DISTRIBUTED SENSOR SELECTION FOR FIELD ESTIMATION

Sijia Liu[†] *Sundeep Prabhakar Chepuri*[‡] *Geert Leus*[‡] *Alfred O. Hero III*[†]

[†]Department of EECS, University of Michigan, Ann Arbor, MI 48109, USA
[‡]Delft University of Technology, The Netherlands
[†]{lsjxjtu, hero}@umich.edu, [‡]{s.p.chepuri, g.j.t.leus}@tudelft.nl

***Abstract*—We study the sensor selection problem for field estimation, where a best subset of sensors is activated to monitor a spatially correlated random field. Different from most commonly used centralized selection algorithms, we propose a decentralized architecture where sensor selection can be carried out in a distributed way and by the sensors themselves. A decentralized approach is essential since each sensor has access only to the information (e.g., correlation) in its neighborhood. To make distributed optimization possible, we decompose the global cost function into local cost functions that require only the information in local neighborhoods of sensors. We then employ the alternating direction method of multipliers (ADMM) to solve the proposed sensor selection problem. In our algorithm, each sensor solves small-scale optimization problems, and communicates directly only with its immediate neighbors. Numerical results are provided to show the effectiveness of our approach.**

***Keywords*—Sensor selection, field estimation, distributed optimization, sparsity, alternating direction method of multipliers.**

I. Introduction

In this paper, we study the problem of distributed sensor selection for field estimation, where a spatially correlated random field is monitored by distributed sensors. In the task of field estimation, an unknown field intensity, such as temperature [1] or precipitation [2], is interpolated by sensor measurements. Due to limited sensor battery power, it is desirable to select only a subset of sensors for field monitoring such that a tradeoff can be achieved between estimation accuracy and sensor activations [3]. However, the existing works on sensor selection [4]–[12] mainly focus on the development of centralized sensor selection algorithms, where the complete knowledge about the system model is given in advance, and the computation is performed at a central processing unit, namely, the fusion center (FC). Different from the literature [4]–[12], we aim to develop a decentralized architecture where sensor selection can be carried out in a distributed way and by the sensors themselves. More importantly, due to the assumption that every sensor only has access to its neighbors' information, i.e., partial (local) information instead of complete (global) information, the need for a distributed algorithm is well motivated.

The problem of distributed sensor selection has attracted significant attention [13]–[16]. In [13], a distributed online greedy algorithm was proposed to maximize the utility of the activated sensors by assuming that the utility function is submodular (with diminishing returns property). However, for most commonly used estimation error metrics, the objective function used for sensor selection is not submodular in general [17]. In [14], [15], a dual subgradient method was employed to solve the distributed sensor selection problem under a linear Gaussian measurement model. Those papers considered distributed sensor selection for estimating deterministic unknowns with no or weak noise correlation, whereas we consider a more general problem. In [16], the distributed sensor selection problem was studied for a network with two leader nodes. The use of leader nodes makes the sensor selection algorithm not fully distributed. Different from [13]–[16], we propose a general framework of distributed sensor selection for field estimation, where both the sensor activation scheme, and the estimator gain matrix used to combine sensor measurements, are jointly optimized. This is based on the fact that a sensor being in the off state is represented by the corresponding column of the estimator gain matrix being identically zero.

Besides [13]–[16], our work is also related to, but different from, the problem of consensus-based distributed estimation [18]–[26]. The success of most approaches to distributed estimation is primarily due to the decomposability of the global cost (e.g., sum-of-squares cost in [22], [23]) into a sum of local cost functions accessed by sensors. The main contribution of our paper is a new framework for designing the optimal sensor selection scheme for reconstructing the spatially-correlated random field, wherein the objective function used for sensor selection is not separable. To circumvent this difficulty, we propose an approximate decomposition of the global sensor selection cost into local cost functions that require only the information in neighborhoods of each of the sensors. We then employ alternating direction method of multipliers (ADMM) [27] to solve the resulting distributed optimization problem. Although subgradient-based algorithms [28]–[30] are applicable for distributed optimization, ADMM has much faster empirical convergence as demonstrated in many applications [22], [27], [31]. Moreover, our novel ADMM algorithm allows us to split the original complex problem into simple subproblems, each of which can be solved analytically and in a distributed manner.

II. Problem Statement

We consider a generic system, where sensors are deployed to monitor a spatially correlated random field. The task of field estimation is to estimate (or interpolate) the unknown field intensity from the measurements of the activated sensors, where the design of optimal sensor activations is to be carried out in a distributed way and by the sensors themselves.

A. Field estimation

We denote by $\phi(\mathbf{a})$ the field intensity at a 2D location $\mathbf{a} \in \mathbb{R}^2$. Without loss of generality, we assume that $\mathbb{E}[\phi(\mathbf{a})] = 0$ and $\mathbb{E}[\phi^2(\mathbf{a})] = \sigma_\phi^2$. The measurement model at the ith sensor is given by $y_i = \phi(\mathbf{s}_i) + v_i$ for $i \in [m]$, where $\mathbf{s}_i$ is the location of sensor i, v_i is the i.i.d. measurement noise with $\mathbb{E}[v_i] = 0$ and $\mathbb{E}[v_i^2] = \sigma_v^2$, and $[m]$ denotes the integer set $\{1, 2, \ldots, m\}$. Here we assume that v_i is uncorrelated with $\phi(\mathbf{s}_i)$. Upon defining the vectors of length M $\boldsymbol{\phi} := [\phi(\mathbf{s}_1), \phi(\mathbf{s}_2), \ldots, \phi(\mathbf{s}_m)]^T$ and $\mathbf{v} := [v_1, v_2, \ldots, v_m]^T$, the measurement model is written as

$$\mathbf{y} = \boldsymbol{\phi} + \mathbf{v}. \tag{1}$$

Let $\boldsymbol{\theta} := [\phi(\mathbf{u}_1), \phi(\mathbf{u}_2), \ldots, \phi(\mathbf{u}_n)]^T$ be the vector of field intensities to be estimated at n *unobserved* locations $\{\mathbf{u}_i\}_{i=1}^n$. A

