Sparsity-Aware Sensor Selection for Correlated Noise

Hadi Jamali-Rad, Andrea Simonetto, Geert Leus
Faculty of EEMCS
Delft University of Technology (TU Delft)
2628CD Delft, The Netherlands
emails:{h.jamalirad, a.simonetto, g.j.t.leus}@tudelft.nl

Xiaoli Ma
School of ECE
Georgia Institute of Technology (GaTech)
Atlanta, Georgia 30332–0250, USA
email: xiaoli@gatech.edu

***Abstract*—The selection of the minimum number of sensors within a network to satisfy a certain estimation performance metric is an interesting problem with a plethora of applications. We have recently explored the sparsity embedded within this problem and have proposed a relaxed sparsity-aware sensor selection (SparSenSe) approach as well as a distributed version of it. In this paper, we generalize our recently proposed sensor selection paradigm to be able to operate even in cases where the measurement noise experienced by the sensors is correlated. We derive the related centralized and distributed algorithms and analyze them in terms of their computational and communication complexities. We also provide general remarks on the convergence of our proposed distributed algorithm. Our simulation results corroborate our claims and illustrate a promising performance for the proposed centralized and distributed algorithms.**

***Index Terms*—Distributed estimation, sensor selection, sparse reconstruction.**

I. Introduction

We study the problem of selecting the minimum number of sensors among a network of sensor nodes in order to estimate a vector of interest so that a given mean squared error (MSE) is satisfied. This problem is of great interest in several practical application domains including robotics, target tracking, and energy efficient network management, to name a few (see for instance [1] and references therein). A straightforward method to solve such a problem is a combinatorial approach considering all possible combinations of all possible sizes of candidate sensors to satisfy the constraint, which is numerically intractable for a large number of sensors and thus motivates a more intelligent and structured approach. The problem becomes even more challenging when a distributed context is considered.

A related sensor selection problem has been studied in [1] where elegant convex relaxations are designed for primal and dual problems. However, instead of optimizing different performance metrics and fixing the number of sensors as in [1], we minimize the number of sensors given a performance constraint, which is generally more practical. Interestingly, this enables us to exploit the sparsity embedded within the problem. From this angle, our approach is closer to what is proposed in [2] for selecting reliable sensors, also called "robust sensing". However, we consider a different constraint than the one in [2], and we do not need the sensors to take measurements for solving the selection problem; we only need them to know their regression coefficients. Also, in both [1] and [2], a distributed approach has not been considered.

A decentralized implementation of [1] is proposed in [3]; however, the heuristic assumption of two "leader" nodes violates the classical definition of a distributed approach. In [4], two distributed implementations of [1] based on a truncated Newton algorithm are proposed. In [5] we have explored the sparsity embedded within the problem and have propose a relaxed sparsity-aware sensor selection approach called SparSenSe. We have also presented a reasonably low-complexity and elegant distributed version of SparSenSe, called DisParSense, such that each sensor can decide itself whether it should contribute to the estimation or not. Compared to [5], the work of [4] deals with a slightly different problem and also requires the private sensor information to be broadcast whereas the proposed approach in [5] avoids that. Moreover, the distributed approach of [5] is considerably more efficient in terms of complexity compared to [4]. Finally, another relevant problem, but of a different nature, is considered in [6], where a distributed algorithm is designed to identify the sensors containing relevant information by a sparsity-aware decomposition of the measurement covariance matrix.

In [5], we have only considered the case where the noise experienced by the sensors is uncorrelated. This might be a justifiable assumption in some cases, but in general the experienced noise of the sensors can be correlated. Particularly, as it is pointed out in [7], [8], since the measurement noises of different sensors may depend on a common "estimatee" (as is the case in our problem formulation), the sensors can observe correlated noise. Another example occurs where the estimatee is observed by sensors in a common noisy environment, such as noise generated by a jammer. In such cases the measurement noises of the sensors are often correlated. This motivates us to extend our previously proposed algorithms to be able to operate in a more practical (and more general) framework of correlated noise.

The rest of this paper is organized as follows. In Section II,

This work was supported in part by NWO-STW under the VICI program (10382) and in part by STW under the D2S2 project from the ASSYS program (Project 10561). X. Ma was supported in part by NSF Grant No. ECCS-1202286.

the problem of interest is defined and our goal is stated. In Section III, the centralized problem is explained and a relaxed convexified solution is provided. Section IV is devoted to a distributed derivation of the centralized problem where a convergence analysis is briefly summarized. In Section V, the computational and communication costs involved in the proposed algorithms are investigated and compared. Finally, in Section VII, some important features of the proposed distributed algorithm are discussed and the paper is concluded by highlighting our future research directions.

