Distributed Compressed Estimation Based on Compressive Sensing

Songcen Xu, Rodrigo C. de Lamare, *Senior Member, IEEE,* and H. Vincent Poor, *Fellow, IEEE*

***Abstract*—This letter proposes a novel distributed compressed estimation scheme for sparse signals and systems based on compressive sensing techniques. The proposed scheme consists of compression and decompression modules inspired by compressive sensing to perform distributed compressed estimation. A design procedure is also presented and an algorithm is developed to optimize measurement matrices, which can further improve the performance of the proposed distributed compressed estimation scheme. Simulations for a wireless sensor network illustrate the advantages of the proposed scheme and algorithm in terms of convergence rate and mean square error performance.**

***Index Terms*—Distributed compressed estimation, compressive sensing, measurement matrix optimization, sensor networks.**

I. INTRODUCTION

DISTRIBUTED signal processing algorithms are of great importance for statistical inference in wireless networks and applications such as wireless sensor networks (WSNs) [1], [2], [3], [4]. Distributed processing techniques deal with the extraction of information from data collected at nodes that are distributed over a geographic area [1]. In this context, for each node a set of neighbor nodes collect and process their local information, and transmit their estimates to a specific node. Then, each specific node combines the collected information together with its local estimate to generate improved estimates.

In many scenarios, the unknown parameter vector to be estimated can be sparse and contain only a few nonzero coefficients. Many algorithms have been developed in the literature for sparse signal estimation [5], [6], [7], [8], [9], [10], [11], [12]. However, these techniques are designed to take into account the full dimension of the observed data, which increases the computational cost, slows down the convergence rate and degrades mean square error (MSE) performance.

Compressive sensing (CS) [13], [14] has recently received considerable attention and been successfully applied to diverse fields, e.g., image processing [15], wireless communications [16] and MIMO radar [17]. The theory of CS states that an S–sparse signal $\boldsymbol{\omega}_0$ of length M can be recovered exactly with high probability from $\mathcal{O}(S\log M)$ measurements. Mathematically, the vector $\bar{\boldsymbol{\omega}}_0$ with dimension $D\times 1$ that carries sufficient information about $\boldsymbol{\omega}_0$ ($D \ll M$) can be obtained via a linear model [14]

$$\bar{\boldsymbol{\omega}}_0 = \boldsymbol{\Phi}\boldsymbol{\omega}_0 \tag{1}$$

where $\boldsymbol{\Phi} \in R^{D\times M}$ is the measurement matrix.

The application of CS to WSNs has been recently investigated in [16], [18], [19], [20]. A compressive wireless sensing scheme was developed in [16] to save energy and bandwidth, where CS is only employed in the transmit layer. In [18], a greedy algorithm called precognition matching pursuit was developed for CS and used at sensors and the fusion center to achieve fast reconstruction. However, the sensors are assumed to capture the target signal perfectly with only measurement noise. The work of [19] introduced a theory for distributed CS based on jointly sparse signal recovery. However, in [19] CS techniques are only applied to the transmit layer, whereas distributed CS in the estimation layer has not been widely investigated. A sparse model that allows the use of CS for the online recovery of large data sets in WSNs was proposed in [20], but it assumes that the sensor measurements could be gathered directly, without an estimation procedure. In summary, prior work has focused on signal reconstruction algorithms in a distributed manner but has not considered both compressed transmit strategies and estimation techniques.

In this work, we focus on the design of an approach that exploits lower dimensions, reduces the required bandwidth, and improves the convergence rate and the MSE performance. Inspired by CS, we introduce a scheme that incorporates compression and decompression modules into the distributed estimation procedure. In the compression module, we compress the unknown parameter $\boldsymbol{\omega}_0$ into a lower dimension. As a result, the estimation procedure is performed in a compressed dimension. After the estimation procedure is completed, the decompression module recovers the compressed estimator into its original dimension using an orthogonal matching pursuit (OMP) algorithm [21], [22], [23]. We also present a design procedure and develop an algorithm to optimize the measurement matrices, which can further improve the performance of the proposed scheme. Specifically, we derive an adaptive s-tochastic gradient recursion to update the measurement matrix. Simulation results illustrate the performance of the proposed scheme and algorithm against existing techniques.

This paper is organized as follows. Section II describes the system model. In Section III, the proposed distributed compressed estimation scheme is introduced. The proposed measurement matrix optimization is illustrated in Section IV. Simulation results are provided in Section V. Finally, we conclude the paper in Section VI.