This work was partially supported by the US Army Research Office under grant W911NF-15-1-0479 and the KAUST-MIT-TUD consortium under grant OSR-2015-Sensors-2700.

linear estimator for $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = \mathbf{W}\mathbf{y}, \tag{2}$$

where $\mathbf{W} \in \mathbb{R}^{n\times m}$ is the unknown estimator gain matrix determined by the minimum mean squared error (MMSE) criterion. The estimation distortion for the estimator (2) is given by

$$f(\mathbf{W}) := \mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})] = \operatorname{tr}(\mathbf{W}\mathbf{Q}\mathbf{W}^T) - 2\operatorname{tr}(\mathbf{R}\mathbf{W}), \tag{3}$$

where $\mathbf{Q} \in \mathbb{R}^{m\times m}$ and $\mathbf{R} \in \mathbb{R}^{m\times n}$ are defined as

$$\mathbf{Q} := \mathbb{E}[\mathbf{y}\mathbf{y}^T] = \mathbb{E}[\boldsymbol{\phi}\boldsymbol{\phi}^T] + \sigma_v^2\mathbf{I} \tag{4}$$

$$\mathbf{R} := \mathbb{E}[\mathbf{y}\boldsymbol{\theta}^T] = \mathbb{E}[\boldsymbol{\phi}\boldsymbol{\theta}^T]. \tag{5}$$

In (3), we have ignored a constant term $\mathbb{E}[\boldsymbol{\phi}^T\boldsymbol{\phi}] = n\sigma_\phi^2$ since it does not affect the minimum of $f(\mathbf{W})$. We will call $f(\mathbf{W})$ a global sensor selection cost, which relies on information about the entire field correlation. Specifically, the matrix $\mathbf{Q}$ in (4) corresponds to the spatial correlation of the field points at sensor locations $\{\mathbf{s}_i\}_{i=1}^m$, and the matrix $\mathbf{R}$ in (5) corresponds to the spatial correlation between the observed field points at $\{\mathbf{s}_i\}_{i=1}^m$ and the field points to be estimated at $\{\mathbf{u}_i\}_{i=1}^n$.

B. Centralized sensor selection

The estimator (2) implies a one-to-one correspondence between every sensor and a column of $\mathbf{W}$. Suppose, for example, if only the ith sensor reports a measurement, then it follows from (2) that $\mathbf{W}\mathbf{y}$ reduces to $\mathbf{w}_i y_i$, where $\mathbf{w}_i$ is the ith column of $\mathbf{W}$. Therefore, the design of an optimal sensor selection scheme can be accomplished by striking a balance between minimizing the estimation distortion (3) and promoting the column-sparsity of $\mathbf{W}$, namely,

$$\underset{\mathbf{W}}{\text{minimize}} \quad f(\mathbf{W}) + \gamma\sum_{i=1}^m \|\mathbf{w}_i\|_2, \tag{6}$$

where we have an implicit constraint $\mathbf{w}_i = \mathbf{W}\mathbf{e}_i$ with $\mathbf{e}_i$ being the basis vector with 1 at the ith coordinate and zeros elsewhere, the ℓ_2 norm $\|\mathbf{w}_i\|_2$ is introduced to measure the column-sparsity of $\mathbf{W}$ [32], and γ is a regularization parameter that governs the relative importance of achieving a good estimation performance versus activating a small number of sensors.

It has been shown in [31]–[33] that problem (6) can be solved at the FC. However, in the *absence* of a FC or the full knowledge about the spatial correlation, it is critical to compute the sensor selection scheme in a distributed manner that is based on information about correlation only in the local neighborhoods of each sensor. Throughout the paper, the optimal solution of problem (6) will be used as a benchmark to compare against the proposed distributed sensor selection schemes.

C. Distributed sensor selection

A graph yields a succinct description of the local (one hop) connectivity of the considered sensor network. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph, where $\mathcal{V}$ is the set of m vertices (namely, sensors), and $\mathcal{E}$ is the set of edges (namely, communication links among sensors). We assume that $\mathcal{G}$ is connected, namely, there exists a path that joins any pair of sensors in the network. Each sensor can communicate directly only with its immediate neighbors, where the 1-hop neighborhood of sensor i is given by $\mathcal{N}_i = \{j \in \mathcal{V} \,|\, (i,j) \in \mathcal{E}\} \cup \{i\}$.

