past state estimate. This paper, on the other hand, deals with general non-linear models, without an explicit linearization. Sensor selection for detection problems is studied in [23]. In [24], reliable sensor selection based on the actual measurements to identify the outliers is presented. A different problem, yet related to sensor selection, is the problem of identifying source-informative sensors, which is studied in [25].

B. Contributions

We consider general scenarios where the measurements of the unknown parameter follow a non-linear model (unlike [9] for instance). Non-linear measurement models are frequently encountered in applications like source localization, tracking, field estimation, or phase retrieval, to list a few. The error covariance matrix for non-linear models is not always available in closed form, and more importantly it depends on the unknown parameter. Our first contribution in the context of sensor selection is to leverage the additive property of the inverse Cramér-Rao bound (CRB) or the Fisher information matrix (FIM) for independent observations, and thus to express the performance requirement as a convex set. The CRB is a rigorous performance measure for optimality, and it generalizes very well for non-linear measurement models (not necessarily in additive Gaussian noise). Moreover, we do not need the actual measurements, and hence, our framework is also well-suited for solving offline design problems. In addition to this, the number of sensors that have to be selected, i.e., K , is generally not known in practice. Hence, instead of fixing K as in (1), we pose sensor selection alternatively

as a cardinality minimization problem that provides the number of selected sensors as a byproduct.

The proposed sensor selection framework is generic and can be applied to any non-linear estimation problem (linear being a special case). The selection problem is formulated as the design of a sparse selection vector, which is an ℓ_0 -(quasi) norm non-convex Boolean optimization problem. The nonconvex sensor selection problem is relaxed using standard convex relaxation techniques that can be efficiently solved in polynomial time.

A sparsity-enhancing concave surrogate for the ℓ_0 -(quasi) norm is also proposed for sensor selection as an alternative to the traditional best convex relaxation. This is particularly advantageous when there are multiple (nearly) identical sensor measurements.

To cope with large-scale problems, we further present a projected subgradient algorithm. It is worth mentioning that the projected subgradient algorithm allows a very easy distributed implementation. In essence, we seek a sparse vector (i.e., a vector with many zeros and a few non-zero entries) that determines the sensing pattern. Sparse sensing leads to energy-efficient sampling schemes. We illustrate the sensor selection problem using examples of sensor placement for source localization.

C. Outline and Notations

The remainder of the paper is organized as follows. In Section II, we present the non-linear measurement model. In Section III, we show the problem formulation, and in Section IV we present the algorithms that solve the relaxed optimization problem. In Section V, we derive the Lagrangian dual problem, and provide some extensions. In Section VI, the proposed framework is applied to a number of different models related to localization. The paper finally concludes with Section VII.

The notation used in this paper can be described as follows. Upper (lower) bold face letters are used for matrices (column vectors). $(\cdot)^T$ denotes transposition. $\text{diag}(\cdot)$ refers to a block diagonal matrix with the elements in its argument on the main diagonal. $\mathbf{1}_N$ ($\mathbf{0}_N$) denotes the $N \times 1$ vector of ones (zeros). \mathbf{I}_N is an identity matrix of size N . $\mathbb{E}\{\cdot\}$ denotes the expectation operation. $\text{tr}\{\cdot\}$ is the matrix trace operator. $\det\{\cdot\}$ is the matrix determinant. $\lambda_{\min}\{\mathbf{A}\}$ ($\lambda_{\max}\{\mathbf{A}\}$) denotes the minimum (maximum) eigenvalue of a symmetric matrix \mathbf{A} . $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is a positive semidefinite matrix. \mathbb{S}^N (\mathbb{S}_+^N) denotes the set of symmetric (symmetric positive semi-definite) matrices of size $N \times N$. $|\mathcal{U}|$ denotes the cardinality of the set \mathcal{U} .

II. NON-LINEAR MEASUREMENT MODEL

In this paper, we consider a generic non-linear measurement model

$$y_m = h_m(\boldsymbol{\theta}, n_m), \quad m = 1, 2, \dots, M, \quad (2)$$

where y_m is the m th spatial or temporal sensor measurement, $\boldsymbol{\theta} \in \mathbb{R}^N$ is the unknown parameter, n_m for $m = 1, 2, \dots, M$, is the noise process, and the regressors h_m for $m = 1, 2, \dots, M$, are (in general) non-linear functionals. Let the vector $\mathbf{y} = [y_1, y_2, \dots, y_M]^T \in \mathbb{R}^M$ collect the measurements. The likelihood of the measurements $p(\mathbf{y}; \boldsymbol{\theta})$ is the probability

density function (pdf) of \mathbf{y} parameterized by the unknown vector $\boldsymbol{\theta}$.

We make the following assumptions:

a1. *Regularity conditions*: The log-likelihood of the measurements satisfies the regularity condition $\mathbb{E} \left\{ \frac{\partial \ln p(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\} = 0$. This is a well-known condition for the CRB to exist [26].

a2. *Independent observations*: The measurements y_m for $m = 1, 2, \dots, M$, are a sequence of independent random variables.

The proposed framework for sensor selection is valid as long as the above two assumptions hold.

Assuming (a1) holds, the covariance of any unbiased estimate $\hat{\boldsymbol{\theta}} \in \mathbb{R}^N$ of the unknown parameter satisfies the well-known inequality [26]

$$\mathbb{E} \left\{ (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \right\} \geq \mathbf{C}(\boldsymbol{\theta}) = \mathbf{F}^{-1}(\boldsymbol{\theta}),$$

where the Fisher information matrix (FIM) is given by [26]

$$\begin{aligned} \mathbf{F}(\boldsymbol{\theta}) &= -\mathbb{E} \left\{ \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \left(\frac{\ln p(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right\} \\ &= \mathbb{E} \left\{ \left(\frac{\partial \ln p(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ln p(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right\} \in \mathbb{R}^{N \times N}, \end{aligned} \quad (3)$$

and $\mathbf{C}(\boldsymbol{\theta})$ is the CRB matrix. An important property of the Fisher information is that it is *additive* for independent observations, which follows from the fact that

$$\ln p(\mathbf{y}; \boldsymbol{\theta}) = \ln \prod_{m=1}^M p(y_m; \boldsymbol{\theta}) = \sum_{m=1}^M \ln p(y_m; \boldsymbol{\theta}), \quad (4)$$

where we assume that condition (a2) holds. Using (4) in (3), the FIM $\mathbf{F}(\boldsymbol{\theta})$ can be alternatively expressed as

$$\mathbf{F}(\boldsymbol{\theta}) = \sum_{m=1}^M -\mathbb{E} \left\{ \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \left(\frac{\ln p(y_m; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right\}$$

which can be further simplified to

$$\mathbf{F}(\boldsymbol{\theta}) = \sum_{m=1}^M \mathbf{F}_m(\boldsymbol{\theta}), \quad (5)$$

where

$$\begin{aligned} \mathbf{F}_m(\boldsymbol{\theta}) &= -\mathbb{E} \left\{ \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \left(\frac{\ln p(y_m; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right\} \\ &= \mathbb{E} \left\{ \left(\frac{\partial \ln p(y_m; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ln p(y_m; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right\}, \end{aligned} \quad (6)$$

is the $N \times N$ FIM of the m th measurement. In other words, (5) means that every independent measurement contributes to the information measure. Note that the FIM for non-linear models depends on the unknown vector $\boldsymbol{\theta}$.

Assume for instance that the observations belong to the family of exponential distributions. The log-likelihood of the observations can then be expressed in the form

$$\ln p(y_m; \boldsymbol{\theta}) = \ln r(y_m) + a_m(\boldsymbol{\theta})b(y_m) - c(\boldsymbol{\theta}), \quad (7)$$

where $r(y_m)$ and $b(y_m)$ are known functions of the observations only, while $a_m(\boldsymbol{\theta})$ and $c(\boldsymbol{\theta})$ depend only on the unknown parameter. The regularity conditions in general hold for observations that belong to the family of exponential pdfs, and it already includes a large number of distributions.

One specific example that often occurs in practice is the case where the observations y_m , $m = 1, 2, \dots, M$, are related through the following additive Gaussian non-linear model

$$y_m = h_m(\boldsymbol{\theta}) + n_m, \quad m = 1, 2, \dots, M, \quad (8)$$

where $h_m(\cdot)$ is a non-linear function, and n_m is a zero-mean Gaussian random variable with variance σ_m^2 . The log-likelihood of y_m is then given by (7) with

$$\begin{aligned} r(y_m) &= \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp \left(-\frac{1}{2\sigma_m^2} y_m^2 \right), \\ b(y_m) &= \frac{y_m}{\sigma_m^2}, \\ a_m(\boldsymbol{\theta}) &= h_m(\boldsymbol{\theta}), \\ \text{and } c(\boldsymbol{\theta}) &= \frac{1}{2\sigma_m^2} h_m^2(\boldsymbol{\theta}). \end{aligned}$$

Assuming (a2) holds, it is then easy to verify that (6) simplifies to

$$\mathbf{F}_m(\boldsymbol{\theta}) = \frac{1}{\sigma_m^2} \left(\frac{\partial h_m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial h_m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T.$$

Remark 1 (Additive Gaussian Linear Model): As a special case, when the measurement process is linear, we have $y_m = \mathbf{h}_m^T \boldsymbol{\theta} + n_m$, $m = 1, 2, \dots, M$, i.e., $h_m(\boldsymbol{\theta}, n_m) := \mathbf{h}_m^T \boldsymbol{\theta} + n_m$ with $\mathbf{h}_m \in \mathbb{R}^N$ being the regressor. The computation of the FIM for a linear model is straightforward, and is given by

$$\mathbf{F} = \sum_{m=1}^M \frac{1}{\sigma_m^2} \mathbf{h}_m \mathbf{h}_m^T.$$

The CRB for linear models in additive Gaussian noise is equal to the mean squared error (MSE), and more importantly it is *independent* of the unknown vector $\boldsymbol{\theta}$.

III. PROBLEM FORMULATION

Our goal is now to select the best subset ($\geq N$) of the M available (or candidate) sensors such that a certain accuracy on the estimate $\hat{\boldsymbol{\theta}}$ is guaranteed. We next mathematically formulate this sensor selection problem.

A. Sparse Sensing

We model the sensing operation using a selection vector

$$\mathbf{w} = [w_1, w_2, \dots, w_M]^T \in \{0, 1\}^M,$$

where $w_m = 1(0)$ indicates that the m th sensor measurement is (not) selected. The measurement model including the selection parameter is given as

$$y_m = w_m h_m(\boldsymbol{\theta}, n_m), \quad m = 1, 2, \dots, M, \quad (9)$$

where the selection vector modifies the log-likelihood of the selected measurements¹ as

$$\ln \prod_{m=1}^M p(y_m; \boldsymbol{\theta})^{w_m} = \sum_{m=1}^M w_m \ln p(y_m; \boldsymbol{\theta}).$$

Furthermore, the additive property of the FIM matrix allows us to introduce the selection variable in (5) to obtain

$$\mathbf{F}(\mathbf{w}, \boldsymbol{\theta}) = \sum_{m=1}^M w_m \mathbf{F}_m(\boldsymbol{\theta}). \quad (10)$$

In other words, we use the Boolean selection parameter to choose the most informative sensors (or measurements). The entries of \mathbf{w} are designed to be as sparse as possible such that the sensing cost is minimized, yet guaranteeing a certain estimation quality. Hence, the resulting sensing scheme is referred to as *sparse sensing*.

B. Performance Measures

We do not restrict ourselves to any specific estimator, however, we use the CRB as a performance measure. The motivation behind using the CRB is as follows:

- 1) The CRB is a measure for the (local) identifiability of the problem [27]. More specifically, a non-singular FIM implies (local) solvability and a unique estimate of $\boldsymbol{\theta}$, however, the converse is not necessarily true. The sensor selection problem presented in this paper seeks a subset of sensors for which the FIM has full rank in some domain such that the solvability of the problem in that domain is always ensured.
- 2) Typically, the subset of selected sensors that yields a lower CRB also yields a lower MSE, and thus improves the performance of any practical system.

The CRB also has a very attractive mathematical structure resulting in a selection problem that can be efficiently solved using convex optimization techniques.

