Proceedings of the 35th Chinese Control Conference
July 27-29, 2016, Chengdu, China

Compressive Distributed Adaptive Filtering

Siyu Xie[1], Lei Guo[1]

1. Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, CAS, Beijing 100190, P. R. China
E-mail: xiesiyu7@gmail.com, lguo@amss.ac.cn

Abstract: The paper proposes a class of compressive distributed adaptive filtering algorithms, aiming to estimate unknown high-dimensional and sparse parameters in sensor networks, based on compressive sensing (CS) method. The algorithms first use compression estimation to obtain the compressed unknown parameters, then apply decompression algorithms to obtain the desired estimates. In the paper, we focus on compressive distributed least mean square (CDLMS) algorithms and show that the algorithms can fulfil the estimation or tracking tasks under a compressed information condition, which is weaker than the information condition for distributed LMS algorithm in [1].

Key Words: Sparse, compressive sensing, the CDLMS algorithms, compressed information condition, stochastic stability

1 Introduction

One fundamental signal processing problem in sensor networks is to estimate and track an unknown dynamic process of interest from noisy measurements cooperatively. In distributed estimation algorithms, each sensor will construct a local estimate of the unknown parameter based on the noisy measurements or estimates from its own and other neighbouring sensors, e.g. [1]-[5], where practical motivations, various implementations and theoretical analysis for distributed filtering are presented. For big data problems, the dimensions of the unknown parameters may be incredibly high, which is difficult to obtain the estimation or tracking results. In some practical situations like radar systems, multi propagation, etc., the unknown high-dimensional parameters can be sparse on some basis, which means that only a few elements of the parameters are non-zero in the domain.

There are many researches about sparse parameter estimation or tracking problems and one important class is influenced by compressive sensing (CS) theory [6][7]. Further, the existing literature can be roughly classified into two categories depending on the dimensions considered in the estimation or tracking process: One is the case where full dimension is considered. In this scenario, the error function is regularized by adding another function which takes into account the sparsity of unknown parameters, for example, Chen, Gu and Hero proposed a new approach to adaptive system identification when system model is sparse and applied l_1 relaxation in cost function to speed up convergence and reduce MSE in [8], Lorenzo, Barbarossa and Sayed established diffusion LMS techniques, which can exploit sparsity in the model, for distributed estimation over sensor networks in [9] and so on. Another category considers the reduced dimension. Bajwa et al. [10] and Baron et al. [11] both applied compressive sensing in sensor networks, while the CS techniques are used in the transit layer. However, Hosseini and Shayesteh [12] proposed a new method in which they identified the sparse system in the compressed domain, which means that the CS method can be applied in estimation layer, i.e. the proposed algorithm may estimate the compressed parameter instead of the original one. In addition, Xu et al. [13] considered a novel diffusion compressed estimation scheme for sparse signals in compressed state, which can reduce the bandwidth and increase the convergence. In our work, we consider extremely high-dimensional sparse parameters and we also study the second case.

However, the second category in the literature lacks theoretical analysis for stability. Both the individual and distributed compressive algorithms verify the effectiveness of CS applications in parameter estimation through the simulation results instead of the theoretical analysis. In contrast, the results that we are going to present in this paper will give the stability analysis for the CDLMS algorithms under a compressed information condition. We should notice that the stability analysis will be based on the stability results of [1], in which we have obtained the exponential stability and upper bound of the tracking error under a weakest possible information condition without independency and stationarity considerations. In addition, we will show that under the compressed information condition, the CDLMS algorithms can accomplish the estimation task, while the information condition in [1] may not be satisfied so that the distributed LMS algorithm can not estimate the unknown high-dimensional sparse parameters.

This work is organized as follows. After introducing the problem formulation in Section 2, we will present stability analysis of CDLMS algorithms in Section 3. In Section 4, we give some simulation results and some conclusions are made in Section 5.

Notations: In the sequel, $\mathbb{R}^{m\times n}$ denotes the set of $m \times n$ matrices and $\mathbb{R}^m$ denotes the set of m-dimensional column vectors. I_m denotes an unit diagonal matrix with m rows and m columns and $O_{m\times n}$ denotes a zero matrix with m rows and n columns. $\lambda_{max}(\cdot)$ and $\lambda_{min}(\cdot)$ denote the largest and the smallest eigenvalues of matrix $(\cdot)$ respectively. The norm $\|x\|_{l_0}$ of vector $x \in \mathbb{R}^m$ is defined as the number of the non-zero coefficients and the norm $\|x\|_{l_1}$ of vector $x \in \mathbb{R}^m$ is defined as $\sum_{i=1}^m |x_i|$, where $x_i (i = 1, \dots, m)$ is the ith element of x. For any matrix $X \in \mathbb{R}^{m\times n}$, its norm is defined as $\| X \| = \{\lambda_{max}(XX^\tau)\}^{\frac{1}{2}}$ and for any random matrix A_k, its L_p-norm is defined as $\| A_k \|_{L_p} = \{\mathbb{E} \| A_k \|^p\}^{\frac{1}{p}}$.

2 Problem Formulation

2.1 Overview of Compressive Sensing

Compressive sensing (CS) [6][7][14] is a new type of sampling theory appeared in the beginning of the 21st century, which predicts that sparse signals can be reconstructed from what was previously believed to be incomplete measurements. CS has attracted considerable attention in many research areas, e.g., medical imaging [15], geological exploration [16], image processing [17] and so on. Compared with Nyqvist sampling theory, CS makes the additional assumption that signal is sparse or can be sparse on some orthonormal basis. A vector $x \in \mathbb{R}^m$ is called s-sparse ($s << m$) if at most s of its coordinates are non-zero. The vector of observation is a $d(s \leq d << m)$-dimensional vector y defined as

$$y = Mx, \tag{1}$$

where $M \in \mathbb{R}^{d\times m}$ is a measurement matrix and each of its d rows can be considered as a basic vector, usually orthogonal. x is thus down sampled to an d-dimensional vector y. The main goal of CS is to recovery x from y by properly choosing the measurement matrix M and by properly solving the reconstruction problem.