We assume that each sensor has access to local information about correlation in its local neighborhood, namely, $\mathbf{Q}_{\mathcal{N}_i,\mathcal{N}_i} \in \mathbb{R}^{|\mathcal{N}_i|\times|\mathcal{N}_i|}$ for $i \in [m]$, where $\mathbf{Q}_{\mathcal{N}_i,\mathcal{N}_i}$ is the submatrix of $\mathbf{Q}$ in (4) with rows and columns indexed by $\mathcal{N}_i$, and $|\mathcal{N}_i|$ is the cardinality of $\mathcal{N}_i$. It is known from [34] that $\mathbf{Q}_{\mathcal{N}_i,\mathcal{N}_i}$ can be estimated by a local sample covariance matrix. We also assume that each sensor has access to the spatial correlation between itself and the field points to be estimated, namely, $\mathbf{r}_i \in \mathbb{R}^n$ for $i \in [m]$, where $\mathbf{r}_i$ is the ith row vector of $\mathbf{R}$ in (5). It has been suggested by [35] that $\mathbf{r}_i$ can be learnt from local data using a parametric correlation model (e.g., a power exponential model).

Given $\{\mathbf{Q}_{\mathcal{N}_i,\mathcal{N}_i}\}_{i=1}^m$ and $\{\mathbf{r}_i\}_{i=1}^m$, we pose the optimization problem for distributed sensor selection as follows

$$\underset{\mathbf{W}}{\text{minimize}} \quad \sum_{i=1}^m f_i(\mathbf{W}, \mathbf{Q}_{\mathcal{N}_i,\mathcal{N}_i}, \mathbf{r}_i) + \gamma\sum_{i=1}^m \|\mathbf{w}_i\|_2, \tag{7}$$

where f_i denotes a local cost function available to sensor i. Compared to problem (6), we may ask the following questions: What is the appropriate choice of f_i? Is it possible to decompose the global cost function f into local cost functions $\{f_i\}$? In the next section, we will answer these questions in the affirmative.

III. Local Cost versus Global Cost

Given information about local correlation $\{\mathbf{Q}_{\mathcal{N}_i,\mathcal{N}_i}\}_{i=1}^m$, we note that it is difficult to reconstruct the entire correlation matrix $\mathbf{Q}$ based on such information. For example, consider a 1D field with $m = 4$ sensors in Fig. 1. It is clear that the correlation Q_{14} between sensor 1 and sensor 4 is not included in $\{\mathbf{Q}_{\mathcal{N}_i,\mathcal{N}_i}\}_{i=1}^4$. Therefore, there is an information mismatch between $\{\mathbf{Q}_{\mathcal{N}_i,\mathcal{N}_i}\}_{i=1}^m$ and $\mathbf{Q}$. Consequently, recovering the global cost function f from local cost functions $\{f_i\}$ is intractable in general. However, Proposition 1 shows that there exist special cases in which f can be exactly decomposed into $\sum_{i=1}^m f_i$.

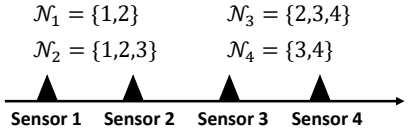


Fig. 1: 1D field with 4 sensors.

Proposition 1: Given $\{\mathbf{Q}_{\mathcal{N}_i,\mathcal{N}_i}\}_{i=1}^m$ and $\{\mathbf{r}_i\}_{i=1}^m$, if $Q_{ij} = 0$ when $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$, the global cost function (3) becomes

$$\begin{cases} f(\mathbf{W}) = \sum_{i=1}^m f_i(\mathbf{W}_{\mathcal{N}_i}) \\ f_i(\mathbf{W}_{\mathcal{N}_i}) = \operatorname{tr}[\mathbf{W}_{\mathcal{N}_i}(\mathbf{C}_i \circ \mathbf{Q}_{\mathcal{N}_i,\mathcal{N}_i})\mathbf{W}_{\mathcal{N}_i}^T] - 2\mathbf{r}_i^T\mathbf{w}_i, \end{cases} \tag{8}$$

where Q_{ij} is the (i,j)th entry of $\mathbf{Q}$, $\mathbf{W}_{\mathcal{N}_i} \in \mathbb{R}^{n\times|\mathcal{N}_i|}$ is the submatrix of $\mathbf{W}$ with columns indexed by $\mathcal{N}_i$, $\mathbf{C}_i \in \mathbb{R}^{|\mathcal{N}_i|\times|\mathcal{N}_i|}$ is a given coefficient matrix, and $\circ$ is the entrywise product. Let $\mathcal{N}_i = \{j_1, j_2, \ldots, j_{|\mathcal{N}_i|}\}$, the (p,q)th entry of $\mathbf{C}_i$ is given by

$$[\mathbf{C}_i]_{p,q} = \begin{cases} \frac{1}{|\mathcal{N}_{j_p}|} & \text{if } p = q \\ \frac{1}{|\mathcal{N}_{j_p} \cap \mathcal{N}_{j_q}|} & \text{if } p \neq q, \end{cases} \tag{9}$$

for $p, q = 1, 2, \ldots, |\mathcal{N}_i|$.

Proof: The proof is reported in Appendix A of [36]. ■

Proposition 1 implies that it is possible to decompose the global cost function into local cost functions when $\mathbf{Q}$ is sparse (specifically, $Q_{ij} = 0$ if sensor i is at least a 3-hop neighbor of sensor j). Note that a similar 3-hop condition has been shown in [34], where the maximum likelihood estimate of the precision matrix (i.e., inverse of the covariance matrix) could be computed in a distributed manner over the graphical model that defines the precision matrix sparsity structure.