II. Problem Definition

We consider m sensor nodes distributed over an area of interest in $\mathbb{R}^d$, with $d \leq m$, which are supposed to estimate the unknown vector $\mathbf{x} \in \mathbb{R}^n$. The sensor nodes are equipped with (limited) computational and communication capabilities and each of them measures

$$y_i = \mathbf{a}_i^T \mathbf{x} + \eta_i, \ i = 1, \ldots, m, \tag{1}$$

where the $\mathbf{a}_i$'s $\in \mathbb{R}^n$ span $\mathbb{R}^n$ ($m \gg n$) and the η_i's are the noise experienced by different sensors which are considered to be zero-mean and "*correlated*". Note that, considering the spatial distribution of the sensors, we assume that the $\mathbf{a}_i$'s are different so that we can distinguish the sensors based on their regressors. Here, we are interested in selecting *a priori* the minimum number of sensors (namely, measurements) so that the mean squared error (MSE) of estimating $\mathbf{x}$ is smaller than a desired value γ. Furthermore, we are interested in algorithms that would enable the sensors themselves to decide their own active/inactive status, without a centralized collection of the $\mathbf{a}_i$ vectors, i.e., we are interested in *distributed* algorithms.

III. Centralized Optimization Problem

In a centralized setup, all $\mathbf{a}_i$'s are available in a central unit which permits us to define the matrix $\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_m]^T$. Now, we can construct $\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\eta}$, where $\mathbf{y} = [y_1, \cdots, y_m]^T$, and $\boldsymbol{\eta} = [\eta_1, \cdots, \eta_m]^T$. Note that, by considering correlated noise $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \ \mathbf{C})$, where $\mathbf{C}$ by definition is a symmetric and positive semidefinite (PSD) matrix [9]. For the linear measurement model (1) and the maximum likelihood estimator, the MSE can be expressed as

$$\text{MSE} = \mathbb{E}\left[\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2\right] = \text{tr}\left((\mathbf{A}^T \, \mathbf{C}^{-1} \, \mathbf{A})^{-1}\right),$$

where tr(.) stands for the trace operator. Notably, different from the case of uncorrelated noise in [5], here $\mathbf{C}$ is not diagonal and can even be a full matrix due to the correlated noise assumption. The non-diagonal elements $[\mathbf{C}]_{ij}$, $i \neq j$, should also be incorporated within our selection procedure. In order to handle the non-diagonal elements, we define a symmetric PSD selection matrix $\mathbf{W} = \mathbf{w}\mathbf{w}^T$, where $\mathbf{w} = [w_1, \ldots, w_m]^T$ is the selection vector (similar to [5]) and the variable $w_i \in \{0, 1\}$ encodes whether the i-th sensor (measurement) is to be used. Notice that based on this new definition of $\mathbf{W}$, $[\mathbf{C}]_{ij}$ will only be incorporated if both w_i and w_j are non-zero at the same time. The associated selection constraint on the MSE can then be stated as

$$\text{tr}\left((\mathbf{A}^T[\mathbf{W} \odot \mathbf{C}^{-1}]\mathbf{A})^{-1}\right) \leq \gamma, \tag{2}$$

where $\odot$ stands for the Hadamard product. In practice, only a few sensors should be activated to satisfy the MSE constraint, which triggers the idea of exploiting the sparsity embedded within the problem. Note that since $w_i \in \{0, 1\}$, we have $\text{diag}(\mathbf{W}) = \mathbf{w}$, where diag(.) returns a vector containing the diagonal elements. Thus, the problem can be cast as the following optimization program

$$\underset{\mathbf{W}, \mathbf{u}}{\text{minimize}} \quad \|\text{diag}(\mathbf{W})\|_0 \tag{3a}$$

$$\text{s.t.} \quad \left[\begin{array}{c|c} \mathbf{A}^T[\mathbf{W} \odot \mathbf{C}^{-1}]\mathbf{A} & \mathbf{e}_j \\ \hline \mathbf{e}_j^T & u_j \end{array}\right] \succeq 0, \ j = 1, \ldots, n, \tag{3b}$$

$$||\mathbf{u}||_1 \leq \gamma, \ u_j \geq 0, \ j = 1, \ldots, n, \tag{3c}$$

$$[\mathbf{W}]_{i,j} \in \{0, 1\}, \ \mathbf{W} \in \mathbb{S}_+^m, \ \text{rank}(\mathbf{W}) = 1, \tag{3d}$$

where $\mathbf{u} = [u_1, \ldots, u_n]^T$ is a vector of auxiliary variables, $\mathbf{e}_j$ is the j-th column of the $n \times n$ identity matrix $\mathbf{I}_n$, and $\mathbb{S}_+^m = \{\mathbf{X} \in \mathbb{R}^{m \times m} | \mathbf{X} \succeq 0, \mathbf{X}^T = \mathbf{X}\}$ denotes the set of symmetric PSD matrices. The constraints (3b) and (3c) are a more suitable representation of the original constraint (2), obtained using the Schur complement [10]. We denote the solution to (3) as $(\mathbf{W}^*, \mathbf{u}^*)$. Clearly, (3) is non-convex due to its objective (ℓ_0 norm), and the first and the third terms in (3d) (finite-alphabet constraint on the elements of $\mathbf{W}$ and rank-1 constraint, respectively). In order to define a convex problem, we relax these three non-convex terms as

$$\underset{\mathbf{W}, \mathbf{u}}{\text{minimize}} \quad \text{tr}(\mathbf{W}) \tag{4a}$$

