This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

Distributed Sparsity-Aware Sensor Selection

Hadi Jamali-Rad*, *Student Member, IEEE*, Andrea Simonetto*, *Member, IEEE*,
Xiaoli Ma§, *Senior Member, IEEE*, and Geert Leus*, *Fellow, IEEE*

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. The problem becomes even more interesting in a distributed configuration when each sensor has to decide itself whether it should contribute to the estimation or not. In this paper, we explore the sparsity embedded within the problem and propose a *sparsity-aware* sensor selection paradigm for both uncorrelated and correlated noise experienced at different sensors. We also present reasonably low-complexity and elegant *distributed* algorithms in order to solve the centralized problems with convergence guarantees within a bounded error. Furthermore, we analytically quantify the complexity of the distributed algorithms compared to centralized ones. Our simulation results corroborate our claims and illustrate a promising performance for the proposed centralized and distributed algorithms.

EDICS: OPT-DOPT, NET-APPL.

Index Terms

Distributed parameter estimation, sensor selection, sparsity.

I. Introduction

We consider a typical sensor network estimation problem, where the sensors are supposed to estimate a vector of interest in a linear measurement model. For such a network, we study the problem of selecting

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. A portion of this work is published in IEEE Signal Processing Letters (SPL) and has been presented in the SPL track of the IEEE Conf. on Acoustic, Speech and Sig. Proc. (ICASSP), May 2014.

*The authors are with the Faculty of EEMCS, Delft University of Technology, 2628CD Delft, The Netherlands. e-mails:{h.jamalirad, a.simonetto, g.j.t.leus}@tudelft.nl. Corresponding author: Hadi Jamali-Rad, phone: (+31)152786280, fax: (+31)152786190, e-mail: h.jamalirad@tudelft.nl.

§The author is with the School of ECE, Georgia Institute of Technology, Atlanta, GA 30332-0250, USA. e-mail: xiaoli@gatech.edu.

the minimum number of sensors within the network, so that a given mean squared error (MSE) estimation performance is satisfied. This generic problem is of great interest in several practical application domains including radar and target tracking [1], event detection [2], and energy-efficient network management [3], to name a few. A straightforward solution to this 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, where each sensor should itself decide about its state of being selected (active) or not (inactive).

A related sensor selection problem has been studied in [4] where elegant convex relaxations are designed for primal and dual problems. Also, in [2] the same problem with a different optimality (selection) constraint is considered for event detection in sensor networks. However, instead of optimizing different performance metrics and fixing the number of sensors as in [2] and [4], we minimize the number of sensors given a performance constraint, which is generally more practical from a design perspective. Interestingly, this enables us to exploit the sparsity embedded within the problem and propose sparsity-aware solutions. From this angle, our approach is closer to what is proposed in [5] for selecting reliable sensors, also called "robust sensing". However, we consider a different constraint from the one in [5], and we do not need the sensors to be activated and take measurements for solving the selection problem; we only need them to know their regression coefficients. Worthy of being mentioned, is the work of [1], wherein both ideas (minimizing the number of sensors and minimizing the performance constraint) are considered for a multiple-radar localization architecture. The problem is formulated within a combinatorial optimization framework as a knapsack problem. Notably, in all the aforementioned studies, a distributed approach has not been considered.

The problem of distributed sensor selection is of crucial importance because in many practical network configurations, it is impossible to establish a central processing unit to gather all the information and make centralized decisions. Even if possible, this centralized process may drain significantly on the communication and energy resources [6], [7]. The alternative approach is to make decisions using in-network distributed processing [6]. A decentralized implementation of [4] is proposed in [8]; however, the heuristic assumption of two "leader" nodes violates the classical definition of a distributed approach. In [9], two distributed implementations of [4] based on a truncated Newton algorithm are proposed. In [10], we have explored the sparsity embedded within the problem and have proposed a relaxed sparsity-aware sensor selection approach. We have also presented a reasonably low-complexity distributed implementation of the centralized algorithm such that each sensor can decide itself whether it should

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

contribute to the estimation or not. Compared to [10], the work of [9] deals with a slightly different problem and also requires the private sensor information to be broadcast whereas the proposed approach in [10] avoids that. Moreover, the distributed approach of [10] is considerably more efficient in terms of complexity compared to [9]. Finally, another relevant problem, but of a different nature, is considered in [11], where a distributed algorithm is designed to identify the sensors containing relevant information by a sparsity-aware decomposition of the measurement covariance matrix.

In [10], 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 noise experienced by the sensors can be correlated. Particularly, as it is pointed out in [12] and [13], since the measurement noises of different sensors may depend on a common parameter (as is the case in our problem formulation), the sensors can observe correlated noise. Another example occurs when the unknown vector of interest 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.

In [14], a modified version of [4] has been proposed to handle correlated noise; however, as we discussed earlier, the problem formulation in these works is different than ours. A particular case where we can handle correlated noise in our problem is when we consider clusters of sensors with correlated noise and assume that the inter-cluster noise correlation is negligible, as we have proposed in [15]. This intuitive approach necessitates considering some sensors as "cluster heads" with higher processing power. In practice, such clusters (with zero inter-cluster correlations) can not always be defined. Furthermore, cluster heads impose extra constraints and violate the homogeneity of the sensor network. Thus, we would like to develop a generalized approach by dropping the cluster assumption. In this paper, we extend our basic idea in [10] by presenting the following main contributions.

i) First, we modify the proposed distributed approach for uncorrelated noise in [10] by introducing a novel consensus weighting and conducting a double-consensus, which results in better convergence properties and robustness against the choice of regressors.

ii) Second, we formulate the centralized problem for the case of correlated noise, as well as propose an elegant low-complexity distributed implementation of the problem, where we have no clusters and cluster heads.

iii) Further, we analyze and quantify the convergence behavior of all our proposed distributed algorithms and prove that we have convergence guarantees (with a bounded error) to the centralized algorithms.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

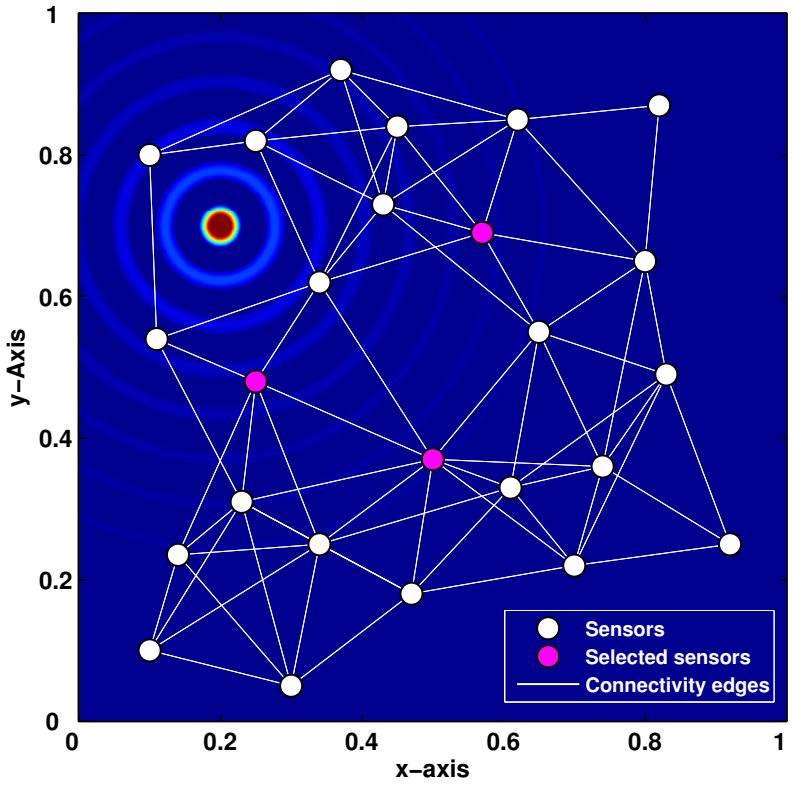


Fig. 1: Schematic view of 2-D sensor selection

iv) Finally, we investigate the computational and communication complexities involved in the proposed centralized and distributed algorithms, and promote that it is wise to employ the proposed distributed approaches.

The rest of this paper is organized as follows. In Section II, we define the problem under consideration. Section III describes our proposed centralized and distributed algorithms for the case of uncorrelated noise. Section IV is devoted to our proposed algorithms in order to handle correlated noise. In Section V, the computational and communication costs involved in the proposed algorithms are investigated and compared. Numerical results are illustrated in Section VI, and the paper is concluded in Section VII.

II. Problem Definition

We consider a network estimation problem where m sensor nodes distributed over an area of interest in $\mathbb{R}^d$ ($d \leqslant m$) are supposed to estimate an unknown vector $\mathbf{x} \in \mathbb{R}^n$. The elements of $\mathbf{x}$ can for instance represent the contribution of a physical phenomenon in different dimensions within the area of interest. A schematic view of such a network deployed in order to estimate a wave field in a 2-D area is shown in Fig. 1. The sensor nodes are equipped with computational and communication capabilities. Note that the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

computational load of a sensor will be less than what would be required for a fusion center. Each node measures $y_i = \mathbf{a}_i^T\mathbf{x}+\eta_i,\ i=1,\ldots,m$, where the regressors $\mathbf{a}_i$'s $\in \mathbb{R}^n$ are assumed known (or measured) and they should span $\mathbb{R}^n$ ($m \gg n$). The η_i's are the additive noise experienced by different sensors, for which we need to know (or estimate) their second-order statistics. Note that, given the spatial distribution of the sensors, it practically makes sense 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* (without measuring the y_i's) the minimum number of sensors 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. The next two sections of this paper, which explain our proposed algorithms, are respectively derived based on the assumptions that the noise experienced by the sensors is uncorrelated or correlated.

III. Sensor Selection for Uncorrelated Noise

In this section, we develop a sparsity-aware sensor selection paradigm, by considering uncorrelated noise. This is normally the case when the sensors are placed far apart. We derive centralized and distributed algorithms and investigate the convergence properties of the distributed algorithm.

A. Centralized Optimization Problem

In a centralized setup, all $\mathbf{a}_i$'s are transmitted to a central processing unit which allows us to define the matrix $\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_m]^T$. Now, we can construct

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\eta}, \tag{1}$$

where $\mathbf{y} = [y_1, \cdots, y_m]^T$, and $\boldsymbol{\eta} = [\eta_1, \cdots, \eta_m]^T$. We consider $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0},\ \mathbf{C})$, where the covariance matrix of the measurement noise $\mathbf{C}$ is by definition a positive definite matrix. For the centralized linear measurement model (1) and the maximum likelihood (ML) estimator, the MSE can be expressed as [16,

$$\begin{aligned}\mathcal{L}(\mathbf{w},\mathbf{u},\mathbf{G}) &= \sum_{i=1}^m w_i - \sum_{j=1}^n \text{tr}\left(\left[\begin{array}{c|c} \sum_{i=1}^m w_i\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^T & \mathbf{e}_j \\ \hline \mathbf{e}_j^T & u_j \end{array}\right]\mathbf{G}_j\right) = \sum_{i=1}^m w_i - \sum_{j=1}^n\sum_{i=1}^m \text{tr}\left(\left[\begin{array}{c|c} w_i\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^T & \mathbf{e}_j/m \\ \hline \mathbf{e}_j^T/m & u_j/m \end{array}\right]\mathbf{G}_j\right) \\ &= \sum_{i=1}^m\left(w_i - \sum_{j=1}^n \text{tr}\left(\left[\begin{array}{c|c} w_i\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^T & \mathbf{e}_j/m \\ \hline \mathbf{e}_j^T/m & u_j/m \end{array}\right]\mathbf{G}_j\right)\right) = \sum_{i=1}^m \mathcal{L}_i(w_i,\mathbf{u},\mathbf{G}),\end{aligned} \tag{9}$$

p. 186, eq. 7.47]

