ADAPTIVE DISTRIBUTED COMPRESSED ESTIMATION BASED ON RECURSIVE LEAST SQUARES WITH SENSING MATRIX DESIGN

Huang Bai[1], *Songcen Xu*[2], *Sheng Li*[1], *Rodrigo C. de Lamare*[2], *Xiongxiong He*[1], *H. Vincent Poor*[3]

[1]College of Information Engineering, Zhejiang University of Technology, Hangzhou, China.
[2]CETUC / PUC-Rio, Brazil and Department of Electronics, University of York, U.K.
[3] Department of Electrical Engineering, Princeton University, Princeton, USA

ABSTRACT

In this paper, a distributed compressed estimation (DCE) scheme is presented based on a distributed recursive-least squares algorithm for sparse signals and systems along with a sensing matrix design procedure based on compressive sensing techniques. The DCE scheme consists of compression and decompression modules inspired by compressive sensing to perform distributed compressed estimation. A design procedure is developed under the DCE framework and a novel algorithm is developed to optimize the sensing matrix, which can further improve the performance of the proposed DCE and distributed adaptive algorithms. Simulations for a wireless sensor network show 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, sensing matrix design, sensor networks.

1. INTRODUCTION

Distributed signal processing techniques extract information from data collected at nodes that are distributed over a geographic area [1]. With these distributed approaches, instead of generating an estimate at one node, an improved estimate is obtained by a specific node that collects and processes the local information from a set of neighbor nodes and then combines the information provided by the neighbor nodes with its local estimate. In many scenarios, the unknown parameter vector to be estimated can be considered as a sparse vector that contains only a few nonzero coefficients. Many algorithms have been developed for sparse signal estimation [2, 3, 4, 5]. However, these techniques are designed with 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) [6, 7] is an emerging research area in the signal processing community. It has been successfully applied to diverse fields, such as image processing [8], wireless communications [9] and MIMO radar [10]. The application of CS to wireless sensor networks (WSNs) has been recently investigated in [9], [11] and [12]. In [11], 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. A sparse model that allows the use of CS for the online recovery of large data sets in WSNs was proposed in [13], but it assumes that the sensor measurements could be gathered directly, without an estimation procedure. In Xu and de Lamare [12] compressed transmit strategies and estimation techniques were considered using a distributed compressed estimation (DCE) scheme along with stochastic gradient algorithms for parameter estimation and adjustment of the *sensing matrix*.

In CS problems, we consider $\boldsymbol{v} \in \Re^{M\times 1}$ with v_k being its kth element. The zero-norm $\|\boldsymbol{v}\|_0$ is defined to count the number of nonzero elements in $\boldsymbol{v}$. For a matrix $\boldsymbol{\Phi} \in \Re^{D\times M}$, the *spark* is defined as the smallest number of columns in $\boldsymbol{\Phi}$ that are linearly dependent. It has been shown in [14] that any S-sparse vector $\boldsymbol{\omega}$ can be exactly recovered from the projection $\bar{\boldsymbol{\omega}}$ via

$$\min_{\boldsymbol{\omega}} \ \|\boldsymbol{\omega}\|_0 \quad s.t. \quad \bar{\boldsymbol{\omega}} = \boldsymbol{\Phi}\boldsymbol{\omega} \tag{1}$$

which can be solved using orthogonal matching pursuit (OMP) [15] as long as the spark of the *sensing matrix* $\boldsymbol{\Phi}$ is greater than $2S$. Moreover, it has been pointed out that a larger spark of $\boldsymbol{\Phi}$ leads to a larger signal space among which the sparse vector can be exactly recovered. However, calculating the spark of a matrix is very complex. Alternative properties that guarantee the recovery are then required for designing a suitable sensing matrix $\boldsymbol{\Phi}$.