The CS theory shows that the restricted isometry property (RIP) constraint on the measurement matrix M can guarantee the high quality of y [7][18][19]. We recall that if for any s-sparse signal $x \in \mathbb{R}^m$, there exists a constant $\delta_s \in [0, 1)$ such that

$$(1 - \delta_s)\|x\|^2 \leq \|Mx\|^2 \leq (1 + \delta_s)\|x\|^2, \tag{2}$$

then the measurement matrix $M \in \mathbb{R}^{d\times m}$ satisfies the sth order RIP and δ_s is the sth RIP constant. Some random matrices, e.g., Gaussian measurements, Bernoulli measurements, Fourier measurements and Incoherent measurements, can satisfy RIP for any sparse vector with high possibility [18]. In our work, we employ Gaussian measurements for our simulations, i.e., the entries of measurement matrix M are independently sampled from a Gauss distribution with zero mean and variance $1/d$ [20]. About random matrices, we have the following theorem.

Theorem 2.1 [14]: Suppose the measurement matrix $M \in \mathbb{R}^{d\times m}$ is a Gaussian or Bernoulli random matrix. If

$$s \leq C_1 d/\log(m/s), \tag{3}$$

then the measurement matrix M satisfies condition (2) with probability not less than

$$1 - exp(-C_2 d),$$

where C_1 and C_2 are constants which only depend on sth RIP constant δ_s.

Remark 2.1: In fact, [21] gives a simple proof for *Theorem 2.1* and shows the connection between C_1 and C_2, in which constant C_1 should be small enough to ensure $C_2 > 0$, see *Theorem 5.2* in [21] for details. From [21], we may have $C_1 = C_0\delta_s^2$ where C_0 is a bounded constant. Then we can obtain

$$d = O(s\log(m/s)/\delta_s^2).$$

with probability not less than $1 - exp(-C_2 d)$.

In order to design a signal reconstruction algorithm to obtain the sparse solution of (1), it is natural for us to add the following sparse constraint

$$P_0 : \quad \min_{x\in\mathbb{R}^m} \|x\|_{l_0} \quad s.t.\ Mx = y, \tag{4}$$

Unfortunately, l_0 norm is not convex and the computation of (4) is $\mathcal{NP}$ hard [22]. To overcome the difficulty, we consider another kind of constraint, i.e., the l_1 norm

$$P_1 : \quad \min_{x\in\mathbb{R}^m} \|x\|_{l_1} \quad s.t.\ Mx = y, \tag{5}$$

We denote the decoding result as $\Delta_1(Mx)$ and from [23], we know that RIP is sufficient to guarantee the high property of $\Delta_1(Mx)$. When x is s-sparse, we have the following result.

Theorem 2.2 [14]: If $M \in \mathbb{R}^{d\times m}$ satisfies $2s$th RIP and RIP constant $\delta_{2s} \leq \sqrt{2} - 1$, then for all s-sparse signal x,

$$\Delta_1(Mx) = x. \tag{6}$$

2.2 Compressive Distributed LMS Algorithms

In sensor networks, we consider a set of n sensors and the network connections are usually modeled as a weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ with the set of nodes $\mathcal{V} = \{1, 2, ..., n\}$, the set of edges $\mathcal{E}$ where $(i, j) \in \mathcal{E}$ if and only if node j is a neighbor of node i and the weighted adjacency matrix $\mathcal{A} = (a_{ij})_{n\times n}$ with $0 \leq a_{ij} \leq 1, \sum_{j=1}^n a_{ij} = 1, \forall i, j = 1, \ldots, n$. We assume that the graph $\mathcal{G}$ is undirected, then $a_{ij} = a_{ji}$. Node i denotes the ith sensor and the set of neighbors of sensor i is denoted as

$$\mathcal{N}_i = \{l \in \mathcal{V}|(i, l) \in \mathcal{E}\}.$$

The Laplacian matrix $\mathcal{L}$ of the graph $\mathcal{G}$ is defined as $\mathcal{L} = \mathcal{D} - \mathcal{A}$ [24], where $\mathcal{D} = diag\{d_1, d_2, ..., d_n\}$ and $d_i = \sum_{j=1}^n a_{ij} = 1$, i.e., $\mathcal{L} = I_{n\times n} - \mathcal{A}$. From [24], we know that the Laplacian matrix $\mathcal{L}$ of graph $\mathcal{G}$ has at least one zero eigenvalue, with other eigenvalues positive and not more than 2. Furthermore, if the graph $\mathcal{G}$ is connected, $\mathcal{L}$ has only one zero eigenvalue.

At every time instant k, every sensor $i(i = 1, \ldots, n)$ can take a measurement according to the following time-varying stochastic linear regression model

$$y_k^i = (\varphi_k^i)^\tau \theta_k + v_k^i, \quad k \geq 0, \tag{7}$$

where y_k^i, v_k^i are scalar observation and scalar noise, φ_k^i is the $m \times 1$-dimensional stochastic regressor of sensor i and θ_k is an unknown $m \times 1$-dimensional parameter vector which can be estimated or tracked in real-time through local interactions between connected sensors. In this paper, we assume that θ_k is a sparse vector with $s(s << m)$ non-zero coefficients and we can denote the iteration of θ_k as follows

$$\theta_k = \theta_{k-1} + \gamma\omega_k, \quad k \geq 1, \tag{8}$$

where γ is a scalar value and ω_k is an yet undefined variable which is the same sparse as θ_k. When $\gamma = 0$, θ_k is time-invariant. Otherwise, it is time-varying. In addition, we also assume that $\varphi_k^i(i = 1, \ldots, n)$ are s-sparse.

