Analysis of compressed distributed adaptive filters☆

Siyu Xie [a], Lei Guo [b],*

[a] *Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI 48202, USA*
[b] *Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China*

ARTICLE INFO

ABSTRACT

In order to estimate an unknown high-dimensional sparse signal in the network, we present a class of compressed distributed adaptive filtering algorithms based on consensus strategies and compressive sensing (CS) method. This class of algorithms is designed to first use the compressed regressor data to obtain an estimate for the compressed unknown signal, then apply some signal reconstruction algorithms to obtain a high-dimensional estimate for the original unknown signal. Here we consider the compressed consensus normalized least mean squares (NLMS) algorithm, and show that even if the traditional non-compressed distributed algorithm cannot fulfill the estimation or tracking task due to the sparsity of the regressors, the compressed algorithm introduced in this paper can be used to estimate the unknown high-dimensional sparse signal under a compressed information condition, which is much weaker than the cooperative information condition used in the existing literature, without such stringent conditions as independence and stationarity for the system signals.

1. Introduction

For problems of estimating parameters or tracking signals in sensor networks, there are basically three approaches for constructing solutions, i.e., centralized, distributed, and the combination of both. Since the centralized solutions may lack robustness at the fusion center, and need strong communication capability and high energy consumption over the networks, distributed ones have attracted more and more research attention in recent years. In the traditional single sensor case, parameter estimation algorithms are usually investigated under the background of adaptive filters, which include the well-known least mean squares (LMS), recursive least squares (RLS) and Kalman filtering (KF) algorithms (Guo, 1994; Widrow & Stearns, 1985). In the distributed adaptive filtering algorithms, there are multiple sensors and each sensor constructs a local estimate of the unknown parameter based on the noisy measurements and estimates from its own and other neighboring sensors (see e.g., Carli, Chiuso, Schenato, & Zampieri, 2008; Chen & Sayed, 2015a; Piggott & Solo, 2016; Rastegarnia, Tinati, & Khalili, 2010; Sayed, 2014; Schizas, Mateos, & Giannakis, 2009; Xie & Guo, 2018a, 2018b, 2018c). Various motivations, practical implementations and theoretical

analyses for different distributed algorithms have been presented therein. Roughly speaking, three different types of distributed strategies have been investigated, i.e., the incremental strategies (Rastegarnia et al., 2010), consensus strategies (Carli et al., 2008; Schizas et al., 2009; Xie & Guo, 2018a, 2018b); and diffusion strategies (Chen & Sayed, 2015a; Piggott & Solo, 2016; Xie & Guo, 2018c).

Although the distributed estimation algorithms are able to estimate some unknown signals, some *priori* knowledge about the system can be helpful for performance improvement. A common *priori* information is sparsity. In some practical situations like radar systems, multi propagation, channel estimation in Ultra Wideband communication systems (Khalifa, Abdelhafiz, & Zerguine, 2013) etc., the unknown high-dimensional signals can be sparse on some basis, which means that only a few elements of the signals are non-zero in the domain. There are many algorithms which have been developed in the literature for sparse signal estimation problems.

One direction for sparse signal estimation is called sparse optimization. In this scenario, the error function is regularized by adding another item which takes into account the sparsity of the unknown parameter. For example, Chen, Gu, and Hero (2009) and Shi and Shi (2010) presented an approach to adaptive system identification based on LMS when the system model is sparse, and they applied ℓ_1-norm penalty in the cost function to speed up convergence and reduce the mean squares error (MSE). Moreover, several other algorithms for sparse adaptive filtering have also been proposed based on RLS (Angelosante, Bazerque, & Giannakis, 2010; Babadi, Kalouptsidis, & Tarokh, 2010) and

projection-based algorithms (Kopsinis, Slavakis, & Theodoridis, 2011). Furthermore, Chouvardas, Slavakis, Kopsinis, and Theodoridis (2011), Liu, Li, and Zhang (2012), Liu, Liu, and Li (2014), Lorenzo and Sayed (2013) and Sayin and Kozat (2014) studied distributed adaptive filtering algorithms, which can exploit sparsity in the underlying system model, for distributed estimation over sensor networks. Some non-convex penalties were also proposed as relaxations of the ℓ_1-norm (Candès, Wakin, & Boyd, 2008; Chartrand, 2007; Fan & Li, 2001; Zeng, Lin, & Xu, 2016).

Another direction is influenced by the compressive sensing (CS) theory (Candès, Romberg, & Tao, 2006; Donoho, 2006). The CS theory is a new type of sampling theory that appeared in the beginning of the 21st century, which provides a robust framework that can reduce the number of measurements required to estimate a sparse signal. The CS theory has attracted considerable attention in many research areas, e.g., medical imaging (Li, Zhang, Shen, Shi, & Shen, 2011), geological exploration (Herrmann, Friedlander, & Yilmaz, 2012), image processing (Romberg, 2008) and so on. Compared with Nyqvist sampling theory, the CS theory makes the additional assumption that the signal is sparse or can be sparse on some orthonormal basis. Bajwa, Haupt, Sayeed, and Nowak (2006) and Baron, Duarte, Wakin, Sarvotham, and Baraniuk (2009) both applied CS in sensor networks, while the CS techniques are only used in the transit layer. Hosseini and Shayesteh (2012) 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, de Lamare, and Poor (2015) considered a linear regression model formed directly by the compressed regressor, and proposed a novel diffusion compressed estimation scheme for sparse signals in compressed state, which can reduce the bandwidth and increase the convergence. However, most of the existing literature only provide numerical simulations for the effectiveness of the CS-based adaptive filtering algorithms, and lack of comprehensive theoretical analysis.

In a preliminary work (Xie & Guo, 2016), we have proposed a class of compressed consensus NLMS algorithms, and established the stability result under a compressed cooperative information condition without independent and stationary signal assumptions. However, there are some difficulties in the signal reconstruction algorithm used in Xie and Guo (2016), since the decompressed estimate may no longer have the same sparsity as the original unknown parameter.

