

Sparsity-Aware Sensor Selection: Centralized and Distributed Algorithms

Hadi Jamali-Rad, Andrea Simonetto, and Geert Leus

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 explore the sparsity embedded within the problem and propose a relaxed sparsity-aware sensor selection approach which is equivalent to the unrelaxed problem under certain conditions. We also present a reasonably low-complexity and elegant distributed version of the centralized problem with convergence guarantees such that each sensor can decide itself whether it should contribute to the estimation or not. 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

Manuscript received August 19, 2013; revised December 23, 2013; accepted December 27, 2013. Date of publication January 02, 2014; date of current version January 13, 2014. This work was supported in part by NWO-STW under the VICI program (Project 10382), and in part by STW under the D2S2 project from the ASSYS program (Project 10561). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Alireza Seyedi.

The authors are with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2826 CD Delft, The Netherlands (e-mail: h.jamali-rad@tudelft.nl; a.simonetto@tudelft.nl; g.j.t.leus@tudelft.nl).

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

Digital Object Identifier 10.1109/LSP.2013.2297419

to exploit the sparsity embedded within the problem. From this angle, our look 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. Another relevant problem, but of a different nature, is considered in [4], where a distributed algorithm is designed to identify the sensors containing relevant information by a sparsity-aware decomposition of the measurement covariance matrix. Finally, in [5], two distributed implementations of [1] based on a truncated Newton algorithm are proposed. Compared to our work, first, [5] deals with a slightly different problem. Second, it considers a log-barrier and truncated Hessian approximate of the relaxed problem with no convergence (error) guarantees. Third, private sensor information has to be broadcast in this approach whereas we avoid that. Finally, our complexity is not a function of the number of sensors but of the number of sensed dimensions, and hence, it is considerably lower.

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

$$y_i = \mathbf{a}_i^T \mathbf{x} + \eta_i, \quad i = 1, \dots, m, \quad (1)$$

where the \mathbf{a}_i 's $\in \mathbb{R}^n$ span \mathbb{R}^n ($m \gg n$) and η_i is an additive zero-mean white measurement noise. Notably, 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 $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_m]^T$. Now, we can construct $\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\eta}$, where $\mathbf{y} = [y_1, \dots, y_m]^T$, and $\boldsymbol{\eta} = [\eta_1, \dots, \eta_m]^T$. For the linear measurement model (1), the MSE can be expressed as $\text{MSE} = \mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2] =$

$\text{tr}(\mathbf{A}^T \mathbf{C}^{-1} \mathbf{A})^{-1}$, where $\text{tr}(\cdot)$ stands for the trace operator and \mathbf{C} is the covariance matrix of the noise vector η [6]. Let the noise at the sensor nodes be uncorrelated, i.e., $\mathbb{E}[\eta_i \eta_j^T] = \sigma_i^2 \delta(i-j)$ with $\delta(\cdot)$ denoting the Kronecker delta, and thus $\mathbf{C} = \text{diag}([\sigma_1^2, \dots, \sigma_m^2])$. Based on this assumption, the MSE can be reformulated as

$$\text{MSE} = \text{tr} \left(\sum_{i=1}^m \tilde{\mathbf{a}}_i \tilde{\mathbf{a}}_i^T \right)^{-1}, \quad (2)$$

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

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

where the variable $w_i \in \{0, 1\}$ encodes whether the i -th sensor (measurement) is to be used. In practice, only a few sensors should be activated to satisfy the MSE constraint. Therefore, the problem can be cast as the following optimization program

$$\underset{\mathbf{w} \in \{0,1\}^m, \mathbf{u}}{\text{minimize}} \quad \|\mathbf{w}\|_0 \quad (4a)$$

$$\text{s.t.} \quad \left[\begin{array}{c|c} \sum_{i=1}^m w_i \tilde{\mathbf{a}}_i \tilde{\mathbf{a}}_i^T & \mathbf{e}_j \\ \hline \mathbf{e}_j^T & u_j \end{array} \right] \succeq 0, \quad j = 1, \dots, n, \quad (4b)$$

$$\|\mathbf{u}\|_1 \leq \gamma, \quad u_j \geq 0, \quad j = 1, \dots, n, \quad (4c)$$