Notation: We use boldface uppercase letters to denote matrices and boldface lowercase letters to denote vectors. We use $(\cdot)^{-1}$ to denote the inverse operator, $(\cdot)^H$ for conjugate transposition and $(\cdot)^*$ for complex conjugate.

II. SYSTEM MODEL AND PROBLEM STATEMENT

A wireless sensor network (WSN) with N nodes, which have limited processing capabilities, is considered with a partially

This work has been funded by CNPq, FAPERJ, the University of York and U. S. National Science Foundation under Grant CCF-1420575.

S. Xu is with the Department of Electronics, University of York, YO10 5DD York, U.K. (e-mail: songcen.xu@york.ac.uk).

R. C. de Lamare is with CETUC / PUC-Rio, Brazil and Department of Electronics, University of York, U.K. (e-mail: rodrigo.delamare@york.ac.uk).

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton NJ 08544 USA (e-mail: poor@princeton.edu).

connected topology. A diffusion protocol is employed although other strategies, such as incremental [24] and consensus [25] could also be used. A partially connected network means that nodes can exchange information only with their neighbors as determined by the connectivity topology. In contrast, a fully connected network means that, data broadcast by a node can be captured by all other nodes in the network [26]. At every time instant i, the sensor at each node k takes a scalar measurement $d_k(i)$ according to

$$d_k(i) = \boldsymbol{\omega}_0^H \boldsymbol{x}_k(i) + n_k(i), \quad i = 1, 2, \ldots, \mathrm{I}, \tag{2}$$

where $\boldsymbol{x}_k(i)$ is the $M \times 1$ input signal vector with zero mean and variance $\sigma^2_{x,k}$, $n_k(i)$ is the noise at each node with zero mean and variance $\sigma^2_{n,k}$. From (2), we can see that the measurements for all nodes are related to an unknown parameter vector $\boldsymbol{\omega}_0$ with size $M \times 1$ that should be estimated by the network. We assume that $\boldsymbol{\omega}_0$ is a sparse vector with $S \ll M$ non-zero coefficients. The aim of such a network is to compute an estimate of $\boldsymbol{\omega}_0$ in a distributed fashion, which minimizes the cost function

$$J_\omega(\boldsymbol{\omega}) = \sum_{k=1}^{N} \mathbb{E}\{|d_k(i) - \boldsymbol{\omega}^H \boldsymbol{x}_k(i)|^2\}, \tag{3}$$

where $\mathbb{E}\{\cdot\}$ denotes expectation. Distributed estimation of $\boldsymbol{\omega}_0$ is appealing because it provides robustness against noisy measurements and improved performance as reported in [1], [24], [25]. To solve this problem, a cost-effective technique is the adapt–then–combine (ATC) diffusion strategy [1]

$$\begin{cases} \boldsymbol{\psi}_k(i) = \boldsymbol{\omega}_k(i) + \mu_k \boldsymbol{x}_k(i)\big[d_k(i) - \boldsymbol{\omega}_k^H(i)\boldsymbol{x}_k(i)\big]^*, \\ \boldsymbol{\omega}_k(i+1) = \sum\limits_{l \in \mathcal{N}_k} c_{kl} \boldsymbol{\psi}_l(i), \end{cases} \tag{4}$$

where $\mathcal{N}_k$ indicates the set of neighbors for node k, $\boldsymbol{\psi}_k(i)$ is the local estimator of node k, $|\mathcal{N}_k|$ denotes the cardinality of $\mathcal{N}_k$ and c_{kl} is the combination coefficient, which is calculated with respect to the Metropolis rule

$$\begin{cases} c_{kl} = \frac{1}{\max(|\mathcal{N}_k|, |\mathcal{N}_l|)}, & \text{if } k \neq l \text{ are linked} \\ c_{kl} = 0, & \text{for } k \text{ and } l \text{ not linked} \\ c_{kk} = 1 - \sum\limits_{l \in \mathcal{N}_k / k} c_{kl}, & \text{for } k = l \end{cases} \tag{5}$$

and should satisfy

$$\sum_{l} c_{kl} = 1, l \in \mathcal{N}_k \forall k. \tag{6}$$

Existing distributed sparsity-aware estimation strategies, e.g., [5], [6], [7], are designed using the full dimension signal space, which reduces the convergence rate and degrades the MSE performance. In order to improve performance, reduce the required bandwidth and optimize the distributed processing, we incorporate at each node of the WSN the proposed distributed compressed estimation scheme based on CS techniques, together with a measurement matrix optimization algorithm.