The *mutual coherence* of $\boldsymbol{\Phi}$, which is defined as

$$\mu(\boldsymbol{\Phi}) \ \triangleq \ \max_{1\le i\ne j\le L} \frac{|\boldsymbol{\phi}_i^{\mathcal{T}}\boldsymbol{\phi}_j|}{\|\boldsymbol{\phi}_i\|_2\|\boldsymbol{\phi}_j\|_2} \tag{2}$$

can be employed to provide recovery guarantees [14], [16]. Note that $\mathcal{T}$ denotes the transpose operator and $\boldsymbol{\phi}_k$ is the kth atom of $\boldsymbol{\Phi}$. As shown in [14], an S-sparse signal can be successfully recovered from the measurement if

$$S < 1/2[1+\mu^{-1}(\boldsymbol{\Phi})] \tag{3}$$

The system will allow a wider set of candidate signals to be exactly recovered with a smaller $\mu(\boldsymbol{\Phi})$.

In this work, a distributed CS-based approach based on the DCE scheme with a distributed recursive least-squares (DRLS) algorithm that incorporates compression and decompression modules along with a novel sensing matrix design is developed for distributed estimation problems. In the compression module, the unknown parameter $\boldsymbol{\omega}_0$ is compressed into a lower dimension followed by a DRLS algorithm performed in a compressed dimension. This approach reduces the required bandwidth and improves the MSE performance with a faster convergence rate. Then the decompression module recovers the compressed estimator into its original dimension using the OMP algorithm [15]. A novel sensing matrix design approach is

This research was supported in part by the National Natural Science Foundation of China under Grants 61503339, 61273195 and 61473262, and in part by the U.S. National Science Foundation under Grants CNS-1456793 and ECCS-1343210.

proposed to minimize the mutual coherence and is incorporated into the DCE scheme. With a closed-form solution, the designed sensing matrix can be pre-calculated and does not require any computational load to update its parameter during the operation of the system. Simulation results illustrate the performance of the proposed scheme and algorithm against existing techniques.

This paper is organized as follows. Section 2 describes the system model and the problem statement. In Section 3, we review the DCE scheme and its operation. The proposed sensing matrix design is presented in Section 4. Simulation results are provided in Section 5, whereas the conclusions are given in Section 6.

2. SYSTEM MODEL AND PROBLEM STATEMENT

Let us consider a WSN with N nodes in a partially connected topology which employs a diffusion protocol. A partially connected network means that nodes can exchange information only with their neighbors as determined by the connectivity topology. At every time instant i, the scalar measurement $d_k(i)$ received by the sensor at each node k can be expressed as

$$d_k(i) = \boldsymbol{\omega}_0^H \boldsymbol{x}_k(i) + n_k(i), \quad i = 1, 2, \ldots, \mathrm{I}, \tag{4}$$

where $\boldsymbol{x}_k(i)$ is the M-dimensional input signal vector, $n_k(i)$ is the additive noise at each node with zero mean and variance $\sigma_{n,k}^2$. The M-dimensional unknown parameter vector $\boldsymbol{\omega}_0$ should be estimated by the network in a distributed fashion. We assume that $\boldsymbol{\omega}_0$ is a sparse vector with only $S \ll M$ non-zero coefficients. The aim of such a network is to estimate the parameter vector that minimizes the cost function given by

$$J(\boldsymbol{\omega}) = \sum_{k=1}^{N} \sum_{j=1}^{i} \lambda^{i-j} \{|d_k(j) - \boldsymbol{\omega}^H \boldsymbol{x}_k(j)|^2\}, \tag{5}$$

where λ is a positive real-valued forgetting factor that is less than one. Distributed estimation of $\boldsymbol{\omega}_0$ is appealing because it can bring robustness against noisy measurements, reduce complexity and communication overheads, and improve performance. In order to obtain a distributed estimator, the cost-effective adapt–then–combine (ATC) diffusion strategy can be employed [1], which firstly updates the local estimators $\boldsymbol{\psi}_k(i)$ with $\boldsymbol{\omega}_k(i)$, and then computes the improved estimator as

$$\boldsymbol{\omega}_k(i+1) = \sum_{l \in \mathcal{N}_k} c_{kl} \boldsymbol{\psi}_l(i), \tag{6}$$