where $\mathbf{w} = [w_1, \dots, w_m]^T$ is the selection vector, $\mathbf{u} = [u_1, \dots, u_n]^T$ 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 (4b) and (4c) represented by Ω_γ are a more suitable representation of the original constraint (3), obtained using the Schur complement [7]. We denote the solution to (4) as $(\mathbf{w}^*, \mathbf{u}^*)$. Since $\|\mathbf{w}\|_0$ in the cost function of (4) and the finite-alphabet constraint on the w_i 's are both non-convex, we consider the following relaxed version of the problem called sparsity-aware sensor selection (SparSenSe)

$$(\hat{\mathbf{w}}, \hat{\mathbf{u}}) := \arg \min_{\mathbf{w} \in [0,1]^m, \mathbf{u}} \{ \|\mathbf{w}\|_1, \text{s.t.} (\mathbf{w}, \mathbf{u}) \in \Omega_\gamma \}. \quad (5)$$

IV. EQUIVALENCE THEOREM

In this section, we present an equivalence result, i.e., we prove that provided some conditions, the number of selected sensors is the same for both the original problem (4) and the relaxed version (5). To this aim, the following simplifying assumptions will be employed in this section.

Assumption 1: Only one element of $\tilde{\mathbf{a}}_i$ is non-zero, i.e., there exists a single j for which $\tilde{a}_{i,j} \neq 0$.

This assumption implies that the sensors can only sense one element (dimension) of the vector of interest. Let $\mathcal{I}_j, \forall j$ be the set containing the indices of the sensors which sense the j -th dimension, i.e., $\mathcal{I}_j = \{i \in \{1, 2, \dots, m\} | \tilde{a}_{i,j} \neq 0\}$ and \mathcal{V}_j be the set containing the corresponding values, i.e., $\mathcal{V}_j = \{\tilde{a}_{i,j} | i \in \mathcal{I}_j\}$.

Assumption 2: For each \mathcal{V}_j there exists an $i \in \mathcal{I}_j$ such that $|\tilde{a}_{i,j}| > |\tilde{a}_{k,j}|, \forall k \in \mathcal{I}_j, k \neq i$. We denote this $\tilde{a}_{i,j}$ as v_j^* .

This assumption states that for each dimension j there exists a unique dominant sensor. Based on this, the following proposition and its proof are in place.

Proposition 1: Under Assumption 1 and Assumption 2, there exists a lower bound $\gamma^* = \max_j 1/|v_j^*| \sum_{j=1}^n 1/|v_j^*|$, such that

if $\gamma \geq \gamma^*$, then $\|\hat{\mathbf{w}}\|_0 = \|\mathbf{w}^*\|_0 = n$. In addition, in this case, the solution of the relaxed version (5) is unique and corresponds to activating the sensors with v_j^* in the regressors.

Proof: The solution of the original non-convex problem has a cardinality of at least n , i.e., $\|\mathbf{w}^*\|_0 \geq n$. This is because we need to activate at least n sensors to attain a finite MSE in (3). Furthermore, in general, $\|\hat{\mathbf{w}}\|_0 \geq \|\mathbf{w}^*\|_0$. In the following, we will show that under Assumption 1 and Assumption 2, $\|\hat{\mathbf{w}}\|_0 = n$, and therefore our claim holds. The core idea is that under the aforementioned assumptions we can analytically compute $(\hat{\mathbf{w}}, \hat{\mathbf{u}})$ as explained in the following. The linear matrix inequality constraint (4b) can be written as

$$\mathbf{e}_j^T \left(\sum_{i=1}^m w_i \tilde{\mathbf{a}}_i \tilde{\mathbf{a}}_i^T \right)^{-1} \mathbf{e}_j \leq u_j, \quad j = 1, \dots, n. \quad (6)$$

Under Assumption 1, we can write

$$\sum_{i=1}^m w_i \tilde{\mathbf{a}}_i \tilde{\mathbf{a}}_i^T = \text{diag} \left(\sum_{i \in \mathcal{I}_1} w_i a_{i,1}^2, \dots, \sum_{i \in \mathcal{I}_n} w_i a_{i,n}^2 \right),$$

which is not singular (and therefore the MSE is finite) when we select at least one sensor per dimension. This means that (6) yields

$$\sum_{i \in \mathcal{I}_j} w_i \tilde{a}_{i,j}^2 \geq u_j^{-1}, \quad j = 1, \dots, n. \quad (7)$$