In this paper, we will study the estimation problem of the unknown high-dimensional sparse signal based on consensus strategies and the CS theory. We will first use the consensus NLMS algorithm to construct a low-dimensional estimate for the compressed unknown signal by using the compressed regressor data. Then we will use a suitable signal reconstruction algorithm to obtain a high-dimensional sparse estimate for the original unknown signal. In our theoretical analysis, we will first provide an upper bound for the compressed estimation error of the compressed algorithm under a compressed information condition, without statistical independent and stationary assumptions for the regressors, and then give an upper bound for the desired high-dimensional estimation error. We would like to remark that in the case where the traditional non-compressed algorithm cannot estimate the unknown high-dimensional sparse signal due to lack of sufficient cooperative information, the "compression-estimation-reconstruction (CER) algorithm" investigated in this paper can accomplish the estimation task under a compressed cooperative information condition, because it is much weaker than the cooperative information condition for the non-compressed algorithm as introduced in Xie and Guo (2018b).

The remainder of the paper is organized as follows. Section 2 presents the data model, the compressed consensus adaptive filters and some useful definitions. In Section 3, we present the main theoretical results. Then we give some simulations and some concluding remarks in Sections 4 and 5, respectively. Some preliminaries on CS theory and signal reconstruction algorithms used in this paper are provided in Appendix.

2. Problem formulation

2.1. Data model

In sensor networks, a set of n sensors are taken as the nodes and the network connections are usually modeled as a weighted undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, with the set of nodes $\mathcal{V} = \{1, 2, \ldots, n\}$, 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 $\mathcal{A} = \{a_{ij}\}_{n \times n}$ where $0 \leq a_{ij} \leq 1, \sum_{j=1}^{n} a_{ij} = 1, \forall i, j = 1, \ldots, n$. Since the graph $\mathcal{G}$ is undirected, then we have $a_{ij} = a_{ji}$. The set of neighbors of sensor i is denoted as $\mathcal{N}_i = \{\ell \in \mathcal{V} | (i, \ell) \in \mathcal{E}\}$. The Laplacian matrix $\mathcal{L}$ of the graph $\mathcal{G}$ is defined as $\mathcal{L} = I_n - \mathcal{A}$, where I_n denotes an $n \times n$-dimensional identity matrix. By the graph theory (see e.g., Agaev & Chebotarev, 2006), 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, each sensor $i(i = 1, \ldots, n)$ can take a measurement according to the following time-varying stochastic linear regression model

$$y_k^i = (\varphi_k^i)^T \theta_k + v_k^i, \qquad k \geq 0, \tag{1}$$

where y_k^i, v_k^i are scalar observation and noise, respectively, $\varphi_k^i \in \mathbb{R}^m$ is the stochastic regressor of node i at time k, $\theta_k \in \mathbb{R}^m$ is an unknown parameter vector, and $(\cdot)^T$ denotes the transpose operator. In this paper, we assume that for any $k \geq 0$, θ_k is s-sparse (i.e., θ_k has at most s non-zero elements). We can denote the variation of θ_k as follows

$$\gamma \omega_k \stackrel{\triangle}{=} \theta_k - \theta_{k-1}, \qquad k \geq 1, \tag{2}$$

where γ is a scalar value, and ω_k is an yet undefined vector. When $\gamma = 0$, θ_k is time-invariant. Otherwise, it is time-varying. In addition, we assume that $\varphi_k^i(i = 1, \ldots, n)$ are $3s$-sparse. In some practical situations, the regressors and unknown parameter are sparse, for example, high-dimensional data classification problems in machine learning, field monitoring, channel estimation, cooperative spectrum sensing for cognitive radio networks and so on (Paredes, Arce, & Wang, 2007).

In sensor networks, if sensor i has access only to the information from its neighbors $\{\ell \in \mathcal{N}_i\}$, we can adopt the consensus NLMS algorithm to estimate θ_k (Sayed, 2014; Xie & Guo, 2018b). Since the regressors and the unknown parameter vector are high-dimensional, the consensus NLMS algorithm needs to exchange high-dimensional data over the sensor networks. Moreover, since φ_k^i is sparse, the cooperative information condition introduced in Xie and Guo (2018b) for the stability of the consensus NLMS algorithm is hard to be satisfied. Thus, the usual adaptive algorithms are not able to estimate or track unknown high-dimensional sparse signals accurately, and this motivates the introduction of CS method in the paper.

2.2. Compressed consensus adaptive filters

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 the CS

method to the estimation of $\boldsymbol{\theta}_k$, 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 $\boldsymbol{\varphi}_k^i$, then with the help of measurement matrix $M \in \mathbb{R}^{d \times m}(s \le d \ll m)$, we can obtain the compressed $d \times 1$-dimensional regressor $\boldsymbol{\psi}_k^i = M\boldsymbol{\varphi}_k^i$. Instead of estimating $m \times 1$-dimensional parameter $\boldsymbol{\theta}_k$, the proposed method estimates the compressed $d \times 1$-dimensional parameter $\boldsymbol{\zeta}_k = M\boldsymbol{\theta}_k$. Note that $\boldsymbol{\zeta}_k = \boldsymbol{\zeta}_{k-1} + \gamma \bar{\boldsymbol{\omega}}_k (k \ge 1)$, where $\bar{\boldsymbol{\omega}}_k = M\boldsymbol{\omega}_k$. Then we have

$$
\begin{aligned}
y_k^i &= (\boldsymbol{\varphi}_k^i)^T \boldsymbol{\theta}_k + v_k^i \\
&= (\boldsymbol{\psi}_k^i)^T \boldsymbol{\zeta}_k + (\boldsymbol{\varphi}_k^i)^T \boldsymbol{\theta}_k - (\boldsymbol{\psi}_k^i)^T \boldsymbol{\zeta}_k + v_k^i \\
&= (\boldsymbol{\psi}_k^i)^T \boldsymbol{\zeta}_k + (\boldsymbol{\varphi}_k^i)^T [I_m - M^T M] \boldsymbol{\theta}_k + v_k^i \\
&= (\boldsymbol{\psi}_k^i)^T \boldsymbol{\zeta}_k + \bar{v}_k^i, \quad k \ge 0,
\end{aligned}
\tag{3}
$$

where the new "noise" term $\bar{v}_k^i \triangleq (\boldsymbol{\varphi}_k^i)^T [I_m - M^T M] \boldsymbol{\theta}_k + v_k^i$.