Based on [34], the 1-hop and 2-hop neighbors of each sensor can be interpreted as its protected and buffer nodes, respectively. If $\mathbf{Q}$ is not sparse, the right hand side of (8) also provides a fine approximation of the global cost function, since the spatial correlation of two field points often decays fast as their distance increases [35].

Based on (8), problem (7) becomes

$$\underset{\mathbf{W}}{\text{minimize}} \quad \sum_{i=1}^{m}[\text{tr}(\mathbf{W}_{\mathcal{N}_i}\mathbf{S}_i\mathbf{W}_{\mathcal{N}_i}^T) - 2\mathbf{r}_i^T\mathbf{w}_i] + \gamma\sum_{i=1}^{m}\|\mathbf{w}_i\|_2, \quad (10)$$

where $\mathbf{S}_i = \mathbf{C}_i \circ \mathbf{Q}_{\mathcal{N}_i,\mathcal{N}_i} \in \mathbb{R}^{|\mathcal{N}_i|\times|\mathcal{N}_i|}$. Here the index set $\mathcal{N}_i$ may overlap with each other, hence $\{\mathcal{N}_1,\ldots,\mathcal{N}_m\}$ is a cover rather than a partition of $[m]$. In this sense, problem (10) can be regarded as a general consensus problem [37], where the objective function is a sum of local objective functions but with overlapping optimization variables. Another key observation in (10) is that $\mathbf{S}_i$ is not necessarily a positive semidefinite matrix. The indefiniteness of $\mathbf{S}_i$ makes it challenging to solve problem (10) in a distributed manner.

IV. Decentralized Optimization Method

We begin by reformulating problem (10) as

$$\begin{array}{ll}\underset{\mathbf{W},\{\mathbf{Z}_i\}}{\text{minimize}} & \sum_{i=1}^{m}[\text{tr}(\mathbf{S}_i\mathbf{Z}_i) - 2\mathbf{r}_i^T\mathbf{w}_i] + \gamma\sum_{i=1}^{m}\|\mathbf{w}_i\|_2 \\ \text{subject to} & \mathbf{Z}_i = \mathbf{W}_i^T\mathbf{W}_i,\ i\in[m],\end{array} \quad (11)$$

where $\mathbf{Z}_i \in \mathbb{R}^{|\mathcal{N}_i|\times|\mathcal{N}_i|}$ is an auxiliary optimization variable, and the variable $\mathbf{W}_{\mathcal{N}_i} \in \mathbb{R}^{n\times|\mathcal{N}_i|}$ is replaced with $\mathbf{W}_i$ for notational simplicity. Problem (11) contains a convex objective function but introduces nonconvex equality constraints. It is known from [38] that the nonconvex constraints of problem (11) can be relaxed to convex positive semidefinite constraints $\mathbf{Z}_i \succeq \mathbf{W}_i^T\mathbf{W}_i$ for $i\in[m]$ (such a convex relaxation is known as semidefinite relaxation), where $\mathbf{X}\succeq\mathbf{Y}$ signifies that $\mathbf{X}-\mathbf{Y}$ is positive semidefinite.

Problem (11) is relaxed to

$$\begin{array}{ll}\underset{\mathbf{W},\{\mathbf{Z}_i\}}{\text{minimize}} & \sum_{i=1}^{m}[\text{tr}(\mathbf{S}_i\mathbf{Z}_i) - 2\mathbf{r}_i^T\mathbf{w}_i] + \gamma\sum_{i=1}^{m}\|\mathbf{w}_i\|_2 \\ \text{subject to} & \begin{bmatrix}\mathbf{Z}_i & \mathbf{W}_i^T\\ \mathbf{W}_i & \mathbf{I}\end{bmatrix}\succeq\mathbf{0},\ i\in[m],\end{array} \quad (12)$$

where the linear matrix inequality is a reformulation of $\mathbf{Z}_i \succeq \mathbf{W}_i^T\mathbf{W}_i$ using the Schur complement. Problem (12) is not equivalent to problem (11) in general. However, Proposition 2 shows that the matrix $\mathbf{Z}_i - \mathbf{W}_i^T\mathbf{W}_i$ at the solution of problem (12) is of low rank. The low-rank property together with our numerical results in Fig. 2 suggest that the semidefinite relaxation performs well and could provide near-optimal sensor selection schemes.

Proposition 2: Let $\{(\mathbf{Z}_i^*, \mathbf{W}_i^*)\}_{i=1}^m$ be the solution of problem (12), we have $\text{rank}\left(\mathbf{Z}_i^* - (\mathbf{W}_i^*)^T\mathbf{W}_i^*\right) < |\mathcal{N}_i|$.

Proof: The proof is reported in Appendix B of [36]. ■

In what follows, ADMM is used as the decentralized message sharing protocol. The core advantage of ADMM is that it allows us to split problem (12) into subproblems, each of which can be solved analytically based on only local data passage between neighboring sensors. We refer interested readers to [27] for a survey on ADMM.