III. Proposed Distributed Compressed Estimation Scheme

In this section, we detail the proposed distributed compressed estimation (DCE) scheme based on CS. The proposed scheme, depicted in Fig. 1, employs compression and decompression modules inspired by CS techniques to perform distributed compressed estimation. In the proposed scheme, at each node, the sensor first observes the $M \times 1$ vector $\boldsymbol{x}_k(i)$, then with the help of the $D \times M$ measurement matrix obtains the compressed version $\bar{\boldsymbol{x}}_k(i)$, and performs the estimation of $\boldsymbol{\omega}_0$ in the compressed domain. In other words, the proposed scheme estimates the $D \times 1$ vector $\bar{\boldsymbol{\omega}}_0$ instead of the $M \times 1$ vector $\boldsymbol{\omega}_0$, where $D \ll M$ and the D–dimensional quantities are designated with an overbar. At each node, a decompression module employs a $D \times M$ measurement matrix $\boldsymbol{\Phi}_k$ and a reconstruction algorithm to compute an estimate of $\boldsymbol{\omega}_0$. One advantage for the DCE scheme is that fewer parameters need to be transmitted between neighbour nodes.

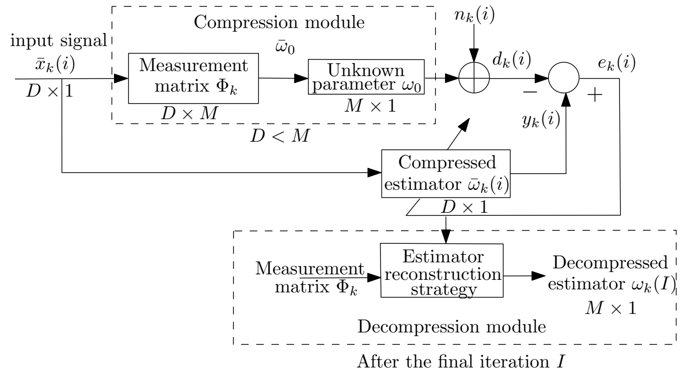

Fig. 1. Proposed Compressive Sensing Modules

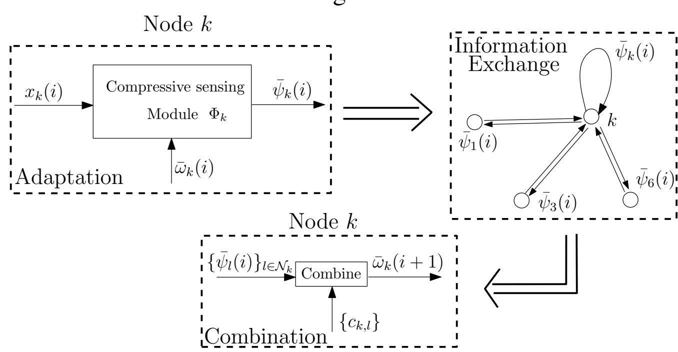

Fig. 2. Proposed DCE Scheme

We start the description of the proposed DCE scheme with the scalar measurement $d_k(i)$ given by

$$d_k(i) = \bar{\boldsymbol{\omega}}_0^H \bar{\boldsymbol{x}}_k(i) + n_k(i), \quad i = 1, 2, \ldots, \mathrm{I}, \tag{7}$$

where $\bar{\boldsymbol{\omega}}_0 = \boldsymbol{\Phi}_k \boldsymbol{\omega}_0$ and $\bar{\boldsymbol{x}}_k(i)$ is the $D \times 1$ input signal vector. This operation is depicted in Fig. 1 as the compression module.

Fig. 2 illustrates the proposed DCE scheme. The scheme can be divided into three steps:

- Adaptation

In the adaptation step, at each time instant i=1,2, . . . , I, each node k=1,2, …, N, generates a local compressed estimator $\bar{\boldsymbol{\psi}}_k(i)$ through

$$\bar{\boldsymbol{\psi}}_k(i) = \bar{\boldsymbol{\omega}}_k(i) + \mu_k(i) e_k^*(i) \bar{\boldsymbol{x}}_k(i), \tag{8}$$

where $e_k(i) = d_k(i) - \bar{\boldsymbol{\omega}}_k^H(i)\bar{\boldsymbol{x}}_k(i)$ and $\mu_k(i) = \frac{\mu_0}{\bar{\boldsymbol{x}}_k^H(i)\bar{\boldsymbol{x}}_k(i)}$.