We next use the consistency assumption of the estimator to derive thresholds for the performance measures. We constrain the estimation error $\boldsymbol{\epsilon} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ to be within an origin-centered circle of radius R_e with a probability higher than P_e , i.e.,

$$\Pr(\|\boldsymbol{\epsilon}\|_2 \leq R_e) \geq P_e, \quad (11)$$

where $\Pr(\cdot)$ denotes probability, and the values of R_e and P_e define the accuracy required and are assumed to be known. A higher accuracy level is obtained by reducing R_e and/or increasing P_e . This metric is used in several occasions as an accuracy measure (e.g., see [4], [28], [29]). We next discuss two popular performance measures that satisfy the above requirement.

1) *Trace Constraint*: A sufficient condition to satisfy the accuracy requirement in (11) is (see Appendix A)

$$\text{tr}\{\mathbf{C}(\mathbf{w}, \boldsymbol{\theta})\} = \text{tr}\left\{\left(\sum_{m=1}^M w_m \mathbf{F}_m(\boldsymbol{\theta})\right)^{-1}\right\} \leq \lambda_{\text{tr}} = (1-P_e)R_e^2.$$

¹The pdf of the selected measurements is of reduced dimension, i.e., it does not include the measurements that are set to zero.

This measure is related to the *A-optimality* or the average-variance criterion, which restricts the sum of the semi-axes of the confidence ellipsoid to λ_{tr} .

2) *Minimum Eigenvalue Constraint*: Another popular sufficient condition that also satisfies the accuracy requirement in (11) is

$$\lambda_{\min}\{\mathbf{F}(\mathbf{w}, \boldsymbol{\theta})\} \geq \lambda_{\text{eig}} = \frac{N}{R_e^2} \left(\frac{1}{1-P_e} \right),$$

where λ_{eig} is derived in [29] (see also Appendix A). This measure is related to the *E-optimality* or the worst-case error, which restricts the semi-major axis of the confidence ellipsoid to λ_{eig} . The inequality constraint $\lambda_{\min}\{\mathbf{F}\} \geq \lambda_{\text{eig}}$ can be equivalently expressed as the following linear matrix inequality (LMI):

$$\sum_{m=1}^M w_m \mathbf{F}_m(\boldsymbol{\theta}) - \lambda_{\text{eig}} \mathbf{I}_N \succeq \mathbf{0}_N. \quad (12)$$

In other words, we put a lower bound on each eigenvalue of the matrix \mathbf{F} . The solution set of \mathbf{w} satisfying this LMI is convex as $\mathbf{F}_m(\boldsymbol{\theta}) \in \mathbb{S}^N$, $m = 1, 2, \dots, M$, and $\lambda_{\text{eig}} \mathbf{I}_N \in \mathbb{S}^N$ [11, p. 38].

The trace constraint has a larger feasible set as compared to the minimum eigenvalue constraint. However, although the trace constraint is a sufficient condition, the resulting sensor selection problem is computationally less attractive compared to the minimum eigenvalue constraint (as we show later on in Section V-A). Moreover, LMIs can be used to also represent the trace constraint. For these reasons, we focus on the minimum eigenvalue (LMI) constraint from now on. However, either one of the two performance constraints can be used.

The above performance measures depend on the unknown parameter $\boldsymbol{\theta}$. In practice, the unknown parameter $\boldsymbol{\theta}$ has a physical meaning and takes values within a certain domain denoted by \mathcal{U} . For example, in the case of direction-of-arrival estimation, \mathcal{U} is the sector where the source is expected or for target localization it is the surveillance area where the target resides. Since the FIM for non-linear models depends on the unknown $\boldsymbol{\theta}$, we propose to constrain every point within the domain \mathcal{U} .

C. Problem Statement

Having introduced the selection vector as well as the performance measure we can now formally state the problem.

Problem (Sensor Selection): Given the likelihoods of the measurements, $p(y_m; \boldsymbol{\theta})$, $m = 1, 2, \dots, M$, and a desired inference performance λ_{eig} , find a vector $\mathbf{w} \in \{0, 1\}^M$ that selects the minimum number of most informative sensors satisfying the performance measure $\sum_{m=1}^M w_m \mathbf{F}_m(\boldsymbol{\theta}) - \lambda_{\text{eig}} \mathbf{I}_N \succeq \mathbf{0}_N$, $\forall \boldsymbol{\theta} \in \mathcal{U}$.

In order to reduce the hardware costs, storage, processing, and communications overhead, we minimize the number of selected sensors. This can be achieved by minimizing the number of non-zero entries of the selection vector \mathbf{w} . Mathematically, the sensor selection problem is formulated as the following optimization problem

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\mathbf{w}\|_0 \quad (13a)$$

$$\text{s.t.} \quad \sum_{m=1}^M w_m \mathbf{F}_m(\boldsymbol{\theta}) - \lambda_{\text{eig}} \mathbf{I}_N \succeq \mathbf{0}_N, \quad \forall \boldsymbol{\theta} \in \mathcal{U}, \quad (13b)$$

$$\mathbf{w} \in \{0, 1\}^M, \quad (13c)$$

where the ℓ_0 -(quasi) norm refers to the number of non-zero entries in \mathbf{w} , i.e., $\|\mathbf{w}\|_0 := |\{m : w_m \neq 0\}|$. The threshold λ_{eig} imposes the accuracy requirement. The threshold λ_{eig} is also the sparsity-inducing parameter, where $\lambda_{\text{eig}} \rightarrow 0$ implies a sparser solution. Alternatively, the sensor selection problem can also be expressed as in (1) when K is known.

Suppose the domain \mathcal{U} consists of D points, obtained by gridding the entire domain at a certain resolution. The resulting multiple LMI constraints can be stacked together as a single LMI constraint. Let us consider the domain $\mathcal{U} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_D\}$ with $|\mathcal{U}| = D$. The constraints in (13b) can then be equivalently expressed as a single LMI constraint written as $\sum_{m=1}^M w_m \mathbf{F}_m - \lambda_{\text{eig}} \mathbf{I}_{DN} \succeq \mathbf{0}_{DN}$, where

$$\mathbf{F}_m = \text{diag}(\mathbf{F}_m(\boldsymbol{\theta}_1), \mathbf{F}_m(\boldsymbol{\theta}_2), \dots, \mathbf{F}_m(\boldsymbol{\theta}_D)) \in \mathbb{S}^{DN}$$

for $m = 1, 2, \dots, M$. Note that the FIM after gridding is independent of $\boldsymbol{\theta}$. Henceforth, we denote this simply by \mathbf{F}_m (not explicitly as a function of $\boldsymbol{\theta}$). The computational complexity of the resulting solvers depends on the number of grid points, which is due to the fact that we do not exactly know where the true parameter is located and because we are dealing with a non-linear system model. We make the following remarks to indicate some scenarios where the computational burden due to gridding can be reduced.

Remark 2 (Worst-Case Constraints): If for every $\boldsymbol{\theta} \in \mathcal{U}$, there exists some $\tilde{\boldsymbol{\theta}} \in \mathcal{U}_{\text{worst}} \subset \mathcal{U}$ such that

$$\lambda_{\min}\{\mathbf{F}(\mathbf{w}, \tilde{\boldsymbol{\theta}})\} \leq \lambda_{\min}\{\mathbf{F}(\mathbf{w}, \boldsymbol{\theta})\}, \forall \mathbf{w} \in \{0, 1\}^M,$$

then it is sufficient to constrain the performance for only the worst-case set $\mathcal{U}_{\text{worst}}$ instead of \mathcal{U} . This property can be used as a guideline for gridding.

Remark 3 (Bayesian CRB Constraint): In a Bayesian setting, when *prior* information of the unknown parameter $\boldsymbol{\theta}$ is available, this additional knowledge can be incorporated in the CRB. The related information matrix is often called the Bayesian information matrix (BIM) [30], and it is independent of the unknown parameter (hence, gridding is not needed). The BIM is given by $\mathbf{F}_B(\mathbf{w}) = \mathbf{J}_p + \mathbb{E}_{\boldsymbol{\theta}}\{\mathbf{F}(\mathbf{w}, \boldsymbol{\theta})\}$, where \mathbf{J}_p is the prior information matrix $\mathbf{J}_p = -\mathbb{E}_{\boldsymbol{\theta}}\left\{\frac{\partial}{\partial \boldsymbol{\theta}}\left(\frac{\ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right\}$ with the (log) prior $\ln p(\boldsymbol{\theta})$, and the expectation $\mathbb{E}_{\boldsymbol{\theta}}\{\cdot\}$ is under the pdf $p(\boldsymbol{\theta})$. The LMI constraint in (12) for the Bayesian setting will then be

$$\mathbf{J}_p + \sum_{m=1}^M w_m \mathbb{E}_{\boldsymbol{\theta}}\{\mathbf{F}_m(\boldsymbol{\theta})\} \succeq \lambda_{\text{eig}} \mathbf{I}_N. \quad (14)$$

The prior information typically comes from the dynamics, previous measurements, or combining other available measurements.

Note that in order to optimize the Bayesian CRB, we need to know the distribution of the unknown parameter.

IV. SENSOR SELECTION SOLVERS

It is well known that the ℓ_0 -(quasi) norm optimization is NP-hard and nonconvex. More specifically, the original sensor selection problem in (13) is NP-hard. The Boolean constraint in (13c) is non-convex and incurs a combinatorial complexity.

It requires a brute-force evaluation over all the 2^M choices. For example, with $M = 100$ available potential sensors, there are in the order of 10^{30} possible choices whose direct enumeration is clearly impossible. We next present a number of solvers for which the relaxed convex problem can be solved efficiently in polynomial time.

A. Convex Approximation Based on ℓ_1 -Norm

A computationally tractable (suboptimal) solution is to use the traditional best convex surrogate for the ℓ_0 -(quasi) norm, namely the ℓ_1 -norm heuristic. The ℓ_1 -norm is known to represent an efficient heuristic for the ℓ_0 -(quasi) norm optimization with convex constraints especially when the solution is sparse [31]. Such relaxations are well-studied for problems with linear constraints in the context of compressed sensing (CS) and sparse signal recovery [32]. The non-convex Boolean constraint in (13c) is further relaxed to the convex box constraint $[0, 1]^M$.

The relaxed optimization problem is given as the following SDP problem

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{R}^M} \|\mathbf{w}\|_1 \quad (15a)$$

$$\text{s.t.} \quad \sum_{m=1}^M w_m \mathbf{F}_m - \lambda_{\text{eig}} \mathbf{I}_{DN} \succeq \mathbf{0}_{DN}, \quad (15b)$$

$$0 \leq w_m \leq 1, \quad m = 1, 2, \dots, M, \quad (15c)$$

where $\|\mathbf{w}\|_1 = \sum_{m=1}^M |w_m|$ denotes the ℓ_1 -norm. Due to the positivity constraint, the objective function $\|\mathbf{w}\|_1$ will simply be an affine function $\mathbf{1}_M^T \mathbf{w}$. The optimization problem in (15) is a standard SDP problem in the inequality form, which can be efficiently solved in polynomial time using interior-point methods [11]. An implementation of the interior-point method for solving SDP problems in the inequality form is typically based on Newton's method using an approximating barrier function. A brief description of the *projected Newton's method* is provided in Appendix C, which is used to analyze the computational complexity of the relaxed sensor selection problem.

Remark 4 (Complexity per Iteration): The computational cost involved during each iteration is as follows [11, p. 619]. The matrices \mathbf{F}_m , $m = 1, 2, \dots, M$, have a block-diagonal structure with D blocks. Forming the matrix $\mathbf{S} = \sum_{m=1}^M w_m \mathbf{F}_m - \lambda_{\text{eig}} \mathbf{I}_{DN}$ costs $O(DMN^2)$ flops; computing $\mathbf{S}^{-1} \mathbf{F}_i$ via Cholesky factorization costs $O(MDN^3)$ flops; the Hessian matrix is computed via the inner product of the matrices $\mathbf{S}^{-1} \mathbf{F}_i$ and $\mathbf{S}^{-1} \mathbf{F}_j$, which costs $O(DM^2N^2)$ $\forall i, j$. Finally, the Newton step is computed via Cholesky factorization costing $O(DM^3)$ flops, and the projection costs $O(M)$ flops. Assuming that $M \gg N$, the overall computational complexity per iteration of the projected Newton's algorithm is then $O(DM^3)$.