A. Distributed optimization via ADMM

We express problem (12) in a way that lends itself to the application of ADMM [27], [37]

$$\begin{aligned}\text{minimize} \quad & \sum_{i=1}^{m}\text{tr}(\mathbf{S}_i\mathbf{Z}_i) + \sum_{i=1}^{m}(\gamma\|\mathbf{w}_i\|_2 - 2\mathbf{r}_i^T\mathbf{w}_i) + \sum_{i=1}^{m}\mathcal{I}_+(\mathbf{Y}_i) \\ & + \mathcal{I}_{\mathcal{C}}(\{\mathbf{W}_i\}) && (13a)\\ \text{subject to} \quad & \begin{bmatrix}\mathbf{Z}_i & \mathbf{X}_i^T\\ \mathbf{X}_i & \mathbf{I}\end{bmatrix} = \mathbf{Y}_i,\ i\in[m] && (13b)\\ & \mathbf{X}_i = \mathbf{W}_i,\ i\in[m] && (13c)\\ & \mathbf{X}_i\mathbf{e}_{l_i} = \mathbf{w}_i,\ i\in[m], && (13d)\end{aligned}$$

where $\{\mathbf{Z}_i\}$, $\{\mathbf{W}_i\}$, $\{\mathbf{X}_i\}$, $\{\mathbf{Y}_i\}$ and $\{\mathbf{w}_i\}$ are optimization variables. In (13a), $\mathcal{I}_{\mathcal{C}}(\{\mathbf{W}_i\})$ is an indicator function of the convex set described by consensus constraints $\mathcal{C} = \{\{\mathbf{W}_i\}\,|\,(\mathbf{W}_i)_k = (\mathbf{W}_j)_k,\ i,j\in\mathcal{N}_k\}$ [37], where $\mathcal{I}_{\mathcal{C}}(\{\mathbf{W}_i\}) = 0$ if $\{\mathbf{W}_i\}\in\mathcal{C}$ and 0 otherwise, and $(\mathbf{W}_i)_k$ denotes the kth column of $\mathbf{W}$ that appears in $\mathbf{W}_i$, e.g., $(\mathbf{W}_3)_3 = \mathbf{w}_3$ if $\mathbf{W}_1 = [\mathbf{w}_1, \mathbf{w}_3]$. Moreover, the indicator function $\mathcal{I}_+(\mathbf{Y}_i)$ is used to characterize the positive semidefinite constraint, where $\mathcal{I}_+(\mathbf{Y}_i) = 0$ if $\mathbf{Y}_i\succeq\mathbf{0}$ and ∞ otherwise. In (13d) , since $i\in\mathcal{N}_i$, there exists a column index l_i of $\mathbf{X}_i$ (and thus $\mathbf{W}_i$) such that $\mathbf{w}_i$ is the l_ith column of $\mathbf{W}_i$, e.g., $l_3 = 2$ if $\mathbf{W}_3 = [\mathbf{w}_1, \mathbf{w}_3]$.

Let $\mathscr{X}$ denote the set of variables $\{\mathbf{X}_i\}$ and $\{\mathbf{Z}_i\}$, and $\mathscr{Y}$ denote the set of variables $\{\mathbf{W}_i\}$, $\{\mathbf{Y}_i\}$ and $\{\mathbf{w}_i\}$. Also let $\{\mathbf{\Pi}_i\}$, $\{\mathbf{\Gamma}_i\}$, and $\{\boldsymbol{\mu}_i\}$ denote the dual variables (also known as Lagrangian multipliers) associated with the linear constraints (13b)–(13d), respectively. The dual variables are updated by

$$\mathbf{\Pi}_i(t+1) = \mathbf{\Pi}_i(t) + \rho\begin{bmatrix}\mathbf{Z}_i(t+1) & \mathbf{X}_i(t+1)^T\\ \mathbf{X}_i(t+1) & \mathbf{I}\end{bmatrix} - \rho\mathbf{Y}_i(t+1) \quad (14)$$

$$\mathbf{\Gamma}_i(t+1) = \mathbf{\Gamma}_i(t) + \rho\mathbf{X}_i(t+1) - \rho\mathbf{W}_i(t+1) \quad (15)$$

$$\boldsymbol{\mu}_i(t+1) = \boldsymbol{\mu}_i(t) + \rho\mathbf{X}_i(t+1)\mathbf{e}_{l_i} - \rho\mathbf{w}_i(t+1), \quad (16)$$

for $i\in[m]$, where t is the ADMM iteration, and $\rho > 0$ is a regularization parameter. The most important step of ADMM is to alternatively minimize the augmented Lagrangian [27, Sec. 3] of problem (13) over $\mathscr{X}$ and $\mathscr{Y}$ at each iteration. The expression for the augmented Lagrangian is omitted here for the sake of brevity.

By fixing values of variables in $\mathscr{Y}$, we minimize the augmented Lagrangian of problem (13) with respect to variables in $\mathscr{X}$. This leads to a sequence of quadratic programs

$$\begin{aligned}\underset{\mathbf{X}_i,\mathbf{Z}_i}{\text{minimize}} \quad & \text{tr}(\mathbf{S}_i\mathbf{Z}_i) + \tfrac{\rho}{2}\|\mathbf{X}_i - \mathbf{G}_i(t)\|_F^2 + \tfrac{\rho}{2}\|\mathbf{X}_i\mathbf{e}_{l_i} - \mathbf{h}_i(t)\|_2^2 \\ & + \tfrac{\rho}{2}\|\mathbf{Z}_i - \mathbf{F}_i^{11}(t)\|_F^2 + \rho\|\mathbf{X}_i - \mathbf{F}_i^{21}(t)\|_F^2,\end{aligned} \quad (17)$$