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

$$\begin{cases} c_{kl} = \frac{1}{\max(|\mathcal{N}_k|, |\mathcal{N}_l|)}, & \text{if } k \neq l \text{ are linked} \\ c_{kl} = 0, & \text{for } k \text{ and } l \text{ are not linked} \\ c_{kk} = 1 - \sum\limits_{l \in \mathcal{N}_k / k} c_{kl}, & \text{for } k = l \end{cases} \tag{7}$$

and should satisfy $\sum_l c_{kl} = 1, l \in \mathcal{N}_k, \forall k$. Existing distributed sparsity-aware estimation strategies, e.g., [2], [4] and [5], are designed using the full dimension signal space. In order to improve the MSE performance, reduce the required bandwidth and optimize the distributed processing, the DCE scheme based on CS techniques has been developed for a WSN [12]. The problem we are interested in here is how to design and incorporate an optimized sensing matrix computed offline into the DCE scheme.

3. PROPOSED DCE WITH OPTIMIZED SENSING MATRIX

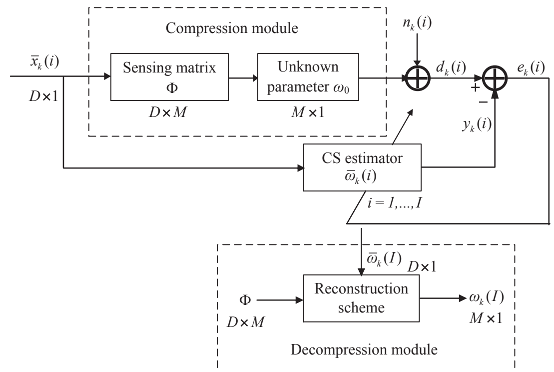

Fig. 1. Proposed Compressive Sensing Modules

In this section, the proposed DCE scheme with optimized sensing matrix will be detailed. As shown in Fig. 1, the proposed scheme employs compression and decompression modules along with an optimized sensing matrix 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 the compressed version $D\times 1$ vector $\bar{\boldsymbol{x}}_k(i)$ is obtained with a $D\times M$ sensing matrix. In what follows, all D–dimensional quantities are designated with an overbar. By introducing the CS based transformation, the estimation of $\boldsymbol{\omega}_0$ will be carried out in the compressed domain using a DRLS algorithm. In other words, the proposed scheme estimates $\bar{\boldsymbol{\omega}}_0$ instead of $\boldsymbol{\omega}_0$; note that D is assumed much smaller than M. The decompression module employs a $D \times M$ sensing matrix $\boldsymbol{\Phi}$ and a reconstruction algorithm to obtain an estimate of $\boldsymbol{\omega}_0$. One advantage for the DCE scheme with optimal sensing matrix is that fewer parameters need to be transmitted between neighbor nodes and the sensing matrix is fixed during the transmission which will reduce the computational load in the nodes.

We start the description of the proposed DCE scheme based on a DRLS algorithm with optimized sensing matrix with the scalar measurement $d_k(i)$ given by

$$d_k(i) = \bar{\boldsymbol{\omega}}_0^H \bar{\boldsymbol{x}}_k(i) + n_k(i), \quad i = 1, 2, \ldots, \mathrm{I}, \tag{8}$$

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.

The DCE scheme with a DRLS algorithm consists of three steps:

- Adaptation of local estimator

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

$$\bar{\boldsymbol{\psi}}_k(i) = \bar{\boldsymbol{\omega}}_k(i-1) + \boldsymbol{K}_k(i) e_k^*(i), \tag{9}$$

where $e_k(i) = d_k(i) - \bar{\boldsymbol{\omega}}_k^H(i-1)\bar{\boldsymbol{x}}_k(i)$ is the error signal, the Kalman gain is given by $\boldsymbol{K}_k(i) = \frac{\bar{\boldsymbol{R}}_k^{-1}(i-1)\bar{\boldsymbol{x}}_k(i)}{\lambda + \bar{\boldsymbol{x}}_k^H(i)\bar{\boldsymbol{R}}_k^{-1}(i-1)\bar{\boldsymbol{x}}_k(i)}$, and the update of the inverse of the covariance matrix is obtained by $\bar{\boldsymbol{R}}_k^{-1}(i) = \frac{1}{\lambda}\left[\bar{\boldsymbol{R}}_k^{-1}(i-1) - \boldsymbol{K}_k(i)\bar{\boldsymbol{x}}_k^H(i)\bar{\boldsymbol{R}}_k^{-1}(i-1)\right]$, where $\bar{\boldsymbol{R}}_k^{-1}(0) = \boldsymbol{I}/\delta$ is the initialization.