Implementations of the interior-point methods are easily available in the form of well-known toolboxes like Yalmip [33], SeDuMi [34], and CVX [35].

B. Projected Subgradient Algorithm

The second-order Newton's method (cf. Appendix C) is typically intractable when the number of candidate sensors is very

large ($M \gg 1000$, for example). To circumvent this problem, we propose a subgradient algorithm. The projected subgradient algorithm is a first-order method which is attractive for large-scale problems as each iteration is much cheaper to process.

The subgradient method is typically used for optimizations involving non-differentiable functions [36], [37]. The subgradient method is a generalization of the gradient method for non-smooth and non-differentiable functions, such as, the ℓ_1 -norm and the minimum eigenvalue constraint functions. We next derive the projected subgradient algorithm.

The relaxed sensor selection problem in (15) can be equivalently expressed as

$$\arg \min_{\mathbf{w}} \quad \mathbf{1}_M^T \mathbf{w} \quad (16a)$$

$$\text{s.t. } f_{\text{eig}}(\mathbf{w}) \geq \lambda_{\text{eig}}, \quad (16b)$$

$$\mathbf{w} \in \mathcal{W}, \quad (16c)$$

where $f_{\text{eig}}(\mathbf{w}) := \lambda_{\min}\{\sum_{m=1}^M w_m \mathbf{F}_m\}$ is the constraint function in (15b), and the set $\mathcal{W} = \{\mathbf{w} \in \mathbb{R}^M | 0 \leq w_m \leq 1, m = 1, 2, \dots, M\}$ denotes the box constraints in (15c).

The objective $\mathbf{1}_M^T \mathbf{w}$ is affine, so a subgradient of the objective is the all-one vector $\mathbf{1}_M$. Let $\mathbf{g}^k \in \partial f_{\text{eig}}(\mathbf{w}^k)$ denote a subgradient of the constraint function $f_{\text{eig}}(\mathbf{w})$ evaluated at $\mathbf{w} = \mathbf{w}^k$. Here, the set $\partial f_{\text{eig}}(\mathbf{w}^k)$ denotes the subdifferential of $f_{\text{eig}}(\mathbf{w})$ towards \mathbf{w} evaluated at $\mathbf{w} = \mathbf{w}^k$. To compute \mathbf{g}^k , we express the constraint function $f_{\text{eig}}(\mathbf{w}^k)$ as

$$f_{\text{eig}}(\mathbf{w}^k) = \inf_{\|\mathbf{v}\| \leq 1} \mathbf{v}^T \left(\sum_{m=1}^M w_m^k \mathbf{F}_m \right) \mathbf{v}.$$

The computation of a subgradient is straightforward, and is given by

$$\mathbf{g}^k = \left[(\mathbf{v}_{\min}^k)^T \mathbf{F}_1 \mathbf{v}_{\min}^k, \dots, (\mathbf{v}_{\min}^k)^T \mathbf{F}_M \mathbf{v}_{\min}^k \right]^T \in \partial f_{\text{eig}}(\mathbf{w}^k),$$

where \mathbf{v}_{\min}^k is the eigenvector corresponding to the minimum eigenvalue $\lambda_{\min}\{\sum_{m=1}^M w_m^k \mathbf{F}_m\}$. The minimum eigenvalue and the corresponding eigenvector can be computed using a low-complexity iterative algorithm, for example, the power method (see Appendix D) or using the standard eigenvalue decomposition [38]. Let the projection of a point onto the set \mathcal{W} be denoted by $\mathcal{P}_{\mathcal{W}}(\cdot)$, which can be expressed elementwise as

$$[\mathcal{P}_{\mathcal{W}}(\mathbf{w})]_m = \begin{cases} 0 & \text{if } w_m \leq 0, \\ w_m & \text{if } 0 < w_m < 1, \\ 1 & \text{if } w_m \geq 1. \end{cases} \quad (17)$$

The projected subgradient algorithm then proceeds as follows:

$$\mathbf{w}^{k+1} = \begin{cases} \mathcal{P}_{\mathcal{W}}(\mathbf{w}^k - \alpha^k \mathbf{1}_M) & \text{if } f_{\text{eig}}(\mathbf{w}^k) \geq \lambda_{\text{eig}}, \\ \mathcal{P}_{\mathcal{W}}(\mathbf{w}^k + \alpha^k \mathbf{g}^k) & \text{if } f_{\text{eig}}(\mathbf{w}^k) < \lambda_{\text{eig}}. \end{cases} \quad (18)$$

In other words, if the current iterate \mathbf{w}^k is feasible (i.e., $f_{\text{eig}}(\mathbf{w}^k) \geq \lambda_{\text{eig}}$), we update \mathbf{w} in the direction of a negative objective subgradient, as if the LMI constraints were absent; If the current iterate \mathbf{w}^k is infeasible (i.e., $f_{\text{eig}}(\mathbf{w}^k) < \lambda_{\text{eig}}$), we update \mathbf{w} in the direction of a subgradient \mathbf{g}^k associated with the LMI constraints. After the update is computed, the iterate is projected onto the constraint set \mathcal{W} using $\mathcal{P}_{\mathcal{W}}(\cdot)$. When the k th

Algorithm 1: Projected subgradient algorithm

1. **Initialize** iteration counter $k = 0$, $\mathbf{w}^k = \mathbf{1}_M$, $\mathbf{g}^k = \mathbf{0}$, k_{\max} , ϵ , and λ_{eig} .
 2. **for** $k = 0$ to k_{\max}
 3. **compute** $f_{\text{eig}}(\mathbf{w}^k) = \lambda_{\min}\{\sum_{m=1}^M w_m^k \mathbf{F}_m\}$
 4. **update**
 5. **if** $f_{\text{eig}}(\mathbf{w}^k) \geq \lambda_{\text{eig}}$
 6. $\mathbf{w}^{k+1} = \mathcal{P}_{\mathcal{W}}(\mathbf{w}^k - (1/\sqrt{k})\mathbf{1}_M)$
 7. **elseif** $f_{\text{eig}}(\mathbf{w}^k) < \lambda_{\text{eig}}$
 8. $\mathbf{w}^{k+1} = \mathcal{P}_{\mathcal{W}}\left(\mathbf{w}^k + \frac{f_{\text{eig}}(\mathbf{w}^k) + \epsilon}{\|\mathbf{g}^k\|_2^2} \mathbf{g}^k\right)$
 9. **end**
 10. **end**
 11. $\hat{\mathbf{w}} = \mathbf{w}^{k_{\max}}$
-

iterate is feasible, a diminishing non-summable step size $\alpha^k = 1/\sqrt{k}$ is used. When the iterate is not feasible Polyak's step size $\alpha^k = \frac{f_{\text{eig}}(\mathbf{w}^k) + \epsilon}{\|\mathbf{g}^k\|_2^2}$ is used, where we adopt the optimal value for $\epsilon := \mathbf{1}_M^T \mathbf{w}^*$ when $\|\mathbf{w}\|_0$ is known (i.e., the number of sensors to be selected is known). If this is not known, then we approximate it with $\epsilon := f_{\text{best}}^k + \gamma$, where $\gamma = 10/(10 + k)$, and $f_{\text{best}}^k = \min\{f_{\text{best}}^{k-1}, \mathbf{1}_M^T \mathbf{w}^k\}$ [36]. The algorithm is terminated after a specified maximum number of iterations k_{\max} . Finally, the estimate is denoted by $\hat{\mathbf{w}} = \mathbf{w}^{k_{\max}}$.

The convergence results of the subgradient method for the constrained optimization (i.e., without the projection step) are derived in [36]. Since the projection onto a convex set is non-expansive [37], it does not affect the convergence. The projected subgradient algorithm is summarized as Algorithm 1.

Remark 5 (Complexity per Iteration): We first form the matrix $\sum_{m=1}^M w_m \mathbf{F}_m$, which costs $O(DMN^2)$ flops. The minimum eigenvalue and the corresponding eigenvector can be computed using the power method at a cost of $O(DN^2)$ flops [38]. Forming the vector \mathbf{g} costs $O(DMN^2)$ flops, computing its norm costs $O(M)$ flops, and the update and projection together cost $O(M)$ flops. Assuming that $M \gg N$ as earlier, the computational cost of the projected subgradient algorithm is $O(DMN^2)$, which is much lower than the complexity of the projected Newton's method.

A distributed implementation of the projected subgradient algorithm is very easy. A simple distributed averaging algorithm (e.g., [39]) can be used to compute the sum of matrices $\sum_{m=1}^M w_m \mathbf{F}_m$. The minimum eigenvalue and the corresponding eigenvector can then be computed using power iterations at each node independently. The update step (18), the subgradient vector \mathbf{g} , and the projection are computed coordinatewise and are already distributed.

Subgradient methods are typically very slow compared to the interior-point method involving Newton iterations, and subgradient methods typically require a few hundred iterations. Newton's method typically requires in the order of ten steps. On the other hand, unlike the projected subgradient method, Newton's method cannot be easily distributed, and incurs a

Algorithm 2: Sparsity-enhancing iterative algorithm

1. **Initialize** the iteration counter $i = 0$, the weight vector $\mathbf{u}[0] = [u_1[0], u_2[0], \dots, u_M[0]]^T = \mathbf{1}_M$, δ , and i_{\max} .

2. **for** $i = 0$ to i_{\max}

3. **solve** the weighted ℓ_1 -norm minimization problem

$$\hat{\mathbf{w}}[i] = \arg \min_{\mathbf{w} \in \mathbb{R}^M} \mathbf{u}[i]^T \mathbf{w} \quad (19a)$$

$$\text{s.t.} \quad \sum_{m=1}^M w_m \mathbf{F}_m - \lambda_{\text{eig}} \mathbf{I}_{DN} \succeq \mathbf{0}_{DN}, \quad (19b)$$

$$0 \leq w_m \leq 1, \quad m = 1, 2, \dots, M. \quad (19c)$$

4. **update** the weight vector $u_m[i+1] = \frac{1}{\delta + \hat{w}_m[i]}$, for each $m = 1, 2, \dots, M$.

5. **end**

6. $\hat{\mathbf{w}} = \hat{\mathbf{w}}[i_{\max}]$.

relatively high complexity per iteration due to the computation and storage of up to second-order derivatives. Depending on the scale of the problem and the resources available for processing one could choose between the subgradient or Newton's algorithm.

C. Concave Surrogate Based on Sum-of-Logarithms

The ℓ_1 -norm is customarily used as the best convex relaxation for the ℓ_0 -norm. However, the intersection of the ℓ_1 -norm ball (or an affine subspace) with the positive semi-definite cone (i.e., the LMI constraint) is not always a unique point as shown in the following Theorem.

Theorem 1 (Uniqueness): The projection of a point $\mathbf{w} \in [0, 1]^M$ onto a convex LMI constraint set $\sum_{m=1}^M w_m \mathbf{F}_m - \lambda_{\text{eig}} \mathbf{I}_{DN} \succeq \mathbf{0}_{DN}$ under the ℓ_1 -norm is not always unique.