- Information exchange

Given the network topology structure, only the local compressed estimator $\bar{\boldsymbol{\psi}}_k(i)$ will be transmitted between node k and all its neighbor nodes. The measurement matrix $\boldsymbol{\Phi}_k$ will be kept locally.

- Combination

At each time instant i=1,2, . . . , I, the combination step starts after the information exchange is finished. Each node will combine the local compressed estimators from its neighbor nodes and itself through

$$\bar{\boldsymbol{\omega}}_k(i+1) = \sum_{l \in \mathcal{N}_k} c_{kl} \bar{\boldsymbol{\psi}}_l(i), \tag{9}$$

to compute the updated compressed estimator $\bar{\boldsymbol{\omega}}_k(i+1)$.

After the final iteration I, each node will employ the OMP reconstruction strategy to generate the decompressed estimator $\boldsymbol{\omega}_k(I)$. Other reconstruction algorithms can also be used. The decompression module described in Fig. 1 illustrates the details. In summary, during the DCE procedure, only the local compressed estimator $\bar{\boldsymbol{\psi}}_k(i)$ will be transmitted over the network resulting in a reduction of the number of parameters to be transmitted from M to D. The proposed DCE scheme is given in Table I.

The computational complexity of the proposed DCE scheme is $O(NDI + ND^3)$, where N is the number of nodes in the WSN and I is the number of time instants. The distributed NLMS algorithm has a complexity $O(NMI)$, while the complexity of the sparse diffusion NLMS algorithm [6] is $O(3NMI)$. For the distributed compressive sensing algorithm of [18], the computational complexity is $O(NMI + ND^3 I)$. In the proposed DCE scheme, only the local compressed estimator $\bar{\boldsymbol{\psi}}_k(i)$ with D parameters will be transmitted through the network, which means the transmission requirement is greatly reduced as compared with the standard schemes that transmit $\boldsymbol{\psi}_k(i)$ with M parameters.

TABLE I
THE PROPOSED DCE SCHEME

```
Initialize: ω̄_k(1)=0
For each time instant i=1,2, . . . , I-1
    For each node k=1,2, ..., N
        ψ̄_k(i) = ω̄_k(i) + μ(i) e*_k(i) x̄_k(i)
        where e_k(i) = d_k(i) − ω̄_k^H(i) x̄_k(i),
        d_k(i) = ω̄_0^H x̄_k(i) + n_k(i) = (Φ_k ω_0)^H x̄_k(i) + n_k(i)
        and Φ_k is the D × M random measurement matrix
    end
    For each node k=1,2, ..., N
        ω̄_k(i+1) = Σ_{l∈N_k} c_kl ψ̄_l(i)
    end
end
After the final iteration I
For each node k=1,2, ..., N
    ω_k(I) = f_OMP{ω̄_k(I)}
    where ω_k(I) is the final decompressed estimator.
end
```

IV. Measurement Matrix Optimization

To further improve the performance of the proposed DCE scheme, an optimization algorithm for the design of the measurement matrix $\boldsymbol{\Phi}_k(i)$, which is now time–variant, is developed here. Unlike prior work [17], [27], this optimization is distributed and adaptive. Let us consider the cost function

$$\begin{aligned}\mathcal{J} &= \mathbb{E}\{|e_k(i)|^2\} = \mathbb{E}\{|d_k(i) - y_k(i)|^2\} \\ &= \mathbb{E}\{|d_k(i)|^2\} - \mathbb{E}\{d_k^*(i) y_k(i)\} - \mathbb{E}\{d_k(i) y_k^*(i)\} \\ &\quad + \mathbb{E}\{|y_k(i)|^2\},\end{aligned} \tag{10}$$

where $y_k(i) = \bar{\boldsymbol{\omega}}_k^H(i)\bar{\boldsymbol{x}}_k(i)$. To minimize the cost function, we need to compute the gradient of $\mathcal{J}$ with respect to $\boldsymbol{\Phi}_k^*(i)$ and equate it to a null vector, i.e., $\nabla \mathcal{J}_{\boldsymbol{\Phi}_k^*(i)} = \boldsymbol{0}$. As a result, only the first three terms in (10) need to be considered. Taking the first three terms of (10) we arrive at