Considering (7) and the fact that we need to minimize $\|\mathbf{w}\|_1$, we have to maximize the u_j 's w.r.t. the constraints $\|\mathbf{u}\|_1 \leq \gamma$ and $u_j \geq 0, \forall j$, which leads to $\|\hat{\mathbf{u}}\|_1 = \gamma$. Given any \mathbf{u} , we can compute $w_i(\mathbf{u})$ analytically since the relaxed problem can now be written as the following linear program (LP)

$$\underset{\mathbf{w} \in [0,1]^m}{\text{minimize}} \quad \|\mathbf{w}\|_1 \quad (8a)$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{I}_j} w_i \tilde{a}_{i,j}^2 \geq u_j^{-1}, \quad j = 1, \dots, n. \quad (8b)$$

The solution of this LP lies on the vertices of the polytope defining the constraints (following Assumption 2) and for $u_j \geq 1/v_j^{*2}$ it is given by

$$w_i(\mathbf{u}) = \begin{cases} (u_j v_j^{*2})^{-1} & \text{if } \tilde{a}_{i,j} = v_j^*, \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

This helps us to rewrite (5) as

$$\hat{\mathbf{u}} = \arg \min_{\mathbf{u}, u_j \geq 1/v_j^{*2}} \left\{ \sum_{j=1}^n (u_j v_j^{*2})^{-1}, \text{s.t.} \|\mathbf{u}\|_1 = \gamma \right\}, \quad (10)$$

which is convex for $u_j \geq 0$. The optimal $\hat{\mathbf{u}}$ has to satisfy the KKT conditions given by

$$\hat{u}_j^{-2} = \lambda v_j^{*2}, \quad j = 1, \dots, n, \quad (11a)$$

$$\|\hat{\mathbf{u}}\|_1 = \gamma, \quad (11b)$$

where λ is the Lagrange multiplier associated with $\|\mathbf{u}\|_1 = \gamma$. From (11a), $\lambda \geq 0$; solving for \hat{u}_j and substituting it into (11b), after some simplifications, leads to

$$\hat{u}_j = \frac{\gamma}{|v_j^*| \sum_{j=1}^n 1/|v_j^*|}, \quad j = 1, \dots, n, \quad (12)$$

which due to the convexity of (10) is the unique optimizer of (10). Substituting (12) back into (9) yields

$$\hat{w}_i = \begin{cases} \frac{\sum_{j=1}^n 1/|v_j^*|}{\gamma |v_j^*|} & \text{if } \tilde{a}_{i,j} = v_j^*, \\ 0 & \text{otherwise,} \end{cases} \quad (13)$$

for $\hat{u}_j \geq 1/v_j^{*2}$, i.e., $\gamma \geq \max_j 1/|v_j^*| \sum_{j=1}^n 1/|v_j^*| = \gamma^*$. Thus, for $\gamma \geq \gamma^*$, $\hat{\mathbf{w}}$ is unique and has cardinality n . ■

V. 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. Let us start with some notations. We call \mathcal{N}_i the neighborhood set of the i -th sensor including i itself, with cardinality $|\mathcal{N}_i| = N_i$ (either given or to be estimated). We also define the following convex sets:

$$\mathcal{W}_i = \{w_i | 0 \leq w_i \leq 1\}, \quad (14)$$

$$\mathcal{U} = \{\mathbf{u} | u_j \geq 0, \sum_{j=1}^n u_j \leq \gamma\}, \quad (15)$$

and form the Lagrangian of the problem (5) given by

$$\begin{aligned} \mathcal{L} &= \sum_{i=1}^m w_i - \sum_{j=1}^n \text{tr} \left(\left[\begin{array}{c|c} \sum_{i=1}^m w_i \tilde{\mathbf{a}}_i \tilde{\mathbf{a}}_i^T & \mathbf{e}_j \\ \hline \mathbf{e}_j^T & u_j \end{array} \right] \mathbf{G}_j \right) \\ &= \sum_{i=1}^m w_i - \sum_{j=1}^n \sum_{i=1}^m \text{tr} \left(\left[\begin{array}{c|c} w_i \tilde{\mathbf{a}}_i \tilde{\mathbf{a}}_i^T & \mathbf{e}_j/m \\ \hline \mathbf{e}_j^T/m & u_j/m \end{array} \right] \mathbf{G}_j \right) \\ &= \sum_{i=1}^m \left(w_i - \sum_{j=1}^n \text{tr} \left(\left[\begin{array}{c|c} w_i \tilde{\mathbf{a}}_i \tilde{\mathbf{a}}_i^T & \mathbf{e}_j/m \\ \hline \mathbf{e}_j^T/m & u_j/m \end{array} \right] \mathbf{G}_j \right) \right) \\ &= \sum_{i=1}^m \mathcal{L}_i(w_i, \mathbf{u}, \mathbf{G}), \end{aligned} \quad (16)$$