In the following, we present the compressed consensus adaptive filter based on the CS method, which consists of two algorithms. The first one is used to obtain a low-dimensional estimate for the compressed unknown signal by using the compressed regressor data, and the second one is used to reconstruct a high-dimensional sparse estimate for the original unknown signal.

Algorithm 1 Consensus NLMS algorithm based on compressed regressor data

---

For any given sensor $i \in \{1 = 1, \ldots, n\}$, begin with an initial estimate $\widehat{\boldsymbol{\zeta}}_0^i \in \mathbb{R}^d$. The algorithm is recursively defined for iteration $k \ge 0$ as follows:

$$
\begin{aligned}
\widehat{\boldsymbol{\zeta}}_{k+1}^i = \widehat{\boldsymbol{\zeta}}_k^i + \mu \Bigg\{ & \frac{\boldsymbol{\psi}_k^i}{1 + \| \boldsymbol{\psi}_k^i \|^2} [y_k^i - (\boldsymbol{\psi}_k^i)^T \widehat{\boldsymbol{\zeta}}_k^i] \\
& - \nu \sum_{\ell \in \mathcal{N}_i} a_{\ell i} (\widehat{\boldsymbol{\zeta}}_k^i - \widehat{\boldsymbol{\zeta}}_k^\ell) \Bigg\},
\end{aligned}
\tag{4}
$$

where μ and ν are step sizes suitably chosen in $(0, 1)$.

In Algorithm 1, $\widehat{\boldsymbol{\zeta}}_k^i$ is an estimate for the compressed unknown signal at sensor i and time instant k. At the kth iteration, each sensor can employ the following signal reconstruction algorithm introduced in the Appendix to obtain a decompressed estimate $\widehat{\boldsymbol{\theta}}_k^i$ as an estimate of $\boldsymbol{\theta}_k$. Note that for a vector $x \in \mathbb{R}^m$, the norm $\|x\|_{\ell_1}$ is defined as $\sum_{i=1}^m |x_i|$, where $x_i(i = 1, \ldots, m)$ is the ith element of x, and for any matrix $X \in \mathbb{R}^{m \times n}$, its Euclidean norm is defined as $\|X\|_{\ell_2} = \{\lambda_{max}(XX^T)\}^{\frac{1}{2}}$, where $\lambda_{max}(\cdot)$ denotes the largest eigenvalue of matrix $(\cdot)$. For simplicity, we will write $\| \cdot \|_{\ell_2}$ as $\| \cdot \|$ in the sequel.

Algorithm 2 Reconstruction of the estimates

At every time instant $k \ge 0$, each sensor $i(i = 1, \ldots, n)$ solves the following reconstruction problem to obtain a decompressed estimate $\widehat{\boldsymbol{\theta}}_k^i$:

$$
\min_{\boldsymbol{\theta} \in \mathbb{R}^m} \|\boldsymbol{\theta}\|_{\ell_1}, \quad s.t. \quad \|M\boldsymbol{\theta} - \widehat{\boldsymbol{\zeta}}_k^i\| \le C.
$$

Here C is the upper bound of the estimation error $\|\widetilde{\boldsymbol{\zeta}}_k^i\|$. By Theorem 3, we could choose $C = \eta c_2^{1-\xi}(\gamma, \delta_{4s})$, where $c_2(\gamma, \delta_{4s})$ is defined in the same theorem, δ_{4s} is the RIP constant of the measurement matrix M which is defined in Assumption 3, and $\eta = \max\{1, 2c_2^{\frac{\xi}{2}}(\gamma, \delta_{4s})\}$, $\xi \in (0, 1)$. There are many algorithms used to solve the above optimization problem in the literature,

e.g., the orthogonal matching pursuit (OMP) (Tropp & Gilbert, 2007; Xu, 2010), the basis pursuit (BP), the basis pursuit denoising (BPDN) (Berg & Friedlander, 2008), the least absolute shrinkage and selection (LASSO), ℓ_1-norm penalty least squares (L1LS) (Figueiredo, Nowak, & Wright, 2007; Kim, Koh, Lustig, Boyd, & Gorinevsky, 2007; Wright, Nowak, & Figueiredo, 2009; Yang & Zhang, 2011) and so on. In practice, we can apply the reconstruction algorithm every $K(K \gg 1)$ iteration to obtain $\widehat{\boldsymbol{\theta}}_K^i, \widehat{\boldsymbol{\theta}}_{2K}^i, \ldots$.

To write the compressed consensus NLMS algorithm into a compact form, we introduce the following notations:

$$
\boldsymbol{Y}_k \triangleq \text{col}\{y_k^1, \ldots, y_k^n\}, \quad \boldsymbol{V}_k \triangleq \text{col}\{v_k^1, \ldots, v_k^n\},
$$

$$
\boldsymbol{\Phi}_k \triangleq \text{diag}\{\boldsymbol{\varphi}_k^1, \ldots, \boldsymbol{\varphi}_k^n\}, \quad \boldsymbol{\Psi}_k \triangleq \text{diag}\{\boldsymbol{\psi}_k^1, \ldots, \boldsymbol{\psi}_k^n\},
$$

$$
\boldsymbol{\Omega}_k \triangleq \text{col}\{\underbrace{\boldsymbol{\omega}_k, \ldots, \boldsymbol{\omega}_k}_{n}\}, \quad \bar{\boldsymbol{\Omega}}_k \triangleq \text{col}\{\underbrace{\bar{\boldsymbol{\omega}}_k, \ldots, \bar{\boldsymbol{\omega}}_k}_{n}\},
$$