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

where $\hat{\mathbf{x}}$ is the ML estimate and tr(.) stands for the trace operator. Given uncorrelated noise, we have $\mathbb{E}[\eta_i\,\eta_j^T] = \sigma_i^2\delta(i-j)$ with $\delta(.)$ denoting the Kronecker delta, and thus $\mathbf{C} = \text{diag}([\sigma_1^2, \cdots, \sigma_m^2])$. Note that $\text{diag}(\mathbf{x})$ returns a diagonal matrix with the elements of $\mathbf{x}$ on its diagonal. Based on this assumption, the MSE can be reformulated as

$$\text{MSE} = \text{tr}\left((\sum_{i=1}^{m} \tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^T)^{-1}\right), \tag{3}$$

where $\tilde{\mathbf{a}}_i = \mathbf{a}_i/\sigma_i = [\tilde{a}_{i,1}, \cdots, \tilde{a}_{i,n}]^T$. The associated selection constraint on the MSE can then be stated as

$$\text{tr}\left((\sum_{i=1}^{m} w_i\,\tilde{\mathbf{a}}_i\,\tilde{\mathbf{a}}_i^T)^{-1}\right) \leqslant \gamma, \tag{4}$$

where the variable $w_i \in \{0,1\}$ encodes whether the i-th sensor has to be activated. 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. Therefore, we cast the problem as the following program

$$\underset{\mathbf{w}\in\{0,1\}^m,\mathbf{u}}{\text{minimize}} \quad \|\mathbf{w}\|_0 \tag{5a}$$

$$\text{s.t.} \quad \left[\begin{array}{c|c} \sum_{i=1}^m w_i\,\tilde{\mathbf{a}}_i\,\tilde{\mathbf{a}}_i^T & \mathbf{e}_j \\ \hline \mathbf{e}_j^T & u_j \end{array}\right] \succeq 0,\ \forall j, \tag{5b}$$

$$\|\mathbf{u}\|_1 \leqslant \gamma,\ \ u_j \geqslant 0,\ j = 1, \ldots, n, \tag{5c}$$

where $\mathbf{w} = [w_1, \ldots, w_m]^T$ is the selection vector, $\mathbf{u} = [u_1, \ldots, u_n]^T$ 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 the constraints (5b) and (5c) are a linear matrix inequality (LMI) representation of the original constraint (4), obtained by using the Schur complement [17]. We denote the global optimizers of (5) as $(\mathbf{w}^*, \mathbf{u}^*)$. Since both the cost $\|\mathbf{w}\|_0$ in (5) and the finite-alphabet constraint on the w_i's are non-convex, we consider the following relaxed version of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

the problem labeled as sparsity-aware sensor selection (SparSenSe)

$$\underset{\mathbf{w}\in[0,1]^m,\mathbf{u}}{\text{minimize}} \quad \|\mathbf{w}\|_1 \tag{6a}$$

$$\text{s.t.} \quad \left[\begin{array}{c|c} \sum_{i=1}^m w_i\,\tilde{\mathbf{a}}_i\,\tilde{\mathbf{a}}_i^T & \mathbf{e}_j \\ \hline \mathbf{e}_j^T & u_j \end{array}\right] \geq 0,\ \forall j, \tag{6b}$$

$$\|\mathbf{u}\|_1 \leqslant \gamma,\ u_j \geqslant 0,\ j=1,\ldots,n, \tag{6c}$$

and we denote its optimizers as $(\hat{\mathbf{w}},\hat{\mathbf{u}})$. In [10], we present a detailed equivalence result, i.e., we prove that in special simplified cases, the number of selected sensors is the same for both the original (5) and the relaxed problem (6). Note that these simplified conditions in [10] only help to obtain closed-form expressions in the equivalence proof, which for all of our simulations do not necessarily hold. Yet, in practice, we always obtain a desirable sparse solution that is close to the optimal one.

B. Distributed Algorithm

In this subsection, we develop a distributed version of SparSenSe based on dual decomposition [18]. First, we construct the Lagrangian of the problem at hand, and then define our dual function. Next, we investigate the decomposability of the dual function w.r.t. the primal variables of interest for a fixed dual variable. With this in place, we look at the Slater's condition for the original problem, and finally, we propose an iterative solution for the distributed algorithm.

Let us start with some notations. We call $\mathcal{N}_i$ the neighborhood set of the i-th sensor, with cardinality $|\mathcal{N}_i| = N_i$ (either given or to be estimated). Similarly, we define $\bar{\mathcal{N}}_i = \mathcal{N}_i \cup i$ with cardinality $|\bar{\mathcal{N}}_i| = \bar{N}_i = N_i + 1$. We also define the following convex sets to simplify our notations

$$\mathcal{W}_i \;=\; \{w_i \mid 0 \leqslant w_i \leqslant 1\}, \tag{7}$$

$$\mathcal{U} \;=\; \{\mathbf{u} \mid u_j \geqslant 0,\ \|\mathbf{u}\|_1 \leqslant \gamma\}, \tag{8}$$

and form the Lagrangian of (6a)-(6b) given by (9) (shown at the bottom of this page), where $\mathbf{G}_j \geq 0,\ \forall j$, are appropriately sized dual variables, and $\mathbf{G} = [\mathbf{G}_1,\ldots,\mathbf{G}_n]$. The dual function of $\mathcal{L}$ can be given by

$$q(\mathbf{G}) = \min_{w_i\in\mathcal{W}_i,\mathbf{u}\in\mathcal{U}} \sum_{i=1}^m \mathcal{L}_i(w_i,\mathbf{u},\mathbf{G}). \tag{10}$$

Given the special linear structure of the Lagrangian $\mathcal{L}$ in (9), the following proposition is in place.

Proposition 1. *The dual function of $\mathcal{L}$ defined in* (10) *is decomposable as*

$$q(\mathbf{G}) = \sum_{i=1}^{m} \left(\min_{w_i \in \mathcal{W}_i, \mathbf{u} \in \mathcal{U}} \mathcal{L}_i(w_i, \mathbf{u}, \mathbf{G}) \right) = \sum_{i=1}^{m} q_i(\mathbf{G}), \tag{11}$$

which can be computed by solving two linear programs (LPs) for each i.

Proof: First, by looking at the structure of the matrix multiplied by $\mathbf{G}_j$ in (9), we partition $\mathbf{G}_j$ as $\mathbf{G}_j = [\mathbf{G}_{j1}, \mathbf{G}_{j12}; \mathbf{G}_{j12}^T, \mathbf{G}_{j2}]$, where $\mathbf{G}_{j1}$, $\mathbf{G}_{j12}$ and $\mathbf{G}_{j2}$ are respectively of size $n \times n$, $n \times 1$, and 1×1 (i.e., a scalar). Therefore, the minimization problem in (10) reads

$$q(\mathbf{G}) = \min_{w_i \in \mathcal{W}_i, \mathbf{u} \in \mathcal{U}} \sum_{i=1}^{m} w_i \left(1 - \sum_{j=1}^{n} \mathrm{tr}(\tilde{\mathbf{a}}_i \tilde{\mathbf{a}}_i^T \mathbf{G}_{j1}) \right) - \sum_{i=1}^{m} \sum_{j=1}^{n} 2\,\mathrm{tr}(\mathbf{e}_j \mathbf{G}_{j12}^T)/m - \sum_{i=1}^{m} \sum_{j=1}^{n} u_j \mathrm{tr}(\mathbf{G}_{j2})/m. \tag{12}$$

Note that the middle term can be dropped as it does not contain any of our optimization variables. Since the w_i's and the auxiliary vector $\mathbf{u}$ are independent, we have two LP's that could be treated separately. The optimization problem is also separable in the w_i's which allows us to write the following LP

$$\min_{w_i \in \mathcal{W}_i} \sum_{i=1}^{m} w_i \left(1 - \sum_{j=1}^{n} \mathrm{tr}(\tilde{\mathbf{a}}_i \tilde{\mathbf{a}}_i^T \mathbf{G}_{j1}) \right) \quad \equiv \quad \sum_{i=1}^{m} \min_{w_i \in \mathcal{W}_i} w_i \left(1 - \sum_{j=1}^{n} \mathrm{tr}(\tilde{\mathbf{a}}_i \tilde{\mathbf{a}}_i^T \mathbf{G}_{j1}) \right). \tag{13}$$

Then, we focus on $\mathbf{u}$, that is

$$\min_{\mathbf{u} \in \mathcal{U}} - \sum_{i=1}^{m} \sum_{j=1}^{n} u_j \mathrm{tr}(\mathbf{G}_{j2})/m. \tag{14}$$

The key point is that since the local problems are in fact *the same* for all i's, we can swap the sum and minimization operators leading to

$$\min_{\mathbf{u} \in \mathcal{U}} - \sum_{i=1}^{m} \sum_{j=1}^{n} u_j \mathrm{tr}(\mathbf{G}_{j2})/m \equiv \sum_{i=1}^{m} \min_{\mathbf{u} \in \mathcal{U}} - \sum_{j=1}^{n} u_j \mathrm{tr}(\mathbf{G}_{j2})/m, \tag{15}$$

which is an LP, and thus the claim follows. ■

Note that in (10) and (11), we try to decompose the global problem into local problems, and to this aim, we reformulate the Lagrangian and corresponding dual function as the summation of local Lagrangians $\mathcal{L}_i(w_i, \mathbf{u}, \mathbf{G})$ and dual functions $q_i(\mathbf{G})$. Given a certain value of $\mathbf{G}$, the functions $q_i(\mathbf{G})$ and their subgradient with respect to (w.r.t.) $\mathbf{G}$, called $\mathbf{Q}$ and defined later on, can be computed locally, for instance using the MATLAB optimization function *linprog*, by solving the resulting LPs in and at each sensor [18].

Whenever γ is large enough so that we expect sparse solutions in terms of $\hat{\mathbf{w}}$, Slater's condition holds

for (6), which can be formulated as the following proposition.

Proposition 2. *Slater's condition holds for* (6)*, for sufficiently large* γ*.*

Proof: For sufficiently large γ, we can always find a pair $(\mathbf{w},\ \mathbf{u})$ that strictly satisfies (6b) - (6c). ■

Therefore, the original ℓ_1-regularization (6) leads to the dual optimization problem

$$\underset{\mathbf{G}_1\geq 0,\ldots,\mathbf{G}_n\geq 0}{\text{maximize}} \sum_{i=1}^{m} q_i(\mathbf{G}), \tag{16}$$

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 proximal-based methods, such as the dual averaging scheme of [19] with a variable step-size, or the simpler dual subgradient of [18] with a fixed step-size. The latter method has the advantage of providing a recovery mechanism for the primal solution, i.e., we recover $\hat{\mathbf{w}}$ while computing $\mathbf{G}$, which is in fact our goal. Furthermore, the subgradient method of [18] has the benefit of employing a fixed step-size which yields a simpler implementation. That is why we opt to employ the dual subgradient method of [18]. In order to implement the dual subgradient of [18], each sensor requires a copy of the subgradient of $q(\mathbf{G})$ w.r.t. $\mathbf{G}_j$, $\forall j$, i.e., each sensor requires a copy of $\mathbf{Q} = [\mathbf{Q}_1, \ldots, \mathbf{Q}_n]$. Given that $\mathcal{W}_i$ and $\mathcal{U}$ are compact and convex, we can define such a subgradient as

$$\mathbf{Q}_j = \sum_{i=1}^{m} \nabla_{\mathbf{G}_j} q_i(\mathbf{G}_j) = -\left[\begin{array}{c|c} \sum_{i=1}^{m} \bar{w}_i \tilde{\mathbf{a}}_i \tilde{\mathbf{a}}_i^T & \mathbf{e}_j \\ \hline \mathbf{e}_j^T & \bar{u}_j \end{array}\right]. \tag{17}$$

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

$$q(\mathbf{G}) = \min_{w_i\in\mathcal{W}_i, \mathbf{u}\in\mathcal{U}} \sum_{i=1}^{m} \mathcal{L}_i(w_i, \mathbf{u}, \mathbf{G}). \tag{18}$$

Note that the dimension of $\mathbf{Q}$ is the same as that of $\mathbf{G}$. The need for this *global* parameter can be circumvented by using the method of [20] where the sensors have different local copies of both $\mathbf{G}$ and $\mathbf{Q}$, say $\mathbf{G}^i$ and $\mathbf{Q}^i$, and they run an inexact consensus procedure for φ times ($\varphi \in \mathbb{N}_+$). In particular, to solve (16), we will consider the following inexact subgradient update. We call the i-th sensor version of $\mathbf{G}$ at iteration t, $\mathbf{G}^{i,t}$. We start with a given initial condition $\mathbf{G}_j^{i,0}$ for each sensor, and then $\forall\, t \geqslant 0$ we have

$$\mathbf{V}_j^{i,\tau=0,t} = \mathbf{G}_j^{i,t} + \alpha\, m\, \mathbf{Q}_j^{i,t},\ \text{for } j = 1, \cdots, n, \tag{19}$$

where α is the step-size. Next, we run φ times a consensus procedure as

$$\mathbf{V}_j^{i,\tau,t} = \sum_{p=1}^{m} [\mathbf{Z}]_{i,p} \mathbf{V}_j^{p,\tau-1,t}, \tag{20}$$

and finally a projection over the cone of positive semidefinite (PSD) matrices as

$$\mathbf{G}_j^{i,t+1} = \mathcal{P}_{\geq 0} \left\{ \mathbf{V}_j^{i,\varphi,t} \right\}, \text{ for } j = 1, \cdots, n. \tag{21}$$

In (20), $\mathbf{Z} \in \mathbb{R}^{m \times m}$ indicates a proper sensor-wise consensus matrix whose weights have been defined using a Metropolis weighting, i.e.,

$$[\mathbf{Z}]_{i,p} = \begin{cases} 1/(\max\{\bar{N}_i, \bar{N}_p\}) & \text{if } p \in \bar{\mathcal{N}}_i \\ 0 & \text{if } p \notin \bar{\mathcal{N}}_i, p \neq i \\ 1 - \sum_{l=1}^{m} [\mathbf{Z}]_{i,l} & \text{if } i = p. \end{cases} \tag{22}$$

If we execute (20) for $\varphi \to \infty$, we recover the procedure of [18], whereas if φ is limited we introduce an additional error in the distributed optimization procedure. Our proposed distributed SparSenSe (called DiSparSenSe) algorithm is summarized in Algorithm 1, where we denote the primal optimizer of DiSparSenSe at iteration t as $\hat{\mathbf{w}}^t$.