$$\text{s.t.} \quad \left[\begin{array}{c|c} \mathbf{A}^T[\mathbf{W} \odot \mathbf{C}^{-1}]\mathbf{A} & \mathbf{e}_j \\ \hline \mathbf{e}_j^T & u_j \end{array}\right] \succeq 0, \ j = 1, \ldots, n, \tag{4b}$$

$$||\mathbf{u}||_1 \leq \gamma, \ u_j \geq 0, \ j = 1, \ldots, n, \tag{4c}$$

$$0 \leq [\mathbf{W}]_{i,j} \leq 1, \ \mathbf{W} \in \mathbb{S}_+^m. \tag{4d}$$

We call this algorithm SparSenSe-C to distinguish it from the algorithm previously developed for uncorrelated noise, and we denote its solution as $(\hat{\mathbf{W}}, \hat{\mathbf{u}})$. A final step to recover $\hat{\mathbf{w}}$ from $\hat{\mathbf{W}}$ would be to apply a Choleskey decomposition and a possible randomization procedure to compensate for the relaxed rank-1 constraint. Alternatively, we can simply consider $\hat{\mathbf{w}} = \text{diag}(\hat{\mathbf{W}})$, which is what we do in this paper.

IV. Distributed algorithm

Triggered by the localized nature of many phenomena of interest in practical applications, in this section, we develop a distributed version of the centralized approach proposed earlier. In practice, the noise correlation follows a regional (localized) pattern. This means $\mathbf{C}$ will not be a full matrix but there will be correlation regions within which the noise experienced by the sensors is correlated and the cross-correlations of the noise experienced by these regions would be orders of magnitude smaller than the correlation within each region. Let

$$\mathcal{L} = \sum_{l=1}^{C}\sum_{i\in\mathcal{M}_l^c} w_{ii} - \sum_{j=1}^{n}\text{tr}\left(\left[\begin{array}{c|c} \sum_{l=1}^{C}\sum_{i\in\mathcal{M}_l^c}\sum_{k\in\mathcal{M}_l^c} w_{ik}c_{ik}^{-1}\mathbf{a}_i\mathbf{a}_k^T & \mathbf{e}_j \\ \hline \mathbf{e}_j^T & u_j \end{array}\right]\mathbf{G}_j\right)$$
$$= \sum_{l=1}^{C}\left(\sum_{i\in\mathcal{M}_l^c} w_{ii} - \sum_{j=1}^{n}\text{tr}\left(\left[\begin{array}{c|c} \sum_{i\in\mathcal{M}_l^c}\sum_{k\in\mathcal{M}_l^c} w_{ik}c_{ik}^{-1}\mathbf{a}_i\mathbf{a}_k^T & \mathbf{e}_j/C \\ \hline \mathbf{e}_j^T/C & u_j/C \end{array}\right]\mathbf{G}_j\right)\right) = \sum_{l=1}^{C}\mathcal{L}_l(\mathbf{W}_l,\mathbf{u},\mathbf{G}). \quad (6)$$

us call these regions "clusters".

Assumption 1: For the sake of simplicity of our next derivations, we assume that the inter-cluster noise cross-correlations are negligible. As a result, we can rearrange the rows (and columns) of $\mathbf{C}$ *so that we have a block-diagonal matrix. Evidently,* $\mathbf{C}$ *will still be symmetric and PSD as each of its blocks will own these properties.*

Let us now define some notations. We consider a total of C clusters and we call the set of sensors in the l-th cluster $\mathcal{M}_l^c$ with cardinality $|\mathcal{M}_l^c| = M_l^c$. Besides, we call the neighborhood set of the l-th cluster including l itself $\mathcal{N}_l^c$, with cardinality $|\mathcal{N}_i^c| = N_i^c$. Furthermore, we denote the (i,j)-th element of $\mathbf{C}$, $\mathbf{C}^{-1}$ and $\mathbf{W}$ by c_{ij}, c_{ij}^{-1} and w_{ij}, respectively. This helps us to expand (2) as

$$\text{tr}\Big(\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}\, c_{ij}^{-1}\,\mathbf{a}_i\,\mathbf{a}_j^T\Big)^{-1} \le \gamma,$$

and form the Lagrangian of (4a)-(4b) given by

$$\mathcal{L} = \sum_{i=1}^{m} w_{ii} - \sum_{j=1}^{n}\text{tr}\left(\left[\begin{array}{c|c} \sum_{i=1}^{m}\sum_{k=1}^{m} w_{ik}c_{ik}^{-1}\mathbf{a}_i\mathbf{a}_k^T & \mathbf{e}_j \\ \hline \mathbf{e}_j^T & u_j \end{array}\right]\mathbf{G}_j\right), \quad (5)$$

where $\mathbf{G}_j \succeq 0,\ \forall j$ are appropriately sized dual variables, and $\mathbf{G} = [\mathbf{G}_1,\ldots,\mathbf{G}_n]$. Note that different from the derivations in [5], we cannot simply decompose (5) with respect to (w.r.t.) w_{ii}'s because there exist some coupling terms $w_{ik},\ i \neq k$.