$$\begin{aligned}&\mathbb{E}\{|d_k(i)|^2\} - \mathbb{E}\{d_k^*(i) y_k(i)\} - \mathbb{E}\{d_k(i) y_k^*(i)\} \\ &= \mathbb{E}\{|\boldsymbol{\omega}_0^H \boldsymbol{\Phi}_k^H(i)\bar{\boldsymbol{x}}_k(i) + n_k(i)|^2\} \\ &\quad - \mathbb{E}\{(\boldsymbol{\omega}_0^H \boldsymbol{\Phi}_k^H(i)\bar{\boldsymbol{x}}_k(i) + n_k(i))^* y_k(i)\} \\ &\quad - \mathbb{E}\{(\boldsymbol{\omega}_0^H \boldsymbol{\Phi}_k^H(i)\bar{\boldsymbol{x}}_k(i) + n_k(i)) y_k^*(i)\} \\ &= \mathbb{E}\{|\boldsymbol{\omega}_0^H \boldsymbol{\Phi}_k^H(i)\bar{\boldsymbol{x}}_k(i)|^2\} + \mathbb{E}\{(\boldsymbol{\omega}_0^H \boldsymbol{\Phi}_k^H(i)\bar{\boldsymbol{x}}_k(i))^* n_k(i)\} \\ &\quad + \mathbb{E}\{(\boldsymbol{\omega}_0^H \boldsymbol{\Phi}_k^H(i)\bar{\boldsymbol{x}}_k(i)) n_k^*(i)\} + \mathbb{E}\{|n_k(i)|^2\} \\ &\quad - \mathbb{E}\{(\boldsymbol{\omega}_0^H \boldsymbol{\Phi}_k^H(i)\bar{\boldsymbol{x}}_k(i))^* y_k(i)\} - \mathbb{E}\{n_k^*(i) y_k(i)\} \\ &\quad - \mathbb{E}\{(\boldsymbol{\omega}_0^H \boldsymbol{\Phi}_k^H(i)\bar{\boldsymbol{x}}_k(i)) y_k^*(i)\} - \mathbb{E}\{n_k(i) y_k^*(i)\}.\end{aligned} \tag{11}$$

Because the random variable $n_k(i)$ is statistically independent from the other parameters and has zero mean, (11) can be further simplified as

$$\begin{aligned}&\mathbb{E}\{|d_k(i)|^2\} - \mathbb{E}\{d_k^*(i) y_k(i)\} - \mathbb{E}\{d_k(i) y_k^*(i)\} \\ &= \mathbb{E}\{|\boldsymbol{\omega}_0^H \boldsymbol{\Phi}_k^H(i)\bar{\boldsymbol{x}}_k(i)|^2\} + \sigma_{n,k}^2 - \mathbb{E}\{(\boldsymbol{\omega}_0^H \boldsymbol{\Phi}_k^H(i)\bar{\boldsymbol{x}}_k(i))^* y_k(i)\} \\ &\quad - \mathbb{E}\{(\boldsymbol{\omega}_0^H \boldsymbol{\Phi}_k^H(i)\boldsymbol{x}_k(i)) y_k^*(i)\}.\end{aligned} \tag{12}$$

Then, we have

$$\nabla \mathcal{J}_{\boldsymbol{\Phi}_k^*(i)} = \boldsymbol{R}_k(i)\boldsymbol{\Phi}_k(i)\boldsymbol{R}_{\boldsymbol{\omega}_0} - \boldsymbol{P}_k(i), \tag{13}$$

where $\boldsymbol{R}_k(i) = \mathbb{E}\{\bar{\boldsymbol{x}}_k(i)\bar{\boldsymbol{x}}_k^H(i)\}$, $\boldsymbol{R}_{\boldsymbol{\omega}_0} = \mathbb{E}\{\boldsymbol{\omega}_0\boldsymbol{\omega}_0^H\}$ and $\boldsymbol{P}_k(i) = \mathbb{E}\{y_k^*(i)\bar{\boldsymbol{x}}_k(i)\boldsymbol{\omega}_0^H\}$. Equating (13) to a null vector, we obtain

$$\boldsymbol{R}_k(i)\boldsymbol{\Phi}_k(i)\boldsymbol{R}_{\boldsymbol{\omega}_0} - \boldsymbol{P}_k(i) = \boldsymbol{0}, \tag{14}$$

$$\boldsymbol{\Phi}_k(i) = \boldsymbol{R}_k^{-1}(i)\boldsymbol{P}_k(i)\boldsymbol{R}_{\boldsymbol{\omega}_0}^{-1}. \tag{15}$$