Remark 1. It is worth highlighting that in this paper we have modified our previously proposed distributed algorithm in [10] in the consensus averaging step from two aspects. First, here we apply a *double-consensus* on both $\mathbf{G}$ and $\mathbf{Q}$ instead of only a consensus on $\mathbf{Q}$ in [10]. Second, instead of a simple consensus averaging in [10], here we propose a symmetric consensus matrix $\mathbf{Z}$. We illustrate in Subsection VI-A that these refinements lead to a smoother and faster convergence of DiSparSenSe.

C. Convergence Properties of DiSparSenSe

We would like to highlight that DiSparSenSe will converge to the solution of SparSenSe with an error floor dependent on α and φ. This can be proven extending the ϵ-subgradient argument discussed in [18] and [20], as is briefly summarized in this subsection and is detailed in Appendix A. We investigate both primal and dual convergence problems in Appendix A. For the latter, we prove that there exists a finite $\bar{\varphi} > 0$ such that if $\varphi \geqslant \bar{\varphi}$ the sequence of dual functions $\{q(\mathbf{G}^{i,t})\}$ generated by DiSparSenSe converges. Based on this, we prove that the convergence of the running average primal sequence $\{\hat{\mathbf{w}}^t\}$ (as defined in step 2 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 Appendix A, (64)-(65), show that the running average primal

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

Algorithm 1 DiSparSenSe

Initialization: $\mathbf{G}_j^{i,0} = \mathbf{I}_{n+1}, \quad w_i^0 = 0, \forall i,\ j.$

Input: $\mathbf{G}_j^{i,t}$, $\hat{w}_i^t$, $\forall i,\ j.$

1- Dual optimization (LP): Compute, in parallel at each sensor i, the value of $q_i(\mathbf{G}^{i,t})$, its derivative $\mathbf{Q}^{i,t}$ using (17), and the related optimal primal variables $\bar{w}_i^t$.

2- Primal recovery: Update method of [18]:

$$\hat{w}_i^{t+1} = t\,\hat{w}_i^t/(t+1) + \bar{w}_i^t/(t+1).$$

3- Consensus:

For $\tau = 1$ to φ

- $\diamond$ Send $\mathbf{G}^{i,t}$ and $\mathbf{Q}^{i,t}$ to the neighboring sensor nodes;
- $\diamond$ Perform, in parallel, one consensus step as

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

which is initialized as in (19).

End

4- Dual recovery: Update each sensor's dual variable as

$$\mathbf{G}_j^{i,t+1} = \mathcal{P}_{\geq 0}\left\{\mathbf{V}_j^{i,\varphi,t}\right\}, \forall j.$$

Output: $\mathbf{G}_j^{i,t+1}$, $\hat{w}_i^{t+1}$, $\forall i,\ j.$

function is upper bounded as

$$\|\hat{\mathbf{w}}^t\|_1 \quad \leqslant \quad \|\hat{\mathbf{w}}\|_1 \ + \ \frac{nG^2}{2t\alpha/m} \ + \ \frac{\alpha m\,(\sqrt{n}\,Q+\tau)^2}{2} \ + \ \tau m\sqrt{n}\,G \ + \ m\psi_2(\alpha,Q,\varphi). \quad (23)$$

and it is lower bounded as

$$\|\hat{\mathbf{w}}^t\|_1 \quad \geqslant \quad \|\hat{\mathbf{w}}\|_1 \ - \ \frac{9nG^2}{2t\alpha/m} \ - \ \frac{\alpha m\,(\sqrt{n}\,Q+\tau)^2}{2} \ - \ \tau m\sqrt{n}\,G \ - \ m\psi_2(\alpha,Q,\varphi), \quad (24)$$

where Q is an upper bound on the norm of the dual subgradient $\mathbf{Q}_j^{i,t}$, and ψ_2 is a non-negative function monotonically increasing with α and decreasing with φ, and τ is a non-negative scalar depending on φ. To sum up, 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 $\|\hat{\mathbf{w}}\|_1$ (the solution to SparSenSe). The width of this region depends on α and φ.

IV. Sensor Selection for Correlated Noise

In this section, we develop a sparsity-aware sensor selection paradigm, by considering a correlated noise. This normally happens for neighboring sensors in a dense network. We derive centralized and distributed algorithms and investigate the convergence properties of the distributed algorithm.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

A. Centralized Algorithm

Similar to the uncorrelated case in Subsection III-A, we can construct (1) and compute the MSE of the ML estimator as in (2). However, given correlated noise, different from the case of uncorrelated noise, $\mathbf{C}$ is not diagonal and can even be a full matrix if all the sensors experience correlated noise. Thus, the non-diagonal elements $[\mathbf{C}]_{i,j}$, $i \neq j$, should also be incorporated within our selection procedure. In order to handle these non-diagonal elements, we define a symmetric PSD selection matrix $\mathbf{W} = \mathbf{w}\mathbf{w}^T$, where $\mathbf{w}$ is our selection vector as defined earlier. Notice that based on this new definition of $\mathbf{W}$, $[\mathbf{C}]_{i,j}$ will only be incorporated if both w_i and w_j are non-zero at the same time. A possible selection constraint can then be stated as

$$\text{tr}\left((\mathbf{A}^T[\mathbf{W} \odot \mathbf{C}^{-1}]\mathbf{A})^{-1}\right) \leqslant \gamma, \tag{25}$$

where $\odot$ stands for the Hadamard product. Note that since $w_i \in \{0, 1\}$, we have $\text{diag}(\mathbf{W}) = \mathbf{w}$, where $\text{diag}(\mathbf{X})$ returns a vector containing the diagonal elements of $\mathbf{X}$. Again, by exploring the sparsity embedded within the problem, it can be cast as the following optimization program

$$\underset{\mathbf{W},\mathbf{u}}{\text{minimize}} \quad \|\text{diag}(\mathbf{W})\|_0 \tag{26a}$$

$$\text{s.t.} \quad \left[\begin{array}{c|c} \mathbf{A}^T[\mathbf{W} \odot \mathbf{C}^{-1}]\mathbf{A} & \mathbf{e}_j \\ \hline \mathbf{e}_j^T & u_j \end{array}\right] \geq 0, \ \forall j, \tag{26b}$$

$$\|\mathbf{u}\|_1 \leqslant \gamma, \ u_j \geqslant 0, \ j = 1, \ldots, n, \tag{26c}$$

$$\mathbf{W} \geq 0, \tag{26d}$$

$$[\mathbf{W}]_{i,j} \in \{0, 1\}, \ \text{rank}(\mathbf{W}) = 1. \tag{26e}$$

Similar to the derivations in Subsection III-A, the constraints (26b) and (26c) are a more suitable representation of the original constraint (25). We denote the global optimizers of (26) as $(\mathbf{W}^*, \mathbf{u}^*)$. Clearly, the problem in (26) is non-convex due to its objective (ℓ_0 norm), and the first and the third terms in (26d) (finite-alphabet constraint on the elements of $\mathbf{W}$ and rank-1 constraint, respectively). Delving

$$\begin{aligned} \mathcal{L}(\mathbf{W}, \mathbf{u}, \mathbf{G}) &= \sum_{i=1}^{m} w_{ii} - \sum_{j=1}^{n} \text{tr}\left(\left[\begin{array}{c|c} \sum_{i=1}^{m}\sum_{k\in\overline{\mathcal{N}}_i} w_{ik}\, \tilde{c}_{ik}^{-1}\, \mathbf{a}_i\, \mathbf{a}_k^T & \mathbf{e}_j \\ \hline \mathbf{e}_j^T & u_j \end{array}\right] \mathbf{G}_j\right) \\ &= \sum_{i=1}^{m}\left(w_{ii} - \sum_{j=1}^{n} \text{tr}\left(\left[\begin{array}{c|c} \sum_{k\in\overline{\mathcal{N}}_i} w_{ik}\, \tilde{c}_{ik}^{-1}\, \mathbf{a}_i\, \mathbf{a}_k^T & \mathbf{e}_j/m \\ \hline \mathbf{e}_j^T/m & u_j/m \end{array}\right] \mathbf{G}_j\right)\right) = \sum_{i=1}^{m} \mathcal{L}_i(\mathbf{w}_i, \mathbf{u}, \mathbf{G}) \end{aligned} \tag{35}$$

deeper in (26) reveals a problem on our way to distribute it in the next subsection, and that is the positive semidefiniteness constraint on $\mathbf{W}$ in (26d). Positive semidefiniteness is a global constraint and cannot be decomposed into corresponding sub-constraints, as we desire in the next subsection. That is why we use the following lemma and replace (26d) with the following sufficient condition

$$\mathbf{W} \in \mathbb{D}^m,\ \mathbf{W}^T = \mathbf{W}, \tag{27}$$

where $\mathbb{D} = \{\mathbf{X} \mid [\mathbf{X}]_{i,i} \geqslant \sum_{j \neq i} [\mathbf{X}]_{i,j}, \forall i\}$ denotes the set of diagonally dominant matrices.

Lemma 1. *A symmetric diagonally dominant matrix with real non-negative entries is PSD.*

Proof: The proof follows from Greshgorin's circle theorem [21]. ■

Finally, we relax the three non-convex terms to obtain

$$\underset{\mathbf{W},\mathbf{u}}{\text{minimize}} \quad \text{tr}(\mathbf{W}) \tag{28a}$$

$$\text{s.t.} \quad \left[\begin{array}{c|c} \mathbf{A}^T[\mathbf{W} \odot \mathbf{C}^{-1}]\mathbf{A} & \mathbf{e}_j \\ \hline \mathbf{e}_j^T & u_j \end{array}\right] \succeq 0,\ \forall j, \tag{28b}$$

$$\|\mathbf{u}\|_1 \leqslant \gamma,\ \ u_j \geqslant 0,\ j = 1, \ldots, n, \tag{28c}$$

$$0 \leqslant [\mathbf{W}]_{i,j} \leqslant 1,\ \mathbf{W} \in \mathbb{D}^m,\ \mathbf{W}^T = \mathbf{W}. \tag{28d}$$

We call this algorithm SparSenSe-C to distinguish it from SparSenSe, and we denote its optimizer as $(\hat{\mathbf{W}}, \hat{\mathbf{u}})$. A final step to recover $\hat{\mathbf{w}}$ from $\hat{\mathbf{W}}$ is to apply a Choleskey decomposition and a possible randomization 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.

B. Distributed Algorithm

In this section, we develop a distributed version of SparSenSe-C. In order to derive a distributed algorithm, we follow the same steps as described in the beginning of Subsection III-B, i.e. constructing the Lagrangian, defining a dual function, investigating its decomposability for a fixed dual variable, and

$$(\hat{\mathbf{w}}_i^{s+1}, \hat{\mathbf{u}}^{s+1}) = \underset{\mathbf{w}_i \in \overline{\mathcal{W}}_i^c, \mathbf{u} \in \mathcal{U}}{\text{argmin}} \mathcal{L}_i(\mathbf{w}_i, \mathbf{u}, \mathbf{G}) + \sum_{k \in \mathcal{N}_i} \lambda_{ik}^s (w_{ik} - \frac{w_{ik}^s + w_{ki}^s}{2}) + \sum_{k \in \mathcal{N}_i} \frac{\rho}{2} \left\| w_{ik} - \frac{w_{ik}^s + w_{ki}^s}{2} \right\|_2^2. \tag{39}$$

finally proposing an overall distributed algorithm.

In regards to the correlated noise, our general approach towards the problem is to decompose it so that the i-th sensor can estimate the i-th row of $\mathbf{W}$. Looking at SparSenSe-C, we clearly observe that the non-diagonal elements of $\mathbf{C}^{-1}$ complicate the derivation of a distributed algorithm compared to the case of DiSparSenSe. The more non-diagonal elements, the more coupling terms are introduced, and thus, the more computational and communication steps are required.