- Information exchange

In the DCE scheme, only the local compressed estimator $\bar{\boldsymbol{\psi}}_k(i)$ will be transmitted between node k and all its neighbor nodes.

- Combination for an improved estimator

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

$$\bar{\boldsymbol{\omega}}_k(i+1) = \sum_{l \in \mathcal{N}_k} c_{kl} \bar{\boldsymbol{\psi}}_l(i), \tag{10}$$

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 shown 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.

4. SENSING MATRIX OPTIMIZATION

In this section, a novel sensing matrix design that minimizes the mutual coherence of the sensing matrix is proposed to further improve the performance of the DCE scheme. As the sensing matrix is designed for some matrix properties, it can be pre-calculated and remains fixed during the operation of the system to save computational complexity at each node.

Let us denote $\boldsymbol{G} = \boldsymbol{\Phi}^{\mathcal{T}}\boldsymbol{\Phi} = \{g_{ij}\}$ and let $\boldsymbol{S}_c$ be the diagonal matrix whose kth element is given by $g_{kk}^{-1/2}$ for $k = 1, 2, \ldots, M$. The Gram matrix of $\bar{\boldsymbol{\Phi}} = \boldsymbol{\Phi}\boldsymbol{S}_c$, denoted as $\bar{\boldsymbol{G}} = \{\bar{g}_{ij}\}$, is then normalized such that $\bar{g}_{kk} = 1,\ \forall k$. Therefore, we have $\mu(\boldsymbol{\Phi}) = \max_{i \neq j} |\bar{g}_{ij}|$. This means that $\mu(\boldsymbol{\Phi})$ measures the maximum linear dependency possibly achieved by any two columns of the matrix $\boldsymbol{\Phi}$. It has been shown in [17] that for a matrix $\boldsymbol{\Phi}$ of dimensions $D \times M$, $\mu(\boldsymbol{\Phi})$ is bounded by

$$\underline{\mu} \ \triangleq\ \sqrt{\frac{M-D}{D(M-1)}} \ \leq\ \mu(\boldsymbol{\Phi}) \ \leq\ 1, \tag{11}$$

with $\underline{\mu}$ being the Welch bound. Furthermore, the author of [16] defined another indicator that is more related to the matrix performance, namely the averaged mutual coherence:

$$\mu_{av}(\boldsymbol{\Phi}) \ \triangleq\ \frac{\sum_{\forall\ (i,j) \in \mathcal{S}_{av}} |\bar{g}_{ij}|}{N_{av}}, \tag{12}$$

where $\mathcal{S}_{av} \ \triangleq\ \{(i,j) :\ \bar{\mu} \leq |\bar{g}_{ij}| < 1\}$, with $0 \leq \bar{\mu} < 1$ a given value and N_{av} is the number of elements in $\mathcal{S}_{av}$.

As seen from the relationship between sparsity and mutual coherence (3), it is natural for us to design the sensing matrix $\boldsymbol{\Phi}$ such that the absolute values of the off-diagonal elements of the corresponding Gram matrix are as small as possible, that is

$$\min_{\boldsymbol{\Phi}} \ \|\boldsymbol{\Phi}^{\mathcal{T}}\boldsymbol{\Phi} - \boldsymbol{I}_M\|_F^2, \tag{13}$$

with $\boldsymbol{I}_M$ denoting the identity matrix of dimension M and $\|\cdot\|_F$ the *Frobenius* norm. In [18] and [19], the closed-form solutions to (13) were derived as

$$\boldsymbol{\Phi} \ =\ \boldsymbol{U}[\boldsymbol{I}_D \ \ \boldsymbol{0}]\boldsymbol{V}^{\mathcal{T}}, \tag{14}$$

where both $\boldsymbol{U}$ and $\boldsymbol{V}$ are arbitrary orthonormal matrices that leave us space to further improve the performance of $\boldsymbol{\Phi}$.