The expression in (15) cannot be solved in closed–form because $\boldsymbol{\omega}_0$ is an unknown parameter. As a result, we employ the previous estimate $\bar{\boldsymbol{\omega}}_k(i)$ to replace $\boldsymbol{\omega}_0$. However, $\bar{\boldsymbol{\omega}}_k(i)$ and $\boldsymbol{\Phi}_k(i)$ depend on each other, thus, it is necessary to iterate (15) with an initial guess to obtain a solution. In particular, we replace the expected values with instantaneous values. Starting from (13), we use instantaneous estimates to compute

$$\hat{\boldsymbol{R}}_k(i) = \bar{\boldsymbol{x}}_k(i)\bar{\boldsymbol{x}}_k^H(i), \tag{16}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{\omega}_0} = \boldsymbol{\omega}_0\boldsymbol{\omega}_0^H \tag{17}$$

and

$$\hat{\boldsymbol{P}}_k(i) = y_k^*(i)\bar{\boldsymbol{x}}_k(i)\boldsymbol{\omega}_0^H. \tag{18}$$

According to the method of steepest descent [28], the updated parameters of the measurement matrix $\boldsymbol{\Phi}_k(i)$ at time $i+1$ are computed by using the simple recursive relation

$$\begin{aligned}\boldsymbol{\Phi}_k(i+1) &= \boldsymbol{\Phi}_k(i) + \eta[-\nabla \mathcal{J}_{\boldsymbol{\Phi}_k^*(i)}] \\ &= \boldsymbol{\Phi}_k(i) + \eta[\hat{\boldsymbol{P}}_k(i) - \hat{\boldsymbol{R}}_k(i)\boldsymbol{\Phi}_k(i)\hat{\boldsymbol{R}}_{\boldsymbol{\omega}_0}] \\ &= \boldsymbol{\Phi}_k(i) + \eta[y_k^*(i)\bar{\boldsymbol{x}}_k(i)\boldsymbol{\omega}_0^H - \bar{\boldsymbol{x}}_k(i)\bar{\boldsymbol{x}}_k^H(i)\boldsymbol{\Phi}_k(i)\boldsymbol{\omega}_0\boldsymbol{\omega}_0^H].\end{aligned} \tag{19}$$

where η is the step size and $\boldsymbol{\omega}_0$ is the $M \times 1$ unknown parameter vector that must be estimated by the network. Then, the parameter vector $\bar{\boldsymbol{\omega}}_k(i)$ is used to reconstruct the estimate of $\boldsymbol{\omega}_0$ as follows

$$\boldsymbol{\omega}_{re_k}(i) = f_{\text{OMP}}\{\bar{\boldsymbol{\omega}}_k(i)\}, \tag{20}$$

where the operator $f_{\text{OMP}}\{\cdot\}$ denotes the OMP reconstruction algorithm. Note that other reconstruction algorithms could also be employed. Replacing $\boldsymbol{\omega}_0$ by $\boldsymbol{\omega}_{re_k}(i)$, we arrive at the expression for updating the measurement matrix described by

$$\begin{aligned}\boldsymbol{\Phi}_k(i+1) = \boldsymbol{\Phi}_k(i) + \eta\big[y_k^*(i)\bar{\boldsymbol{x}}_k(i)\boldsymbol{\omega}_{re_k}^H(i)\\ - \bar{\boldsymbol{x}}_k(i)\bar{\boldsymbol{x}}_k^H(i)\boldsymbol{\Phi}_k(i)\boldsymbol{\omega}_{re_k}(i)\boldsymbol{\omega}_{re_k}^H(i)\big]. \quad (21)\end{aligned}$$

The computational complexity of the proposed scheme with measurement matrix optimization is $O(NDI + ND^3I)$.

V. Simulations

We assess the proposed DCE scheme and the measurement matrix optimization algorithm in a WSN application, where a partially connected network with $N = 20$ nodes is considered. We compare the proposed DCE scheme with uncompressed schemes, including the distributed NLMS (dNLMS) algorithm (normalized version of [1]), sparse diffusion NLMS algorithm [6], sparsity-promoting adaptive algorithm [8], and the distributed compressive sensing algorithm [18], in terms of MSE performance. Note that other metrics such as mean-square deviation (MSD) could be used but result in the same performance hierarchy between the analyzed algorithms.