where $\mathbf{G}_j \succeq 0, \forall j$ are appropriately sized dual variables, and $\mathbf{G} = [\mathbf{G}_1, \dots, \mathbf{G}_n]$. The dual function defined on \mathcal{L} can be given by

$$\begin{aligned} q(\mathbf{G}) &= \min_{w_i \in \mathcal{W}_i, \mathbf{u} \in \mathcal{U}} \sum_{i=1}^m \mathcal{L}_i(w_i, \mathbf{u}, \mathbf{G}) \\ &= \sum_{i=1}^m \left(\min_{w_i \in \mathcal{W}_i, \mathbf{u} \in \mathcal{U}} \mathcal{L}_i(w_i, \mathbf{u}, \mathbf{G}) \right) = \sum_{i=1}^m q_i(\mathbf{G}). \end{aligned} \quad (17)$$

Notably, since both \mathcal{W}_i and \mathcal{U} are convex and compact sets, given a certain value of \mathbf{G} , the functions $q_i(\mathbf{G})$ and their subgradient w.r.t. \mathbf{G} , called \mathbf{Q} and defined later on, can be computed locally (for example using SeDuMi to solve the resulting LPs) at each sensor [8].

Whenever γ is large enough so that we expect sparse solutions in terms of $\hat{\mathbf{w}}$, Slater's condition holds for (4). In fact, in this case, we can always find a pair $(\hat{\mathbf{w}}, \hat{\mathbf{u}})$ that satisfies (4b) and (4c) strictly. Therefore, the original ℓ_1 -regularization (5) leads to the differentiable dual optimization problem

$$\underset{\mathbf{G}_1 \succeq 0, \dots, \mathbf{G}_n \succeq 0}{\text{maximize}} \sum_{i=1}^m q_i(\mathbf{G}), \quad (18)$$

with zero duality gap. This convex optimization program can be solved iteratively in a distributed fashion using a variety of algorithms. For instance, we can use gradient-based methods, such as the dual averaging scheme of [9] with a variable stepsize, or the simpler dual subgradient of [8] with a fixed stepsize. The latter method has the advantage of providing a recovery mechanism for the primal solution (i.e., we recover $\hat{\mathbf{w}}$ as a by-product of the optimal \mathbf{G} , which is in fact our goal). Furthermore, the

subgradient method of [8] has the benefit to employ a fixed stepsize giving explicit trade-offs in terms of accuracy and feasibility of the solution and the number of iterations. In particular, given the number of iterations t , and the stepsize α , we can prove that (see [8, Proposition 1])

$$\text{tr} \left(\sum_{i=1}^m w_i^t \tilde{\mathbf{a}}_i \tilde{\mathbf{a}}_i^T \right)^{-1} \leq \gamma + \frac{c^2}{t\alpha}, \quad (19)$$

where w_i^t is the recovered approximate primal solution for sensor i at iteration t , and c^2 is a positive constant that depends on the problem at hand. This equation tells us *a priori* how many iterations we need to run before we reach a given infeasibility level; or provides us with a bound on how much we should tighten the constraint on γ to guarantee feasibility w.r.t. the MSE constraint for finite t .

In order to implement the dual subgradient of [8], each node requires a copy of $\mathbf{Q} = \sum_{i=1}^m \nabla_{\mathbf{G}^i} q_i(\mathbf{G}^i)$. This can be circumvented by using the method of [10] where the local sensor nodes have different local copies of \mathbf{Q} , say \mathbf{Q}^i , and they run an inexact consensus procedure for φ times (where $\varphi \in \mathbb{N}^+$). If $\varphi \rightarrow \infty$, we recover the procedure of [8], while if φ is limited we introduce an additional error in the distributed optimization procedure. Our proposed distributed sparsity-aware sensor selection (DiSparSenSe) algorithm can be summarized in Algorithm 1.

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 using an ϵ -subgradient argument as discussed in [8] and [10].

Algorithm 1 DiSparSenSe

1: We call the i -th sensor version of \mathbf{G} at iteration t , $\mathbf{G}^{i,t}$. Let an initial value for $\mathbf{G}^{i,t}$ be given at each sensor node (cold start $\mathbf{G}_j^{i,0} = \mathbf{I}$). Initialize the w_i^t 's with $w_i^0 = 0$.