for $i\in[m]$, where $\mathbf{G}_i(t) = \mathbf{W}_i(t) - (1/\rho)\mathbf{\Gamma}_i(t)$, $\mathbf{h}_i(t) = \mathbf{w}_i(t) - (1/\rho)\boldsymbol{\mu}_i(t)$, $\begin{bmatrix}\mathbf{F}_i^{11}(t) & (\mathbf{F}_i^{21}(t))^T\\ \mathbf{F}_i^{21}(t) & \mathbf{F}_i^{22}(t)\end{bmatrix} = \mathbf{Y}_i(t) - (1/\rho)\mathbf{\Pi}_i(t)$, and $\|\cdot\|_F$ is the Frobenius norm of a matrix. The solution of problem (17) is explicitly given by

$$\begin{cases}\mathbf{X}_i(t+1) := (3\mathbf{I} + \mathbf{e}_{l_i}\mathbf{e}_{l_i})^{-1}\left(\mathbf{G}_i(t) + \mathbf{h}_i(t)\mathbf{e}_{l_i}^T + 2\mathbf{F}_i^{21}(t)\right)\\ \mathbf{Z}_i(t+1) := \mathbf{F}_i^{11}(t) - (1/\rho)\mathbf{S}_i,\end{cases} \quad (18)$$

which can be computed at each sensor.

By fixing values of variables in $\mathscr{X}$, we minimize the augmented Lagrangian of problem (13) with respect to variables in $\mathscr{Y}$. This leads to a sequence of subproblems with respect to $\mathbf{w}_i$, $\mathbf{Y}_i$ ($\forall i\in[m]$) and

$\{\mathbf{W}_i\}_{i=1}^m$, respectively. That is,

$$\underset{\mathbf{w}_i}{\text{minimize}} \ \|\mathbf{w}_i\|_2 + 1/(2\hat{\gamma})\|\mathbf{w}_i - \hat{\mathbf{h}}_i(t) - (2/\rho)\mathbf{r}_i\|_2^2, i \in [m] \quad (19a)$$

$$\underset{\mathbf{Y}_i}{\text{minimize}} \ \|\mathbf{Y}_i - \hat{\mathbf{F}}_i(t)\|_F^2, \ \text{subject to} \ \mathbf{Y}_i \succeq \mathbf{0}, \qquad i \in [m] \quad (19b)$$

$$\begin{cases} \underset{\{\mathbf{W}_i\}}{\text{minimize}} & \sum_{k=1}^m \sum_{i\in\mathcal{N}_k} \|(\mathbf{W}_i)_k - (\hat{\mathbf{G}}_i(t))_k\|_2^2 \\ \text{subject to} & (\mathbf{W}_i)_k = (\mathbf{W}_j)_k, \ i,j \in \mathcal{N}_k, \ k \in [m], \end{cases} \quad (19c)$$

where $\hat{\gamma} = \gamma/\rho$, $\hat{\mathbf{h}}_i(t) = \mathbf{X}_i(t+1)\mathbf{e}_{l_i} + (1/\rho)\boldsymbol{\mu}_i(t)$, $\hat{\mathbf{F}}_i(t) = \begin{bmatrix} \mathbf{Z}_i(t+1) & \mathbf{X}_i(t+1)^T \\ \mathbf{X}_i(t+1) & \mathbf{I} \end{bmatrix} + (1/\rho)\boldsymbol{\Pi}_i(t)$, and $\hat{\mathbf{G}}_i(t) = \mathbf{X}_i(t+1) + (1/\rho)\boldsymbol{\Gamma}_i(t)$.

The solution of problem (19a) is given by a block soft thresholding operator [37, Sec. 6.5]

$$\mathbf{w}_i(t+1) := (1 - \hat{\gamma}/\|\hat{\mathbf{h}}_i(t) + (2/\rho)\mathbf{r}_i\|_2)_+(\hat{\mathbf{h}}_i(t) + (2/\rho)\mathbf{r}_i), \quad (20)$$

where $(x)_+ = \max\{x, 0\}$.

The solution of problem (19b) is given by

$$\mathbf{Y}_i(t+1) = \sum_{j=1}^{n+|\mathcal{N}_i|} (\sigma_j(t))_+ \boldsymbol{\omega}_j(t)\boldsymbol{\omega}_j(t)^T, \quad (21)$$

where $\sum_{j=1}^{n+|\mathcal{N}_i|} \sigma_j(t)\boldsymbol{\omega}_j(t)\boldsymbol{\omega}_j(t)^T$ is the eigenvalue decomposition of $\hat{\mathbf{F}}_i(t)$.

Problem (19c) is separate over k and yields an average consensus solution [39] $(\mathbf{W}_i(t+1))_k = \frac{\sum_{j\in\mathcal{N}_k}(\hat{\mathbf{G}}_j(t))_k}{|\mathcal{N}_k|}$. Together with (15), we obtain $\sum_{i\in\mathcal{N}_k}(\boldsymbol{\Gamma}_i(t+1))_k = \mathbf{0}$. Therefore, the solution of problem (19c) becomes

$$[\mathbf{W}_i(t+1)]_k = \frac{1}{|\mathcal{N}_k|}\sum_{j\in\mathcal{N}_k}[\mathbf{X}_j(t+1)]_k, \ k \in [m], \ i \in \mathcal{N}_k. \quad (22)$$

The computation in (20) and (21) is easily performed in a distributed manner since it is separable over sensors. By contrast, inter-sensor communications are required in (22) to compute the average consensus value within a sensor's neighborhood.