However, for a matrix of dimension $D \times M$, the ideal Grammian $\boldsymbol{I}_M$ is unattainable. As indicated in [17] - [19], an equiangular tight frame (ETF) can achieve the minimal mutual coherence which is the Welch bound defined in (11). It is this fact that motivates us to further design the sensing matrix to approximate an ETF under the constraint of (14) with regard to orthonormal matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, that is

$$\min_{\boldsymbol{\Phi}} \ \|\boldsymbol{\Phi} - \boldsymbol{\Phi}_{etf}\|_F^2 \quad s.t. \quad \boldsymbol{\Phi} \ =\ \boldsymbol{U}[\boldsymbol{I}_D \ \ \boldsymbol{0}]\boldsymbol{V}^{\mathcal{T}}, \tag{15}$$

with $\boldsymbol{\Phi}_{etf}$ denoting the target ETF. According to [20] and [21], we construct a relaxed ETF as the target frame in the following way:

- With an initial $\boldsymbol{\Phi}$, obtain the matrix $\bar{\boldsymbol{\Phi}}$ by normalizing the columns of $\boldsymbol{\Phi}$;
- Calculate the Gram matrix $\bar{\boldsymbol{G}} = \bar{\boldsymbol{\Phi}}^{\mathcal{T}}\bar{\boldsymbol{\Phi}} = \{\bar{g}_{ij}\}$ and then apply the following *shrinking* operation

$$\tilde{\boldsymbol{G}}(i,j) = \begin{cases} \bar{g}_{ij}, & i \neq j,\ |\bar{g}_{ij}| \leq \eta \\ sgn[\bar{g}_{ij}]\eta, & i \neq j,\ |\bar{g}_{ij}| > \eta \\ 1, & i = j \end{cases}, \tag{16}$$

  with $sgn[\cdot]$ being the sign function and η a shrinking threshold to bound the absolute values of the off-diagonal elements, producing $\tilde{\boldsymbol{G}}$;
- Apply the singular value decomposition (SVD) to $\tilde{\boldsymbol{G}}$ to set the matrix rank to D, obtaining $\check{\boldsymbol{G}}$, and then compute a $\check{\boldsymbol{\Phi}} \in \Re^{D \times M}$ such that $\check{\boldsymbol{\Phi}}^{\mathcal{T}}\check{\boldsymbol{\Phi}} = \check{\boldsymbol{G}}$;
- Find a new $\boldsymbol{\Phi}$ that is the nearest M/D-tight frame to $\check{\boldsymbol{\Phi}}$, according to $\boldsymbol{\Phi} = \sqrt{M/D}(\check{\boldsymbol{\Phi}}\check{\boldsymbol{\Phi}}^{\mathcal{T}})^{-1/2}\check{\boldsymbol{\Phi}}$.

Repeat the above procedures several times and we can get a relaxed ETF $\boldsymbol{\Phi}_{etf}$ as the target.

Now the design problem can be stated as follows:

$$\min_{\boldsymbol{U} \in \mathcal{S}_o(D),\ \boldsymbol{V} \in \mathcal{S}_o(M)} \ \|\boldsymbol{U}[\boldsymbol{I}_D \ \ \boldsymbol{0}]\boldsymbol{V}^{\mathcal{T}} - \boldsymbol{\Phi}_{etf}\|_F^2, \tag{17}$$

where $\mathcal{S}_o(N)$ denotes the set of $N \times N$ orthonormal matrices. A popular method to handle this multivariate problem is alternating minimization [22, 23, 24, 25], i.e., updating $\boldsymbol{U}$ and $\boldsymbol{V}$ alternately. To this end, let us define the following cost function:

$$\begin{aligned} f(\boldsymbol{U}, \boldsymbol{V}) \ &=\ \|\boldsymbol{U}[\boldsymbol{I}_D \ \ \boldsymbol{0}]\boldsymbol{V}^{\mathcal{T}} - \boldsymbol{\Phi}_{etf}\|_F^2 \\ &=\ \|\boldsymbol{U}\boldsymbol{F}_1 - \boldsymbol{\Phi}_{etf}\|_F^2 \\ &=\ \|\boldsymbol{V}\boldsymbol{F}_2 - \boldsymbol{\Phi}_{etf}^{\mathcal{T}}\|_F^2, \end{aligned}$$

where

$$\boldsymbol{F}_1 \triangleq [\boldsymbol{I}_D \ \ \boldsymbol{0}]\boldsymbol{V}^{\mathcal{T}}, \quad \boldsymbol{F}_2 \triangleq [\boldsymbol{I}_D \ \ \boldsymbol{0}]^{\mathcal{T}}\boldsymbol{U}^{\mathcal{T}}$$

are only related to $\boldsymbol{V}$ and $\boldsymbol{U}$, respectively. So the alternating minimization procedure can be listed as follows:

- *Step I*: Fix $\boldsymbol{V}$, update $\boldsymbol{U}$ with

$$\min_{\boldsymbol{U} \in \mathcal{S}_o(D)} \ \|\boldsymbol{U}\boldsymbol{F}_1 - \boldsymbol{\Phi}_{etf}\|_F^2$$

- *Step II*: Fix $\boldsymbol{U}$, update $\boldsymbol{V}$ with

$$\min_{\boldsymbol{V} \in \mathcal{S}_o(M)} \ \|\boldsymbol{V}\boldsymbol{F}_2 - \boldsymbol{\Phi}_{etf}^{\mathcal{T}}\|_F^2$$

Iterating *Step I* and *Step II* results in our designed sensing matrix.

Considering the general expression:

$$\min_{\boldsymbol{X} \in \mathcal{S}_o(N)} \ \{\|\boldsymbol{X}\boldsymbol{Y} - \boldsymbol{Z}\|_F^2 \triangleq \ \varrho\}. \tag{18}$$

Denote $tr[\cdot]$ as the trace operator, then

$$\varrho = tr[\boldsymbol{Y}^{\mathcal{T}}\boldsymbol{Y}] + tr[\boldsymbol{Z}^{\mathcal{T}}\boldsymbol{Z}] - 2tr[\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Z}^{\mathcal{T}}].$$

Hence

$$(18) \Leftrightarrow \max_{\boldsymbol{X} \in \mathcal{S}_o(N)} tr[\boldsymbol{X}\boldsymbol{W}] \;\; s.t. \;\; \boldsymbol{W} = \boldsymbol{Y}\boldsymbol{Z}^{\mathcal{T}}.$$

Let $\boldsymbol{W} = \boldsymbol{U}_w \boldsymbol{\Sigma}_w \boldsymbol{V}_w^{\mathcal{T}}$ be the SVD of $\boldsymbol{W}$, then

$$tr[\boldsymbol{X}\boldsymbol{W}] = tr[\boldsymbol{\Sigma}_w \boldsymbol{V}_w^{\mathcal{T}} \boldsymbol{X} \boldsymbol{U}_w] \triangleq tr[\boldsymbol{\Sigma}_w \boldsymbol{R}].$$

Note that $\boldsymbol{R} = \boldsymbol{V}_w^{\mathcal{T}} \boldsymbol{X} \boldsymbol{U}_w = \{r_{ij}\}$ is still orthonormal. Thus, we have

$$tr[\boldsymbol{\Sigma}_w \boldsymbol{R}] = \sum_{k=1}^{N} \sigma_k r_{kk} \leq \sum_{k=1}^{N} \sigma_k$$

in which equality holds if $r_{kk} = 1$, $\forall k$, i.e., $\boldsymbol{R}$ is the identity matrix, which leads to

$$\boldsymbol{X} = \boldsymbol{V}_w \boldsymbol{U}_w^{\mathcal{T}}. \tag{19}$$

With the above conclusion, we can solve (17) by optimizing $\boldsymbol{U}$ and $\boldsymbol{V}$ in an alternating fashion, resulting in the solution of (15).