The input signal is generated as $\boldsymbol{x}_k(i) = [x_k(i) \quad x_k(i-1) \quad \ldots \quad x_k(i-M+1)]^T$ and $x_k(i) = u_k(i) + \alpha_k x_k(i-1)$, where α_k is a correlation coefficient and $u_k(i)$ is a white noise process with variance $\sigma^2_{u,k} = 1 - |\alpha_k|^2$, to ensure the variance of $\boldsymbol{x}_k(i)$ is $\sigma^2_{x,k} = 1$. The compressed input signal is obtained by $\bar{\boldsymbol{x}}_k(i) = \boldsymbol{\Phi}_k\boldsymbol{x}_k(i)$. The measurement matrix $\boldsymbol{\Phi}_k$ is an i.i.d. Gaussian random matrix that is kept constant. The noise samples are modeled as complex Gaussian noise with variance $\sigma^2_{n,k} = 0.001$. The unknown $M \times 1$ parameter vector $\boldsymbol{\omega}_0$ has sparsity S, where M=50, D=10 and S=3. The step size μ_0 for the distributed NLMS, distributed compressive sensing, sparse diffusion LMS and the proposed DCE algorithms is 0.45. The parameter that controls the shrinkage in [6] is set to 0.001. For [8], the number of hyperslabs equals 55 and the width of the hyperslabs is 0.01.

Fig. 3 illustrates the comparison between the DCE scheme with other existing algorithms, without the measurement matrix optimization. It is clear that, when compared with the existing algorithms, the DCE scheme has a significantly faster convergence rate and a better MSE performance. These advantages consist in two features: the compressed dimension brought by the proposed scheme and CS being implemented in the estimation layer. As a result, the number of parameters for transmission in the network is significantly reduced.

In the second scenario, we employ the measurement matrix optimization algorithm to in the DCE scheme. The parameter η for the measurement matrix optimization algorithm is set to 0.08 and all other parameters remain the same as in the previous scenario. In Fig. 4, we observe that with the help of the measurement matrix optimization algorithm, DCE can achieve a faster convergence when compared with DCE without the measurement matrix optimization.

In the third scenario, we compare the DCE scheme with the distributed NLMS algorithm with different levels of resolution in bits per coefficient, reduced dimension D and sparsity level S. The x-axis stands for the reduced dimension D and their corresponding sparsity level S can be found in Fig. 5. In Fig. 5, it is clear that with the increase of the sparsity level S the MSE performance degrades. In addition, the MSE performance will increase when the transmission has more bits per coefficient. For the DCE scheme, the total number of bits required for transmission is D times the number of bits per coefficient, whereas for the distributed NLMS algorithm it is M times the number of bits per coefficient. A certain level of redundancy is required between the sparsity level and the reduced dimension due to the error introduced by the estimation procedure.

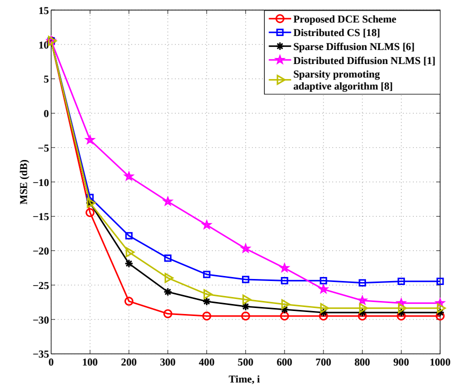

Fig. 3. MSE performance against time

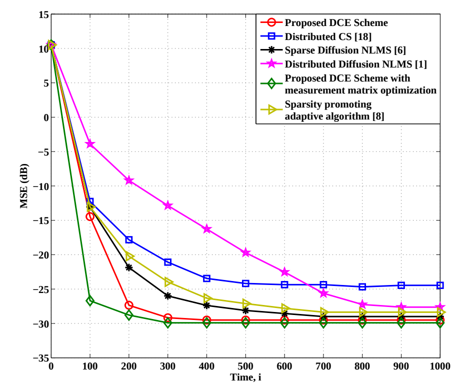

Fig. 4. MSE performance against time with measurement matrix optimization

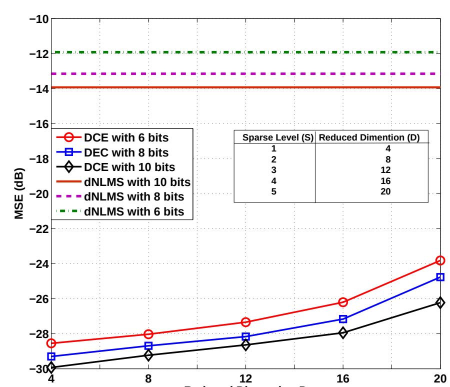

Fig. 5. MSE performance against reduced dimension D for different levels of resolution in bits per coefficient

VI. Conclusions

We have proposed a novel DCE scheme and algorithms for sparse signals and systems based on CS techniques and a measurement matrix optimization. In the DCE scheme, the estimation procedure is performed in a compressed dimension. The results for a WSN application show that the DCE scheme outperforms existing strategies in terms of convergence rate, reduced bandwidth and MSE performance.