Proof: The proof follows from the fact that the ℓ_1 -norm is not strictly convex, and from the linearity of the constraint set. Let us consider an example with $M = 2$ (w.l.o.g.), and $\mathbf{F}_1 = \mathbf{F}_2 \succeq \lambda_{\text{eig}} \mathbf{I}_{DN}$. In other words, the observations are identical. In this case, the extreme points of the ℓ_1 -norm ball, i.e., $\hat{\mathbf{w}}_1 = (1, 0)$ and $\hat{\mathbf{w}}_2 = (0, 1)$ are two example solutions. Moreover, since the solution set of a convex minimization problem is convex, $\tau \hat{\mathbf{w}}_1 + (1 - \tau) \hat{\mathbf{w}}_2$ is also a solution for any $0 < \tau < 1$, which gives an infinite number of solutions to the relaxed optimization problem (15). For such cases, the ℓ_1 -norm relaxation will typically not result in a sparse solution. ■

To improve upon the ℓ_1 -norm solution due to its non-uniqueness following from Theorem 1, we propose an alternative relaxation for the original sensor selection problem which also results in fewer selected sensors. Instead of relaxing the ℓ_0 -(quasi) norm with the ℓ_1 -norm, using a nonconvex surrogate function can yield a better approximation. It is motivated in [40] that the logarithm of the geometric mean of its elements can be used as an alternative surrogate function for linear inverse problems in

CS. Adapting this to our sensor selection problem, we arrive at the optimization problem

$$\arg \min_{\mathbf{w} \in \mathbb{R}^M} \sum_{m=1}^M \ln(w_m + \delta) \quad (20a)$$

$$\text{s.t.} \quad \sum_{m=1}^M w_m \mathbf{F}_m - \lambda_{\text{eig}} \mathbf{I}_{DN} \succeq \mathbf{0}_{DN}, \quad (20b)$$

$$0 \leq w_m \leq 1, \quad m = 1, 2, \dots, M, \quad (20c)$$

where $\delta > 0$ is a small constant that prevents the cost from tending to $-\infty$. The cost (20a) is concave, but since it is smooth w.r.t. \mathbf{w} , iterative linearization can be performed to obtain a local minimum [40]. The first-order approximation of $\ln(w_m + \delta)$ around $w_m[i-1] + \delta$ results in

$$\ln(w_m + \delta) \leq \ln(w_m[i-1] + \delta) + \frac{w_m - w_m[i-1]}{w_m[i-1] + \delta}.$$

Here, i denotes the iteration index. Instead of minimizing the original cost, the majorizing cost (second term on the right-hand side of the above inequality) can be optimized to attain a local minimum. More specifically, the optimization problem (20) can be iteratively driven to a local minimum using the iterations

$$\hat{\mathbf{w}}[i] = \arg \min_{\mathbf{w} \in \mathbb{R}^M} \sum_{m=1}^M \frac{w_m}{\hat{w}_m[i-1] + \delta} \quad (21a)$$

$$\text{s.t.} \quad \sum_{m=1}^M w_m \mathbf{F}_m - \lambda_{\text{eig}} \mathbf{I}_{DN} \succeq \mathbf{0}_{DN}, \quad (21b)$$

$$0 \leq w_m \leq 1, \quad m = 1, 2, \dots, M. \quad (21c)$$

The iterative algorithm is summarized as Algorithm 2. Each iteration in (21) solves a weighted ℓ_1 -norm optimization problem. The weight updates force the small entries of the vector $\hat{\mathbf{w}}[i]$ to zero and avoid inappropriate suppression of larger entries. The parameter δ provides stability, and guarantees that the zero-valued entries of $\hat{\mathbf{w}}[i]$ do not strictly prohibit a nonzero estimate at the next step. Finally, the estimate is given by $\hat{\mathbf{w}} = \hat{\mathbf{w}}[i_{\max}]$, where i_{\max} is the specified maximum number of iterations.

Remark 6 (Sparsity-Enhancing Projected Subgradient Algorithm): The projected subgradient algorithm can be adapted to fit into the sparsity-enhancing iterative algorithm as well. The optimization problem (19) is then replaced with the following update equations:

$$\mathbf{w}^{k+1}[i] = \begin{cases} \mathcal{P}_{\mathcal{W}}(\mathbf{w}^k[i] - \alpha^k \mathbf{u}[i]) & \text{if } f_{\text{eig}}(\mathbf{w}^k[i]) \geq \lambda_{\text{eig}}, \\ \mathcal{P}_{\mathcal{W}}(\mathbf{w}^k[i] + \alpha^k \mathbf{g}^k[i]) & \text{if } f_{\text{eig}}(\mathbf{w}^k[i]) < \lambda_{\text{eig}}, \end{cases}$$

where we solve a number of iterations (inner loop) of the projected subgradient algorithm within the i th iteration (outer loop) of Algorithm 2. Here, the k th iterate of the inner loop in the i th outer loop is denoted as $(\cdot)^k[i]$.

From the solution of the relaxed optimization problem, the approximate Boolean solution to $\mathbf{w} \in \{0, 1\}^M$ can be obtained using randomization techniques, as described next.

Algorithm 3: Randomized rounding algorithm

1. Generate L candidate estimates of the form
 $w_{m,l} = 1$ ($l = 1, 2, \dots, L$) with a probability \hat{w}_m (or
 $w_{m,l} = 0$ with a probability $1 - \hat{w}_m$) for $m = 1, 2, \dots, M$.
2. Define $\mathbf{w}_l = [w_{1,l}, \dots, w_{M,l}]^T$ and the index set of the
candidate estimates satisfying the constraints as

$$\Omega \triangleq \{l \mid \lambda_{\min}\{\mathbf{F}(\mathbf{w}_l, \boldsymbol{\theta})\} \geq \lambda_{\text{eig}}, \forall \boldsymbol{\theta} \in \mathcal{U}, l = 1, 2, \dots, L\}.$$

3. If the set Ω is empty, go back to step 1.
4. The suboptimal Boolean estimate is the solution to the
optimization problem

$$\hat{\mathbf{w}}_{\text{bp}} = \underset{l \in \Omega}{\operatorname{argmin}} \|\mathbf{w}_l\|_1.$$

D. Randomized Rounding

The solution of the relaxed optimization problem is used to compute the suboptimal Boolean solution for the selection problem. A straightforward technique that is often used is the simple rounding technique, in which the Boolean estimate is given by $\text{round}(\hat{w}_m)$, $m = 1, 2, \dots, M$, where we define $\hat{\mathbf{w}} \triangleq [\hat{w}_1, \hat{w}_2, \dots, \hat{w}_M]^T$, and the $\text{round}(\cdot)$ operator rounds its arguments towards the nearest integer. However, there is no guarantee that the Boolean estimates obtained from the rounding technique always satisfy the performance constraints. Hence, we propose a randomized rounding technique, where the suboptimal Boolean estimates are computed based on random experiments guided by the solution from the SDP problem (15) or the iterative version in (21). The randomized rounding technique is summarized as Algorithm 3.

V. EXTENSIONS**A. Scalar Constraints**

In this section, we will discuss the relaxed sensor selection problem based on the optimization criteria related to A-optimality and D-optimality.

1) *Trace Constraint:* The relaxed sensor selection problem with the scalar trace constraint is given as follows

$$\begin{aligned} & \underset{\mathbf{w} \in \mathbb{R}^M}{\operatorname{argmin}} \quad \|\mathbf{w}\|_1 \\ & \text{s.t.} \quad \operatorname{tr} \left\{ \left(\sum_{m=1}^M w_m \mathbf{F}_m(\boldsymbol{\theta}) \right)^{-1} \right\} \leq \lambda_{\text{tr}}, \quad \forall \boldsymbol{\theta} \in \mathcal{U}, \\ & \quad 0 \leq w_m \leq 1, \quad m = 1, 2, \dots, M. \end{aligned} \quad (22)$$

The trace constraint in (22) is convex in \mathbf{w} ; this is easier to verify when the above trace constraint is expressed as an LMI [11, p. 387]. Let $\mathbf{x} = [x_1, x_2, \dots, x_N]^T \in \mathbb{R}^N$ denote the auxiliary

variable and \mathbf{e}_n denote the n th unit vector in \mathbb{R}^N . The optimization problem (22) is a convex problem, and can be cast as an SDP:

$$\begin{aligned} & \underset{\mathbf{w} \in \mathbb{R}^M, \mathbf{x} \in \mathbb{R}^N}{\operatorname{argmin}} \quad \|\mathbf{w}\|_1 \\ & \text{s.t.} \quad \begin{bmatrix} \sum_{m=1}^M w_m \mathbf{F}_m(\boldsymbol{\theta}) & \mathbf{e}_n \\ \mathbf{e}_n^T & x_n \end{bmatrix} \succeq \mathbf{0}_{N+1}, \quad n = 1, 2, \dots, N, \\ & \quad \quad \quad \forall \boldsymbol{\theta} \in \mathcal{U}, \\ & \quad \mathbf{1}_N^T \mathbf{x} \leq \lambda_{\text{tr}}, \\ & \quad x_n \geq 0, \quad n = 1, 2, \dots, N, \\ & \quad 0 \leq w_m \leq 1, \quad m = 1, 2, \dots, M. \end{aligned} \quad (23)$$

In addition to the box constraints, the optimization problem (23) has N LMI constraints (of size $N+1$) for every point in \mathcal{U} and $N+1$ inequality constraints, while (15) has only one LMI constraint (of size N) for every point in \mathcal{U} . Hence, solving (23) is computationally more intense than solving (15).

2) *Determinant Constraint:* Another popular scalar performance measure that quantifies the estimation accuracy is the determinant (product of eigenvalues) constraint. This measure is related to the D-optimality. The relaxed sensor selection problem with the determinant constraint is given as follows

$$\begin{aligned} & \underset{\mathbf{w} \in \mathbb{R}^M}{\operatorname{argmin}} \quad \|\mathbf{w}\|_1 \\ & \text{s.t.} \quad \ln \det \left\{ \sum_{m=1}^M w_m \mathbf{F}_m(\boldsymbol{\theta}) \right\} \geq \lambda_{\text{det}}, \quad \forall \boldsymbol{\theta} \in \mathcal{U}, \\ & \quad 0 \leq w_m \leq 1, \quad m = 1, 2, \dots, M, \end{aligned} \quad (24)$$

where the threshold λ_{det} specifies the mean radius of a confidence ellipsoid (see Appendix A). The log-determinant constraint is a concave function of \mathbf{w} for $w_m \geq 0$, for $m = 1, 2, \dots, M$. Note that although the constraint $\ln \det \left\{ \sum_{m=1}^M w_m \mathbf{F}_m(\boldsymbol{\theta}) \right\} \geq \lambda_{\text{det}}$ is an indication of the performance of the estimator, it is not a sufficient condition for (11).

The relaxed sensor selection problem with the scalar (trace or determinant) constraints can be solved with either one of the two proposed approximations of the cardinality cost, i.e., the ℓ_1 -norm or log-based concave surrogate.

B. The Lagrangian Dual Problem

The Lagrangian dual of the relaxed primal optimization problem (15) has an interesting relation to the diameter of the confidence ellipsoid, and is closely related to the dual of the *E-optimal design* [11, p. 388]. The Lagrangian dual problem of (15) is given as follows

$$\begin{aligned} & (\hat{\mathbf{Z}}, \hat{\boldsymbol{\mu}}) = \underset{\mathbf{Z}, \boldsymbol{\mu}}{\operatorname{argmax}} \quad \lambda_{\text{eig}} \operatorname{tr}\{\mathbf{Z}\} - \mathbf{1}_M^T \boldsymbol{\mu} \\ & \text{s.t.} \quad \mathbb{E}\{\mathbf{s}_m^T \mathbf{Z} \mathbf{s}_m\} \leq 1 + \mu_m, \quad m = 1, 2, \dots, M, \\ & \quad \mathbf{Z} \succeq \mathbf{0}, \quad \mu_m \geq 0, \quad m = 1, 2, \dots, M, \end{aligned} \quad (25)$$

where $\mathbf{Z} \in \mathbb{S}_+^{DN}$ and $\boldsymbol{\mu} \in \mathbb{R}^N$ are the dual variables, and we use $\mathbb{E}\{\mathbf{s}_m^T \mathbf{Z} \mathbf{s}_m\} = \operatorname{tr}\{\mathbf{F}_m \mathbf{Z}\}$ with

$$\mathbf{s}_m = \left[\frac{\partial h_m(\tilde{\boldsymbol{\theta}}_1, n_m)}{\partial \tilde{\boldsymbol{\theta}}_1^T}, \dots, \frac{\partial h_m(\tilde{\boldsymbol{\theta}}_D, n_m)}{\partial \tilde{\boldsymbol{\theta}}_D^T} \right]^T \in \mathbb{R}^{DN}.$$

For a detailed derivation of the Lagrangian dual problem, see Appendix B. The dual problem can be interpreted as the problem of maximizing the (average) diameter of the confidence ellipsoid. If we set $\mu_m = 0$, $m = 1, 2, \dots, M$, the optimal solution $\hat{\mathbf{Z}}$ to the problem (25) is also the solution to the dual of the *E-optimal* design problem [11, p. 388], which maximizes the diameter of the confidence ellipsoid centered around the origin.