A possible solution and the idea behind our next derivations is to look at the problem "*cluster-wise*" and rewrite (5) as (6) (shown on top of this page). Note that $\mathbf{W}$ is rearranged so that the rows (and columns) corresponding to members of each cluster are placed within one subblock. This means $\mathbf{W}_l$ is the subblock of $\mathbf{W}$ corresponding to the l-th cluster, and therefore,

$$\sum_{i\in\mathcal{M}_l^c} w_{ii} = \text{tr}(\mathbf{W}_l).$$

We also define the following convex sets:

$$\mathcal{W}_l = \{\mathbf{W}_l \mid 0 \le w_{l,ik} \le 1, \forall i,k \in \mathcal{M}_l^c, \mathbf{W}_l \in \mathbb{S}_+^m\}, \quad (7)$$

$$\mathcal{U} = \{\mathbf{u} \mid u_j \ge 0,\ \sum_{j=1}^{n} u_j \le \gamma\}, \quad (8)$$

where $w_{l,ik}$ denotes the (i,k)-th element of $\mathbf{W}_l$. Now, the dual function defined on $\mathcal{L}$ can be given by

$$q(\mathbf{G}) = \min_{\mathbf{W}_l\in\mathcal{W}_l,\mathbf{u}\in\mathcal{U}} \sum_{l=1}^{C}\mathcal{L}_l(\mathbf{W}_l,\mathbf{u},\mathbf{G})$$
$$= \sum_{l=1}^{C}\left(\min_{\mathbf{W}_l\in\mathcal{W}_l,\mathbf{u}\in\mathcal{U}} \mathcal{L}_l(\mathbf{W}_l,\mathbf{u},\mathbf{G})\right) = \sum_{l=1}^{C} q_l(\mathbf{G}).$$

Notably, since both $\mathcal{W}_l$ and $\mathcal{U}$ are convex sets, given a certain value of $\mathbf{G}$, the functions $q_l(\mathbf{G})$ and their subgradient w.r.t. $\mathbf{G}$, called $\mathbf{Q}$ and defined later on, can be computed locally, for instance, using CVX [11] to solve the resulting semidefinite program (SDP) at each cluster [12]. This again mandates performing cluster-wise operations, where all the information of the sensors in each cluster should be collected and processed somewhere, for instance, in a "data-gathering node" or "cluster head". We simply consider one of the sensors in each cluster as the cluster head.

Whenever γ is large enough so that we expect sparse solutions in terms of $\hat{\mathbf{w}} = \text{tr}(\hat{\mathbf{W}})$, Slater's condition holds for (4), which can be formulated as the following proposition.

Proposition 1: Slater's condition holds for (4), for sufficiently large γ.

Proof: For sufficiently large γ, we can always find a pair $(\mathbf{W},\ \mathbf{u})$ that satisfies (4b) - (4d) strictly. □

Therefore, the relaxed ℓ_1-regularization (4) leads to the dual optimization problem

$$\underset{\mathbf{G}_1\succeq 0,\ldots,\mathbf{G}_n\succeq 0}{\text{maximize}} \sum_{l=1}^{C} q_l(\mathbf{G}), \quad (9)$$

with *zero* duality gap. This convex optimization program can be solved iteratively in a distributed fashion using a few possible algorithms. For instance, we can use subgradient-based methods, such as the dual averaging scheme of [13] with a variable stepsize, or the simpler dual subgradient of [12] with a fixed stepsize. The latter method has the advantage of providing a recovery mechanism for the primal solution, i.e., we recover $\hat{\mathbf{W}}$ and hence $\hat{\mathbf{w}}$ as a by-product of the optimal $\mathbf{G}$, which is in fact our goal. That is why we opt to employ the dual subgradient method of [12].

However, in order to implement the dual subgradient of [12], each cluster requires a copy of the subgradient of $q(\mathbf{G})$ w.r.t.

$\mathbf{G}_j$, defined as

$$\mathbf{Q}_j = -\sum_{l=1}^{C} \left[\begin{array}{c|c} \sum_{i\in\mathcal{M}_l^c}\sum_{k\in\mathcal{M}_l^c} \bar{w}_{ik} c_{ik}^{-1} \mathbf{a}_i \mathbf{a}_k^T & \mathbf{e}_j/C \\ \hline \mathbf{e}_j^T/C & \bar{u}_j/C \end{array} \right]. \tag{10}$$

where $\bar{w}_{ik}$'s and $\bar{u}_j$'s are the optimizers of

$$q(\mathbf{G}) = \min_{\mathbf{W}_l\in\mathcal{W}_l, \mathbf{u}\in\mathcal{U}} \sum_{l=1}^{C} \mathcal{L}_l(\mathbf{W}_l, \mathbf{u}, \mathbf{G}). \tag{11}$$