In network systems, if sensor i has access only to the information from its neighbors $\{l \in \mathcal{N}_i\}$, we can adopt the following distributed LMS algorithm which is known as the consensus-type algorithm [25][26]

$$\widehat{\theta}_{k+1}^i = \widehat{\theta}_k^i + \mu\Big\{\frac{\varphi_k^i}{1+\parallel \varphi_k^i \parallel^2}[y_k^i - (\varphi_k^i)^\tau\widehat{\theta}_k^i] - \nu\sum_{l\in\mathcal{N}_i} a_{li}(\widehat{\theta}_k^i - \widehat{\theta}_k^l)\Big\}, \quad k \geq 0. \tag{9}$$

where $\mu \in (0, \frac{1}{3}), \nu \in (0, 1]$ are adaptation gains and $\widehat{\theta}_k^i$ is the estimator of sensor i at time instant k.

The CS method provides a robust framework which can reduce the dimension of measurements required to estimate a high-dimensional sparse signal. For this reason, we adopt CS to the estimation of θ_k which can reduce the required bandwidth, accelerate the estimation process and even improve the tracking performance. At every time instant k, sensor $i(i = 1, \ldots, n)$ first observes the $m \times 1$-dimensional stochastic regressor φ_k^i, then with the help of measurement matrix $M \in \mathbb{R}^{d\times m}, (s \leq d << m)$ we can obtain the compressed $d \times 1$-dimensional regressor $\psi_k^i = M\varphi_k^i$. Instead of estimating $m \times 1$-dimensional parameter θ_k, the proposed method estimates the compressed $d \times 1$-dimensional parameter $\vartheta_k = M\theta_k$. Note that $\vartheta_k = \vartheta_{k-1} + \gamma\bar{\omega}_k, k \geq 1$, where $\bar{\omega}_k = M\omega_k$. Then we have

$$\begin{aligned} y_k^i &= (\varphi_k^i)^\tau \theta_k + v_k^i \\ &= (\psi_k^i)^\tau \vartheta_k + (\varphi_k^i)^\tau \theta_k - (\psi_k^i)^\tau \vartheta_k + v_k^i \\ &= (\psi_k^i)^\tau \vartheta_k + (\varphi_k^i)^\tau [I_m - M^\tau M]\theta_k + v_k^i \\ &= (\psi_k^i)^\tau \vartheta_k + \bar{v}_k^i, \quad k \geq 0. \end{aligned} \tag{10}$$

In this case, when we estimate ϑ_k, we consider the latter two items $(\varphi_k^i)^\tau [I_m - M^\tau M]\theta_k + v_k^i$ as a new scalar noise $\bar{v}_k^i$. We can obtain the following CDLMS algorithm

$$\begin{aligned} \widehat{\vartheta}_{k+1}^i = \widehat{\vartheta}_k^i + \mu \Big\{ &\frac{\psi_k^i}{1+ \parallel \psi_k^i \parallel^2} [y_k^i - (\psi_k^i)^\tau \widehat{\vartheta}_k^i] \\ &- \nu \sum_{l\in\mathcal{N}_i} a_{li} (\widehat{\vartheta}_k^i - \widehat{\vartheta}_k^l) \Big\}, \quad k \geq 0, \end{aligned} \tag{11}$$

where $\mu \in (0, \frac{1}{3}), \nu \in (0, 1]$ are adaptation gains, $\widehat{\vartheta}_k^i$ is the compressed estimator of sensor i at time instant k by the CDLMS algorithm. After the iteration, each sensor will employ the reconstruction algorithm to obtain the decompressed estimator $\bar{\theta}_k^i$ as the final estimation of θ_k. In practice, we can apply the reconstruction algorithm every $K(K >> 1)$ iterations to obtain $\bar{\theta}_K^i, \bar{\theta}_{2K}^i, \ldots$.

Remark 2.2: For some signal processing problems, e.g. echo cancellation or adaptive equalizer, we may apply a modified measurement function instead of the original one just like literature [13]. In this situation, the scalar measurement of sensor $i(i = 1, \ldots, n)$ can be modified and given by

$$\bar{y}_k^i = (\psi_k^i)^\tau \vartheta_k + v_k^i = (\bar{\varphi}_k^i)^\tau \theta_k + v_k^i, \quad k \geq 0, \tag{12}$$

where $(\bar{\varphi}_k^i)^\tau = (\varphi_k^i)^\tau M^\tau M \in \mathbb{R}^m$. Then we have the following modified CDLMS algorithm

$$\begin{aligned} \widehat{\vartheta}_{k+1}^i = \widehat{\vartheta}_k^i + \mu \Big\{ &\frac{\psi_k^i}{1+ \parallel \psi_k^i \parallel^2} [\bar{y}_k^i - (\psi_k^i)^\tau \widehat{\vartheta}_k^i] \\ &- \nu \sum_{l\in\mathcal{N}_i} a_{li} (\widehat{\vartheta}_k^i - \widehat{\vartheta}_k^l) \Big\}, \quad k \geq 0. \end{aligned} \tag{13}$$

The main difference between (10) and (12) is the noise, which is $\bar{v}_k^i$ instead of v_k^i.