B. Computational complexity and communication cost

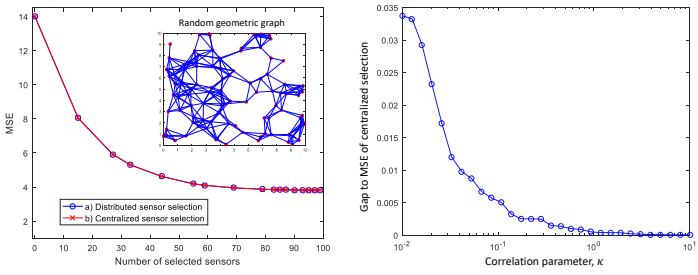
ADMM has a linear convergence rate for general convex optimization problems [40]. At each iteration of the proposed ADMM described above, the computational complexity is dominated by the eigenvalue decomposition in (21), which yields the complexity $O((|\mathcal{N}_i|+n)^{3.5})$ for $i \in [m]$. Since it is often the case that $n \ll m$, $|\mathcal{N}_i| \ll m$ and sensors can compute in parallel, the computational complexity of the distributed algorithm is lower than that of the centralized algorithm with complexity $O((n+m)^{3.5})$. The total number of in-network communications is given by $\sum_{k=1}^m 2(|\mathcal{N}_k|-1)$, where we require $|\mathcal{N}_k| - 1$ hops to compute the average consensus value in (22), and the other $|\mathcal{N}_k| - 1$ communications to broadcast this average in the neighborhood of each sensor.

V. Numerical Results

This section empirically shows the effectiveness of the proposed distributed sensor selection approach. We consider a sensor network with $m = 100$ sensors over a 10×10 square region. The spatial placement of sensors is modeled by a random geometric graph [41]. We estimate the field intensities at $n = 20$ unobserved locations randomly chosen from the square region. The spatial correlation of the random field is given by $\text{cov}(\phi(\mathbf{a}), \phi(\mathbf{b})) = \sigma_\phi^2 e^{-\kappa\|\mathbf{a}-\mathbf{b}\|_2^2}$, where $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ are locations of field points, $\sigma_\phi^2 = 1$, and κ is a parameter to govern the strength of spatial correlation, which increases (or decreases) as κ decreases (or increases). In our numerical examples unless specified otherwise, we set $\kappa = 0.5$ that corresponds to a weak correlation. We obtain the sensor selection scheme under two scenarios: a) the proposed ADMM-based approach to solve problem (10); b) convex optimization to solve the centralized sensor selection problem (6). In our numerical examples, ADMM converges to satisfactory accuracy within 500 iterations, where we set $\rho = 10$ and choose random points drawn from a uniform distribution to initialize ADMM. Given a sensor selection scheme, the estimation performance is measured through the empirical mean squared error (MSE) that is computed over 1000 numerical trials.

In Fig. 2, we compare the MSE of using the proposed distributed sensor selection approach to that of using the centralized sensor selection strategy. To be specific, in Fig. 2-(a), we present the MSE as a function of the number of selected sensors by varying the sparsity-promoting parameter γ in problems (6) and (10). Recall that the column cardinality of the estimator gain matrix gives the number of selected sensors. As we can see, the proposed distributed sensor selection approach achieves almost the same estimation performance as the centralized sensor selection method. As γ decreases, more sensors are activated, and thus the estimation performance improves. This renders the tradeoff between the conflicting objectives of good estimation performance and minimal sensor usage.

In Fig. 2-(b), we present the difference of MSE (averaged over different numbers of selected sensors) between the distributed approach being considered and the centralized selection method as a function of the correlation parameter κ. As we can see, when κ becomes large (corresponding to weak correlation), the performance gap tends to be zero. Conversely, it increases as the correlation becomes strong (namely, κ decreases). This is not surprising, since as indicated by Proposition 1, the proposed distributed sensor selection problem (10) is equivalent to the centralized sensor selection problem (6) when the correlation is weak.



(a) (b)

Fig. 2: Performance evaluation of distributed sensor selection: a) MSE versus number of selected sensors; b) Gap to MSE of using the centralized solution.

VI. Conclusions

In this paper, we proposed a decentralized architecture to design optimal sensor selection schemes for field estimation. We provided a reasonable way to decompose the global selection cost into local cost functions that are amenable for distributed optimization. To solve the resulting optimization problem, we used semidefinite relaxation to circumvent the issue of nonconvexity, and employed ADMM for distributed optimization. In future work, we would like to provide a deeper investigation on the tightness of the used semidefinite relaxation. We also would like to investigate the loss of performance of the proposed myopic formulation of the sensor selection problem relative to optimal multi-stage sensor selection schemes.

REFERENCES

[1] M. C. Vuran, O. B. Akan, and I. F. Akyildiz, "Spatio-temporal correlation: theory and applications for wireless sensor networks," *Computer Networks*, vol. 45, no. 3, pp. 245–259, June 2004.

[2] Phaeton C. Kyriakidis, "A spatial time series framework for modeling daily precipitation at regional scales," *Journal of Hydrology*, vol. 297, no. 4, pp. 236 – 255, Apr. 2004.