The need to this "global" parameter can be circumvented by using the method of [14] where the clusters have different local copies of $\mathbf{G}$ and $\mathbf{Q}$, say $\mathbf{G}^l$ and $\mathbf{Q}^l$, and they run an inexact consensus procedure for φ times (where $\varphi \in \mathbb{N}_+$). In particular, to solve (9), we will consider the following inexact subgradient update. We call the l-th cluster version of $\mathbf{G}$ at iteration t, $\mathbf{G}^{l,t}$. We start with a given initial condition $\mathbf{G}_j^{l,0}$ for each clusters, then for each $t \geq 0$

$$\mathbf{V}_j^{l,\tau=0,t} = \mathbf{G}_j^{l,t} + \alpha C \mathbf{Q}_j^{l,t}, \text{for } j = 1, \cdots, n, \tag{12}$$

where α is the stepsize. Next, we run φ times a consensus procedure as

$$\mathbf{V}_j^{l,\tau,t} = \sum_{p=1}^{C} [\mathbf{Z}]_{lp} \mathbf{V}_j^{p,\tau-1,t}, \tag{13}$$

and we finish with the standard projection over the cone of PSD matrices as

$$\mathbf{G}_j^{l,t+1} = \mathcal{P}_{\succeq 0}\left[\mathbf{V}_j^{l,\varphi,t}\right], \text{for } j = 1, \cdots, n. \tag{14}$$

In (13), $\mathbf{Z} \in \mathbb{R}^{C\times C}$ indicates a proper cluster-wise consensus matrix whose weights have been defined using a Metropolis weighting, i.e.,

$$[\mathbf{Z}]_{lp} = \begin{cases} 1/(\max\{N_l^c, N_p^c\}) & \text{if } p \in \mathcal{N}_l^c \\ 0 & \text{if } p \notin \mathcal{N}_l^c, p \neq l \\ 1 - \sum_{p=1}^{C} [\mathbf{Z}]_{lp} & \text{if } l = p. \end{cases}$$

If we execute (13) for $\varphi \to \infty$, we recover the procedure of [12], whereas if φ is limited we introduce an additional error in the distributed optimization procedure. Our proposed distributed SparSenSe-C (DiSparSenSe-C) algorithm can be summarized in Algorithm 1.

Algorithm 1 DiSparSenSe-C

1: Let an initial value for $\mathbf{G}^{l,t}$ be given at each cluster head. Initialize the $\hat{\mathbf{W}}_l^t$'s with $\hat{\mathbf{W}}_l^0 = 0$.

2: Compute, in parallel at each cluster l, the value of $q_l(\mathbf{G}^{l,t})$ as in (11), its derivative $\mathbf{Q}^{l,t} = \nabla_{\mathbf{G}^{l,t}} q_l(\mathbf{G}^{l,t})$ as in (10), and the related optimal primal variable $\bar{\mathbf{W}}_l^t$. This is actually an SDP problem. Note that the dimension of $\mathbf{Q}^{l,t}$ is the same as that of $\mathbf{G}^{l,t}$.

3: Following the primal recovery method of [12], compute

$$\hat{\mathbf{W}}_l^t = \hat{\mathbf{W}}_l^{t-1}(t-1)/t + \bar{\mathbf{W}}_l^t/t.$$

4: For $\tau = 1$ to φ do

- Send $\mathbf{G}^{l,t}$ and $\mathbf{Q}^{l,t}$ to the neighboring cluster heads;
- Perform, in parallel, one consensus step as

$$\mathbf{V}_j^{l,\tau,t} = \sum_{p=1}^{C} [\mathbf{Z}]_{lp} \mathbf{V}_j^{p,\tau-1,t},$$

which is initialized as in (12).

5: Update each cluster's dual variable

$$\mathbf{G}_j^{l,t+1} = \mathcal{P}_{\succeq 0}\left[\mathbf{V}_j^{l,\varphi,t}\right].$$

6: Go to 2 for the next iteration.

A. Convergence Analysis of DiSparSenSe-C

We would like to highlight that DiSparSenSe-C will converge to the solution of SparSenSe-C with an error floor dependent on α and φ. This can be proven extending the ϵ-subgradient argument discussed in [12] and [14], as is briefly summarized in this subsection. The detailed derivations of the complete proof are omitted for brevity and interested readers are referred to our extended work [15].

In short, since the sets $\mathcal{W}_l$ and $\mathcal{U}$ in (7) and (8) are compact, the subgradient $\mathbf{Q}_j^{l,t}$ is bounded for all l and t. Let us denote such a finite bound by Q, and let $\hat{q}$ be the optimal dual value solution of SparSense-C in (9). In [15], we investigate both primal and dual convergence problems. For the latter, we prove that there exists a finite $\varphi \geq \bar{\varphi} > 0$, for which the sequence of dual functions $\{q(\mathbf{G}^{l,t})\}$ generated by DiSparSenSe-C converges as

$$\limsup_{t\to\infty} q(\mathbf{G}^{l,t}) \geq \hat{q} - C\psi_1(\alpha, Q, \varphi), \ l = 1, \cdots, C,$$

where ψ_1 is a positive function of φ, α, and Q. Besides, we know that under Slater's condition, the dual variables lie in a bounded compact set (comprising the zero element) and,

$$||\mathbf{G}_j^{l,t}|| \leq G, \ j = 1, \cdots, n, \ l = 1, \cdots, C, \ t \geq 0,$$