For convenience of analysis, we introduce the following notations

$$\begin{aligned} &Y_k \triangleq col\{y_k^1, \ldots, y_k^n\}, \ \ \bar{Y}_k \triangleq col\{\bar{y}_k^1, \ldots, \bar{y}_k^n\}, \\ &\Phi_k \triangleq diag\{\varphi_k^1, \ldots, \varphi_k^n\}, \ \ \Psi_k \triangleq diag\{\psi_k^1, \ldots, \psi_k^n\}, \\ &V_k \triangleq col\{v_k^1, \ldots, v_k^n\}, \ \ \bar{V}_k \triangleq col\{\bar{v}_k^1, \ldots, \bar{v}_k^n\}, \\ &\Omega_k \triangleq col\{\underbrace{\omega_k, \ldots, \omega_k}_n\}, \ \ \bar{\Omega}_k \triangleq col\{\underbrace{\bar{\omega}_k, \ldots, \bar{\omega}_k}_n\}, \\ &\Theta_k \triangleq col\{\underbrace{\theta_k, \ldots, \theta_k}_n\}, \ \ \mathit{\Theta}_k \triangleq col\{\underbrace{\vartheta_k, \ldots, \vartheta_k}_n\}, \\ &\widehat{\mathit{\Theta}}_k \triangleq col\{\widehat{\vartheta}_k^1, \ldots, \widehat{\vartheta}_k^n\}, \\ &\widetilde{\mathit{\Theta}}_k \triangleq col\{\widetilde{\vartheta}_k^1, \ldots, \widetilde{\vartheta}_k^n\}, \text{ where } \widetilde{\vartheta}_k^i = \widehat{\vartheta}_k^i - \vartheta_k, \\ &\bar{L}_k \triangleq diag\Big\{\frac{\psi_k^1}{1+ \parallel \psi_k^1 \parallel^2}, \ldots, \frac{\psi_k^n}{1+ \parallel \psi_k^n \parallel^2}\Big\}, \\ &\bar{F}_k \triangleq \bar{L}_k \Psi_k^\tau, \ \ \mathcal{M} = diag\{\underbrace{M, \ldots, M}_n\}, \\ &\bar{G}_k \triangleq \bar{F}_k + \nu(\mathcal{L} \otimes I_d). \end{aligned}$$

By (10), we have

$$Y_k = \Psi_k^\tau \mathit{\Theta}_k + \bar{V}_k, \tag{14}$$

From (11), we have

$$\widehat{\mathit{\Theta}}_{k+1} = \widehat{\mathit{\Theta}}_k + \mu \bar{L}_k (Y_k - \Psi_k^\tau \widehat{\mathit{\Theta}}_k) - \mu\nu(\mathcal{L} \otimes I_d)\widehat{\mathit{\Theta}}_k, \tag{15}$$

where matrix $\mathcal{L}$ is the Laplacian matrix of graph $\mathcal{G}$ and $\otimes$ is the Kronecker Product [27]. Since $\widetilde{\mathit{\Theta}}_k = \widehat{\mathit{\Theta}}_k - \mathit{\Theta}_k$, we can get

$$\begin{aligned} \widetilde{\mathit{\Theta}}_{k+1} =& \widetilde{\mathit{\Theta}}_k - \mu \bar{L}_k \Psi_k^\tau \widetilde{\mathit{\Theta}}_k - \mu\nu(\mathcal{L} \otimes I_d)\widehat{\mathit{\Theta}}_k \\ &+ \mu \bar{L}_k \bar{V}_k - \gamma \bar{\Omega}_{k+1}. \end{aligned}$$

and because $(\mathcal{L} \otimes I_d)\mathit{\Theta}_k = 0$, we have

$$\begin{aligned} \widetilde{\mathit{\Theta}}_{k+1} =& \{I_{dn} - \mu[\bar{F}_k + \nu(\mathcal{L} \otimes I_d)]\}\widetilde{\mathit{\Theta}}_k \\ &+ \mu \bar{L}_k \bar{V}_k - \gamma \bar{\Omega}_{k+1} \\ =& (I_{dn} - \mu \bar{G}_k)\widetilde{\mathit{\Theta}}_k + \mu \bar{L}_k \bar{V}_k - \gamma \bar{\Omega}_{k+1}. \end{aligned} \tag{16}$$

So far, we have obtained the compressive distributed LMS error equation (16). After the estimation iteration, we use the recovery algorithm to obtain the decompressed parameters $\bar{\Theta}_k \triangleq col\{\bar{\theta}_k^1, \ldots, \bar{\theta}_k^n\}$ which will be analyzed in the following section.

Remark 2.3: For (12) and (13), we have the modified compressive distributed LMS error equation

$$\widetilde{\mathit{\Theta}}_{k+1} = (I_{dn} - \mu \bar{G}_k)\widetilde{\mathit{\Theta}}_k + \mu \bar{L}_k V_k - \gamma \bar{\Omega}_{k+1}. \tag{17}$$

2.3 Some Definitions

To proceed with further discussions, we need the following definitions in [1]. Also, $\mathcal{F}_k = \sigma\{\varphi_j^i, \omega_j, v_{j-1}^i, i = 1, \ldots, n, j \leq k\}$.

Definition 2.1 : For a sequence of $l \times l$ random matrices $A = \{A_k, k \geq 0\}$, if it belongs to the following set

$$\begin{aligned} S_p(\lambda) = \Big\{ A : &\Big\| \prod_{j=i+1}^{k} (I - A_j) \Big\|_{L_p} \leq N\lambda^{k-i}, \\ &\forall k \geq i, \forall i \geq 0, \text{ for some } N > 0 \Big\}, \end{aligned} \tag{18}$$

then $I - A$ is called L_p-exponentially stable ($p \geq 0$) with parameter $\lambda \in [0, 1)$.

For convenience of discussions, we introduce the following subclass of $S_1(\lambda)$ for a scalar sequence $a = \{a_k, k \geq 0\}$

$$S^0(\lambda) = \Big\{a : a_k \in [0,1], E\prod_{j=i+1}^{k}(1-a_j) \leq N\lambda^{k-i}, \quad \forall k \geq i, \forall i \geq 0, \text{ for some } N > 0\Big\}, \tag{19}$$

where $\lambda \in [0, 1)$. In addition $S^0 \triangleq \bigcup_{\lambda\in(0,1)} S^0(\lambda)$.