2: Compute, in parallel at each sensor i , the value of $q_i(\mathbf{G}^{i,t})$, its derivative $\mathbf{Q}^{i,t} = \nabla_{\mathbf{G}^{i,t}} q_i(\mathbf{G}^{i,t})$, and the related optimal primal variable \bar{w}_i . The dimension of $\mathbf{Q}^{i,t}$ is the same as that of $\mathbf{G}^{i,t}$. This step requires the solution of an LP problem whose computational complexity is $O(n^3)$.

3: Following the primal recovery method of [8], compute

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

4: For $\tau = 1$ to φ do

- Send $\mathbf{Q}^{i,t}$ to the neighboring sensor nodes. The communication cost involved is of $O(N_i n^3)$;
- Perform, in parallel, one consensus step as

$$\mathbf{Q}^{i,t} \leftarrow \sum_{l \in \mathcal{N}_i} \mathbf{Q}^{l,t} / N_i.$$

5: Update each sensor node's dual variable and store it in local variables \mathbf{G}^i as

$$\mathbf{G}^{i,t+1} = \mathcal{P}_{\succeq 0} [\mathbf{G}^{i,t} + m\alpha \mathbf{Q}^{i,t}],$$

where $\mathcal{P}_{\succeq 0}[\cdot]$ is the projection operator onto the cone of positive semidefinite matrices. This step requires n singular value decompositions (SVDs), each of which has a computational complexity $O(n^3)$.

6: Go to 2 for the next iteration.

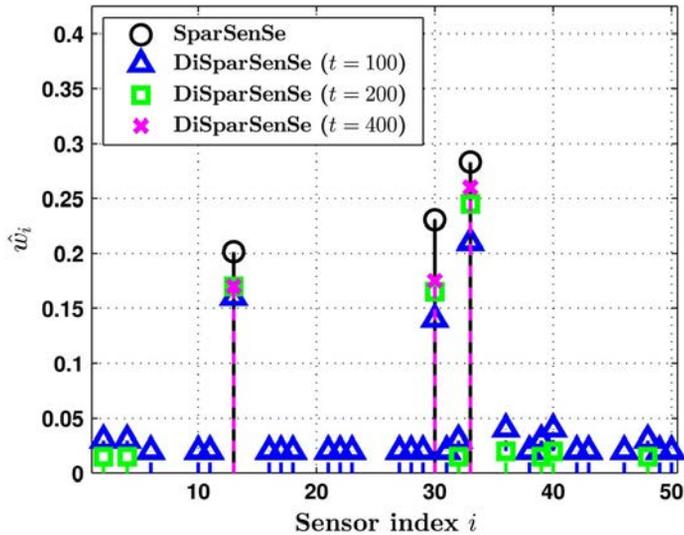


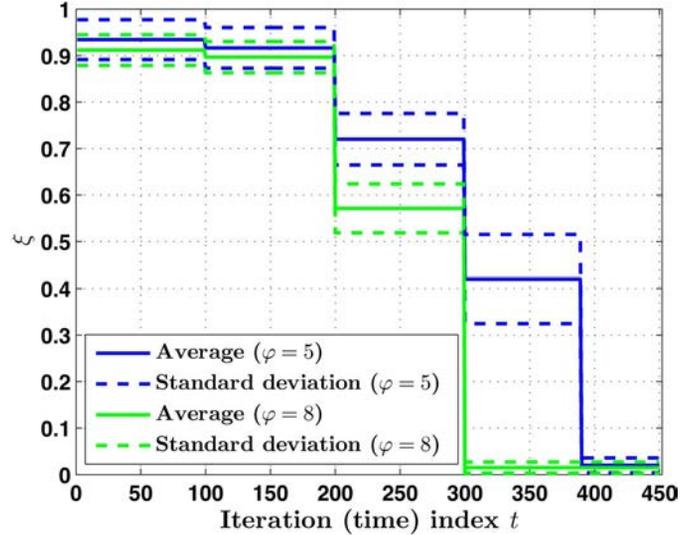
Fig. 1. Centralized versus distributed; selected sensors.