References

[1] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122–3136, July 2008.

[2] S. Xu and R. C. de Lamare, "Distributed conjugate gradient strategies for distributed estimation over sensor networks," *in Proc. Sensor Signal Processing for Defence 2012*, London, UK, 2012.

[3] S. Xu, R. C. de Lamare, and H. V. Poor, "Adaptive link selection strategies for distributed estimation in diffusion wireless networks," *in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, Canada 2013.

[4] F. S. Cattivelli and A. H. Sayed, "Diffusion lms strategies for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 58, pp. 1035–1048, March 2010.

[5] Y. Chen, Y. Gu, and A.O. Hero, "Sparse LMS for system identification," *in Proc. IEEE ICASSP*, pp. 3125–3128, Taipei, Taiwan, May 2009.

[6] P. Di Lorenzo, S. Barbarossa, and A.H. Sayed, "Distributed spectrum estimation for small cell networks based on sparse diffusion adaptation," *IEEE Signal Processing Letters*, vol. 20, no. 12, pp. 1261–1265, Dec 2013.

[7] D. Angelosante, J.A Bazerque, and G.B. Giannakis, "Online adaptive estimation of sparse signals: Where rls meets the ℓ_1–norm," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3436–3447, July 2010.

[8] S. Chouvardas, K. Slavakis, Y. Kopsinis, and S. Theodoridis, "A sparsity promoting adaptive algorithm for distributed learning," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5412–5425, Oct 2012.

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

[10] P. Di Lorenzo and A. H. Sayed, "Sparse distributed learning based on diffusion adaptation," *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1419–1433, March 2013.

[11] Z. Liu, Y. Liu, and C. Li, "Distributed sparse recursive least-squares over networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 6, pp. 1386–1395, March 2014.

[12] M. O. Sayin and S. S. Kozat, "Compressive diffusion strategies over distributed networks for reduced communication load," *IEEE Transactions on Signal Processing*, vol. 62, no. 20, pp. 5308–5323, Oct 2014.

[13] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[14] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," *IEEE Signal. Proc. Mag.*, vol. 25, no. 2, pp. 21–30, March 2008.

[15] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14–20, March 2008.

[16] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proc. IEEE Inform. Process. Sens. Netw.*, Nashville, TN, April 2006.

[17] Y. Yao, A. P. Petropulu, and H. V. Poor, "MIMO radar using compressive sampling," *IEEE Journal on Selected Topics in Signal Processing*, vol. 4, no. 1, pp. 146–163, Feb 2010.

[18] C. Wei, M. Rodrigues, and I. J. Wassell, "Distributed compressive sensing reconstruction via common support discovery," in *IEEE International Conference on Communications*, Kyoto, Japan, June 2011.

[19] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed compressive sensing," in *ECE Department Tech. Report TREE-0612*, Rice University, Houston, TX, November 2006.

[20] G. Quer, R. Masiero, G. Pillonetto, M. Rossi, and M. Zorzi, "Sensing, compression, and recovery for WSNs: Sparse signal modeling and monitoring framework," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3447–3461, October 2012.

[21] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, December 2007.

[22] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," *in Proc. 27th Annu. Asilomar Conf. Signals, Systems,and Computers*, pp. 40–44 vol.1, Pacic Grove, CA, Nov, 1993.

[23] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive Approximation*, vol. 13, no. 1, pp. 57–98, 1997.

[24] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Transactions on Signal Processing*, vol. 48, no. 8, pp. 223–229, Aug 2007.

[25] L. Xie, D.-H. Choi, S. Kar, and H. V. Poor, "Fully distributed state estimation for wide-area monitoring systems," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1154–1169, September 2012.

[26] A. Bertrand and M. Moonen, "Distributed adaptive node–specific signal estimation in fully connected sensor networks–part II: Simultaneous and asynchronous node updating," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5292–5306, 2010.

[27] Y. Yao, A. P. Petropulu, and H. V. Poor, "Measurement matrix design for compressive sensing–based MIMO radar," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5338–5352, November 2011.

[28] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, NJ, USA, third edition, 1996.