Triggered by the localized nature of many phenomena of interest in practical applications, we define the following set of noise covariance matrices

$$\mathcal{C} = \{\mathbf{C} \mid \mathbf{C} \succeq 0, [\mathbf{C}]_{i,j} = 0, \text{ if } [\mathbf{I}_m + \mathcal{A}]_{i,j} = 0\}, \tag{29}$$

where $\mathcal{A}$ is the adjacency matrix associated with the network connectivity graph with zero diagonal elements. This means that we consider the nodes to experience correlated noise with their immediate neighbors. In practice, $\mathbf{C} \in \mathcal{C}$ is a sparse matrix if the network is not highly connected. We can reorder $\mathbf{C}$ using the Cuthill-McKee algorithm [22] to end up with a banded matrix. Note that we need to distribute $\mathbf{C}^{-1}$ as in (28). One solution is to compute the inverse of such a banded matrix in a distributed fashion using only local computations at different sensors via algorithms such as "DICI" in [23]. The alternative solution is to approximate it. Nonetheless, the inverse would not necessarily be a banded matrix. In general, $\mathbf{C} = \mathbf{C}_d + \bar{\mathbf{C}}_d$, where $\mathbf{C}_d$ and $\bar{\mathbf{C}}_d$ respectively stand for the matrices containing the diagonal and non-diagonal elements of $\mathbf{C}$. We can rewrite

$$\mathbf{C}^{-1} = \mathbf{C}_d^{-1} \left(\mathbf{I}_m + \mathbf{C}_d^{-1}\bar{\mathbf{C}}_d\right)^{-1}. \tag{30}$$

Now, for the specific case where the autocorrelation of the noise experienced at each sensor is much larger than the cross-correlation with its neighbors, we have $\|\mathbf{I}_m\|_F \gg \|\mathbf{C}_d^{-1}\bar{\mathbf{C}}_d\|_F$. For such a case, we can use Taylor's expansion as

$$\begin{aligned}\mathbf{C}^{-1} &= \mathbf{C}_d^{-1}\left(\mathbf{I}_m - \mathbf{C}_d^{-1}\bar{\mathbf{C}}_d + \frac{1}{2}(\mathbf{C}_d^{-1}\bar{\mathbf{C}}_d)^T\mathbf{C}_d^{-1}\bar{\mathbf{C}}_d + \cdots\right), \\ &= \mathbf{C}_d^{-1}\left(\mathbf{I}_m - \mathbf{C}_d^{-1}\bar{\mathbf{C}}_d + \frac{1}{2}\mathbf{C}_d^{-1}\bar{\mathbf{C}}_d^2\mathbf{C}_d^{-1} + \cdots\right),\end{aligned} \tag{31}$$

which due to $\bar{\mathbf{C}}_d$, $\bar{\mathbf{C}}_d^2$, and the next powers of $\bar{\mathbf{C}}_d$ mandates single-hop, two-hop, and multi-hop communications. To simplify our next derivations, and without loss of generality (see Remark 2), we can

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

confine ourselves to a first-order approximation as

$$\tilde{\mathbf{C}}^{-1} \approx \mathbf{C}_d^{-1} \left(\mathbf{I}_m - \mathbf{C}_d^{-1} \bar{\mathbf{C}}_d \right), \tag{32}$$

which after reordering is again a banded matrix and easier to be distributed.

To simplify our subsequent notations, let us denote the (i,j)-th element of $\mathbf{C}$, $\tilde{\mathbf{C}}^{-1}$ and $\mathbf{W}$ by c_{ij}, $\tilde{c}_{ij}^{-1}$ and w_{ij}, respectively. We also denote the i-th row of $\mathbf{W}$ by $\mathbf{w}_i$. Next, we define the following convex set

$$\mathcal{W}_i^c = \{\mathbf{w}_i \mid 0 \leqslant w_{ik} \leqslant 1, w_{ii} \geqslant \sum_{j \neq i} w_{ij}, w_{ik} = w_{ki}, \forall k \in \bar{\mathcal{N}}_i, \}. \tag{33}$$

Note that sensor i only estimates $\bar{N}_i$ elements out of m in $\mathbf{w}_i$ because the rest are known to be zeros. The banded property of our newly defined $\tilde{\mathbf{C}}^{-1}$ in (32) helps us to expand (25) as

$$\text{tr} \left(\Big(\sum_{i=1}^{m} \sum_{k \in \bar{\mathcal{N}}_i} w_{ik}\, \tilde{c}_{ik}^{-1}\, \mathbf{a}_i\, \mathbf{a}_k^T \Big)^{-1} \right) \leqslant \gamma, \tag{34}$$

and construct the Lagrangian of (28a)-(28b) as (35) (shown at the bottom of this page). Both $\mathbf{G}_j$ and $\mathbf{G}$ are defined earlier. Now, the dual function of $\mathcal{L}$ can be given by

$$q(\mathbf{G}) = \min_{\mathbf{w}_i \in \mathcal{W}_i^c, \mathbf{u} \in \mathcal{U}} \sum_{i=1}^{m} \mathcal{L}_i(\mathbf{w}_i, \mathbf{u}, \mathbf{G}). \tag{36}$$

Given the structure of $\mathcal{W}_i^c$ the dual function is not immediately separable. This is because the row-wise symmetry constraint within $\mathcal{W}_i^c$, i.e., $w_{ik} = w_{ki}, \forall k \in \bar{\mathcal{N}}_i$, cannot be handled only based on local information available at sensor i; we also need to know the w_{ki}'s. That is why we try to restructure the dual function. An alternative way to write the dual function is to detach and explicitly write the row-wise symmetry constraint $w_{ik} = w_{ki}$. To this aim, we modify $\mathcal{W}_i^c$ as

$$\overline{\mathcal{W}}_i^c = \{\mathbf{w}_i \mid 0 \leqslant w_{ik} \leqslant 1, w_{ii} \geqslant \sum_{j \neq i} w_{ij}, \forall k \in \bar{\mathcal{N}}_i\}, \tag{37}$$

and instead of (36), we write

$$q(\mathbf{G}) = \min_{\mathbf{w}_i \in \overline{\mathcal{W}}_i^c, \mathbf{u} \in \mathcal{U}} \left\{ \sum_{i=1}^{m} \mathcal{L}_i(\mathbf{w}_i, \mathbf{u}, \mathbf{G}) \,|\text{s.t. } w_{ik} = w_{ki}, \forall k \in \bar{\mathcal{N}}_i \right\} \tag{38}$$

Given a certain value of $\mathbf{G}$, the function $q(\mathbf{G})$ and its subgradient w.r.t. $\mathbf{G}$, called $\mathbf{Q}$ and defined later on, can be computed. Bear in mind that (as in Proposition 1) the minimization and the sum operators

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

Algorithm 2 ADMM

Input: ρ, $\mathbf{G}$, λ_{ik}, w_{ik}, $\forall i,\ k$.
For $s = 0$ to $s_{\max} - 1$
1- In parallel at each sensor, solve (39).
2- Each sensor transmits its own estimate $\hat{\mathbf{w}}_i^{s+1}$ to its neighbors.
3- Update λ_{ik}'s as

$$\lambda_{ik}^{s+1} = \lambda_{ik}^{s} + \rho\,\frac{(w_{ik}^{s+1} - w_{ki}^{s+1})}{2}$$

End
Output: $\bar{\mathbf{w}}_i = \hat{\mathbf{w}}_i^{s_{\max}}$ and $\bar{\mathbf{u}} = \hat{\mathbf{u}}^{s_{\max}}$.

can be swapped for $\mathbf{u}$. The point is that (35) is not strongly convex. Besides, we need to decompose (38) and solve it in a distributed fashion at each sensor. A possible solution with good convergence properties under such conditions (in contrast to typical primal or dual decomposition algorithms) is to employ the alternating direction method of multipliers (ADMM) [24]–[26] for a fixed $\mathbf{G}$. An example of using ADMM for sensor scheduling can be found in [27].

Algorithm 2 shows the resulting ADMM, where the symmetry is enforced by the second and third terms of (39) in step 1 of the algorithm. In particular, in step 1, we can readily decompose (39) (shown at the bottom of this page) into an LP for $\mathbf{u}$ and a quadratic program (QP) for $\mathbf{w}_i$. To this goal, it is enough to substitute $\mathcal{L}_i(\mathbf{w}_i, \mathbf{u}, \mathbf{G})$ from (35) in (39), decompose the result into respective parts for $\mathbf{w}_i$ (resulting in a QP) and $\mathbf{u}$ (resulting in an LP). The results can efficiently be solved using the MATLAB optimization functions *linprog* and *quadprog*. This is a straightforward mathematical exercise and details are omitted in favor of limited space and the similarity to the sequence in the proof of Proposition 1.

In practice, as we also discuss in the next subsection, we only need a few iterations to converge ($s_{\max} < 10$). Similar to our analysis in Subsection III-B, the original ℓ_1-regularization (28) leads to the dual optimization problem

$$\underset{\mathbf{G}_1 \succeq 0, \ldots, \mathbf{G}_n \succeq 0}{\text{maximize}} \; q(\mathbf{G}), \tag{40}$$

with *zero* duality gap. Again, we solve (40) using the dual subgradient method of [18], where each sensor requires a copy of the subgradient of $q(\mathbf{G})$ w.r.t. $\mathbf{G}_j$ as before

$$\begin{aligned} \mathbf{Q}_j &= \nabla_{\mathbf{G}_j} q(\mathbf{G}_j) \\ &= -\left[\begin{array}{c|c} \sum_{i=1}^{m} \sum_{k \in \overline{\mathcal{N}}_i} \bar{w}_{ik} \tilde{c}_{ik}^{-1} \mathbf{a}_i \mathbf{a}_k^T & \mathbf{e}_j \\ \hline \mathbf{e}_j^T & \bar{u}_j \end{array}\right], \end{aligned} \tag{41}$$

where the $\bar{w}_{ik}$'s and $\bar{u}_j$'s are the outputs of the ADMM iterations in Algorithm 2. Similar to Subsec-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

Algorithm 3 DiSparSenSe-C

Initialization: $\mathbf{G}_j^{i,0} = \mathbf{I}_{n+1}$, $\mathbf{w}_i^0 = \mathbf{0}$, $\forall i,\ j$.

Input: $\mathbf{G}_j^{i,t}$, $\hat{\mathbf{w}}^t$, $\forall i,\ j$.

1- Dual Optimization (ADMM):

$\diamond$ Initialize Algorithm 2 with ρ, $\lambda_{ik} = 0$, $\forall i$ and $k \in \bar{\mathcal{N}}_i$, $\mathbf{G}^{i,t}$ and $\hat{\mathbf{w}}_i$'s.

$\diamond$ Use outputs $\bar{\mathbf{w}}_i^t = \bar{\mathbf{w}}_i$ and $\bar{\mathbf{u}}^t = \bar{\mathbf{u}}$ to compute $\mathbf{Q}^{i,t}$ using (41).

2- Primal recovery: Update method of [18]:

$$\hat{\mathbf{w}}_i^{t+1} = t\,\hat{\mathbf{w}}_i^t/(t+1) + \bar{\mathbf{w}}_i^t/(t+1).$$

3- Consensus:

For $\tau = 1$ to φ

$\diamond$ Send $\mathbf{G}^{i,t}$ and $\mathbf{Q}^{i,t}$ to the neighboring sensor nodes;

$\diamond$ Perform, in parallel, one consensus step as

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

which is initialized as in (19).

End

4- Dual recovery: Update each sensor's dual variable as

$$\mathbf{G}_j^{i,t+1} = \mathcal{P}_{\geq 0}\left\{\mathbf{V}_j^{i,\varphi,t}\right\}, \forall j.$$

5- Selection: Estimate the selection vector

$$\hat{\mathbf{w}}^{t+1} = \text{diag}\left([(\hat{\mathbf{w}}_i^t)^T, \cdots, (\hat{\mathbf{w}}_i^t)^T]^T\right)$$

Output: $\mathbf{G}_j^{i,t+1}$, $\hat{\mathbf{w}}^{t+1}$, $\forall i,\ j$.

tion III-B, the need for a global knowledge of $\mathbf{Q}$ is circumvented using an inexact consensus procedure. The rest of the steps follow the same trend as in DiSparSenSe except that instead of an LP to solve the dual optimization problem, here we have an extra inner-loop for ADMM. Our proposed algorithm for distributed SparSenSe in case of correlated noise (we call it DiSparSenSe-C) is summarized in Algorithm 3, where we denote the primal optimizer of DiSparSenSe-C at iteration t as $\hat{\mathbf{w}}^t$.

Remark 2. We would like to highlight that our assumption on the structure of $\mathbf{C}^{-1}$ does not limit the generality of the proposed solution, i.e., DiSparSenSe-C. In the most generic case where $\mathbf{C}^{-1}$ is a full matrix, each sensor has to estimate a full row instead of only a few elements (corresponding to its neighbors) in each row of $\mathbf{W}$. This calls for rounds of multi-hop communications if the network is connected. Nonetheless, our proposed approach immediately applies.