$$
\boldsymbol{\Theta}_k \triangleq \text{col}\{\underbrace{\boldsymbol{\theta}_k, \ldots, \boldsymbol{\theta}_k}_{n}\}, \quad \boldsymbol{Z}_k \triangleq \text{col}\{\underbrace{\boldsymbol{\zeta}_k, \ldots, \boldsymbol{\zeta}_k}_{n}\},
$$

$$
\widehat{\boldsymbol{Z}}_k \triangleq \text{col}\{\widehat{\boldsymbol{\zeta}}_k^1, \ldots, \widehat{\boldsymbol{\zeta}}_k^n\},
$$

$$
\widetilde{\boldsymbol{Z}}_k \triangleq \text{col}\{\widetilde{\boldsymbol{\zeta}}_k^1, \ldots, \widetilde{\boldsymbol{\zeta}}_k^n\}, \quad \text{where} \quad \widetilde{\boldsymbol{\zeta}}_k^i = \widehat{\boldsymbol{\zeta}}_k^i - \boldsymbol{\zeta}_k,
$$

$$
\bar{\boldsymbol{L}}_k \triangleq \text{diag}\left\{ \frac{\boldsymbol{\psi}_k^1}{1 + \| \boldsymbol{\psi}_k^1 \|^2}, \ldots, \frac{\boldsymbol{\psi}_k^n}{1 + \| \boldsymbol{\psi}_k^n \|^2} \right\},
$$

$$
\bar{\boldsymbol{F}}_k \triangleq \bar{\boldsymbol{L}}_k \boldsymbol{\Psi}_k^T, \quad \mathcal{M} = \text{diag}\{\underbrace{M, \ldots, M}_{n}\},
$$

$$
\mathscr{L} \triangleq \mathcal{L} \otimes I_d, \quad \bar{\boldsymbol{G}}_k \triangleq \bar{\boldsymbol{F}}_k + \nu \mathscr{L},
$$

where $\text{col}\{\cdots\}$ denotes a vector by stacking the specified vectors, $\text{diag}\{\cdots\}$ is used in a non-standard manner which means that $m \times 1$ column vectors are combined "in a diagonal manner" resulting in a $mn \times n$ matrix, and $\otimes$ is the Kronecker product. By (3), we have

$$
\boldsymbol{Y}_k = \boldsymbol{\Psi}_k^T \boldsymbol{Z}_k + \bar{\boldsymbol{V}}_k,
\tag{5}
$$

and

$$
\boldsymbol{Z}_{k+1} = \boldsymbol{Z}_k + \gamma \bar{\boldsymbol{\Omega}}_{k+1}.
\tag{6}
$$

Moreover, Algorithm 1 can be rewritten in the following compact form

$$
\widehat{\boldsymbol{Z}}_{k+1} = \widehat{\boldsymbol{Z}}_k + \mu \bar{\boldsymbol{L}}_k(\boldsymbol{Y}_k - \boldsymbol{\Psi}_k^T \widehat{\boldsymbol{Z}}_k) - \mu \nu \mathscr{L} \widehat{\boldsymbol{Z}}_k,
\tag{7}
$$

Denoting the estimation error as $\widetilde{\boldsymbol{Z}}_k = \widehat{\boldsymbol{Z}}_k - \boldsymbol{Z}_k$, substituting (5) into (7), and noticing (6), $\mathscr{L}\boldsymbol{Z}_k = 0$, and the notation of $\bar{\boldsymbol{G}}_k$, we can get the compressed estimation error equation for Algorithm 1 as follows

$$
\begin{aligned}
\widetilde{\boldsymbol{Z}}_{k+1} &= \widetilde{\boldsymbol{Z}}_k - \mu \bar{\boldsymbol{L}}_k \boldsymbol{\Psi}_k^T \widetilde{\boldsymbol{Z}}_k - \mu \nu \mathscr{L} \widetilde{\boldsymbol{Z}}_k + \mu \bar{\boldsymbol{L}}_k \bar{\boldsymbol{V}}_k - \gamma \bar{\boldsymbol{\Omega}}_{k+1} \\
&= (I_{dn} - \mu \bar{\boldsymbol{G}}_k) \widetilde{\boldsymbol{Z}}_k + \mu \bar{\boldsymbol{L}}_k \bar{\boldsymbol{V}}_k - \gamma \bar{\boldsymbol{\Omega}}_{k+1},
\end{aligned}
\tag{8}
$$

which will be analyzed in Section 3.

2.3. Notations and definitions

Let $\lambda_{min}(\cdot)$ denote the smallest eigenvalue of matrix $(\cdot)$. For a vector $x \in \mathbb{R}^m$, the norm $\|x\|_{\ell_0}$ is defined as the number of the non-zero coefficients, and for any random matrix Y, its L_p-norm is defined as $\|Y\|_{L_p} = \{\mathbb{E}[\|Y\|^p]\}^{\frac{1}{p}}$, where $\mathbb{E}[\cdot]$ denotes the expectation operator. Also, we use $\mathcal{F}_k = \sigma\{\boldsymbol{\varphi}_i^j, \boldsymbol{\omega}_i, v_{i-1}^j, j = 1, \ldots, n, i \le k\}$ to denote the σ-algebra generated by $\{\boldsymbol{\varphi}_i^j, \boldsymbol{\omega}_i, v_{i-1}^j, j = 1, \ldots, n, i \le k\}$, where the definition of σ-algebra together with

that of conditional mathematical expectation operator $\mathbb{E}[\cdot|\mathcal{F}_k]$ to be used later can be found in Chow and Teicher (1978).

To proceed with further discussions, we need the following definitions in Guo (1994).

Definition 1. For a random matrix sequence $\{A_k, k \geq 0\}$ defined on the basic probability space $(\Omega, \mathcal{F}, P)$, if $\sup_{k \geq 0} \mathbb{E}[\|A_k\|^p] < \infty$ holds for some $p > 0$, then $\{A_k\}$ is called L_p-stable.