[3] A. O. Hero and D. Cochran, "Sensor management: Past, present, and future," *IEEE Sensors Journal*, vol. 11, no. 12, pp. 3064–3075, Dec. 2011.

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

[5] Y. Weng, L. Xie, and W. Xiao, "Sensor selection for random field estimation in wireless sensor networks," *Journal of Systems Science and Complexity*, vol. 25, no. 1, pp. 46–59, 2012.

[6] 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.

[7] 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.

[8] S. P. Chepuri and G. Leus, "Sensor selection for estimation, filtering, and detection," in *Proc. of International Conference on Signal Processing and Communications (SPCOM), 2014*, July 2014, pp. 1–5.

[9] V. Roy and G. Leus, "Correlation-aware sparsity-enforcing sensor placement for spatio-temporal field estimation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 2389–2393.

[10] 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.

[11] M. Calvo-Fullana, J. Matamoros, and C. Antón-Haro, "Sensor selection in energy harvesting wireless sensor networks," in *Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2015, pp. 43–47.

[12] S. P. Chepuri and G. Leus, "Sparse sensing for distributed detection," *IEEE Transactions on Signal Processing*, vol. 64, no. 6, pp. 1446–1460, 2015.

[13] D. Golovin, M. Faulkner, and A. Krause, "Online distributed sensor selection," in *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2010, pp. 220–231.

[14] H. Jamali-Rad, A Simonetto, and G. Leus, "Sparsity-aware sensor selection: Centralized and distributed algorithms," *IEEE Signal Processing Letters*, vol. 21, no. 2, pp. 217–220, Feb 2014.

[15] H. Jamali-Rad, A. Simonetto, X. Ma, and G. Leus, "Distributed sparsity-aware sensor selection," *IEEE Transactions on Signal Processing*, vol. 63, no. 22, pp. 5951–5964, Nov 2015.

[16] F. Altenbach, S. Corroy, G. Böcherer, and R. Mathar, "Strategies for distributed sensor selection using convex optimization," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Dec. 2012, pp. 2367–2372.

[17] S. T. Jawaid and S. L. Smith, "Submodularity and greedy algorithms in sensor scheduling for linear dynamical systems," *Automatica*, vol. 61, pp. 282 – 288, 2015.

[18] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 355–369, Jan. 2009.

[19] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 674–690, Aug. 2011.

[20] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "On distributed averaging algorithms and quantization effects," *IEEE Transactions on Automatic Control*, vol. 54, no. 11, pp. 2506–2517, Nov. 2009.

[21] S. Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6217–6234, Dec. 2012.

[22] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad hoc wsns with noisy links – part i: Distributed estimation of deterministic signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 350–364, Jan. 2008.

[23] I. D. Schizas, G. B. Giannakis, S. I. Roumeliotis, and A. Ribeiro, "Consensus in ad hoc wsns with noisy links – part ii: Distributed estimation and smoothing of random signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1650–1666, April 2008.

[24] R. Olfati-Saber, "Distributed kalman filtering for sensor networks," in *Proc. 46th IEEE Conference on Decision and Control*, Dec. 2007, pp. 5492–5498.

[25] S. Pequito, V. Preciado, and G. J. Pappas, "Distributed leader selection," in *Proc. 54th IEEE Conference on Decision and Control (CDC)*, Dec. 2015, pp. 962–967.

[26] M. Doostmohammadian and U. A. Khan, "On the genericity properties in distributed estimation: Topology design and sensor placement," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 2, pp. 195–204, April 2013.

[27] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[28] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Transactions on Automatic Control*, vol. 57, no. 3, pp. 592–606, March 2012.

[29] S. Sundhar Ram, A. Nedić, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of Optimization Theory and Applications*, vol. 147, no. 3, pp. 516–545, 2010.

[30] D. Blatt, A. O. Hero, and H. Gauchman, "A convergent incremental gradient method with a constant step size," *SIAM Journal on Optimization*, vol. 18, no. 1, pp. 29–51, 2007.

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

[32] S. Liu, F. Chen, A. Vempaty, M. Fardad, L. Shen, and P. K. Varshney, "Sparsity-promoting sensor management for estimation: An energy balance point of view," in *Proc. 18th International Conference on Information Fusion (Fusion)*, July 2015, pp. 231–238.

[33] 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. 3948–3952.

[34] Z. Meng, D. Wei, A. Wiesel, and A. O. Hero, "Marginal likelihoods for distributed parameter estimation of gaussian graphical models," *IEEE Transactions on Signal Processing*, vol. 62, no. 20, pp. 5425–5438, Oct. 2014.

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

[36] S. Liu, S. P. Chepuri, G. Leus, and A. O. Hero, "Distributed sensor selection for field estimation," *https://www.researchgate.net/publication/308305528_Distributed_Sensor_Selection_for_Field_Estimation*.

[37] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2013.

[38] 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.

[39] G. Mateos, J. A. Bazerque, and G. B. Giannakis, "Distributed sparse linear regression," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5262–5276, Oct. 2010.

[40] B. He and X. Yuan, "On the $O(1/n)$ convergence rate of the douglas-rachford alternating direction method," *SIAM Journal on Numerical Analysis*, vol. 50, no. 2, pp. 700–709, 2012.

[41] M. Penrose, Random Geometric Graphs, Oxford University Press, New York, 2003.