Remark 3. DiSparSenSe-C can readily be applied to the case of uncorrelated noise. However, if some knowledge about the nature of the experienced noise is available it makes sense to employ the corresponding algorithm, especially from a complexity perspective.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

C. Convergence Properties of DiSparSenSe-C

In this subsection, we first investigate the convergence of the ADMM iterations. To this aim, we compare the solution of ADMM with the corresponding centralized problem at time iteration t for a fixed $\mathbf{G}^{i,t}$. We show that in practice, the proposed ADMM iterations converge to the result of the centralized problem with a modest accuracy, sufficient for our application, within only a few iterations. The related centralized problem is

$$\underset{\mathbf{W},\mathbf{u}}{\text{minimize}} \quad \text{tr}(\mathbf{W}) - \sum_{i=1}^{m}\sum_{j=1}^{n} \text{tr}\left(\left[\begin{array}{c|c} \sum_{k\in\overline{\mathcal{N}}_i} w_{ik}\, \tilde{c}_{ik}^{-1}\, \mathbf{a}_i\, \mathbf{a}_k^T & \frac{\mathbf{e}_j}{m} \\ \hline \frac{\mathbf{e}_j^T}{m} & \frac{u_j}{m} \end{array}\right] \mathbf{G}_j^{i,t} \right) \tag{42a}$$

$$\text{s.t.} \quad \|\mathbf{u}\|_1 \leqslant \gamma,\ u_j \geqslant 0,\ j = 1,\ldots,n, \tag{42b}$$

$$0 \leqslant w_{i,j} \leqslant 1, \forall i,j,\ \mathbf{W} \in \mathbb{D}^m,\ \mathbf{W}^T = \mathbf{W}, \tag{42c}$$

where we denote the solution to the aforementioned problem with $(\hat{\mathbf{W}}^t_{\text{cent.}}, \hat{\mathbf{u}}^t_{\text{cent.}})$. This convergence is illustrated in Subsection VI-B for our simulation setup. As a result of this fast dual optimization convergence, given that the major difference between DiSparSenSe and DiSparSenSe-C is the dual optimization part, the convergence proof of DiSparSenSe-C follows the same path as the one of DiSparSenSe. Therefore, we can prove similar expressions as (64)-(65), for DiSparSenSe-C. The formal proof is almost identical to the one of DiSparSenSe, and thus, we omit it in this paper due to space limitations.

V. Complexity Analysis

Let us investigate the computational and communication complexities of the proposed distributed algorithms (DiSparSenSe and DiSparSenSe-C) compared to the centralized ones (SparSenSe and SparSenSe-C). A deeper look into the steps of Algorithm 1 reveals that step 2 requires the solution of an LP problem whose computational complexity is $O(n^3)$, where $O(.)$ denotes the order of complexity. Besides, the communication cost involved in step 3 is $O(\varphi N_i n^3)$ because n square matrices of size $n+1$ are broadcast to N_i neighbors for φ times. Furthermore, step 4 requires n singular value decompositions, each of which requires a computational complexity of $O(n^3)$. Thus, the total computational complexity of DiSparSenSe is $O(n^4)$ per sensor per iteration which is considerably lower compared to the computational complexity of SparSenSe which is $O(m^3)$ ($m \gg n$). The communication cost of DiSparSenSe is $O(\varphi N_i n^3)$ per sensor per iteration which is reasonably low as it is independent of m.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

TABLE I: Complexity Order Comparison

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	$O(n^4 + s_{\max}(N_i^3 + n^3))$	$O(N_i(\varphi\, n^3 + s_{\max}))$

DiSparSenSe-C involves almost the same computational and communication costs as compared to DiSparSenSe. The main difference is step 1 of Algorithm 3, i.e., ADMM, which requires the solution of $s_{\max}$ local QP problems with dimension N_i and local LP problems with dimension n, resulting in a total complexity of $O(s_{\max}(N_i^3 + n^3))$. ADMM also calls for an extra communication cost of $O(s_{\max}\, N_i)$ because of step 2 of Algorithm 2. Thus, the total computational cost of DiSparSenSe-C is $O(n^4 + s_{\max}(N_i^3 + n^3))$ and its communication cost is $O(N_i(\varphi\, n^3 + s_{\max}))$, both per sensor per iteration. Table I summarizes the discussed complexities of both centralized and distributed algorithms. From the table, we observe that the computational and communication complexities of DiSparSenSe-C are relatively larger than those of DiSparSenSe due to replacing a simple LP with ADMM iterations in order to handle the correlated noise.

VI. Numerical Results

In this section, we investigate the performance of the proposed algorithms. First, we would like to see whether SparSenSe and SparSenSe-C actually select a few sensors (i.e., a sparse solution) which satisfies the MSE constraint. Then, we consider these centralized algorithms as our selection *performance metric* beyond which we cannot perform, and investigate whether their corresponding distributed algorithms (namely, DiSparSenSe and DiSparSenSe-C) select the same sensors or not.

To this objective, we consider a medium-scale network with $m = 50$ sensors to estimate a parameter of interest $\mathbf{x}$ of dimension $n = 2$. The 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 and DiSparSenSe-C we assume that the sensors are connected based on a random connectivity graph $\mathcal{G}$ with average node degree of 5. Further, we set the SNR to 10dB and $\gamma = 0.1$. We can consider a sensor as active by defining thresholds based on our estimation error floor (a complicated function of α and φ) coming from our convergence analysis in Appendix A. A simpler alternative, a rule of thumb, would be to consider a sensor as active if $w_i > \alpha/10$, and this is what we consider in our simulations.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

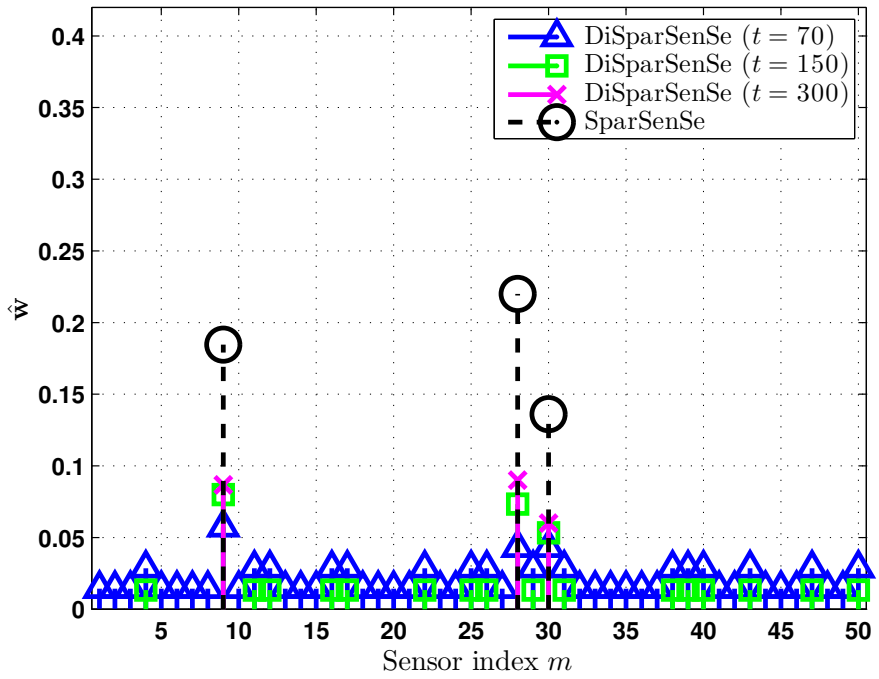


Fig. 2: Selected sensors for the uncorrelated case

In order to quantitatively assess the performance of the distributed algorithms, we define an equivalence metric to investigate the normalized level of similarity between the selected sensor sets by the centralized and distributed algorithms. To this aim, we define $\mathcal{S}_c$ as the set of indices of the selected sensors by the centralized algorithms and $\mathcal{S}_d$ as the set for the corresponding distributed algorithms. This helps us to define an equivalence metric between the distributed and centralized algorithms as

$$\xi = 1 - |\mathcal{S}_c \cap \mathcal{S}_d| / \max\{|\mathcal{S}_c|, |\mathcal{S}_d|\}, \tag{43}$$

which means that if $\mathcal{S}_c \equiv \mathcal{S}_d$, then $\xi = 0$. The centralized problems are solved using CVX [28], and for the distributed ones we implement them using the aforementioned MATLAB optimization functions.

A. Case of Uncorrelated Noise

In case of uncorrelated noise, for the sake of simplicity of our simulations, we assume that the noise experienced at different sensors has the same $\sigma = 1/\sqrt{\text{SNR}}$.

In the first simulation, depicted in Fig. 2, we plot $\hat{\mathbf{w}}$ estimated by SparSenSe and $\hat{\mathbf{w}}^t$ estimated by DiSparSenSe for $\varphi = 5$. As can be seen, only 3 sensors (out of 50) are activated by SparSenSe to satisfy our MSE constraint, which verifies the fact that $\hat{\mathbf{w}}$ is sparse. Note that for $t = 70$ many different sensors

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

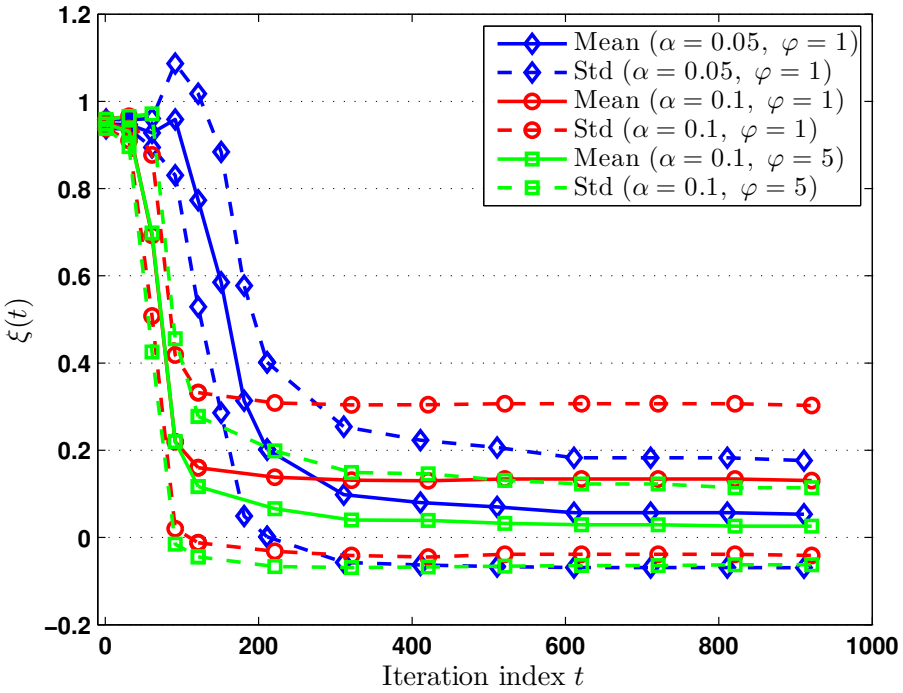


Fig. 3: Equivalence metric $\xi(t)$

are activated by DiSparSenSe. However, as expected, by increasing the number of iterations (from $t = 70$ to $t = 300$), the same sensors as for SparSenSe are activated by DiSparSenSe and the magnitude of the related $\hat{w}_i^t$'s gradually gets closer to the values estimated by SparSenSe. However, as is clear from the figure, it is not necessary to attain the magnitudes estimated by SparSenSe to be able to make a decision about the selected sensors. This result illustrates the fact that our distributed implementation (as expected from our convergence analysis) converges to the solution of the centralized algorithm.

The next simulation result which is illustrated in Fig. 3, investigates the convergence of DiSparSenSe over 100 independent Monte Carlo realizations of $\mathbf{A}$ (leading to 100 different subsets of sensors to be selected) for $\varphi = 5$, and $\alpha = 0.1$ and 0.05. We also plot the standard deviation (std) of our estimates with dashed lines. As can be seen, for both values of α we converge to the correct solution with an error floor. The convergence is faster for $\alpha = 0.1$ as is expected from our convergence analysis, (64)-(65), because the second terms on the right-hand-side of both expressions (the ones $\propto 1/t$) vanish faster with a larger α. Fig. 3 also illustrates the effect of varying φ for $\alpha = 0.1$, where reducing φ from 5 down to 1 leads to a larger error floor. This can also be justified using our explanations in Appendix A.