The dual formulation is often easier to solve and can be solved using Yalmip, SeDuMi, or CVX as earlier.

VI. EXAMPLES: SENSOR PLACEMENT FOR LOCALIZATION

Localization is an important and extensively studied topic in wireless sensor networks (WSNs). Target localization can be performed using a plethora of algorithms [4]–[6] (and references therein), which exploit inter-sensor measurements like time-of-arrival (TOA), time-difference-of-arrival (TDOA), angle-of-arrival (AOA), or received signal strength (RSS). The performance of any location estimator depends not only on the algorithm but also on the placement of the anchors (sensors with known locations). Sensor placement is a key challenge in localization system design, as certain sensor constellations not only deteriorate the performance, but also result in ambiguity or identifiability issues [1].

The sensor placement problem can be interpreted as the problem where we divide a specific sensor area \mathcal{S} into M grid points and select the best subset of locations from these grid points. Here, the selected sensors are deemed the best, if they guarantee a certain prescribed accuracy on the location estimates for a target within a specific target area \mathcal{U} . We consider a two-dimensional network with one target located in the target area \mathcal{U} and M possible sensors located at the M grid points.

The absolute positions of the sensor grid points are known, hence, the considered sensors are commonly referred to as anchor nodes. Let the coordinates of the target and the m th anchor be denoted by the 2×1 vectors $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$ and $\mathbf{a}_m = [a_{m,1}, a_{m,2}]^T$, respectively. Here, $\boldsymbol{\theta}$ is assumed to be unknown, but known to be within \mathcal{U} . We next illustrate the developed theory with a number of examples based on different measurement models.

A. Distance Measurements

Let the pairwise distance between the target and the m th anchor be denoted by $d_m = \|\boldsymbol{\theta} - \mathbf{a}_m\|_2$. In practice, the pairwise distances are obtained by ranging and they are noisy. The range measurements typically follow an additive Gaussian non-linear model, as given by

$$y_m = d_m + n_m, \quad m = 1, 2, \dots, M, \quad (26)$$

where $n_m \sim \mathcal{N}(0, \sigma_m^2)$ is the noise with $\sigma_m^2 = \sigma^2/d_m^{-\eta}$. Here, σ^2 is the nominal noise variance, and η is the path-loss exponent. Using (10), we can now write the FIM for the localization problem as $\mathbf{C}^{-1} = \mathbf{F}(\mathbf{w}, \boldsymbol{\theta}) = \sum_{m=1}^M w_m \mathbf{F}_m(\boldsymbol{\theta})$, where using (6) we can compute

$$\mathbf{F}_m(\boldsymbol{\theta}) := \frac{(\boldsymbol{\theta} - \mathbf{a}_m)(\boldsymbol{\theta} - \mathbf{a}_m)^T}{\sigma_m^2 \|\boldsymbol{\theta} - \mathbf{a}_m\|_2^2}.$$

B. Received Signal Strength

Received signal strength is the voltage measured by a sensor's received signal strength indicator circuit. Received signal strength is often reported as the measured power. The ensemble mean received power at the m th sensor can be expressed as

$$\bar{y}_m = y_0 - 10\eta \ln \frac{d_m}{d_0}, \quad m = 1, 2, \dots, M,$$

where y_0 is the received power (dBm) at a reference distance d_0 . However, due to shadowing, the difference between a measured received power and its ensemble average is random. The randomness due to shadowing is typically modeled as a log-normal process, which is Gaussian if expressed in decibels [6]. More specifically, the received power (dBm) y_m at the m th sensor follows a Gaussian distribution, i.e., $p(y_m; \boldsymbol{\theta}) \sim \mathcal{N}(\bar{y}_m, \sigma_{r,\text{dB}}^2)$. The FIM related to the m th measurement is then given by

$$\mathbf{F}_m(\boldsymbol{\theta}) := \frac{50\eta^2}{\sigma_{r,\text{dB}}^2 d_m^4 \ln 10} (\boldsymbol{\theta} - \mathbf{a}_m)(\boldsymbol{\theta} - \mathbf{a}_m)^T.$$

C. Bearing Measurements

Another popular target localization technique is based on bearing measurements from a set of direction finding (DF) sensors [41]. The bearing measurement of the m th DF sensor is given by

$$y_m = \arctan \left(\frac{\Delta a_{m,2}}{\Delta a_{m,1}} \right) + n_m, \quad m = 1, 2, \dots, M,$$

where $\Delta a_{m,2} = \theta_2 - a_{2,m}$, $\Delta a_{m,1} = \theta_1 - a_{1,m}$, and $n_m \sim \mathcal{N}(0, \sigma_b^2)$ is the noise. Defining a 2×2 permutation matrix

$$\mathbf{P} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

we can then compute the FIM contribution from the m th DF sensor as

$$\mathbf{F}_m = \frac{1}{\sigma_b^2 d_m^4} \mathbf{P}(\boldsymbol{\theta} - \mathbf{a}_m)(\boldsymbol{\theta} - \mathbf{a}_m)^T \mathbf{P}^T.$$

D. Field Estimation

In applications related to field estimation, (active/passive) radar, and sonar, it is important to estimate the location of a point source that emits or reflects energy. Suppose the sensors measure the energy generated by a point source. The measurements are given as

$$y_m = \sqrt{e} h_m(\boldsymbol{\theta}) + n_m, \quad m = 1, 2, \dots, M, \quad (27)$$

where e is the known field intensity emitted or reflected by the point source, the known propagation function for some gain $\beta \geq 0$ is modeled as an isotropic exponential attenuation $h_m(\boldsymbol{\theta}) = \beta/(\beta + d_m^2)$, and $n_m \sim \mathcal{N}(0, \sigma_e^2)$ is the noise. The FIM related to the m th measurement is then given by

$$\mathbf{F}_m(\boldsymbol{\theta}) := \frac{4e\beta^2}{\sigma_e^2 (\beta + d_m^2)^2} (\boldsymbol{\theta} - \mathbf{a}_m)(\boldsymbol{\theta} - \mathbf{a}_m)^T.$$

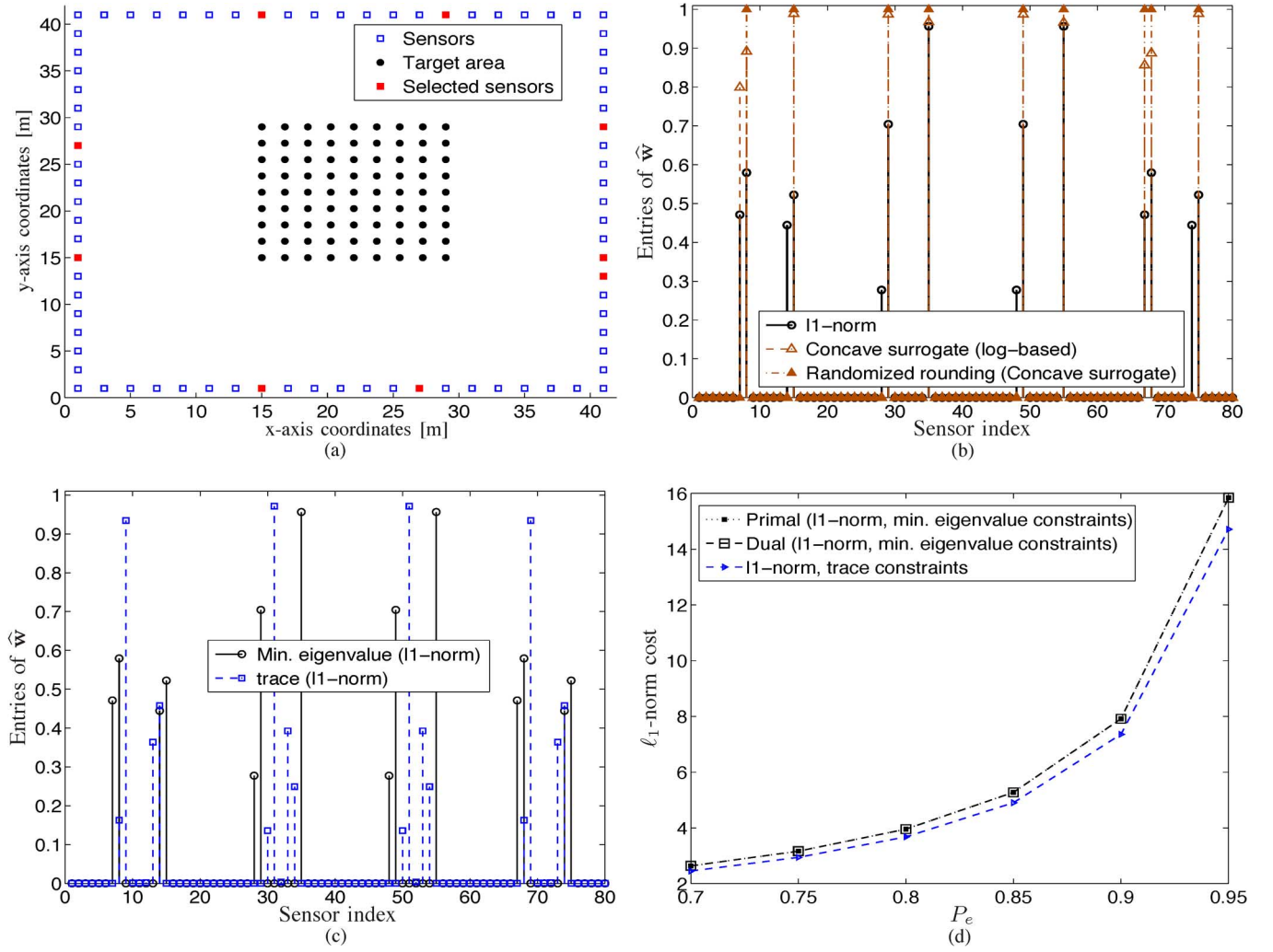


Fig. 1. Sensor placement for target localization with $M = 80$ candidate sensor positions. The thresholds are computed using $R_e = 20$ cm, and $P_e = 0.9$ (except for (d)). (a) Selection based on sparsity-enhancing iterations with minimum eigenvalue constraints. The Boolean solution is recovered using randomized rounding. (b) Minimum eigenvalue constraints with ℓ_1 -norm and concave surrogate based relaxations. Randomized rounding is applied on the concave surrogate based solution. (c) ℓ_1 -norm based selection with the trace constraints. (d) ℓ_1 -norm cost function for different values of P_e and a fixed $R_e = 20$ cm.

Although the examples illustrated so far are related to passive sensing, we underline that the proposed sensor selection framework is not limited to passive sensors.

Remark 7 (Active Sensor Selection): The sensor selection problem can also be formulated for active sensing. In active sensing, the sensors transmit probing signals (e.g., radar, sonar, remote sensing). The selection parameter w_m for active sensing is a soft parameter used for joint selection and resource allocation [1], i.e., $w_m \in [0, 1]$ is a resource (e.g., transmit energy) normalized to the maximum prescribed value, and hence, it is dimensionless. The relaxed active sensor selection problem takes the same form as in (15). In fact, minimizing the ℓ_1 -norm in active sensor selection minimizes the overall network resources (e.g., overall transmit energy).

E. Simulations

We apply the proposed sensor selection problem to sensor placement design for target localization. To test the proposed algorithms, we use CVX [35]. CVX internally calls SeDuMi [34], a MATLAB implementation of the second-order interior-point methods.

We consider the scenario shown in Fig. 1(a) with $M = 80$ sensors to illustrate the sensor selection problem. Recall that the problem here is to choose the best sensor positions out of $M = 80$ available ones, such that a certain specified localization accuracy is achieved. The domain \mathcal{U} for this example will be the target (or surveillance) area where the target resides, and to avoid having infinitely many constraints the area \mathcal{U} consists of grid points at a certain resolution. We grid the target area of 15×15 m² uniformly with a resolution of 1.75 m along both horizontal and vertical directions as shown in Fig. 1(a).