for a certain finite positive constant G. Based on this, we prove in [15] that the convergence of the running average primal sequence $\{\hat{\mathbf{W}}^t, \hat{\mathbf{u}}^t\}$ (as defined in step 3 of the algorithm) can be formulated in terms of a constraint violation, and an upper and lower bound on the primal function. The results in [15] show that the running average primal function is bounded above as

$$\text{tr}(\hat{\mathbf{W}}^t) \leq \text{tr}(\hat{\mathbf{W}}) + \frac{(nCG)^2}{2\alpha C t} + \frac{\alpha C^3}{2} Q^2 n^2,$$

and it is bounded below as

$$\text{tr}(\hat{\mathbf{W}}^t) \geq \text{tr}(\hat{\mathbf{W}}) - nG\left(\frac{G}{\alpha t} + \frac{\psi_2(\alpha, Q, \mathbf{Z}, \varphi)}{\alpha}\right),$$

where ψ_2 is a positive function of φ, α, $\mathbf{Z}$, and Q. These lower and upper bounds on the primal function indicate a convergence rate of $O(1/t)$ for the running average primal sequence to a bounded region around the optimal primal cost $\text{tr}(\hat{\mathbf{W}})$. The width of this region depends on α and φ.

TABLE I: Complexity Order Comparision

Algorithm	Comp. complexity	Comm. complexity
SparSenSe	$O(m^3)$	–
SparSenSe-C	$O(m^3)$	–
DiSparSenSe	$O(n^4)$	$O(\varphi N_i n^3)$
DiSparSenSe-C	$O((M_l^c)^3 n^4)$	$O(\varphi N_l^c n^3)$

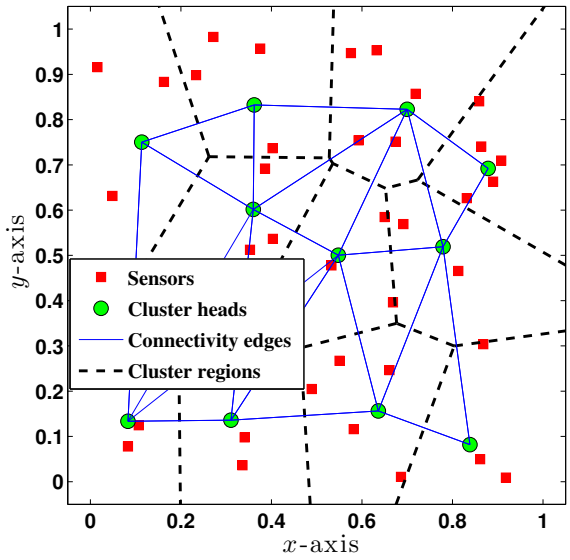

Fig. 1: Schematic view of the network and clusters

V. Complexity Analysis

Let us investigate the computational and communication complexities of the proposed distributed algorithm (DiSparSenSe-C) compared to the centralized one (SparSenSe-C). A deeper look into the steps of Algorithm 1 reveals that step 2 requires the solution of an SDP problem whose computational complexity is $O((M_l^c)^3)$. Besides, the communication cost involved in step 4 is $O(\varphi N_l^c n^3)$. Furthermore, step 5 requires n singular value decompositions (SVDs), each of which requires a computational complexity $O(n^3)$. Thus, all in all, the total computational complexity of DiSparSenSe-C is $O((M_l^c)^3 n^4)$ per node per iteration which is considerably lower compared to the computational complexity of SparSenSe-C $O(m^3)$ ($m \gg n, m \gg M_l^c$). The communication cost of DiSparSenSe-C is $O(\varphi N_l^c n^3)$ per node per iteration which is reasonably low as it is independent of m. Table I summarizes the discussed complexities of both centralized and distributed algorithms and their predecessors proposed in [5]. Note that according to [5], N_i denotes the number of sensors in the neighborhood of the i-th sensor including i itself. From the table, we observe that the computational complexity of DiSparSenSe-C is increased by a factor $(M_l^c)^3$ compared to DiSparSenSe, whereas its communication cost is relatively decreased by a factor of N_l^c/N_i ($N_l^c < N_i$).