Definition 2.2 : A random sequence $x = \{x_k\}$ is called an element of the set $\mathcal{M}_p(p \geq 1)$, if there exists a constant C_p^x depending only on p and the distribution of $\{x_k\}$ such that for $k \geq 0$,

$$\Big\|\sum_{i=k+1}^{k+h} x_i\Big\|_{L_p} \leq C_p^x h^{\frac{1}{2}}, \quad \forall h \geq 1. \tag{20}$$

Remark 2.4 : It is known that the martingale difference, zero mean ϕ- and α-mixing sequences, and the linear process driven by white noises are all in $\mathcal{M}_p$ [28].

Definition 2.3 : Let $\{A_k\}$ be a matrix sequence and $b_k, \forall k \geq 0$ be a positive scalar sequence. Then by $A_k = O(b_k)$ we mean that there exists a constant $N < \infty$ such that

$$\|A_k\| \leq N b_k, \quad \forall k \geq 0. \tag{21}$$

3 The Main Results

To proceed with the stability analysis, we need the following conditions.

Condition 3.1(**Connectivity**) : The graph $\mathcal{G}$ is connected.

Condition 3.2(**Compressed Information Condition**) : Let $\{\psi_k^i, \mathcal{F}_k, k \geq 0\}(i = 1, \ldots, n)$ be n adapted sequences and there exists an integer $h > 0$ such that $\{\lambda_k, k \geq 0\} \in S^0(\lambda)$ for some $\lambda \in (0, 1)$, where λ_k is defined by

$$\lambda_k \triangleq \lambda_{min}\Big\{E\Big[\frac{1}{n(h+1)}\sum_{i=1}^{n}\sum_{j=k+1}^{k+h}\frac{\psi_j^i(\psi_j^i)^\tau}{1+\|\psi_j^i\|^2}\Big|\mathcal{F}_k\Big]\Big\}. \tag{22}$$

Remark 3.1 : The main difference between *Condition 3.2* and the information condition in [1] is that here we use the compressed regressors ψ_j^i instead of the original ones φ_j^i. Note that the dimension of ψ_j^i is much less than m, so *Condition 3.2* is much easier to be satisfied in practical situations. Obviously, when the original regressors are high-dimensional and very sparse, the information condition may not be satisfied. In other words, the distributed LMS algorithm (9) can not fulfill the estimation tasks, while the CDLMS algorithms may stably obtain the compressed estimation results.

We are in the position to give our first main result.

Theorem 3.1: Consider the model (14) and the estimation error equation (16). Suppose that $M \in \mathbb{R}^{d\times m}$ satisfies $2s$th RIP with the RIP constant $\delta_{2s} \leq \sqrt{2} - 1$ and we use l_1 norm reconstruction algorithm. Assume that *Conditions 3.1* and *3.2* are satisfied. Then for any $\mu \in (0, \frac{1}{3}), \nu \in (0, 1)$, $\{I_{mn} - \mu\bar{G}_k, k \geq 1\}$ is L_p-exponentially stable ($p \geq 1$). Furthermore, if for some $p \geq 1$ and $\beta > 1$, $\sigma_p \triangleq \sup_k \|\xi_k log^\beta(e + \xi_k)\|_{L_p} < \infty, \|\widetilde{\Theta}_0\|_{L_p} < \infty$ hold where $\xi_k = 3\delta_{2s}\|\Phi_k\| \cdot \|\Theta_k\| + \|V_k\| + \|\bar{\Omega}_{k+1}\|$, then

$$\limsup_{k\to\infty} \|\bar{\Theta}_k - \Theta_k\|_{L_p} \leq \frac{c}{1-2\delta_{2s}}[\sigma_p log(e + \sigma_p^{-1})],$$

where c is a positive constant.

Proof. By [1] and *Condition 3.1, 3.2*, we know that $\{I_{mn} - \mu\bar{G}_k, k \geq 1\}$ is L_p-exponentially stable ($p \geq 1$).

Since φ_k^i and θ_k are both s-sparse, we denote the positions of their non-zero entries as $i_1, \ldots, i_s$ and $j_1, \ldots, j_s$, respectively. Then we retain all the above $2s$ entries of φ_k^i, θ_k and the $2s$ columns of matrix M and we remove all other entries. Then denote them as $\varphi_{k,2s}^i, \theta_{k,2s}$ and M_{2s} and since M satisfies $2s$th RIP, we know that all the eigenvalues of $M_{2s}^\tau M_{2s}$ are in set $[1 - \delta_{2s}, 1 + \delta_{2s}]$ [14] and