Definition 2. For a sequence of $\ell \times \ell$-dimensional random matrices $A = \{A_k, k \geq 0\}$, if it belongs to the following set with $p \geq 0$,

$$S_p(\lambda) = \left\{ A : \left\| \prod_{j=i+1}^{k} (I - A_j) \right\|_{L_p} \leq N\lambda^{k-i}, \right.$$

$$\left. \forall k \geq i+1, \forall i \geq 0, \text{ for some } N > 0 \right\}, \tag{9}$$

then $\{I - A_k, k \geq 0\}$ is called L_p-exponentially stable with parameter $\lambda \in [0, 1)$.

For convenience of discussions, we introduce the following subclass of $S_1(\lambda)$ for a scalar sequence $a = \{a_k, k \geq 0\}$ with $\lambda \in [0, 1)$,

$$S^0(\lambda) = \left\{ a : a_k \in [0, 1], \mathbb{E}\left[\prod_{j=i+1}^{k} (1 - a_j) \right] \leq N\lambda^{k-i}, \right.$$

$$\left. \forall k \geq i+1, \forall i \geq 0, \text{ for some } N > 0 \right\}. \tag{10}$$

Definition 3. A random sequence $x = \{x_k, k \geq 0\}$ is called an element of the weakly dependent 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 any $k \geq 0$ and $h \geq 1$,

$$\left\| \sum_{i=k+1}^{k+h} x_i \right\|_{L_p} \leq C_p^x h^{\frac{1}{2}}. \tag{11}$$

Remark 1. 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 $\mathcal{M}_p$ (Guo & Ljung, 1995).

Definition 4. Let $\{A_k, k \geq 0\}$ be a matrix sequence and $\{b_k, k \geq 0\}$ be a positive scalar sequence. Then by $A_k = O(b_k)$ we mean that there exists a positive constant $C < \infty$ such that

$$\|A_k\| \leq Cb_k, \quad \forall k \geq 0. \tag{12}$$

3. The main results

To proceed with the theoretical analyses, we need the following two assumptions.

Assumption 1. The graph $\mathcal{G}$ is connected

Remark 2. The compressed distributed adaptive filtering algorithms may not be able to achieve consensus due to the existence of isolated nodes in the network and the lack of necessary information at individual sensors, so it is natural to require the connectivity of the sensor network

Assumption 2 (*Compressed Cooperative Information Condition*). For the adapted sequences $\{\psi_k^i, \mathcal{F}_k, k \geq 0\}(i = 1, \ldots, n)$, there

exists an integer $h > 0$ such that $\{\lambda_k, k \geq 0\} \in S^0(\lambda)$ for some $N \geq 1, \lambda \in (0, 1)$, where λ_k is defined by

$$\lambda_k \triangleq \lambda_{min} \left\{ \mathbb{E}\left[\frac{1}{n(h+1)} \sum_{i=1}^{n} \sum_{j=kh+1}^{(k+1)h} \frac{\psi_j^i (\psi_j^i)^T}{1 + \|\psi_j^i\|^2} \Big| \mathcal{F}_{kh} \right] \right\},$$

and where $\mathbb{E}[\cdot|\mathcal{F}_{kh}]$ is the conditional mathematical expectation operator, and $\mathcal{F}_{kh} = \sigma\{\varphi_j^i, \omega_j, v_{j-1}^i, i = 1, \ldots, n, j \leq kh\}$.

Remark 3. The main difference between Assumption 2 and the cooperative information condition in Xie and Guo (2018b) 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 d, which is much less than m, so Assumption 2 is much easier to be satisfied than the traditional cooperative information condition, see the simulation results in Section 4 for example. Obviously, when the original regressors are high-dimensional and sparse, the cooperative information condition may not be satisfied. In other words, even in the case where the non-compressed consensus NLMS algorithm cannot fulfill the estimation tasks, the compressed consensus NLMS algorithm may still get the compressed estimation results stably.

Assumption 3. The measurement matrix $M \in \mathbb{R}^{d \times m}$ satisfies $4s$th RIP with δ_{4s} be its RIP constant, where RIP is defined in Appendix.

Assumption 3 indicates that s should not be larger than $m/4$. We are now in a position to give our first main theorem on the compressed estimation error, which will play a basic role in the subsequent results and analyses. This theorem contains two parts: the first is the exponential stability of the homogeneous part of the compressed error equation, and the second is the upper bound for the compressed estimation error.

Theorem 1. *Consider the model* (5) *and the compressed estimation error Eq.* (8). *Let Assumptions* 1–3 *be satisfied. For any* $p \geq 1$, *and any* $\mu \in (0, 1)$ *and* $\nu \in (0, 1)$ *satisfying* $\mu(1 + 2\nu) \leq 1$, $\{I_{dn} - \mu\bar{G}_k, k \geq 1\}$ *is L_p-exponentially stable. Furthermore, if for some* $p \geq 1$ *and* $\beta > 1$, $\|\widehat{Z}_0\|_{L_p} < \infty$, *and*

$$\sigma_p \triangleq \sup_k \|\eta_k \log^\beta(e + \eta_k)\|_{L_p} < \infty, \tag{13}$$

hold, where $\eta_k = \frac{3\delta_{4s}}{1-\delta_{4s}}\|\Theta_k\| + \|V_k\| + \gamma(1 + \delta_{4s})\|\Omega_{k+1}\|$, *then* $\{\widehat{Z}_k - Z_k, k \geq 1\}$ *is L_p-stable and*

$$\limsup_{k \to \infty} \|\widehat{Z}_k - Z_k\|_{L_p} \leq c[\sigma_p \log(e + \sigma_p^{-1})], \tag{14}$$

where c is a positive constant depending on β, μ, ν, p, *the matrix* $\mathcal{L}$, *and parameters in Assumption* 2, *i.e.,* N, λ, h.