The original non-convex sensor selection problem is relaxed to an ℓ_1 -norm optimization problem. Alternatively, a concave surrogate function can be used to further enhance the sparsity. The optimization problem with the concave surrogate cost function is iteratively solved by affinely scaling the objective based on the solution from the previous iteration. For the sparsity-enhancing iterative Algorithm 2, we use $i_{\max} = 10$ and $\delta = 10^{-8}$. The number of randomizations used in the randomized rounding Algorithm 3 is $L = 100$. As observed in the simulations, a solution is typically found in the first batch itself, and a few tens of candidate entries are sufficient. We use the following param-

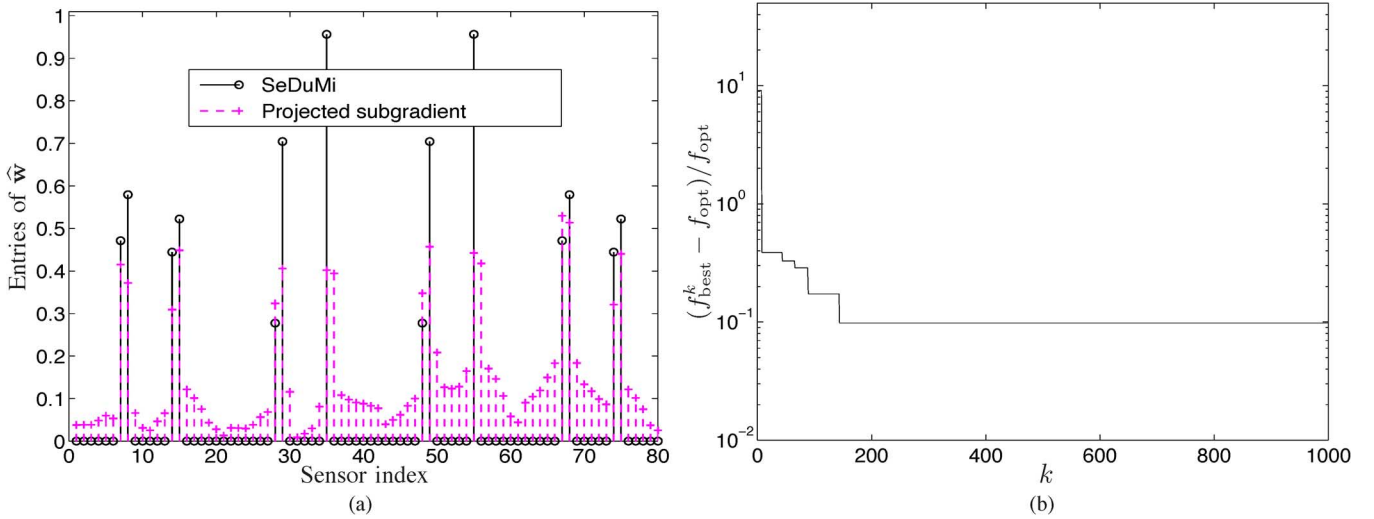


Fig. 2. (a) The projected subgradient algorithm used to solve (15). (b) Performance of the projected subgradient algorithm. The thresholds are computed using $R_e = 20$ cm and $P_e = 0.9$.

eters for the simulations: $\eta = 2$, $\sigma^2 = 2 \times 10^{-5}$, $\sigma_b^2 = 2 \times 10^{-5}$ square-degrees, $\sigma_{r,\text{dB}} = 2$ dB, $\sigma_e^2 = -20$ dBm, $e = 1$, $\beta = 1$, and $P_e = 0.9$ (except in Fig. 1(d)).

Fig. 1 shows the sensor selection for the distance (range) measurement model. The thresholds are computed with $R_e = 20$ cm and $P_e = 0.9$. The selection shown in Fig. 1(a) is based on Algorithm 2 with randomized rounding to recover the approximate Boolean solution. The selection results based on the ℓ_1 -norm cost with the minimum eigenvalue constraint is shown in Fig. 1(b). Fig. 1(b) also shows that the solution based on the concave surrogate cost function with the minimum eigenvalue constraint leads to a sparser solution. The selection results based on the trace constraint obtained by solving (22) are illustrated in Fig. 1(c). The sensors from the same region (close to the red filled boxes in Fig. 1(a)) are selected with either one of the two constraints. Fig. 1(d) shows a zero duality gap (gap between the cost of the primal problem in (15) and the dual problem in (25)) for different values of P_e . Larger values of P_e result in a larger $\|\mathbf{w}\|_1$, and subsequently more sensors are selected. The sufficient trace constraint has a larger feasible set compared to the stronger sufficient minimum eigenvalue constraint. As a result, for the considered scenario, the minimum eigenvalue constraint leads to a slightly larger ℓ_1 -norm compared to the trace constraint.

The optimization problem (15) is also solved using the projected subgradient method summarized in Algorithm 1 with $k_{\max} = 1000$ iterations. The solution of the projected subgradient is shown in Fig. 2(a). The performance of the projected subgradient algorithm is compared to the solution of the interior-point methods (implemented using SeDuMi) denoted by f_{opt} (obtained via SeDuMi), i.e., the convergence $(f_{\text{best}}^k - f_{\text{opt}})/f_{\text{opt}}$ is shown in Fig. 2(b). Even though the convergence of the projected subgradient algorithm is very slow, the estimated support after a few hundred iterations can be used along with randomized rounding to further refine the solution. The computation time on the same computer for the projected subgradient algorithm that solves (15) is around 8.84

seconds for 1000 iterations while SeDuMi takes around 4.03 seconds to solve the SDP problem in (15).

A practical estimator does not meet the CRB in some cases (for instance at low SNRs or finite data records). Therefore, the sensors selected with a specific R_e would lead to an underestimate of the desired MSE. We can account for this gap by choosing R_e appropriately. To this end, we give the entire solution path of the selected sensors for different values of R_e in Fig. 3(a). The solution path can be efficiently computed by increasing R_e . The sensors corresponding to some R_e can then be used to meet the desired MSE requirement. The non-linear model in (26) is solved in the least-squares sense iteratively using Gauss-Newton's method with 10 iterations [26]. The maximum root-MSE (RMSE), maximum root-CRB, average RMSE, and average root-CRB of the location estimates of a target within the target area using the selected sensors (as shown in the solution path) for different values of R_e are shown in Fig. 3(b). For the considered scenario, both the maximum and average root-CRB satisfy the performance constraint which is given by the inequality in (28). The performance constraint is shown as a red solid line in Fig. 3(b). The maximum RMSE does not satisfy the accuracy requirement specified by a certain R_e , and this can be corrected by using an appropriate (lower) R_e . Moreover, for the considered scenario, the gap between the average RMSE and the performance constraint is still reasonable. We also show the maximum RMSE on top of Fig. 3(a).

The proposed framework can be applied to a variety of data models as long as (a1) and (a2) are valid. To illustrate this we next consider a few more measurement models. The sensor selection based on bearing measurements are illustrated in Fig. 4. Here, we use $R_e = 25$ cm and $P_e = 0.9$. The selection results for the RSS based measurement model are shown in Fig. 5. We use $\sigma_{r,\text{dB}}^2 = 2$ dB, $P_e = 0.9$, and $R_e = 5$ m. Sensor selection results based on energy measurements are shown in Fig. 6, where we use $R_e = 10$ cm, and $P_e = 0.9$. An illustration of the field generated by a point source at location $[25, 25]^T$ m with unit amplitude is also shown here.

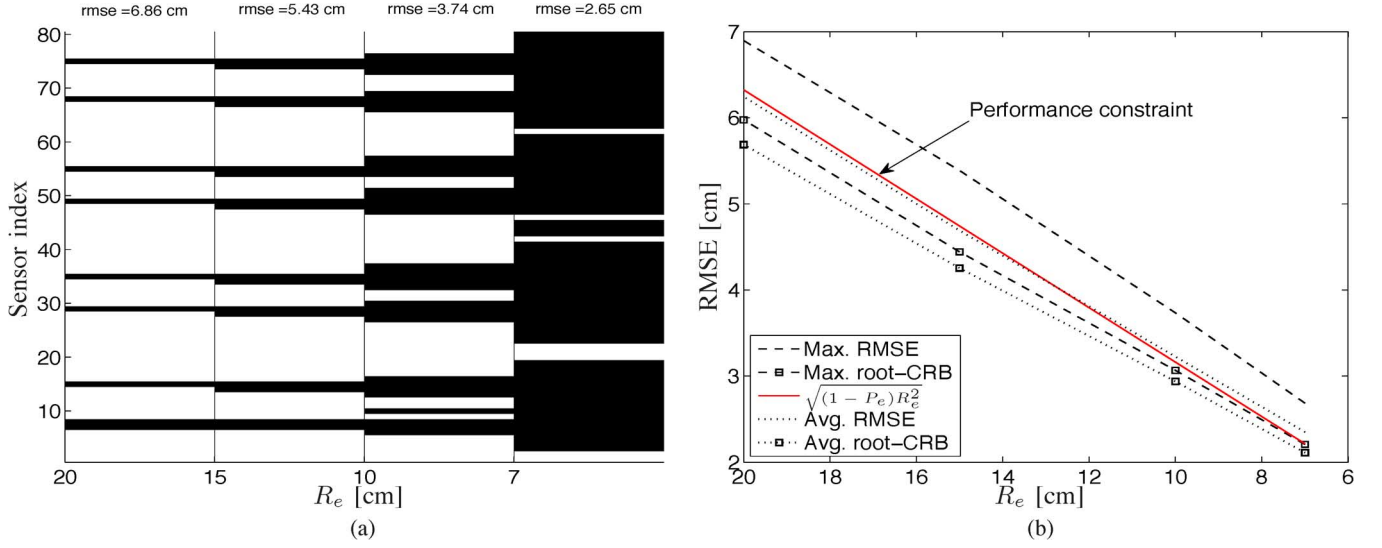


Fig. 3. (a) Solution path of the sensors selected for different values of R_e and $P_e = 0.9$. Maximum RMSE based on selected sensors can be seen on the top of this plot. (b) Maximum and average RMSE of the location estimates based on Gauss-Newton's method, the corresponding maximum and average root-CRB, and the performance constraint in (28) for different values of R_e , and $P_e = 0.9$.

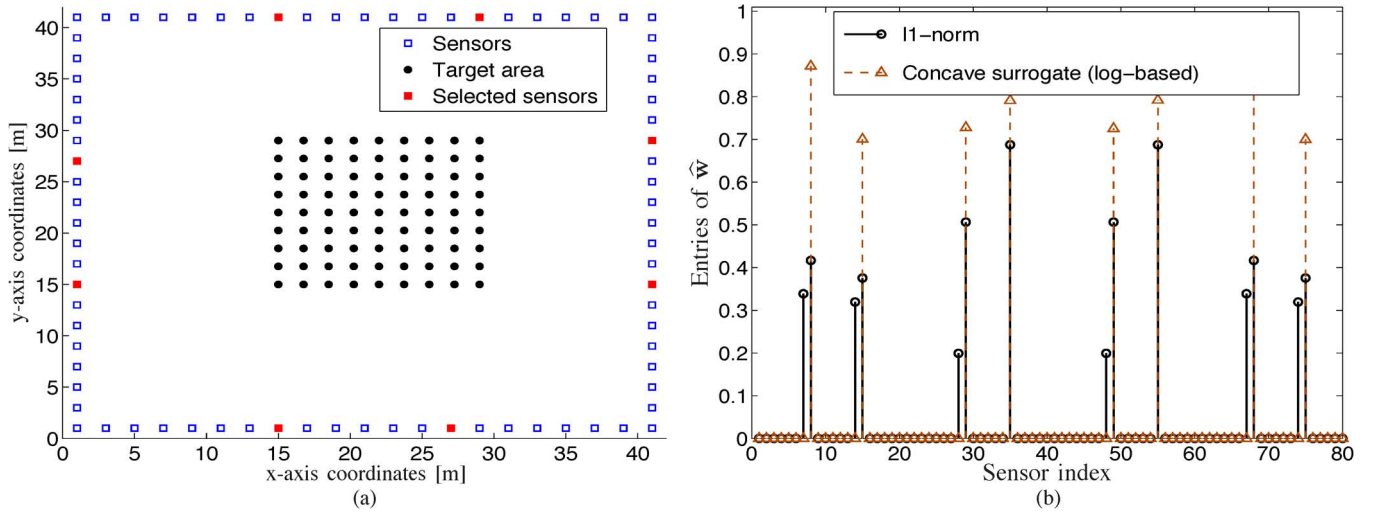


Fig. 4. (a) Sensor selection based on bearing measurements with $M = 80$ available sensors. (b) Sensor selection solved with minimum eigenvalue constraints using ℓ_1 -norm and log-based heuristics. The thresholds are computed using $R_e = 25$ cm, and $P_e = 0.9$. The noise variance is $\sigma_n^2 = 2 \times 10^{-5}$ square-degrees.