$$\begin{aligned}
&\|(\varphi_k^i)^\tau[I_m - M^\tau M]\theta_k\| \\
=&\|(\varphi_{k,2s}^i)^\tau[I_{2s} - M_{2s}^\tau M_{2s}]\theta_{k,2s}\| \\
\leq&\|(\varphi_{k,2s}^i)^\tau[(1+\delta_{2s})I_{2s} - M_{2s}^\tau M_{2s}]\theta_{k,2s}\| \\
&+ \delta_{2s}\|(\varphi_{k,2s}^i)^\tau\theta_{k,2s}\| \\
\leq&\|(\varphi_{k,2s}^i)^\tau[(1+\delta_{2s})I_{2s} - M_{2s}^\tau M_{2s}]^{\frac{1}{2}}\| \\
&\cdot \|[(1+\delta_{2s})I_{2s} - M_{2s}^\tau M_{2s}]^{\frac{1}{2}}\theta_{k,2s}\| \\
&+ \delta_{2s}\|(\varphi_{k,2s}^i)^\tau\theta_{k,2s}\| \\
\leq&\sqrt{(\varphi_{k,2s}^i)^\tau \cdot 2\delta_{2s} \cdot \varphi_{k,2s}^i} \cdot \sqrt{\theta_{k,2s}^\tau \cdot 2\delta_{2s} \cdot \theta_{k,2s}} \\
&+ \delta_{2s}\|(\varphi_{k,2s}^i)^\tau\theta_{k,2s}\| \\
\leq&2\delta_{2s}\|\varphi_{k,2s}^i\| \cdot \|\theta_{k,2s}\| + \delta_{2s}\|\varphi_{k,2s}^i\| \cdot \|\theta_{k,2s}\| \\
=&3\delta_{2s}\|\varphi_{k,2s}^i\| \cdot \|\theta_{k,2s}\| \\
\leq&3\delta_{2s}\|\varphi_k^i\| \cdot \|\theta_k\|.
\end{aligned}$$

Denote $\bar{V}_k = col\{\bar{v}_k^1, \ldots, \bar{v}_k^n\} = col\{(\varphi_k^1)^\tau[I_m - M^\tau M]\theta_k + v_k^1, \ldots, (\varphi_k^n)^\tau[I_m - M^\tau M]\theta_k + v_k^n\}$. Then we have

$$\|\bar{V}_k\| \leq 3\delta_{2s}\|\Phi_k\| \cdot \|\Theta_k\| + \|V_k\|.$$

By [1], we know that

$$\limsup_{k\to\infty} \|\widetilde{\Theta}_k\|_{L_p} \leq c[\sigma_p log(e + \sigma_p^{-1})],$$

holds. Since $\bar{\theta}_k^i$ is s-sparse and by *Theorem 2.1*, we know that the recovery process is exact and s-sparse. By RIP condition, we have

$$\begin{aligned}
&(1-\delta_{2s})\|\bar{\theta}_k^i - \theta_k\|_{L_p} \\
\leq&\|M\bar{\theta}_k^i - M\theta_k\|_{L_p} \\
=&\|\widehat{\vartheta}_k^i - \vartheta_k^i\|_{L_p},
\end{aligned} \tag{23}$$

Then we can obtain

$$\begin{aligned}
\limsup_{k\to\infty} \|\bar{\Theta}_k - \Theta_k\|_{L_p} &\leq \frac{1}{1-\delta_{2s}}\limsup_{k\to\infty} \|\widehat{\Theta}_k - \Theta_k\|_{L_p} \\
&\leq \frac{c}{1-\delta_{2s}}[\sigma_p log(e + \sigma_p^{-1})].
\end{aligned}$$

□

Remark 3.2: From *Remark 2.1*, we know that

$$\delta_s^2 = O(s \log(m/s)/d).$$

When the order of d is larger than $\log(m/s)$, for example $d = O(\log^{\alpha}(m/s))(\alpha > 1)$, we have $\delta_s \to 0$ as $m \to \infty$ with high probability when sparsity s increases slower than m, e.g. s is a fixed integer. This clearly demonstrates the meaningfulness of *Theorem 3.1*.

Remark 3.3: Since the noise is V_k instead of $\bar{V}_k$ in (12) and (13), we have the following similar results for (17).

Theorem 3.2: Consider the model (14) and the estimation error equation (17). Suppose that $M \in \mathbb{R}^{d\times m}$ satisfies $2s$th RIP, the RIP constant $\delta_{2s} \le \sqrt{2} - 1$ and we use l_1 norm reconstruction algorithm. Assume that *Conditions 3.1* and *3.2* are satisfied. For any $\mu \in (0, \frac{1}{3}), \nu \in (0, 1)$, then $\{I_{mn} - \mu\bar{G}_k, k \ge 1\}$ is L_p-exponentially stable ($p \ge 1$). Furthermore, if for some $p \ge 1$ and $\beta > 1$, $\sigma_p' \triangleq \sup_k \|\xi_k' log^{\beta}(e + \xi_k')\|_{L_p} < \infty, \|\widetilde{\Theta}_0\|_{L_p} < \infty$ hold where $\xi_k' = \|V_k\| + \|\bar{\Omega}_{k+1}\|$, then

$$\limsup_{k\to\infty} \|\bar{\Theta}_k - \Theta_k\|_{L_p} \le \frac{c}{1-\delta_{2s}}[\sigma_p' log(e + \sigma_p'^{-1})],$$

where c is a positive constant.

Like the distributed LMS algorithm in [1], we may add the following condition to obtain a better upper bound for estimation $\|\bar{\Theta}_k - \Theta_k\|_{L_p}$.

Condition 3.3 : For some $p \ge 1$, the initial estimation error is bounded, i.e. $\|\widetilde{\Theta}_0\|_{L_{2p}} < \infty$. Furthermore, let $\{\bar{L}_k V_k\} \in \mathcal{M}_{2p}$ and $\{\Omega_k\} \in \mathcal{M}_{2p}$.

Corollary 3.1: Consider the model (14) and the estimation error equation (17). Suppose that $M \in \mathbb{R}^{d\times m}$ satisfies $2s$th RIP, the RIP constant $\delta_{2s} \le \sqrt{2} - 1$ and we use l_1 norm reconstruction algorithm. Assume that *Conditions 3.1, 3.2* and *3.3* are all satisfied, then the tracking errors $\{\bar{\Theta}_k - \Theta_k\}$ have a upper bound, i.e. for any $k \ge 0$ and $\mu \in (0, \frac{1}{3}), \nu \in (0, 1)$, we have

$$\|\bar{\Theta}_{k+1} - \Theta_{k+1}\|_{L_p} \le \frac{1}{1-\delta_{2s}} O\left(\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}} + (1-\alpha\mu)^{k+1}\right), \tag{24}$$

where $\alpha \in (0, 1)$ is a constant.