Proof. By *Theorem 4.1* in Xie and Guo (2018b), we know that for any $p \geq 1$, and any $\mu \in (0, 1)$ and $\nu \in (0, 1)$ satisfying $\mu(1 + 2\nu) \leq 1$,

$$\left\| \prod_{j=i+1}^{k} (I_{dn} - \mu\bar{G}_k) \right\|_{L_p} \leq N_p' \pi_p^{k-i}, \forall k \geq i+1, \forall i \geq 0,$$

holds, where

$$N_p' = \begin{cases} \sqrt{2}N^{\frac{1}{8(1+h)^2}}, & 1 \leq p \leq 2, \\ \sqrt{2}N^{\frac{1}{4(1+h)^2p}}, & p > 2, \end{cases} \tag{15}$$

$$\pi_p = \begin{cases} \lambda^{\frac{\ell_2\nu\mu}{8h(1+h)^3(2+\nu\ell_2)}}, & 1 \leq p \leq 2, \\ \lambda^{\frac{\ell_2\nu\mu}{4h(1+h)^3p(2+\nu\ell_2)}}, & p > 2, \end{cases} \tag{16}$$

and N, h, λ are defined in Assumption 2, ℓ_2 is the second smallest eigenvalue of Laplacian matrix $\mathcal{L}$, μ and ν are step size parameters. Then by Definition 2, it is obvious that $\{I_{dn} - \mu \bar{\mathbf{G}}_k, k \geq 1\}$ is L_p-exponentially stable ($p \geq 1$), and the stability parameters depending on $N, h, \lambda, \mu, \nu, p$, and the matrix $\mathcal{L}$.

Since $\boldsymbol{\varphi}_k^i$ is $3s$-sparse and $\boldsymbol{\theta}_k$ is s-sparse, we denote the positions of their non-zero entries as $i_1, \ldots, i_{3s}$ and $j_1, \ldots, j_s$, respectively. Then we retain all the above $4s$ entries of $\boldsymbol{\varphi}_k^i$ and $\boldsymbol{\theta}_k$, the corresponding $4s$ columns of the matrix M, and we remove all other entries. (When $\boldsymbol{\varphi}_k^i$ and $\boldsymbol{\theta}_k$ have some non-zero entries with the same positions, we should keep less than $4s$ entries of the vectors and matrices. Since the analysis is similar, we only consider the $4s$ situation here.) Then denote them as $\boldsymbol{\varphi}_{k,4s}^i, \boldsymbol{\theta}_{k,4s}$ and M_{4s}. By Assumption 3, M satisfies $4s$th RIP, then we know that all the eigenvalues of $M_{4s}^T M_{4s}$ are in set $[1 - \delta_{4s}, 1 + \delta_{4s}]$ (see Appendix) and

$$
\begin{aligned}
& \|(\boldsymbol{\varphi}_k^i)^T [I_m - M^T M] \boldsymbol{\theta}_k\| \\
=& \|(\boldsymbol{\varphi}_{k,4s}^i)^T [I_{4s} - M_{4s}^T M_{4s}] \boldsymbol{\theta}_{k,4s}\| \\
\leq& \|(\boldsymbol{\varphi}_{k,4s}^i)^T [(1 + \delta_{4s}) I_{4s} - M_{4s}^T M_{4s}] \boldsymbol{\theta}_{k,4s}\| \\
& + \delta_{4s} \|(\boldsymbol{\varphi}_{k,4s}^i)^T \boldsymbol{\theta}_{k,4s}\| \\
\leq& \|(\boldsymbol{\varphi}_{k,4s}^i)^T [(1 + \delta_{4s}) I_{4s} - M_{4s}^T M_{4s}]^{\frac{1}{2}}\| \\
& \cdot \|[(1 + \delta_{4s}) I_{4s} - M_{4s}^T M_{4s}]^{\frac{1}{2}} \boldsymbol{\theta}_{k,4s}\| + \delta_{4s} \|(\boldsymbol{\varphi}_{k,4s}^i)^T \boldsymbol{\theta}_{k,4s}\| \\
\leq& \sqrt{(\boldsymbol{\varphi}_{k,4s}^i)^T \cdot 2\delta_{4s} \cdot \boldsymbol{\varphi}_{k,4s}^i} \cdot \sqrt{\boldsymbol{\theta}_{k,4s}^T \cdot 2\delta_{4s} \cdot \boldsymbol{\theta}_{k,4s}} \\
& + \delta_{4s} \|(\boldsymbol{\varphi}_{k,4s}^i)^T \boldsymbol{\theta}_{k,4s}\| \\
\leq& 2\delta_{4s} \|\boldsymbol{\varphi}_{k,4s}^i\| \cdot \|\boldsymbol{\theta}_{k,4s}\| + \delta_{4s} \|\boldsymbol{\varphi}_{k,4s}^i\| \cdot \|\boldsymbol{\theta}_{k,4s}\| \\
=& 3\delta_{4s} \|\boldsymbol{\varphi}_{k,4s}^i\| \cdot \|\boldsymbol{\theta}_{k,4s}\| \\
\leq& 3\delta_{4s} \|\boldsymbol{\varphi}_k^i\| \cdot \|\boldsymbol{\theta}_k\|.
\end{aligned}
\tag{17}
$$

By the above equation and (27), we know that

$$
\begin{aligned}
& \left\| \frac{\boldsymbol{\psi}_k^i}{1 + \|\boldsymbol{\psi}_k^i\|^2} \cdot (\boldsymbol{\varphi}_k^i)^T [I_m - M^T M] \boldsymbol{\theta}_k \right\| \\
\leq& 3\delta_{4s} \frac{\|\boldsymbol{\psi}_k^i\|}{1 + \|\boldsymbol{\psi}_k^i\|^2} \cdot \|\boldsymbol{\varphi}_k^i\| \cdot \|\boldsymbol{\theta}_k\| \\
=& \frac{3\delta_{4s} \|\boldsymbol{\theta}_k\|}{1 - \delta_{4s}} \cdot \frac{\|M\boldsymbol{\varphi}_k^i\| \cdot \|M\boldsymbol{\varphi}_k^i\|}{1 + \|M\boldsymbol{\varphi}_k^i\| \cdot \|M\boldsymbol{\varphi}_k^i\|} \\
\leq& \frac{3\delta_{4s} \|\boldsymbol{\theta}_k\|}{1 - \delta_{4s}}.
\end{aligned}
\tag{18}
$$