5. SIMULATIONS

In this section, the proposed DCE scheme with a DRLS algorithm and the sensing matrix optimization algorithm are considered in a WSN application, where a partially connected network with $N = 20$ nodes is considered. The contrasts are denoted as least mean square (LMS)-RSM, LMS-OSM, RLS-RSM, and RLS-OSM for the cases that represent the combinations of LMS with random sensing matrix (which is also the main idea of [12]), LMS with optimized sensing matrix, RLS with random sensing matrix, and RLS with optimized sensing matrix, respectively.

We set $M = 50$, $D = 10$, and $S = 3$. The variance of the noise is 0.001, while the variance of the input signal is 1. For the ETF construction method, $\eta = 1.5\underline{\mu}$. The numbers of iterations for constructing ETF, alternately updating $\boldsymbol{U}$ and $\boldsymbol{V}$ are both fixed to 100. The MSE performance and the distance between the unknown parameter vector $\boldsymbol{\omega}_0$ and its estimate are shown for comparison. All results are obtained by averaging 1000 independent runs.

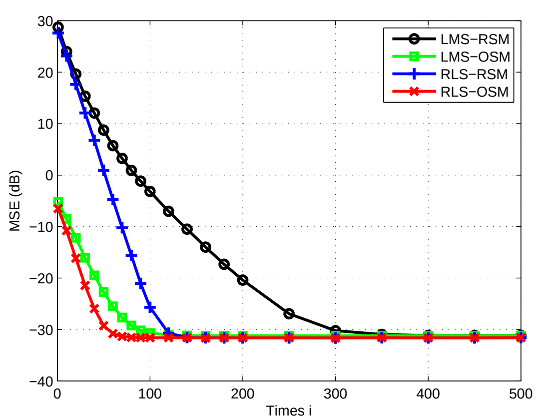

Fig. 2. MSE versus time instant with $D = 10$. The step-size for LMS is $\mu_0 = 0.45$, the parameters for RLS are $\delta = 0.1$ and $\lambda = 0.9$.

As shown in Fig. 2, the MSE performance of the proposed D-CE with different sensing matrices and local estimator updating algorithms are compared. The RLS versions outperform the LMS versions while the proposed sensing matrix design provides DCE a faster convergence when compared with DCE without the sensing matrix optimization.

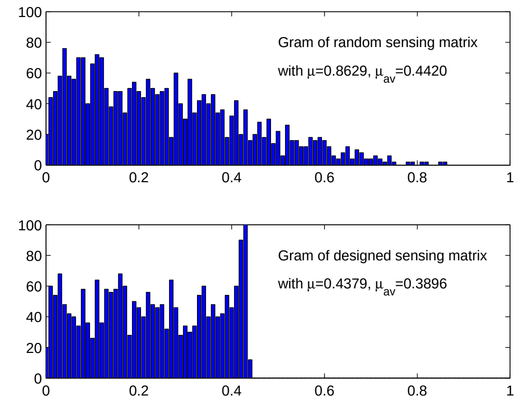

Fig. 3. Gram matrices of different sensing matrices with $D = 10$.

The corresponding Gram matrices of random sensing matrix and designed sensing matrix are depicted in Fig. 3. We set $\bar{\mu} = \underline{\mu}$ when calculating μ_{av}. It can be found that the designed sensing matrix has a smaller mutual coherence and a smaller average mutual coherence. These properties lead to a better reconstruction and hence improve the performance of the system.

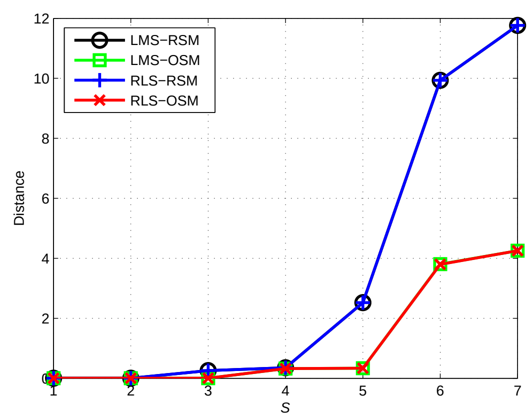

Fig. 4. Distance between the $\boldsymbol{\omega}_0$ and its estimate versus sparsity S.