Since the proofs of the two results are similar to [1] and *Theorem 3.1*, we can omit them here.

4 Simulation Results

We consider a partially connected network with $n = 20$ sensors and all the sensors will estimate or track an unknown 50-dimensional 3-sparse signal θ_k and all the coefficients of θ_k are zero except the first 3 coefficients. Let $\gamma = 0$ (when θ_k is time-invariant) or $\gamma = 0.1$ (when θ_k is time-varying) and the first 3 coordinates of variation $\omega_k \sim N(0, 0.1, 3, 1)$ (Gaussian distribution) while the rest of its coordinates are all zero. We also assume that the observation noises $\{v_k^i, k \ge 1, i = 1, \ldots, 20\}$ are temporally and spatially independently distributed with $v_k^i \sim N(0, 0.1)$ in (10).

Let the regressors $\varphi_k^i \in \mathbb{R}^{50} (i = 1, \ldots, 20)$ be generated by the following equation

$$\begin{cases} x_k^i = & A_i x_{k-1}^i + B_i \xi_k^i, \\ \varphi_k^i = & C_i x_k^i, \end{cases}$$

where $\{\xi_k^i, k \ge 1, i = 1, \ldots, 20\}$ are temporally and spatially independently distributed with $\xi_k^i \sim U(-1, 1, 3, 1)$ (uniform distribution in $(-1, 1)$ with 3 dimensions) and $x_k^i \in \mathbb{R}^{50}$, $A_i \in \mathbb{R}^{50\times 50}$, $B_i \in \mathbb{R}^{50\times 3}$ and $C_i \in \mathbb{R}^{50\times 50} (i = 1, \ldots, 20)$. In the following simulations, we compare three different kinds of algorithms to illustrate the advantage of the CDLMS algorithms, i.e. the individual LMS algorithm in [29][30][31], distributed LMS algorithm (9) and compressive distributed LMS algorithm (11).

In this experiment, we choose a special series of A_i, B_i and C_i in order to satisfy *condition 3.2* of this paper and exclude the information condition in [1]. Let $A_i (i = 1, \ldots, 20)$ are all diagonal matrices with the diagonal elements equal to $1/3$ and let

$$B_1 = B_3 = B_5 = B_7 = O_{50\times 3}, \tag{25}$$

and other B_i, C_i be sparse enough to ensure the sparsity 3 of regressors. It can be verified that both of the information condition in [1] and the individual condition in [31] are not satisfied. However, *condition 3.2* in this paper can be satisfied for compressed regressors $\psi_k^i (i = 1, \ldots, 20)$. Let $x_0^1 = \cdots = x_0^{20} = (1, \ldots, 1)_{1\times 50}^{\tau}$ and the first s coordinates of variation $\theta_0 \sim 0.1 * N(0, 1, 3, 1)$ (Gaussian distribution with 3 dimensions), while the rest of its coordinates are all zero. since the sparsity is 3, $d = 10$ is enough for compression and decompression algorithm and we choose $\hat{\theta}_0^i = (0.1, \ldots, 0.1)_{1\times 50}^{\tau} (i = 1, \ldots, 20), \hat{\vartheta}_0^i = (0.1, \ldots, 0.1)_{1\times 10}^{\tau} (i = 1, \ldots, 20), \mu = 0.3, \nu = 1$. For decompression part, we apply an commonly used greedy algorithm, i.e., orthogonal matching pursuit (OMP) algorithm [32][33] to generate the recovery signal every 100 iterations. [34] extends *Theorem 2.1* to OMP algorithm and has a similar result, i.e., when RIP constant satisfies some constraints, we can recovery sparse signal x exactly by OMP algorithm.

For the individual LMS algorithm and the distributed LMS algorithm, we plot the mean estimation or tracking errors over the sensor network for every 100 iterations, i.e., $\frac{1}{2000k}\sum_{i=1}^{20}\sum_{j=1}^{100k}\|\hat{\theta}_j^i - \theta_j\|^2 (k = 1, \ldots, 20)$. Then for the CDLMS algorithm, we plot $\frac{1}{2000k}\sum_{i=1}^{20}\sum_{j=1}^{100k}\|\bar{\theta}_j^i - \theta_j\|^2 (k = 1, \ldots, 20)$. In the upper one of Fig.1, θ_k is time-invariant, mean estimation errors of the individual LMS algorithm and distributed LMS algorithm both keep large because all the sensors do not satisfy the information condition in [31] and the whole sensor network can not satisfy the information condition in [1]. However, mean estimation errors of the compressive one are smaller than other two situations. When θ_k is time-varying, we have the similar result.

5 Conclusions

In this paper, we have investigated a novel class of compressive distributed adaptive filtering algorithms and proved that the CDLMS algorithms can fulfil the estimation task under a compressed information condition, even when the original information condition can not be satisfied. In other words, when the unknown parameters are high-dimensional and sparse, the CDLMS algorithms can fulfil the estimation task while the distributed LMS can not. Of

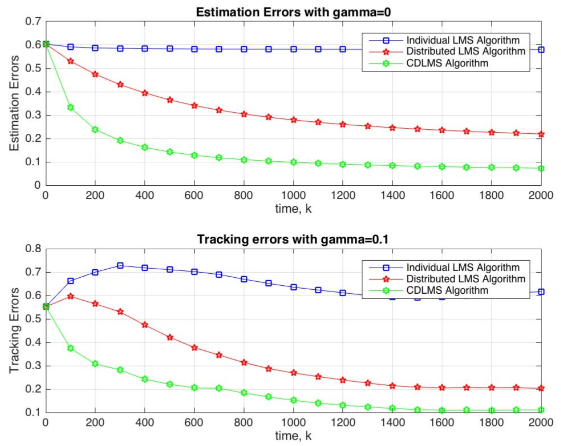

Fig. 1: Estimation errors with $\gamma = 0$ and tracking errors with $\gamma = 0.1$

course, there are still a number of interesting problems for further research, for example, to give a better performance analysis of the CDLMS algorithms and to conduct some real applications of the algorithms.