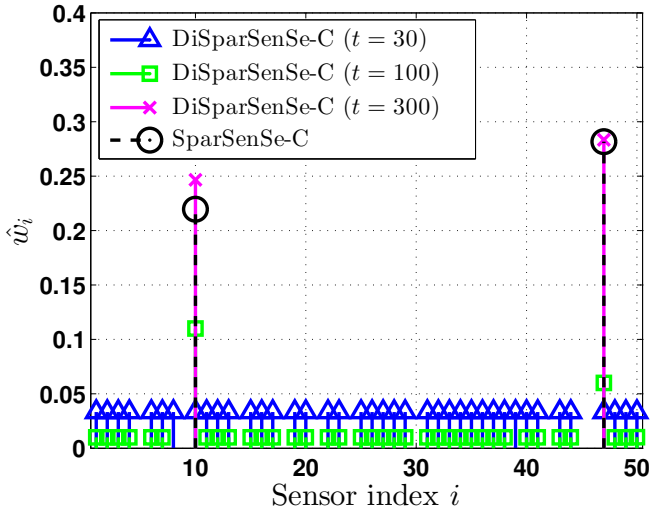

Fig. 2: Centralized vs. distributed; selected sensors

VI. Simulation Results

In this section, we investigate the performance of the proposed algorithms to see if SparSenSe-C actually selects a few sensors to satisfy the MSE constraint as well as to illustrate that DiSparSenSe-C selects the same sensors as SparSenSe-C. To this aim, we consider $m = 50$ sensors to estimate a parameter of interest $\mathbf{x}$ of dimension $n = 2$. The measurement (regression) matrix $\mathbf{A} \in \mathbb{R}^{50\times 2}$ is drawn from a zero-mean unit-variance Gaussian distribution $\mathcal{N}(0, 1)$. For DiSparSenSe we assume that the sensors are connected based on a random connectivity graph $\mathcal{G}$. We define $C = 11$ clusters and consider one of the sensors within each cluster as the cluster head as is depicted in a schematic view in Fig. 1. In order to account for the noise experienced by the sensors, we generate a block-diagonal symmetric PSD matrix $\mathbf{C}$. Further, we set the number of consensus steps to $\varphi = 2$, the MSE constraint to $\gamma = 0.1$, and the SNR to 10dB. Notably, for SparSenSe-C, we consider a sensor as active if $\hat{w}_i > 0$, whereas for DiSparSenSe-C, due to the fixed stepsize error floor, we consider a sensor as active if $\hat{w}_i^t > \alpha/10$.

In the first simulation, depicted in Fig. 2, we plot the estimated $\hat{\mathbf{w}}$ by SparSenSe-C and $\hat{\mathbf{w}}^t$ by DiSparSenSe-C for $\alpha = 0.05$. As can be seen, only 2 sensors (out of 50) are activated by SparSenSe-C to satisfy our MSE constraint which corroborates the fact that $\hat{\mathbf{w}}$ is indeed sparse. Note that for $t = 30$ many different sensors are activated by DiSparSenSe-C. However, as expected, by increasing the number of iterations (from $t = 30$ to $t = 300$), the same sensors as for SparSenSe-C are activated by DiSparSenSe-C, and the magnitude of the related $\hat{w}_i^t$'s gets closer to the values estimated by SparSenSe-C. This illustrates the fact that our distributed implementation (as expected) converges to the centralized algorithm.

In order to be able to quantitatively assess the performance of the distributed algorithm, we define two performance metrics. The first metric quantifies the primal convergence of the

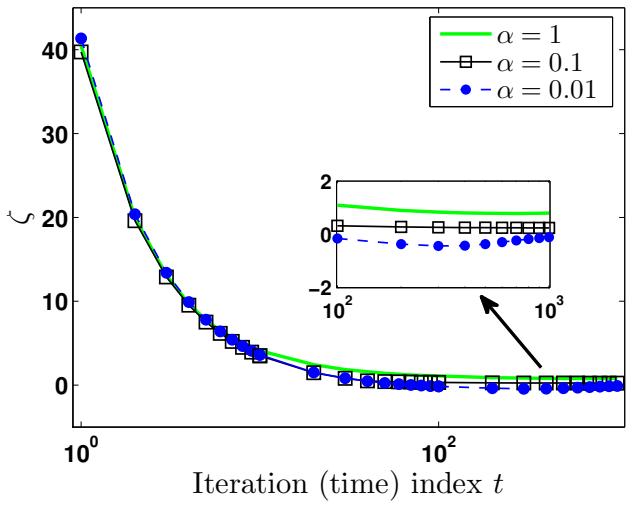

Fig. 3: Primal convergence ζ vs. t for $\alpha = 1$, 0.1, and 0.01

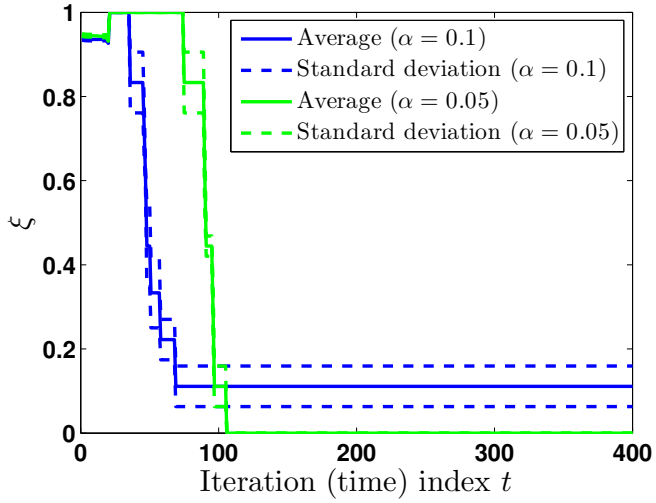

Fig. 4: Equivalence metric ξ vs. t for $\alpha = 0.1$ and 0.05

distributed algorithm and is defined by

$$\zeta(t) = \|\hat{\mathbf{w}}^t\|_1 - \|\hat{\mathbf{w}}\|_1.$$

The result is plotted in Fig. 3 for different stepsize values $\alpha = 1$, 0.1, and 0.01 and they are averaged over 20 independent Monte Carlo trials. As is clear from the figure, DiSparSenSe is converging in the primal sense as ζ goes down to the error floor with increasing t, as explained in our convergence analysis. The notable observation is that with $\alpha = 0.1$ we have the best convergence in the primal sense. Going further down to $\alpha = 0.01$ leads to some oscillations as is shown in the zoom-in plot.