The FIM for all the considered measurement models has a common structure, and it decreases as the distance d_m increases. However, the rate at which it decreases is different for different models. As a result of this decrease, the optimization problem leads to a sensor placement that is close to the target area (in the Euclidean distance sense) for all the considered models.

VII. CONCLUSIONS

Sensor selection can be described as the problem of choosing the best subset of sensors that guarantees a certain desired inference performance. Sensor selection enables space-time sensing design, which is crucial for gathering only the most informative data. We solved the sensor selection problem by designing a sparse sensing operator that leads to possible reductions in the hardware, communications, and other processing costs. In particular, we focussed on observations that follow a non-linear

model. We used a number of scalar functions related to the FIM to determine the sensor selection. The original nonconvex optimization problem is relaxed using convex relaxation techniques which can then be efficiently solved in polynomial time. To handle large-scale problems, we have also presented a projected subgradient algorithm. The proposed framework is applied to sensor placement design for a number of different models related to localization.

APPENDIX A PERFORMANCE THRESHOLDS

Trace and Minimum Eigenvalue Constraints

We can relate the accuracy requirement and CRB using Chebyshev's inequality [28]

$$\Pr(\|\epsilon\|_2 \geq R_e) \leq \frac{\text{tr}\{\mathbf{C}\}}{R_e^2}$$

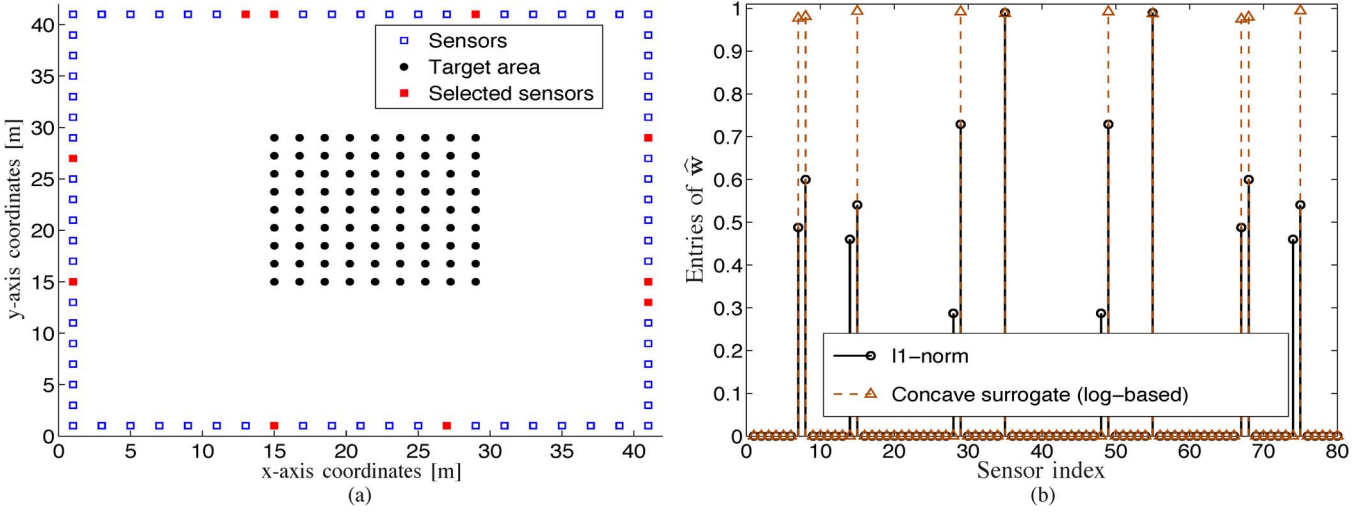


Fig. 5. (a) Sensor selection based on RSS with $M = 80$ available sensors. (b) Sensor selection solved with minimum eigenvalue constraints using ℓ_1 -norm and log-based heuristics. The thresholds are computed using $R_e = 5$ m, and $P_e = 0.9$. We use $\sigma_{r,\text{dB}} = 2$ dB.

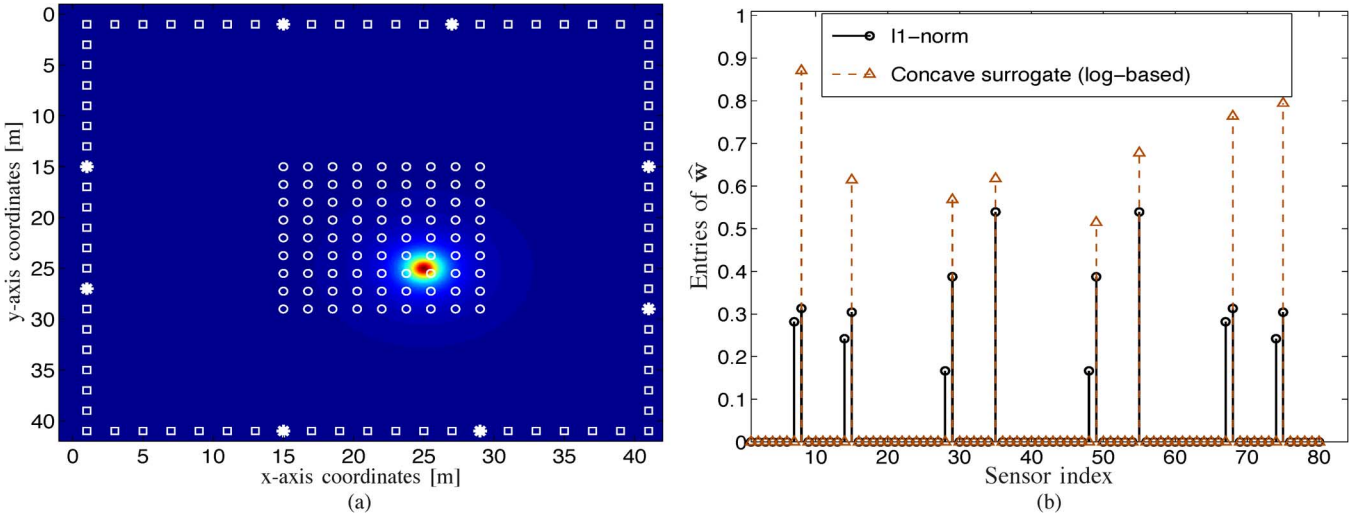


Fig. 6. Sensor selection for field estimation. (a) Illustration of a field generated by a unit amplitude point source at location $\theta = [25, 25]^T$ m according to (27). Out of $M = 80$ available sensors (\square), 16 sensors indicated by (*) are selected. The source domain is indicated by (o). (b) Sensor selection solved with minimum eigenvalue constraints using ℓ_1 -norm and log-based heuristics. The thresholds are computed using $R_e = 10$ cm, and $P_e = 0.9$. We use $e = 1$, $\beta = 1$ and $\sigma_e^2 = 10^{-5}$.

which can be equivalently expressed as $\Pr(\|\epsilon\|_2 \leq R_e) \leq 1 - \text{tr}\{\mathbf{C}\}/R_e^2$. Combining this inequality together with $\Pr(\|\epsilon\|_2 \leq R_e) \geq P_e$ in (11) results in the following sufficient condition

$$\text{tr}\{\mathbf{C}\} \leq \lambda_{\text{tr}} = (1 - P_e)R_e^2. \quad (28)$$

Each eigenvalue of \mathbf{C}^{-1} is greater than $\lambda_{\min}(\mathbf{F})$, and as a result, $\text{tr}\{\mathbf{C}\} \leq N\lambda_{\min}^{-1}(\mathbf{F})$. Hence, a stronger sufficient condition (with a smaller feasible set) is $N\lambda_{\min}^{-1}(\mathbf{F}) \leq (1 - P_e)R_e^2$, or equivalently [29]

$$\lambda_{\min}(\mathbf{F}) \geq \lambda_{\text{eig}} = \frac{N}{R_e^2} \left(\frac{1}{1 - P_e} \right).$$

Determinant Constraint

The determinant constraint is related to the volume or the mean radius of the confidence ellipsoid that contains $\epsilon = \theta -$

$\hat{\theta}$ with probability P_e . Such a confidence ellipsoid can be expressed as

$$\mathcal{E} = \{\epsilon | \epsilon^T \mathbf{F}^{-1} \epsilon \leq \xi\},$$

where ξ is a constant that depends on P_e . Assuming \mathbf{F} has ordered eigenvalues $\lambda_{\max} \geq \lambda_2 \geq \dots \geq \lambda_{\min}$, the length of the n th semi-axis of the ellipsoid \mathcal{E} will be $\sqrt{\xi/\lambda_n}$. The geometric mean radius of the confidence ellipsoid \mathcal{E} given by

$$\bar{R}_e = \frac{\sqrt{\xi}}{(\det\{\mathbf{F}\})^{\frac{1}{2N}}},$$

which gives a quantitative measure of how informative the observations are. For the estimates to be within the confidence ellipsoid \mathcal{E} , we use the constraint

$$\ln \det\{\mathbf{F}\} \geq 2N \ln \frac{\sqrt{\xi}}{\bar{R}_e} = \lambda_{\det},$$

where \bar{R}_e and $\sqrt{\xi}$ specify the required accuracy, and are assumed to be known. A typical choice for ξ is $\xi = F_{\chi_N^2}^{-1}(P_e)$. Here, $F_{\chi_N^2}^{-1}$ is the cumulative distribution function of a chi-squared random variable with N degrees of freedom. This performance measure is related to the D -optimality.

APPENDIX B

DERIVATION OF THE LAGRANGIAN DUAL PROBLEM

Consider the optimization problem (15) as the primal problem. To this problem, we then associate the following dual variables or Lagrangian multipliers: $\mathbf{Z} \in \mathbb{S}^{DN}$ with the LMI constraint; $\nu_m \in \mathbb{R}$ and $\mu_m \in \mathbb{R}$ with the $w_m \geq 0$ and $w_m \leq 1$ constraints, respectively. The Lagrangian is

$$\begin{aligned} \mathcal{L}(\mathbf{w}, \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\nu}) &= \mathbf{1}_M^T \mathbf{w} - \sum_{m=1}^M w_m \text{tr}\{\mathbf{F}_m \mathbf{Z}\} - \text{tr}\{-\lambda_{\text{eig}} \mathbf{Z}\} \\ &\quad + \boldsymbol{\mu}^T (\mathbf{w} - \mathbf{1}_M) - \boldsymbol{\nu}^T \mathbf{w} \\ &= \sum_{m=1}^M w_m (1 - \text{tr}\{\mathbf{F}_m \mathbf{Z}\} + \mu_m - \nu_m) \\ &\quad - \mathbf{1}_M^T \boldsymbol{\mu} - \text{tr}\{-\lambda_{\text{eig}} \mathbf{Z}\}. \end{aligned}$$

The Lagrange dual function is given as

$$\begin{aligned} \phi(\mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\nu}) &= \inf_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\nu}) \\ &= \begin{cases} -\text{tr}\{-\lambda_{\text{eig}} \mathbf{Z}\} - \mathbf{1}_M^T \boldsymbol{\mu} & \text{if } \text{tr}\{\mathbf{F}_m \mathbf{Z}\} + \nu_m = 1 + \mu_m, \\ & m = 1, 2, \dots, M, \\ -\infty & \text{otherwise.} \end{cases} \end{aligned} \quad (29)$$

The dual problem which is also an SDP can therefore be expressed as

$$\begin{aligned} \arg \max_{\mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\nu}} \quad & \lambda_{\text{eig}} \text{tr}\{\mathbf{Z}\} - \mathbf{1}_M^T \boldsymbol{\mu} \\ \text{s.t.} \quad & \text{tr}\{\mathbf{F}_m \mathbf{Z}\} + \nu_m = 1 + \mu_m, \quad m = 1, 2, \dots, M, \\ & \mathbf{Z} \succeq \mathbf{0}, \quad \mu_m \geq 0, \quad \nu_m \geq 0, \quad m = 1, 2, \dots, M. \end{aligned}$$

where $\mathbf{Z} \in \mathbb{S}_+^{DN}$ (we use the fact that \mathbb{S}_+^{DN} is self-dual), $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_M]^T \in \mathbb{R}^M$ and $\boldsymbol{\nu} = [\nu_1, \nu_2, \dots, \nu_M]^T \in \mathbb{R}^M$ are the dual variables. By eliminating ν_m , the dual problem is simplified to

$$\arg \max_{\mathbf{Z}, \boldsymbol{\mu}} \quad \lambda_{\text{eig}} \text{tr}\{\mathbf{Z}\} - \mathbf{1}_M^T \boldsymbol{\mu} \quad (30a)$$

$$\text{s.t.} \quad \text{tr}\{\mathbf{F}_m \mathbf{Z}\} \leq 1 + \mu_m, \quad m = 1, 2, \dots, M, \quad (30b)$$

$$\mathbf{Z} \succeq \mathbf{0}, \quad \mu_m \geq 0, \quad m = 1, 2, \dots, M. \quad (30c)$$

APPENDIX C

PROJECTED NEWTON'S METHOD

In order to analyze the complexity of the interior point methods, we briefly describe the projected Newton's method. The Newton's method for an SDP problem in the inequality form is adapted to suit our problem [11, p. 619].