Acknowledgements

The authors would like to acknowledge Professor Zhiqiang Xu for this helpful suggestion of [21] on the RIP.

References

[1] S. Y. Xie and L. Guo, "Necessary and Sufficient Information Conditions for Stability of Distributed Filters," *Submitted to IEEE Trans. on Signal Processing*, December, 2015.

[2] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Trans. on Signal Process.*, 55(8): 4064–4077, 2007.

[3] C. G. Lopes and A. H. Sayed, "Distributed processing over adaptive networks," in *Proc. Adaptive Sensor Array Process. Workshop*, MIT Lincoln Laboratory, Cambridge, MA, USA, 1–5, 2006.

[4] R. Carli, A. Chiuso, L. Schenato, et al., "Distributed Kalman filtering based on consensus strategies," *Selected Areas in Communications, IEEE Journal on*, 26(4): 622–633, 2008.

[5] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squres over adaptive networks: formulation and performance analysis," *IEEE Trans. on Signal Process.*, 56(7): 3122–3136, 2008.

[6] D.L. Donoho, "Compressed sensing," *IEEE Trans. on Inf. Theory*, 52(4): 1289–1306, 2006.

[7] E. Candes, J. Romberg and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

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

[9] P.D. Lorenzo, S. Barbarossa and A.H. Sayed, "Sparse diffusion LMS for distributed adaptive estimation," *International Conference on Acoustics*, 22(10): 3281–3284, 2012.

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

[11] D. Baron, M.F. Duarte, M.B. Wakin, S. Sarvotham and R.G. Baraniuk, *Distributed compressive sensing*, ECE Dept., Rice Univ., Houston, TX, USA, Tech. Rep., Nov, 2006.

[12] S.H. Hossein and M.G. Shayesteh, "Compressed sensing for denoiseing in adaptive system identification," *Conference on Electrical Engineering*, 1238–1242, 2012.

[13] S.C. Xu, R.C. Damare and H.V. Poor, "Distributed compressed estimation based on compressive sensing," *IEEE Signal Processing Letters*, 22(9): 1311–1315, 2015.

[14] Q.Z. Xu, "Compressed sensing: a survy (in Chinese)," *Sci Sin Math*, 42(9), 865–877, 2012.

[15] D.L. Li, Q. Zhang, H.H .Shen, J.B. .Shi and M.F. Shen, "The application of compressive sensing based on wavelet in the reconstruction of ultrasonic medical image," *Second International Conference on Mechanic Automation and Control Engineering (MACE)*, 5350–5353, 2011.

[16] F.J. Herrmann, M.P. Friedlander and O. Yilmaz, "Fighting the curse of dimensionality: compressive sensing in exploration seismology," *IEEE trans. on Signal Processing*, 29(29): 88–100, 2012.

[17] J. Romberg, "Imaging via compressive sampling," *IEEE Trans. on Signal Processing*, 25(2): 14–20, 2008.

[18] E. Candes, "Compressive sampling," in *Proc. Int. Congr. Math.*, 3: 1433–1452, Madrid, Spain, Jul. 2006.

[19] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Tans. on Inform. Theory*, 51: 4203–4215, 2005.

[20] R.G. Baraniuk, "Compressive sensing," *IEEE Trans. on Signal Process*, 24(4): 118–122, 2007.

[21] R.G. Baraniuk, M. Davenport, R.A. DeVore and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, 28(3): 253–263, 2008.

[22] B.K Natarajan, "Sparse approximate solutions to linear systems," *SIAM J Comput*, 24: 227–234, 1995.

[23] D.L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization," *Proc. Natl. Acad. Sci.*, 100:2197–2202, USA, 2003.

[24] R. P. Agaev and P. Y. Chebotarev, "The matrix of maximum out forests of a digraph and its applications," *Autom. Remote Control*, vol.61, no.9, pp.1424–1450, 2000.

[25] A. H. Sayed, "Adaptive networks," *Proceedings of the IEEE*, 102(4), 460–497, 2014.

[26] I. D. Schizas, G. Mateos and G. B. Giannakis, "Stability analysis of the consensus-based distributed LMS algorithm," *IEEE International Conference on Acoustics, Speech and Signal Processing*, 3289-3292, 2008.

[27] A. J. Laub, Matrix analysis for scientists and Engineers, Philadelphia, PA, USA: SIAM, 2005.

[28] L. Guo and L. Ljung, "Performance analysis of general tracking algorithms," *IEEE Trans. on Automatic Control*, 40(8): 1388-1402, 1995.

[29] L. Ljung and T. Soderstrom, Theory and practice of recursive identification, MIT Press, Cambridge, MA, 1983.

[30] B. Widerow and S. Stearns, Adaptive signal processing, Prentice-Hall Englewood Cliffs NJ, 1985.

[31] L. Guo, "Stability of recursive stochastic tracking algorithms," *SIAM J.on Control and Optimization*, 32(5): 1195–1225,1994.

[32] J.A. Tropp and A.C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. on Inf. Theory*, 53(12): 4655–4666, 2007.

[33] 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*, 1:40–44, Pacific Grove, CA, USA, Nov. 1993.

[34] Z.Xu, "A remark about orthogonal matching pursuit algorithm," ArXiv: 1005.3093.