Notice that Fig. 3 depicts a smoother convergence compared to our initial results in [10]. As we

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

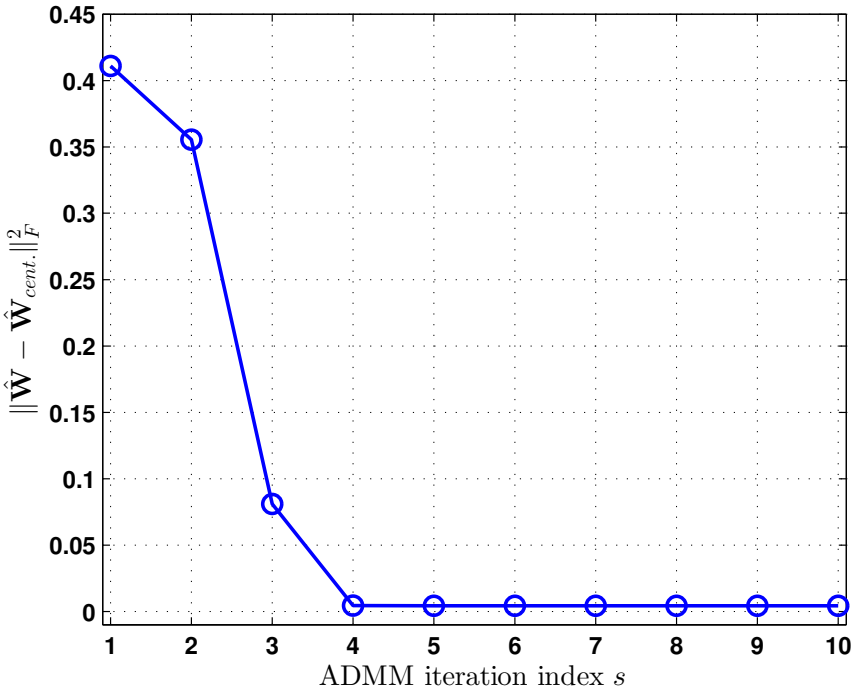


Fig. 4: Convergence of ADMM iterations

discussed earlier in Subsection III-B (Remark 1), this is due to our modified consensus weighting and the double-consensus. We also observe in our simulations that these modifications bring about a more robust performance against the choice of $\mathbf{A}$.

B. Case of Correlated Noise

In case of correlated noise, similar to the previous subsection, we assume that the noise experienced at different sensors has the same $\sigma = 1/\sqrt{\text{SNR}}$ and on top of that a certain level of correlation with the neighbors defined by the following correlation coefficient

$$\mu = \|\mathbf{I}_m\|_F / \|\mathbf{C}_d^{-1}\bar{\mathbf{C}}_d\|_F. \tag{44}$$

For the next simulations, except the last one, we consider 5% correlation with the neighbors per sensor, i.e., $\mu = 0.05$. For the ADMM algorithm, we set ρ to 0.1 and initialize the λ_{ik}'s with zeros.

Let us start by investigating the convergence of ADMM, based on our explanations in Subsection IV-C. The result is illustrated in Fig. 4 where we plot the primal convergence norm $\|\hat{\mathbf{W}}^t - \hat{\mathbf{W}}^t_{\text{cent.}}\|_F^2$ vs. s (the number of ADMM iterations) averaged over 50 iterations t. Here, $\|\cdot\|_F$ stands for the Frobenius norm. As

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

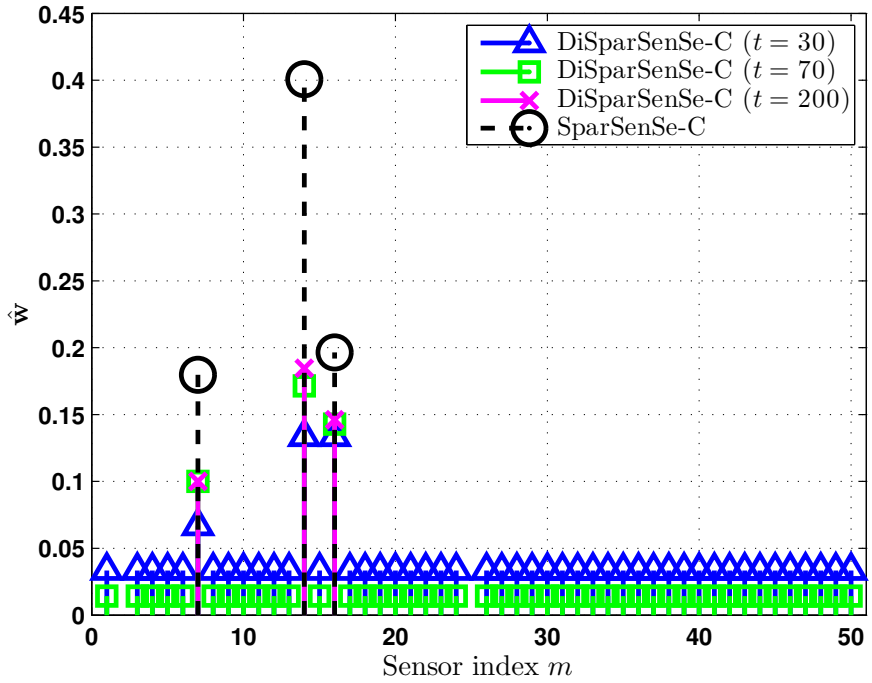


Fig. 5: Selected sensors for the correlated case

can be seen from the figure, in practice, ADMM converges relatively fast within only a few ($s_{\max} < 10$) iterations. Note that this is partly due to the fact that the solution of DiSparSenSe-C is actually sparse, and hence, for many i and $k \in \bar{\mathcal{N}}_i$, $w_{ik} = w_{ki} = 0$. This means $\hat{\mathbf{W}}^t$ is almost automatically symmetric and only a few ADMM iterations would suffice to converge to a feasible solution.

In the next simulation, we plot $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}^t$ respectively estimated by SparSenSe-C and DiSparSenSe-C for $\varphi = 10$ in Fig. 5. As can be seen, only 3 sensors (out of 50) are activated by SparSenSe-C to satisfy our MSE constraint. Similar to the case of DiSparSenSe, by increasing the number of iterations from $t = 30$ to $t = 200$, the same sensors as for SparSenSe-C are activated by DiSparSenSe-C and the magnitude of the related $\hat{w}_i^t$'s gradually gets closer to the values estimated by SparSenSe-C. This result clarifies the fact that our distributed implementation (as expected from our convergence analysis) converges to the solution of the centralized algorithm within a bounded error.

The simulation results depicted in Fig. 6 investigate the convergence of DiSparSenSe-C over 100 independent Monte Carlo trials for $\varphi = 10$, and $\alpha = 0.01$ and 0.005. As can be seen from the figure, for both values of α we converge with an error floor. Similar to the case of DiSparSenSe in the previous subsection, the convergence is faster for the larger $\alpha = 0.01$, as is also expected from our convergence

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

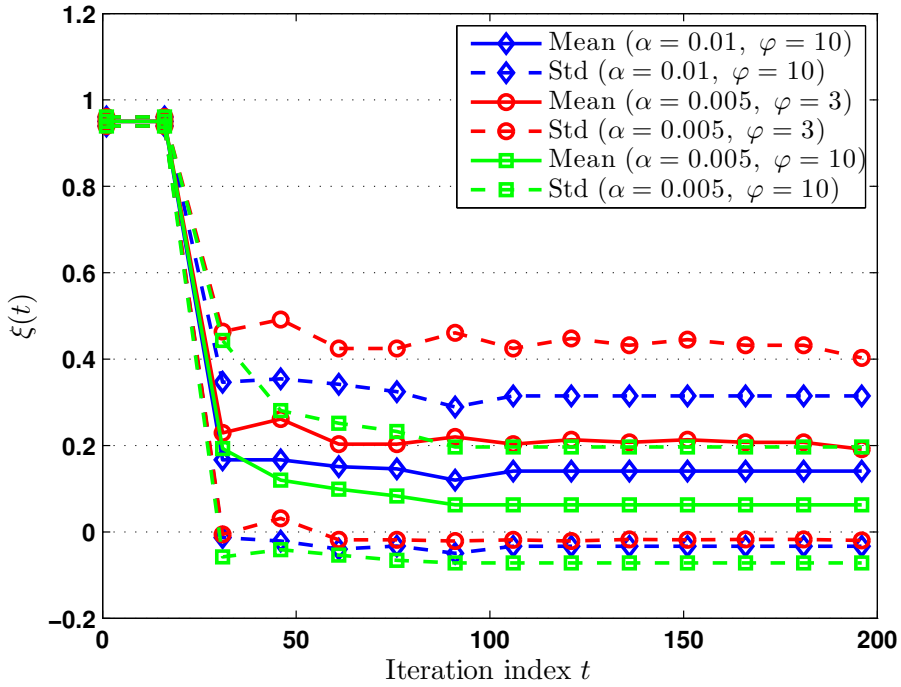


Fig. 6: Equivalence metric $\xi(t)$

analysis. However, we observe here that with $\alpha = 0.005$ we also get a better equivalence performance compared to $\alpha = 0.01$. Fig. 6 also illustrates the effect of varying φ for $\alpha = 0.005$ where reducing φ from 10 to 3 leads to a larger error floor. This can be justified using our convergence results, similar to our explanations for DiSparSenSe in the previous subsection.

Finally, we attempt to investigate the effect of our assumptions (on approximation of $\tilde{\mathbf{C}}^{-1}$ and the diagonally dominant surrogate of positive semidefiniteness) for different correlation regimes (different μ). To this aim, we consider the following four problems. P_0: the original problem in (26) (fed with $\mathbf{C}^{-1}$) which is non-convex and should be solved via exhaustive search, P_1: the relaxed problem SparSenSe-C in (28) (fed with $\tilde{\mathbf{C}}^{-1}$), P_2: the relaxed problem SparSenSe-C in (28) (fed with $\mathbf{C}^{-1}$), P_3: the relaxed problem with PSD constraint replaced with diagonal dominance (fed with $\mathbf{C}^{-1}$). Notably, we have already shown that our distributed algorithms converge to the same solution as the centralized ones, and that is why for this simulation we do not plot them.

The results for the number of selected sensors are depicted in Fig. 7 for $\mu = 0$ to 0.7, and for 100 independent Monte Carlo trials. Note that, $\mu = 0$ corresponds to the uncorrelated case (SparSenSe), and $\mu = 0.7$ represents a highly correlated case. As can be seen, increasing the amount of correlation leads

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

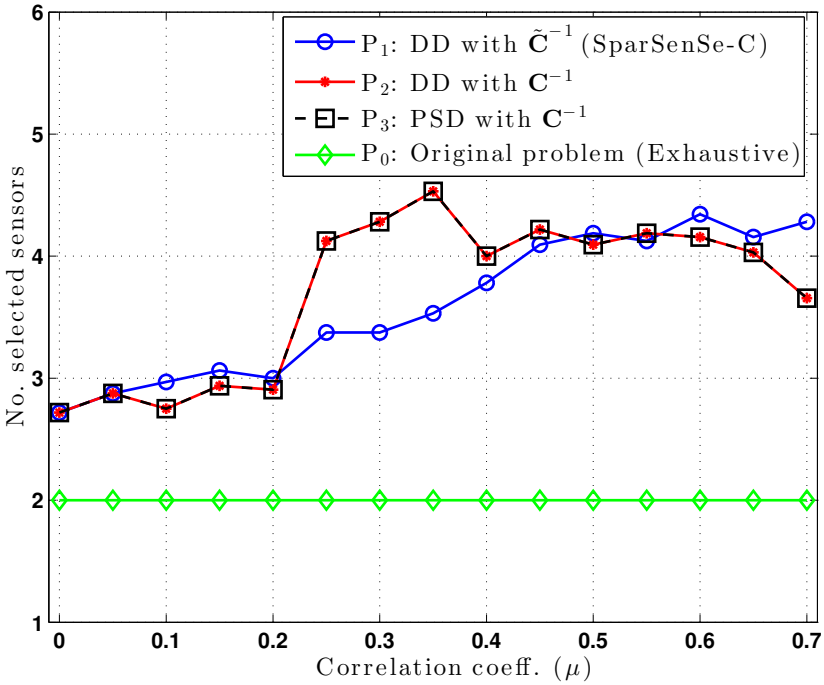


Fig. 7: Performance vs. correlation. DD stands for diagonally dominant $\mathbf{W}$.

to selecting more sensors in all the relaxed problems P_1 to P_3, which is due to the ℓ_1-norm relaxation. On the other hand, the original centralized problem P_0, keeps on selecting only 2 sensors for $\gamma = 0.1$. Remember that decreasing γ leads to selecting more sensors by P_0, and increases the sensitivity to the correlation effect. An important observation to highlight is that the overall gap between P_0 and the other three algorithms is due to the ℓ_1-norm relaxation.

Another observation is that the results of P_2 and P_3 coincide even in terms of the indices of the selected sensors, which is omitted in this figure. This confirms the validity of our proposition on replacing positive semidefiniteness with diagonal dominance. More interestingly, even our proposal to adopt $\tilde{\mathbf{C}}^{-1}$ (leading to the transition from P_2 to P_1) does not considerably affect the performance of the selection algorithm. To be more specific, P_1 (SparSenSe-C) still selects a sparse number of sensors, and more importantly, this number is almost the same as for the other related problems (P_2 and P_3).