The second metric is an equivalence metric and investigates the normalized level of similarity between the selected sensor sets by the centralized and distributed algorithms. To this aim, we define $\mathcal{A}$ as the set of indices of the selected sensors by SparSenSe-C and $\mathcal{B}$ as the corresponding set for DiSparSenSe-C. This helps us to define an equivalence metric between the distributed and centralized algorithms as

$$\xi = 1 - |\mathcal{A} \cap \mathcal{B}| / \max\{|\mathcal{A}|, |\mathcal{B}|\},$$

which means that if $\mathcal{A} \equiv \mathcal{B}$, then $\xi = 0$. The result for this metric is depicted in Fig. 4 for $\alpha = 0.1$ and 0.05 and for 20 independent Monte Carlo trials. As can be seen, we clearly observe from the average of the Monte Carlo trials (the solid line) that with increasing t an equivalence is acquired for $\alpha = 0.05$ as ξ goes to zero. However, for $\alpha = 0.1$ the solid line shows that we cannot attain any better than $\xi = 0.1$ due to the finite error floor.

VII. Discussions and Future Work

We would like to conclude the paper by highlighting that our newly proposed algorithms (compared to their predecessors in [5]) are generalized to operate in the scenarios where the measurement noise experienced by the sensors is correlated. Besides, the "*private*" information contained in the w_i's is not broadcast in the whole network but only within each cluster. Instead, their "*encoded*" versions, i.e., $\mathbf{Q}^{l,t}$ and $\mathbf{G}^{l,t}$ are communicated to reach convergence, which is an important advantage in terms of network security.

As we discussed earlier, in this work for the sake of simplicity of our derivations, we have made Assumption 1 which considered zero inter-cluster noise correlations. We would like to generalize this assumption to the case where such inter-cluster noise correlations exist, which basically leads to overlapping clusters. This would be one of our future research directions. Besides, detailed convergence analysis of all the proposed distributed algorithms and a conclusive comparison (on the convergence rate, etc.) is another part of an extended version of this work. Furthermore, currently we omit the possibility of imposing different "budget constraints" (such as power budget, etc.) on the sensors. We do expect that involving such constraints into our optimization problems would lead to the selection of different subsets of sensors, which is worthy of investigating. This will be yet another extension of this work.

References

[1] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. on Sig. Proc.*, vol. 57, no. 2, pp. 451–462, 2009.

[2] V. Kekatos and G. Giannakis, "From sparse signals to sparse residuals for robust sensing," *IEEE Trans. on Sig. Proc.*, vol. 59, pp. 3355–3368, July 2011.

[3] F. Altenbach, S. Corroy, G. Böcherer, and R. Mathar, "Strategies for distributed sensor selection using convex optimization," *Proc. of IEEE GLOBECOM*, pp. 2367–2372, Dec. 2012.

[4] D. Bickson and D. Dolev, "Distributed sensor selection using a truncated Newton method." submitted for publication, available on arXiv:0907.0931v2 [cs.IT].

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

[6] I. Schizas, "Distributed informative-sensor identification via sparsity-aware matrix decomposition," *IEEE Trans. on Sig. Proc.*, vol. 61, pp. 4610–4624, Sep. 2013.

[7] S. Roy and R. A. Iltis, "Decentralized Linear Estimation in Correlated Measurement Noise," *IEEE Trans. on Aerospace and Elec. Sys.*, vol. 27, no. 6, pp. 939 – 941, 1991.

[8] X. R. Li, Y. Zhu, J. Wang, and C. Han, "Optimal Linear Estimation Fusion-Part I: Unified Fusion Rules," *IEEE Trans. on Info. Theory*, vol. 49, no. 9, pp. 2192 – 2208, 1991.
[9] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[11] CVX Research Inc., "CVX: Matlab software for disciplined convex programming, version 2.0 beta." http://cvxr.com/cvx, Sept. 2012.
[12] A. Nedić and A. Ozdaglar, "Approximate primal solutions and rate analysis for dual subgradient methods," *SIAM J. on Opt.*, vol. 19, no. 4, pp. 1757–1780, 2009.
[13] J. C. Duchi, A. Agarwal, and M. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Trans. on Auto. Cont.*, vol. 57, no. 3, pp. 592–606, 2012.
[14] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson, "Subgradient methods and consensus algorithms for solving convex optimization problems," *Proc. of IEEE CDC*, pp. 4185–4190, Dec. 2008.
[15] H. Jamali-Rad, A. Simonetto, X. Ma, and G. Leus, "Distributed Sparsity-Aware Sensor Selection." to be submitted to IEEE Trans. on Sig. Proc.