Since $\bar{\boldsymbol{V}}_k = \text{col}\{\bar{v}_k^1, \ldots, \bar{v}_k^n\} = \text{col}\{(\boldsymbol{\varphi}_k^1)^T [I_m - M^T M] \boldsymbol{\theta}_k, \ldots, (\boldsymbol{\varphi}_k^n)^T [I_m - M^T M] \boldsymbol{\theta}_k\} + \boldsymbol{V}_k$, then we have

$$
\begin{aligned}
\|\bar{\boldsymbol{L}}_k \bar{\boldsymbol{V}}_k\| \leq& \left\| \text{col}\left\{ \frac{\boldsymbol{\psi}_k^1}{1 + \|\boldsymbol{\psi}_k^1\|^2} \cdot (\boldsymbol{\varphi}_k^1)^T [I_m - M^T M] \boldsymbol{\theta}_k, \right. \right. \\
& \left. \left. \ldots, \frac{\boldsymbol{\psi}_k^n}{1 + \|\boldsymbol{\psi}_k^n\|^2} \cdot (\boldsymbol{\varphi}_k^n)^T [I_m - M^T M] \boldsymbol{\theta}_k \right\} \right\| \\
& + \|\bar{\boldsymbol{L}}_k\| \cdot \|\boldsymbol{V}_k\| \\
\leq& \sqrt{\sum_{i=1}^n \left\| \frac{\boldsymbol{\psi}_k^i}{1 + \|\boldsymbol{\psi}_k^i\|^2} \cdot (\boldsymbol{\varphi}_k^i)^T [I_m - M^T M] \boldsymbol{\theta}_k \right\|^2} \\
& + \|\boldsymbol{V}_k\| \\
\leq& \frac{3\delta_{4s}}{1 - \delta_{4s}} \|\boldsymbol{\Theta}_k\| + \|\boldsymbol{V}_k\|.
\end{aligned}
\tag{19}
$$

Moreover, since $\gamma \|\bar{\boldsymbol{\omega}}_k\| = \gamma \|M\boldsymbol{\omega}_k\| \leq \gamma(1 + \delta_{4s}) \|\boldsymbol{\omega}_k\|$, we have $\gamma \|\bar{\boldsymbol{\Omega}}_{k+1}\| \leq \gamma(1 + \delta_{4s}) \|\boldsymbol{\Omega}_{k+1}\|$. By the proof of *Theorem 4.2* in Guo (1994), we know that $\{\hat{\boldsymbol{Z}}_k - \boldsymbol{Z}_k, k \geq 1\}$ is L_p-stable and (14) holds.

Remark 4. (i) The bound on $\boldsymbol{\Theta}_k$, i.e., (13), excludes the random walk model where both the trajectory and the variance are unbounded. However, this assumption still includes a wide range of interesting situations, e.g., when the parameters are generated from a stable dynamic model. Moreover, by Theorem 1, one can see that the accuracy of the estimation error depends on the property of the compression in the sense that the smaller the RIP parameter δ, the smaller the parameter estimation error. To see when one may have a smaller value of δ, we know by Remark 8 in the Appendix that if the measurement matrix $M \in \mathbb{R}^{d \times m}$ is a Gaussian or Bernoulli random matrix, the original parameter dimension m, the compressed parameter dimension d, the sparse value s, and the RIP parameter δ should satisfy $d \geq 120s \log(m/s)/\delta^3$. From this, one can see that when the order of d is larger than $\log(m/s)$, e.g., $d \geq 120s \log^\alpha(m/s)(\alpha > 1)$, we may have $\delta \to 0$ 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 1.

(ii) Moreover, Assumption 1 is easy to be satisfied if the network topology can be designed. In addition, Assumption 2 is not difficult to check in some cases. For example, when the regressor processes are deterministic, or independent, or ϕ-mixing, or generated by a linear stable filter driving by white noises, see some examples in Guo (1994) and Xie and Guo (2018a, 2018b). As for Assumption 3, there are many existing methods to make M satisfying the RIP condition in the area of CS, see the discussions in Appendix.

(iii) We further remark that since the observation noise $\boldsymbol{V}_k$ in Theorem 1 is not assumed to be zero mean, it is not expected that the upper bound of the estimation error will tend to zero as μ and the RIP parameter δ_{4s} tend to zero when $\gamma = 0$. The meaningfulness of Theorem 1 lies in the fact that when the magnitudes of both the observation noise $\boldsymbol{V}_k$ and the parameter variation $\boldsymbol{\Omega}_k$ are small, the upper bound of the estimation error will be small provided that the RIP constant δ_{4s} is also small. Moreover, if we further assume that the observation noise $\boldsymbol{V}_k$ satisfies more statistic properties including zero mean, then by Theorem 2, it is easy to see that when $\gamma = 0$, the upper bound of the estimation error will tend to zero as μ and the RIP parameter δ_{4s} tend to zero.

We may improve the upper bound of the compressed estimation error $\|\hat{\boldsymbol{Z}}_k - \boldsymbol{Z}_k\|_{L_p}$ in Theorem 1 by introducing additional conditions. The following assumption simply implies that both the noises and parameter variations are weakly dependent.

Assumption 4. For some $p \geq 1$, the initial estimation error is L_{2p}-bounded, i.e., $\|\tilde{\boldsymbol{Z}}_0\|_{L_{2p}} = M_{2p}^{\tilde{\boldsymbol{Z}}_0} < \infty$, and the unknown signal is L_{2p}-bounded, i.e., $\sup_k \|\boldsymbol{\Theta}_k\|_{L_{2p}} = M_{2p}^{\boldsymbol{\Theta}} < \infty$. Furthermore, $\{\bar{\boldsymbol{L}}_k \boldsymbol{V}_k\} \in \mathcal{M}_{2p}$ with the parameter $C_{2p}^{\boldsymbol{LV}}$, and $\{\boldsymbol{\Omega}_k\} \in \mathcal{M}_{2p}$ with the parameter $C_{2p}^{\boldsymbol{\Omega}}$

Under this additional assumption, we are able to get a refined upper bound for the compressed estimation error in the following theorem.