VI. NUMERICAL RESULTS

In this section, we investigate the performance of the proposed algorithms to see if SparSenSe actually selects a few sensors to satisfy the MSE constraint as well as to illustrate that DiSparSenSe selects the same sensors as SparSenSe. 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 $\mathcal{N}(0, 1)$. The noise experienced at different sensors has the same $\sigma = 1/\sqrt{\text{SNR}}$. For DiSparSenSe we assume that the sensors are connected based on a random connectivity graph \mathcal{G} with average node degree of 9. Further, we set the number of consensus steps to $\varphi = 5$ or 8, the step-size to $\alpha = 0.01$ and the SNR to 10 dB. Notably, for SparSenSe, we consider a sensor as active if $w_i > 0$, whereas for DiSparSenSe, due to the fixed step-size error floor, we consider a sensor as active if $w_i > \alpha$.

In the first simulation, depicted in Fig. 1, we plot $\hat{\mathbf{w}}$ estimated by SparSenSe and DiSparSenSe for $\gamma = 1$ and $\varphi = 5$. As can be seen, only 3 sensors (out of 50) are activated by SparSenSe to satisfy our MSE constraint which corroborates the fact that $\hat{\mathbf{w}}$ is sparse. Note that for $t = 100$ many different sensors are activated by DiSparSenSe. However, as expected, by increasing the number of iterations (from $t = 100$ to $t = 400$), the same sensors as for SparSenSe are activated by DiSparSenSe and the magnitude of the related \hat{w}_i 's gets closer to the values estimated by SparSenSe. 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, we also define \mathcal{C} as the set of indices of the selected sensors by SparSenSe and \mathcal{D} as the corresponding set for DiSparSenSe. This helps us to define an equivalence metric between the distributed and centralized algorithms as $\xi = 1 - |\mathcal{C} \cap \mathcal{D}| / \max\{|\mathcal{C}|, |\mathcal{D}|\}$ (i.e., if $\xi = 0$ then $\mathcal{D} \equiv \mathcal{C}$). Again, $\gamma = 1$, and we run 50 independent Monte Carlo trials. The result is shown in Fig. 2, where we clearly observe from the average of the Monte Carlo trials (the solid line) that with increasing t an equivalence is acquired as ξ goes to zero. Finally, the convergence is faster in the case of $\varphi = 8$ compared to $\varphi = 5$.

Fig. 2. Equivalence metric ξ versus t .

VII. DISCUSSION

We would like to conclude this letter by emphasizing the following points. First, note that based on (3) even after rounding the w_i 's to 1 our MSE metric is certainly satisfied. Second, in our distributed algorithm, each sensor itself decides about its status of being active or inactive. More importantly, the “private” information contained in w_i is not broadcast, but instead an “encoded” version $\mathbf{Q}^{i,t}$ is communicated to reach convergence. Furthermore, based upon our earlier explanations, the total computational complexity of DiSparSenSe is $O(n^4)$ per node per iteration which is considerably lower compared to the computational complexity of SparSenSe $O(m^3)$ ($m \gg n$). The communication cost of DiSparSenSe is $O(\varphi N_i n^3)$ per node per iteration which is reasonably low as it is independent of m . Quite a few interesting topics such as a more elaborate equivalence proof and developing centralized and distributed algorithms for the case of correlated noise are left for future work.

REFERENCES

- [1] S. Joshi and S. Boyd, “Sensor selection via convex optimization,” *IEEE Trans. Signal Process.*, 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. Signal Process.*, vol. 59, pp. 3355–3368, Jul. 2011.
- [3] F. Altenbach, S. Corroy, G. Böcherer, and R. Mathar, “Strategies for distributed sensor selection using convex optimization,” in *Proc. IEEE GLOBECOM*, Dec. 2012, pp. 2367–2372.
- [4] I. Schizas, “Distributed informative-sensor identification via sparsity-aware matrix decomposition,” *IEEE Trans. Signal Process.*, vol. 61, pp. 4610–4624, Sep. 2013.
- [5] D. Bickson and D. Dolev, *Distributed Sensor Selection Using a Truncated Newton Method* [Online]. Available: arXiv:0907.0931v2 [cs.IT], submitted for publication
- [6] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
- [7] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [8] A. Nedić and A. Ozdaglar, “Approximate primal solutions and rate analysis for dual subgradient methods,” *SIAM J. Optim.*, vol. 19, no. 4, pp. 1757–1780, 2009.
- [9] J. C. Duchi, A. Agarwal, and M. Wainwright, “Dual averaging for distributed optimization: Convergence analysis and network scaling,” *IEEE Trans. Automat. Contr.*, vol. 57, no. 3, pp. 592–606, 2012.
- [10] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson, “Subgradient methods and consensus algorithms for solving convex optimization problems,” in *Proc. IEEE CDC*, Dec. 2008, pp. 4185–4190.