VII. Conclusions

We have proposed a framework for sparsity-aware sensor selection in centralized and distributed fashions for cases where the noise experienced by different sensors is either uncorrelated or correlated. In favor of the limited space, we have omitted the possibility of imposing different budget constraints

(such as power budget) on the sensors. Our initial results show that involving such constraints into our optimization problems would lead to the selection of different subsets of sensors. Another direction to be investigated is the case of time-varying regressors. We are currently considering dynamic sparse reconstruction algorithms to handle this problem.

APPENDIX A
CONVERGENCE ANALYSIS OF DISPARSENSE

In this appendix, we analyze both primal and dual convergence properties of DiSparSenSe[1]. First of all, since the sets $\mathcal{W}_i$ and $\mathcal{U}$ in (7) and (8) are compact, the subgradient $\mathbf{Q}_j^{i,t}$ is bounded by a certain finite bound Q [20] as

$$\|\mathbf{Q}_j^{i,t}\|_F \leqslant Q,\ j=1,\cdots,n,\ i=1,\cdots,m,\ t\geqslant 0. \tag{45}$$

For the consensus matrix $\mathbf{Z}$ in (22), we can show that

$$\mathbf{Z}=\mathbf{Z}^T,\ \mathbf{Z}\mathbf{1}_m=\mathbf{1}_m,\ \rho\Big(\mathbf{Z}-\frac{\mathbf{1}_m\mathbf{1}_m^T}{m}\Big)\leqslant \nu<1, \tag{46}$$

where $\rho(\cdot)$ returns the spectral radius and ν is an upper bound on the value of the spectral radius. In the following, we assume that the dual variable estimates $\mathbf{G}_j^{i,t}$'s are bounded by a convex compact set (comprising the zero element) as

$$\|\mathbf{G}_j^{i,t}\|_F \leqslant G,\ j=1,\cdots,n,\ i=1,\cdots,m,\ t\geqslant 0, \tag{47}$$

for a certain finite positive constant G. Nonetheless, if this is not the case, we can always project them into such a bounded set, which will not considerably affect our subsequent convergence analysis [18].

A. Dual Objective Convergence

Let us start our convergence analysis in the dual sense by the following theorem.

Theorem 1. *Let $\hat{q}$ be the optimal dual value of SparSenSe* (10)*, i.e.,*

$$\hat{q}=\max_{\mathbf{G}_1\geq 0,\ldots,\mathbf{G}_n\geq 0}\ \sum_{l=1}^{m} q_l(\mathbf{G}). \tag{48}$$

[1]A generalization of this proof for a wider class of optimization problems can be found in [29].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

Then, there exists a finite $\bar{\varphi} > 0$ such that if $\varphi \geqslant \bar{\varphi}$ the sequence of dual functions $\{q(\mathbf{G}^{i,t})\}$ generated by DiSparSenSe converges as

$$\limsup_{t\to\infty} q(\mathbf{G}^{i,t}) \geqslant \hat{q} - m\psi_1(\alpha, Q, \varphi), \;\; i = 1, \cdots, m, \tag{49}$$

where ψ_1 is a non-negative function of φ, α, and Q.

Proof: The proof is based on Theorem 2 of [20]. The trick is to first rewrite the steps of DiSparSenSe in a more compact way. To this aim, we define the vectors $\mathbf{g}_j^{i,t} = \text{vec}(\mathbf{G}_j^{i,t})$ and $\mathbf{h}_j^{i,t} = \text{vec}(\mathbf{Q}_j^{i,t})$. Next, we define the convex set $\mathcal{G}$ as

$$\mathcal{G} = \{\mathbf{g}^{i,t} | \mathbf{G}_j^{i,t} \succeq 0,\, j = 1, \cdots, n\}. \tag{50}$$

This helps us to rewrite the updates in DiSparSenSe (21) as

$$\mathbf{g}^{t+1} = \mathcal{P}_{\mathcal{G}} \left[(\mathbf{Z}\mathbf{I}_{n(n+1)})^{\varphi} (\mathbf{g}^t + \alpha m \mathbf{h}^t) \right], \tag{51}$$

where

$$\mathbf{g}^t = [(\mathbf{g}_1^{1,t})^T, \cdots, (\mathbf{g}_n^{1,t})^T, \cdots, (\mathbf{g}_1^{m,t})^T, \cdots, (\mathbf{g}_n^{m,t})^T]^T, \tag{52}$$

$$\mathbf{h}^t = [(\mathbf{h}_1^{1,t})^T, \cdots, (\mathbf{h}_n^{1,t})^T, \cdots, (\mathbf{h}_1^{m,t})^T, \cdots, (\mathbf{h}_n^{m,t})^T]^T, \tag{53}$$

where $\otimes$ stands for the Kronecker product. Note that now we can see DiSparSenSe as a subgradient method to minimize the function $-q(\mathbf{g})$, exactly as the recursion (8) in [20], and apply Theorem 2 of [20]. To do so, we first have to make sure their main assumptions hold. Assumptions 1, 2, and 4 of [20] hold in our case since the subgradient is bounded, the consensus matrix $\mathbf{Z}$ verifies the properties of Assumption 2, and Assumption 4 holds given that $\|\mathbf{G}_j^{i,t}\|_F \leqslant G, \;\forall i, j$. Let us also define $\mathbf{v}_j^{i,t} = \text{vec}(\mathbf{V}_j^{i,\varphi,t})$. Now, the term

$$\left\| \mathbf{v}_j^{i,0} - \frac{1}{m} \sum_{p=1}^{m} \mathbf{v}_j^{p,0} \right\| \tag{54}$$

is bounded since we initialize the algorithm with a fixed $\mathbf{G}_j^{i,0} = \mathbf{G}^0, \;\forall i, j$, and also because the subgradient is bounded. Given this, Theorem 2 of [20] yields the claim. ■

Notice that due to optimality, $q(\mathbf{G}^{i,t}) = \sum_{l=1}^{m} q_l(\mathbf{G}^{i,t})$ cannot be greater than $\hat{q}$, and therefore $\liminf_{t\to\infty} |q(\mathbf{G}^{i,t}) - \hat{q}| \leqslant m\psi_1(\alpha, Q, \varphi), \;\; i = 1, \cdots, m$, which guarantees convergence of the dual function to a bounded error floor around its optimal value. The requirement $\varphi \geqslant \bar{\varphi}$ is not too restrictive, as explained in [20].

Based on the definition of $\mathbf{g}_j^{i,t}$ and the definition of $\mathbf{v}_j^{i,t}$, let us define the two average vectors $\bar{\mathbf{g}}^t$ and

$\bar{\mathbf{v}}^t$, defined as

$$\bar{\mathbf{g}}^t = \Big[\frac{1}{m}\sum_{i=1}^{m}(\mathbf{g}_1^{i,t})^T, \ldots, \frac{1}{m}\sum_{i=1}^{m}(\mathbf{g}_n^{i,t})^T\Big]^T, \tag{55}$$

$$\bar{\mathbf{v}}^t = \Big[\frac{1}{m}\sum_{i=1}^{m}(\mathbf{v}_1^{i,t})^T, \ldots, \frac{1}{m}\sum_{i=1}^{m}(\mathbf{v}_n^{i,t})^T\Big]^T, \tag{56}$$

as well as the following two supporting sequences

$$\mathbf{y}^t = \mathcal{P}_{\mathcal{G}}[\bar{\mathbf{v}}^{t-1}], \quad \mathbf{d}^t = \bar{\mathbf{g}}^t - \mathbf{y}^t. \tag{57}$$

For the supporting sequence $\mathbf{y}^t$ the following lemmas hold.

Lemma 2. *The sequence $\mathbf{y}^t$ is updated with an ϵ-subgradient method [30] to maximize $q(\mathbf{y})$, that is*

$$\mathbf{y}^{t+1} = \mathcal{P}_{\mathcal{G}}\Big[\mathbf{y}^t + \frac{\alpha}{m}\tilde{\mathbf{h}}^t\Big], \tag{58}$$

where the vector

$$\tilde{\mathbf{h}}^t = \sum_{i=1}^{m}(\mathbf{h}^{i,t} + \mathbf{d}^t/\alpha), \tag{59}$$

with

$$\mathbf{h}^{i,t} = [(\mathbf{h}_1^{i,t})^T, \ldots, (\mathbf{h}_n^{i,t})^T]^T, \tag{60}$$

is an ϵ-subgradient of $q(\mathbf{y})$ and $\epsilon = m\psi_2(\alpha, Q, \varphi)$. Notably, $\mathcal{P}_{\mathcal{G}}[.]$ stands for projection onto the convex set $\mathcal{G}$ and ψ_2 is a positive function of φ, α, and Q. Furthermore, $\mathbf{d}^t/\alpha$ is bounded, i.e., $\|\mathbf{d}^t/\alpha\| \leqslant \tau$, for a certain non-negative scalar τ, and

$$q_i(\mathbf{y}) \leqslant q(\mathbf{y}^t) + (\tilde{\mathbf{h}}^t)^T(\mathbf{y} - \mathbf{y}^t) + \psi_2(\alpha, Q, \varphi), \ \forall \mathbf{y} \in \mathcal{G}. \tag{61}$$

Proof: The proof follows from the definition of the supporting sequences $\mathbf{y}^t$ and $\mathbf{d}^k$ in (57) and, in particular, directly from [20, Lemma 5 and Theorem 2] applied to our update sequence (51). ■

Lemma 3. *For the supporting sequence $\mathbf{y}^t$ the followings hold.*

(a)

$$-\sum_{k=1}^{t}(\tilde{\mathbf{h}}^k)^T\mathbf{y}^k \leqslant \frac{\|\mathbf{y}^1\|^2}{2\alpha/m} + t\,\frac{\alpha m\,(\sqrt{n}\,Q + \tau)^2}{2}. \tag{62}$$

(b)

$$+ \quad \sum_{k=1}^{t} (\tilde{\mathbf{h}}^k)^T \hat{\mathbf{y}} \qquad \leqslant \qquad \frac{\|\mathbf{y}^1 - 2\hat{\mathbf{y}}\|^2}{2\alpha/m} \quad + \quad t\,\frac{\alpha m\,(\sqrt{n}\,Q+\tau)^2}{2} \quad + \quad t\,\psi_2(\alpha, Q, \varphi), \quad (63)$$

where $\hat{\mathbf{y}}$ is the optimal dual variable.

Proof: The result is rather standard and applies to any ϵ-subgradient method. A concise proof for the case $\epsilon = 0$, can be found in [18, Proposition 3-(a)]; extending it to any $\epsilon \geqslant 0$ is straightforward. ■

B. Primal Objective Convergence

In this subsection, we investigate the convergence of the running average cost $\|\hat{\mathbf{w}}^t\|_1$ to the optimal value of the primal cost $\|\hat{\mathbf{w}}\|_1$. Our analysis is formulated in the following theorem.

Theorem 2. *Convergence of the primal running average sequence $\{\hat{\mathbf{w}}^t, \hat{\mathbf{u}}^t\}$ can be formulated as follows.*

(a) *The running average cost is upper bounded as*

$$\|\hat{\mathbf{w}}^t\|_1 \quad \leqslant \quad \|\hat{\mathbf{w}}\|_1 \;+\; \frac{nG^2}{2t\alpha/m} \;+\; \frac{\alpha m\,(\sqrt{n}\,Q+\tau)^2}{2} \;+\; \tau m\sqrt{n}\,G \;+\; m\psi_2(\alpha, Q, \varphi). \quad (64)$$

(b) *The running average cost is lower bounded as*

$$\|\hat{\mathbf{w}}^t\|_1 \quad \geqslant \quad \|\hat{\mathbf{w}}\|_1 \;-\; \frac{9nG^2}{2t\alpha/m} \;-\; \frac{\alpha m\,(\sqrt{n}\,Q+\tau)^2}{2} \;-\; \tau m\sqrt{n}\,G \;-\; m\psi_2(\alpha, Q, \varphi). \quad (65)$$

where $\psi_2(\alpha, Q, \nu, \varphi)$ is a positive function, monotonically increasing with α and decreasing with φ. The non-negative scalar τ is defined in Lemma 2.