The optimization problem in (15) can be approximated using the log-determinant barrier function which is given as

$$\arg \min_{\mathbf{w} \in [0,1]^N} \psi(\mathbf{w}) = t \mathbf{1}_M^T \mathbf{w} - \ln \det \left\{ \sum_{m=1}^M w_m \mathbf{F}_m - \lambda_{\text{eig}} \mathbf{I}_{DN} \right\},$$

where $t > 0$ is a parameter to tune the approximation. The projected Newton's update equation is given by

$$\mathbf{w}^{k+1} = \mathcal{P}_{\mathcal{W}} \left(\mathbf{w}^k - \alpha^k \mathbf{H}_{\psi}^{-1}(\mathbf{w}^k) \mathbf{g}_{\psi}(\mathbf{w}^k) \right), \quad (31)$$

where the (i, j) th entry of the Hessian matrix is given by

$$[\mathbf{H}_{\psi}(\mathbf{w}^k)]_{i,j} = \frac{\partial^2 \psi(\mathbf{w})}{\partial w_i \partial w_j} \bigg|_{\mathbf{w}=\mathbf{w}^k} = \text{tr}\{\mathbf{S}^{-1} \mathbf{F}_i \mathbf{S}^{-1} \mathbf{F}_j\},$$

and the i th entry of the gradient vector is given by

$$[\mathbf{g}_{\psi}(\mathbf{w}^k)]_i = \frac{\partial \psi(\mathbf{w})}{\partial w_i} \bigg|_{\mathbf{w}=\mathbf{w}^k} = t + \text{tr}\{\mathbf{S}^{-1} \mathbf{F}_i\}.$$

Here, we have introduced the matrix $\mathbf{S} = \sum_{m=1}^M w_m \mathbf{F}_m - \lambda_{\text{eig}} \mathbf{I}_{DN}$, and recall the projector operator $\mathcal{P}_{\mathcal{W}}(\cdot)$ defined in (17). The step-length α^k is chosen by line-search.

APPENDIX D

POWER ITERATIONS FOR COMPUTING THE MINIMUM EIGENVALUE

We briefly describe the power iterations [38] to compute the minimum eigenvalue of a matrix $\mathbf{F} \in \mathbb{S}^N$. Assuming \mathbf{F} has ordered eigenvalues $\lambda_{\max} \geq \lambda_2 \geq \dots \geq \lambda_{\min}$, the power iterations

$$\mathbf{v}^{k+1} = \frac{\mathbf{F} \mathbf{v}^k}{\|\mathbf{F} \mathbf{v}^k\|_2}, \quad \text{and} \quad \lambda^{k+1} = \frac{(\mathbf{v}^{k+1})^T \mathbf{F} \mathbf{v}^{k+1}}{\|\mathbf{v}^{k+1}\|_2^2},$$

converge to the eigenvector corresponding to the maximum eigenvalue \mathbf{v}_{\max} , and the maximum eigenvalue λ_{\max} , respectively, as $k \rightarrow \infty$. Here, we use $\mathbf{v}^0 = [1, \mathbf{0}_{N-1}^T]^T$. By forming a matrix $\bar{\mathbf{F}} = \lambda_{\max} \mathbf{I}_N - \mathbf{F}$ which has the dominant eigenvalue $\lambda_{\max} - \lambda_{\min}$, we can apply the above power iterations on $\bar{\mathbf{F}}$ to compute $\lambda_{\max} - \lambda_{\min}$ and \mathbf{v}_{\min} , and thus the minimum eigenvalue of \mathbf{F} and its corresponding eigenvector.

REFERENCES

- [1] S. P. Chepuri, G. Leus, and A.-J. van der Veen, "Sparsity-exploiting anchor placement for localization in sensor networks," in *Proc 21st Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2013, pp. 1–5.
- [2] G. Mateos and K. Rajawat, "Dynamic network cartography: Advances in network health monitoring," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 129–143, May 2013.
- [3] G. B. Giannakis, V. Kekatos, N. Gatsis, S.-J. Kim, H. Zhu, and B. F. Wollenberg, "Monitoring and optimization for power grids: A signal processing perspective," *IEEE Signal Process. Mag.*, vol. 30, no. 5, pp. 107–128, Sept. 2013.
- [4] F. Gustafsson and F. Gunnarsson, "Mobile positioning using wireless networks: Possibilities and fundamental limitations based on available wireless network measurements," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 41–53, Jul. 2005.
- [5] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: A look at positioning aspects for future sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 70–84, Jul. 2005.
- [6] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, III, R. L. Moses, and N. S. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 54–69, Jul. 2005.
- [7] T. van Waterschoot and G. Leus, "Static field estimation using a wireless sensor network based on the finite element method," in *Proc. 4th IEEE Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, 2011, pp. 369–372.

- [8] H. Zhang, J. M. F. Moura, and B. Krogh, "Dynamic field estimation using wireless sensor networks: Tradeoffs between estimation error and communication cost," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2383–2395, Jun. 2009.
- [9] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.
- [10] F. Pukelsheim, *Optimal Design of Experiments*. Philadelphia, PA, USA: SIAM, 1993, vol. 50.
- [11] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.
- [12] V. Kekatos, G. B. Giannakis, and B. Wollenberg, "Optimal placement of phasor measurement units via convex relaxation," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1521–1530, Aug. 2012.
- [13] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies," *J. Machine Learn. Res.*, vol. 9, pp. 235–284, Feb. 2008.
- [14] A. Krause and C. Guestrin, "Near-optimal observation selection using submodular functions," in *AAAI*, 2007, vol. 7, pp. 1650–1654.
- [15] A. Krause, H. B. McMahan, C. Guestrin, and A. Gupta, "Robust submodular observation selection," *J. Mach. Learn. Res.*, vol. 9, no. 12, 2008.
- [16] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: Leveraging submodularity," in *Proc. 49th IEEE Conf. Dec. Control (CDC)*, 2010, pp. 2572–2577, IEEE.
- [17] L. Yao, W. A. Sethares, and D. C. Kammer, "Sensor placement for on-orbit modal identification via a genetic algorithm," *AIAA J.*, vol. 31, no. 10, pp. 1922–1928, 1993.
- [18] E. Masazade, M. Fardad, and P. K. Varshney, "Sparsity-promoting extended Kalman filtering for target tracking in wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 845–848, 2012.
- [19] A. Carmi, "Sensor scheduling via compressed sensing," in *Proc. 13th Conf. Inf. Fusion (FUSION)*, 2010, pp. 1–8.
- [20] Y. Fu, Q. Ling, and Z. Tian, "Distributed sensor allocation for multi-target tracking in wireless sensor networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 4, pp. 3538–3553, 2012.
- [21] J. Ranieri, A. Chebira, and M. Vetterli, "Near-optimal sensor placement for linear inverse problems," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1135–1146, Mar. 2014.
- [22] I. Ford, D. M. Titterton, and C. P. Kitsos, "Recent advances in nonlinear experimental design," *Technometrics*, vol. 31, no. 1, pp. 49–60x, 1989.
- [23] D. Bajovic, B. Sinopoli, and J. Xavier, "Sensor selection for event detection in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4938–4953, Oct. 2011.
- [24] V. Kekatos and G. B. Giannakis, "From sparse signals to sparse residuals for robust sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3355–3368, 2011.
- [25] I. D. Schizas, "Distributed informative-sensor identification via sparsity-aware matrix decomposition," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4610–4624, 2013.
- [26] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
- [27] T. J. Rothenberg, "Identification in parametric models," *Econometrica: J. Econometr. Soc.*, pp. 577–591, 1971.
- [28] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2012.
- [29] T. Wang, G. Leus, and L. Huang, "Ranging energy optimization for robust sensor positioning based on semidefinite programming," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4777–4787, Dec. 2009.
- [30] H. L. Van Trees, *Detection, Estimation, Modulation Theory, Optimum Array Processing*. New York, NY, USA: Wiley, 2004.
- [31] B. Polyak, M. Khlebnikov, and P. Shcherbakov, "An LMI approach to structured sparse feedback design in linear control systems," in *Proc. Eur. Control Conf. (ECC)*, 2013, pp. 833–838.
- [32] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1 -norm solution is also the sparsest solution," *Commun. Pure Appl. Math.*, vol. 59, no. 6, pp. 797–829, 2006.
- [33] J. Lofberg, "Yalmip: A toolbox for modeling and optimization in MATLAB," presented at the CACSD Conf., Taipei, Taiwan, 2004.
- [34] J. F. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optim. Methods Softw.*, vol. 11, no. 1–4, pp. 625–653, 1999.
- [35] M. Grant and S. Boyd, CVX: Matlab Software for Disciplined Convex Programming. ver. 2.0 Beta, Sep. 2012 [Online]. Available: <http://cvxr.com/cvx>
- [36] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," 2003 [Online]. Available: http://www.stanford.edu/class/ee364b/notes/subgrad_method_notes.pdf
- [37] D. P. Bertsekas, *Nonlinear Programming*, ser. Athena Scientific Optimization and Computation Series. Belmont, MA, USA: Athena Scientific, 1999.
- [38] G. H. Golub and C. F. Van Loan, *Matrix Computations*, ser. Johns Hopkins Studies in the Mathematical Sciences. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1996.
- [39] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, 2004.
- [40] E. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, 2008.
- [41] M. Gavish and A. J. Weiss, "Performance analysis of bearing-only target location algorithms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 817–828, Jul. 1992.



Sundeep Prabhakar Chepuri (S'11) was born in India in 1986. He received his B.E. degree (distinction) from the PES Institute of Technology, Bangalore, India, in 2007, and his M.Sc. degree (cum laude) in electrical engineering from the Delft University of Technology, The Netherlands, in 2011. He has held positions at Robert Bosch, India, during 2007–2009, and Holst Centre/imec-nl, The Netherlands, during 2010–2011. Since September 2011, he has been with the Circuits and Systems group at the Faculty of Electrical Engineering, Mathematics and Computer Science of the Delft University of Technology, The Netherlands, pursuing his Ph.D. degree. His research interest lies in the field of signal processing for communications and networks.



Geert Leus (M'01–SM'05–F'12) received the Electrical Engineering degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Belgium, in 1996 and 2000, respectively. Currently, he is an "Antoni van Leeuwenhoek" Full Professor at the Faculty of Electrical Engineering, Mathematics and Computer Science of the Delft University of Technology, Delft, The Netherlands. His research interests are in the area of signal processing for communications.

Prof. Leus received a 2002 IEEE Signal Processing Society Young Author Best Paper Award and a 2005 IEEE Signal Processing Society Best Paper Award. He 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, and the IEEE SIGNAL PROCESSING LETTERS. Currently, he is a member of the IEEE Sensor Array and Multichannel Technical Committee and serves as the Editor in Chief of the *EURASIP Journal on Advances in Signal Processing*.