In the last simulation, the same setting as to obtain Fig. 2 is employed except for the sparsity S. Fig. 4 shows the distance between the unknown parameter vector $\boldsymbol{\omega}_0$ and its estimate versus sparsity S. The proposed sensing matrix introduces a shorter distance than the random sensing matrix, while in this scenario, LMS and RLS adaptive algorithms for updating the local estimator achieve similar performance.

6. CONCLUSIONS

A novel DCE scheme and adaptive RLS version for calculating local estimators have been proposed for sparse parameter vector estimation based on CS techniques. In addition, the sensing matrix has been optimized to minimize the mutual coherence. In the D-CE 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.

7. REFERENCES

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

[2] S. Xu, R. C. de Lamare, and H. V. Poor, “Adaptive link selection strategies for distributed estimation in diffusion wireless networks,” *in Proc. IEEE ICASSP*, Vancouver, Canada, 2013.

[3] 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.

[4] S. Chouvardas, K. Slavakis, Y. Kopsinis, and S. Theodoridis, “A sparsity promoting adaptive algorithm for distributed learning,” *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5412-5425, Oct. 2012.

[5] P. Di Lorenzo and A. H. Sayed, “Sparse distributed learning based on diffusion adaptation,” *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1419-1433, March 2013.

[6] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, April 2006.

[7] E. J. Candès and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21-30, Mar. 2008.

[8] J. Romberg, “Imaging via compressive sampling,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 14-20, March 2008.

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

[10] 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.

[11] 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.

[12] S. Xu. R. C. de Lamare, and H. V. Poor, “Distributed compressed estimation based on compressive sensing,” *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1311-1315, Sept. 2015.

[13] 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.

[14] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthonormal) dictionaries via l_1 minimization,” *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197-2202, Mar. 2003.

[15] J. A. Tropp, “Greed is good: Algorithmic results for sparse approximation,” *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231-2242, Oct. 2004.

[16] M. Elad, “Optimized projections for compressed sensing,” *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695-5702, Dec. 2007.

[17] T. Strohmer and R. W. Heath, “Grassmannian frames with applications to coding and communication,” *Appl. Comp. Harmonic Anal.*, vol. 14, no. 3, pp. 257-275, May 2003.

[18] G. Li, Z. Zhu, D. Yang, L. Chang, and H. Bai, “On projection matrix optimization for compressive sensing systems,” *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2887-2898, Jun. 2013.

[19] H. Bai, G. Li, S. Li, Q. Li, Q. Jiang, and L. Chang, “Alternating optimization of sensing matrix and sparsifying dictionary for compressed sensing,” *IEEE Trans. Signal Process.*, vol. 63, no. 6, pp. 1581-1594, Mar. 2015.

[20] J. A. Tropp, I. S. Dhillon, R. W. Heath, Jr., and T. Strohmer, “Designing structured tight frame via alternating projection,” *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 188-209, 2005.

[21] E. V. Tsiligianni, L. P. Kondi, and A. K. Katsaggelos, “Construction of incoherent unit norm tight frames with application to compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2319-2330, Apr. 2014.

[22] U. Niesen, D. Shah, and G. W. Wornell, “Adaptive alternating minimization algorithms,” *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1423-1429, Mar. 2009.

[23] R. C. de Lamare and R. Sampaio-Neto, “Reduced-rank adaptive filtering based on joint iterative optimization of adaptive filters,” *IEEE Signal Processing Letters*, vol. 14, no. 12, December 2007.

[24] R. C. de Lamare and R. Sampaio-Neto, “Adaptive reduced-rank processing based on joint and iterative interpolation, decimation, and filtering,” *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2503-2514, July 2009.

[25] R. C. de Lamare and R. Sampaio-Neto, “Reduced-rank space-time adaptive interference suppression with joint iterative least squares algorithms for spread-spectrum systems,” *IEEE Transactions on Vehicular Technology*, vol. 59, no. 3, pp. 1217-1228, March 2010.