Proof: The proof is an adaptation of Proposition 3 in [18]. Let us start with claim (a). From convexity of the primal cost $\|\cdot\|_1$ and the definition of $\bar{w}_i^t$ as a minimizer of the local Lagrangian functions, we have for $t \geqslant 1$,

$$\|\hat{\mathbf{w}}^t\|_1 \leqslant \frac{1}{t}\sum_{k=1}^{t} \|\bar{\mathbf{w}}^k\|_1 = \frac{1}{t}\sum_{k=1}^{t}\sum_{i=1}^{m} \Big(q_i(\mathbf{g}^{i,k}) - (\mathbf{g}^{i,k})^T \mathbf{h}^{i,k}\Big), \quad (66)$$

where

$$\mathbf{g}^{i,k} = [(\mathbf{g}_1^{i,k})^T, \ldots, (\mathbf{g}_n^{i,k})^T]^T. \quad (67)$$

From Lemma 2, since $\mathbf{g}^{i,k} \in \mathcal{G}$, we have

$$q_i(\mathbf{g}^{i,k}) \;-\; q_i(\mathbf{y}^k) \quad \leqslant \quad (\mathbf{h}^{i,k})^T \mathbf{g}^{i,k} \;+\; (\mathbf{d}^k/\alpha)^T \mathbf{g}^{i,k} \;-\; (\mathbf{h}^{i,k} \;+\; \mathbf{d}^k/\alpha)^T \mathbf{y}^k \;+\; \psi_2(\alpha, Q, \varphi). \quad (68)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

Next, by summing over i, we obtain

$$\sum_{i=1}^{m} q_i(\mathbf{g}^{i,k}) \;\leqslant\; q(\mathbf{y}^k) \;+\; \sum_{i=1}^{m}(\mathbf{h}^{i,k})^T\mathbf{g}^{i,k} \;+\; \sum_{i=1}^{m}(\mathbf{d}^k/\alpha)^T\mathbf{g}^{i,k} \;-\; (\tilde{\mathbf{h}}^k)^T\mathbf{y}^k \;+\; m\psi_2(\alpha,Q,\varphi), \quad (69)$$

and thus,

$$\|\hat{\mathbf{w}}^t\|_1 \qquad \leqslant \qquad \frac{1}{t}\sum_{k=1}^{t}\Big(q(\mathbf{y}^k) \quad + \quad \sum_{i=1}^{m}(\mathbf{d}^k/\alpha)^T\mathbf{g}^{i,k} \quad - \quad (\tilde{\mathbf{h}}^k)^T\mathbf{y}^k \quad + \quad m\psi_2(\alpha,Q,\varphi)\Big). \quad (70)$$

We can use Lemma 3-(a) to find an upper bound for the term $-(\tilde{\mathbf{h}}^k)^T\mathbf{y}^k$. Besides, since $\|(\mathbf{d}^k/\alpha)^T\mathbf{g}^{i,k}\| \leqslant \|\mathbf{d}^k/\alpha\|\|\mathbf{g}^{i,k}\|$ and we known from Lemma 2 that $\|\mathbf{d}^k/\alpha\| \leqslant \tau$, together with $\|\mathbf{g}^{i,k}\| \leqslant \sqrt{n}\,G$ from our earlier assumption on bounded dual variable estimates, we obtain $\|(\mathbf{d}^k/\alpha)^T\mathbf{g}^{i,k}\| \leqslant \tau\sqrt{n}\,G$. With this in place, we can rewrite (70) as

$$\|\hat{\mathbf{w}}^t\|_1 \quad \leqslant \quad \frac{1}{t}\sum_{k=1}^{t} q(\mathbf{y}^k) \;+\; \frac{\|\mathbf{y}^1\|^2}{2t\alpha/m} \;+\; \frac{\alpha m\,(\sqrt{n}\,Q+\tau)^2}{2} \;+\; m\tau\sqrt{n}\,G \;+\; m\psi_2(\alpha,Q,\varphi). \quad (71)$$

We known that by optimality $q(\mathbf{y}^k) \leqslant \hat{q}$, by strong duality $\hat{q} = \|\hat{\mathbf{w}}\|_1$, and $\|\mathbf{y}^1\|^2 \leqslant nG^2$. Therefore, the claim (a) follows.

As for claim (b), given any optimal dual solution $\hat{\mathbf{y}}$, we have

$$\|\hat{\mathbf{w}}^t\|_1 = \underbrace{\|\hat{\mathbf{w}}^t\|_1 + (\hat{\mathbf{y}})^T\Big(\frac{1}{t}\sum_{k=1}^{t}\tilde{\mathbf{h}}^k\Big)}_{\omega} - (\hat{\mathbf{y}})^T\Big(\frac{1}{t}\sum_{k=1}^{t}\tilde{\mathbf{h}}^k\Big). \quad (72)$$

We also know that,

$$\begin{aligned}\omega \;&=\; \|\hat{\mathbf{w}}^t\|_1 + (\hat{\mathbf{y}})^T\Big(\frac{1}{t}\sum_{k=1}^{t}\sum_{i=1}^{m}\mathbf{h}^{i,k}\Big) + m(\hat{\mathbf{y}})^T\Big(\frac{1}{t}\sum_{k=1}^{t}\mathbf{d}^k/\alpha\Big)\\ &\geqslant\; \|\hat{\mathbf{w}}^t\|_1 + (\hat{\mathbf{y}})^T\Big(\sum_{i=1}^{m}\mathbf{h}^{i,k}(\hat{\mathbf{w}}^k)\Big) - m\,\sqrt{n}\,G\,\tau. \end{aligned} \quad (73)$$

Furthermore, owing to the saddle point property of the Lagrangian function, i.e.,

$$\mathcal{L}(\hat{\mathbf{w}}^k,\hat{\mathbf{u}}^k,\hat{\mathbf{y}}) \geqslant \mathcal{L}(\hat{\mathbf{w}},\hat{\mathbf{u}},\hat{\mathbf{y}}) = \hat{q} = \|\hat{\mathbf{w}}\|_1, \quad (74)$$

we can write

$$\|\hat{\mathbf{w}}^t\|_1 \;+\; (\hat{\mathbf{y}})^T\Big(\sum_{i=1}^{m}\mathbf{h}^{i,k}(\hat{\mathbf{w}}^k)\Big) \;-\; m\sqrt{n}\,G\tau \;=\; \mathcal{L}(\hat{\mathbf{w}}^k,\hat{\mathbf{u}}^k,\hat{\mathbf{y}}) - m\sqrt{n}\,G\tau \;\geqslant\; \|\hat{\mathbf{w}}\|_1 - m\sqrt{n}\,G\tau. \quad (75)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2015.2460224, IEEE Transactions on Signal Processing

We can now find an upper bound for the term $(\hat{\mathbf{y}})^T\big(\frac{1}{t}\sum_{k=1}^{t}\tilde{\mathbf{h}}^k\big)$ in (72) as in Lemma 3-(b). By substituting this bound in (72) and by combining it with (73)-(75), we obtain

$$\|\hat{\mathbf{w}}^t\|_1 \quad \geqslant \quad \|\hat{\mathbf{w}}\|_1 \;-\; m\tau\sqrt{n}G \;-\; \frac{\|\mathbf{y}^1-2\hat{\mathbf{y}}\|^2}{2t\alpha/m} \;-\; \frac{\alpha m(\sqrt{n}G+\tau)^2}{2} \;-\; m\psi_2(\alpha,Q,\varphi), \quad (76)$$

and since $\|\mathbf{y}^1-2\hat{\mathbf{y}}\|^2 \leqslant 9nG^2$, the claim follows. ■

REFERENCES

[1] H. Godrich, A. P. Petropulu, and H. V. Poor, "Sensor selection in distributed multiple-radar architectures for localization: A knapsack problem formulation," *IEEE Trans. on Sig. Proc.*, vol. 60, no. 1, pp. 6033–6043, 2012.

[2] D. Bojavić and J. X. B. Sinopoli, "Sensor selection for event detection in wireless sensor networks," *IEEE Trans. on Sig. Proc.*, vol. 59, no. 10, pp. 4938–4953, 2011.

[3] E. Pollakis, R. L. G. Renato, and S. Stańczak, "Base station selection for energy efficient network operation with the majorization-minimization algorithm," *Proc. of IEEE Intl. Workshop on Sig. Proc. Advances in Wireless Communications (SPAWC)*, pp. 219–223, June 2012.

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

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

[6] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*. Athena Scientific, 1997.

[7] M. Rabbat and R. Nowak, "Distributed optimization in sensor networks," *Proc. of Information Processing in Sensor Networks (IPSN)*, pp. 20–27, Apr. 2004.

[8] 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.

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

[10] 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.

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

[12] 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.

[13] 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.

[14] E. Rigtrop, "Sensor selection with correlated noise," 2010.

[15] H. Jamali-Rad, A. Simonetto, X. Ma, and G. Leus, "Sparsity-aware sensor selection for correlated noise," *Accepted to appear in Proc. of Intl. Conf. on Information Fusion (FUSION)*, July 2014.

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

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[18] 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.

[19] 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.

[20] 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.

[21] G. H. Golub and C. V. Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 3rd ed., 1996.

[22] E. Cuthill and J. McKee, "Reducing the bandwidth of sparse symmetric matrices," *In Proc. of 24th National Confernce ACM*, pp. 157 – 172, 1969.

[23] U. A. Khan and J. F. Moura, "Distributed iterative-collapse inversion (DICI) algorithm for L-banded matrices," *Proc. of IEEE ICASSP*, pp. 2529 – 2532, Apr. 2008.

[24] 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, 2010.

[25] F. Iutzeler, P. Bianchi, P. Ciblat, and W. Hachem, "Explicit convergence rate of a distributed alternating direction method of multipliers," *IEEE Transactions on Automatic Control (to appear)*, 2015.

[26] E. Ghadimi, *Accelerating Convergence of Large-scale Optimization Algorithms*. PhD thesis, KTH, 2015.

[27] S. Liu, M. Fardad, and E. M. andP. K. Varshney, "Optimal periodic sensor scheduling in networks of dynamical systems," *IEEE Trans. on Sig. Proc.*, vol. 62, pp. 3055–3068, Jun. 2014.

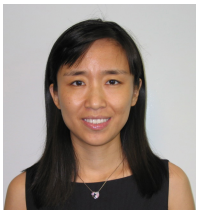
[28] CVX Research Inc., "CVX: Matlab software for disciplined convex programming, version 2.0 beta." http://cvxr.com/cvx, Sept. 2012.

[29] A. Simonetto and H. Jamali-Rad, "Primal recovery from consensus-based dual decomposition for distributed convex optimization," *Journal of Optimization Theory and Applications (to appear)*, 2015.

[30] K. Kiwiel, "Convergence of approximate and incremental subgradient methods for convex optimization," *SIAM Journal of Optimization*, vol. 14, no. 3, pp. 807–840, 2004.

Hadi Jamali-Rad (M'15) received the PhD degree (Cum Laude) from the Delft University of Technology (TU Delft), the Netherlands. He is currently a research scientist at the Shell Global Solutions International, the Netherlands. In 2012, he was a visiting researcher at the signals, identication, system theory and automation (SISTA) division of the Katholieke Universiteit Leuven (KU Leuven), Leuven, Belgium, for which he won a European Erasmus grant. In 2014, he visited the Center for Signal and Information Processing (CSIP) at the Georgia Institute of Technology (Georgia Tech). His general interests are in the area of signal processing for communications, wireless networking, and geophysics. Dr. Jamali-Rad has served as a reviewer for several flagship journals and major conferences.

Andrea Simonetto (M'12) Andrea Simonetto received the M.Sc. degree in space engineering (cum laude) from both Politecnico di Milano and Politecnico di Torino, Italy, in 2008, and the Ph.D. degree in systems and control from the Delft University of Technology, Delft, The Netherlands, in 2012. He is currently a Post-Doctoral Researcher with the Circuits and Systems Group, Electrical Engineering Department, Delft University of Technology. He was a Visiting Researcher with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA and with KTH, Royal Institute of Technology, Stockholm, Sweden. His current research interests include distributed estimation, control, and optimization with applications in sensor networks and mobile robotics. For his Ph.D. work, he was awarded the DISC Ph.D. Thesis award 2012.



Xiaoli Ma (SM'09) received the B.S. degree in Automatic Control from Tsinghua University, Beijing, China in 1998, the M.S. degree in Electrical Engineering from the University of Virginia in 2000, and the Ph.D. degree in Electrical Engineering from the University of Minnesota in 2003. After receiving her Ph.D., Dr. Ma joined the Department of Electrical and Computer Engineering at Auburn University, where she served as an assistant professor until Dec. 2005. Since Jan. 2006, she has been with the School of Electrical and Computer Engineering at Georgia Tech, where she is currently a professor. She was awarded the Lockheed Martin Aeronautics Company Dean's Award for Teaching Excellence by the College of Engineering in 2009, and Outstanding Junior Faculty Award by the School of Electrical and Computer Engineering in 2010, at Georgia Tech. Her research focuses on wireless networking and communications, including network performance analysis, transceiver designs for wireless time- and frequency selective channels, channel estimation and equalization algorithms, carrier frequency synchronization for OFDM systems, performance analysis and cooperative designs for wireless networks. Dr. Ma is a senior member of IEEE, serves as a Senior Area Editor for IEEE Signal Processing Letters since 2014 and Elsevier Digital Signal Processing since June 2012, and has been an Associate Editor for IEEE Signal Processing Letters (2007-2009) and IEEE Transactions on Wireless Communications (2008-2013).

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