Theorem 2. *Consider the model* (5) *and the compressed estimation error Eq.* (8). *Assume that* Assumptions 1–4 *are all satisfied, then for any* $\mu \in (0, 1)$ *and* $\nu \in (0, 1)$ *satisfying* $\mu(1 + 2\nu) \leq 1$, *we have for all* $k \geq 0$,

$$
\begin{aligned}
\|\hat{\boldsymbol{Z}}_{k+1} - \boldsymbol{Z}_{k+1}\|_{L_p} \leq& c_{1,2p} \sqrt{\mu} + c_{2,2p} \frac{\gamma}{\sqrt{\mu}} + c_{3,2p} \frac{\delta_{4s}}{1 - \delta_{4s}} \\
& + c_{4,2p} (1 - \mu \alpha_{2p})^{k+1},
\end{aligned}
\tag{20}
$$

where $c_{1,2p}, c_{2,2p}, c_{3,2p}, c_{4,2p}, \alpha_{2p} \in (0, 1)$ *are positive constants defined in the proof, which do not depend on* μ, γ *and* δ_{4s}.

Proof. Define $\Gamma(\cdot,\cdot)$ as follows

$$\begin{cases} \Gamma(k+1,j) = (I_{dn} - \mu\bar{G}_k)\Gamma(k,j), \\ \Gamma(j,j) = I_{dn}, \forall k \geq j \geq 0. \end{cases}$$

Then by *Lemma 5.6* in Xie and Guo (2018b), we have

$$\|\Gamma(k+1,j+1)\|_{L_{2p}} \leq N'_{2p}(1-\mu\alpha_{2p})^{k-j}, \forall k \geq j \geq 0, \quad (21)$$

where $N'_{2p} = \sqrt{2}N^{\frac{1}{8(1+h)^2 p}}$, $\alpha_{2p} = 1 - \lambda^{\frac{\ell_2 v}{8h(1+h)^3(2+v\ell_2)p}}$, and N, h, λ are defined in Assumption 2, ℓ_2 is the second smallest eigenvalue of Laplacian matrix $\mathcal{L}$.

By $\{\Omega_k\} \in \mathcal{M}_{2p}$, we know that $\{\bar{\Omega}_k\} \in \mathcal{M}_{2p}$ holds. By the definitions of $\Gamma(\cdot,\cdot)$ and $\bar{V}_j$, and (21), we have by (19) and *Lemma A.2* in Guo and Ljung (1995) that

$$\|\widetilde{Z}_{k+1}\|_{L_p}$$
$$= \left\| \Gamma(k+1,0)\widetilde{Z}_0 \right.$$
$$\left. + \sum_{j=0}^{k} \Gamma(k+1,j+1)[\mu\bar{L}_j\bar{V}_j - \gamma\bar{\Omega}_{j+1}] \right\|_{L_p}$$
$$\leq N'_{2p}(1-\mu\alpha_{2p})^{k+1}\|\widetilde{Z}_0\|_{L_{2p}}$$
$$+ \mu\left\| \sum_{j=0}^{k} \Gamma(k+1,j+1)\bar{L}_j\Phi_j(I_{mn} - \mathcal{M}^T\mathcal{M})\Theta_j \right\|_{L_p}$$
$$+ \mu\left\| \sum_{j=0}^{k} \Gamma(k+1,j+1)\bar{L}_j V_j \right\|_{L_p}$$
$$+ \gamma\left\| \sum_{j=0}^{k} \Gamma(k+1,j+1)\bar{\Omega}_{j+1} \right\|_{L_p}$$
$$\leq N'_{2p}(1-\mu\alpha_{2p})^{k+1}\|\widetilde{Z}_0\|_{L_{2p}} + c_{1,2p}\sqrt{\mu} + c_{2,2p}\frac{\gamma}{\sqrt{\mu}}$$
$$+ \frac{3\delta_{4s}}{1-\delta_{4s}}\mu N'_{2p}\sum_{j=0}^{k}(1-\mu\alpha_{2p})^{k-j}\|\Theta_j\|_{L_{2p}}$$
$$\leq c_{1,2p}\sqrt{\mu} + c_{2,2p}\frac{\gamma}{\sqrt{\mu}} + \frac{c_{3,2p}\delta_{4s}}{1-\delta_{4s}} + c_{4,2p}(1-\mu\alpha_{2p})^{k+1},$$

where $c_{1,2p} = N'_{2p}C^{LV}_{2p}\alpha_{2p}^{-2}e^{\alpha_{2p}}$, $c_{2,2p} = N'_{2p}C^{\Omega}_{2p}\alpha_{2p}^{-2}e^{\alpha_{2p}}$, $c_{3,2p} = 3N'_{2p}M^{\Theta}_{2p}\alpha_{2p}^{-1}$, $c_{4,2p} = N'_{2p}M^{\widetilde{Z}_0}_{2p}$, and the constants C^{LV}_{2p}, C^{Ω}_{2p}, M^{Θ}_{2p}, $M^{\widetilde{Z}_0}_{2p}$ are defined in Assumption 4. This completes the proof.

Remark 5. Theorem 2 shows that as δ_{4s} gets smaller, the compressed estimation error will also be smaller. Here we assume that $\gamma > 0$. By minimizing $c_{1,2p}\sqrt{\mu} + c_{2,2p}\gamma/\sqrt{\mu}$, we can get the optimal step-size $\mu^* = \gamma c_{2,2p}(c_{1,2p})^{-1}$ with the corresponding minimum value $2\sqrt{\gamma c_{1,2p}c_{2,2p}}$. Then there exists an integer $K_{0,2p} \geq 0$, such that for any $k \geq K_{0,2p}$, we have $c_{4,2p}(1-\mu\alpha_{2p})^{k+1} \leq \sqrt{\gamma c_{1,2p}c_{2,2p}}$. Hence we can take

$$K_{0,2p} = \max\left\{ \left\lceil \frac{\log(\gamma c_{1,2p}c_{2,2p}) - 2\log(c_{4,2p})}{2\log(1-\mu\alpha_{2p})} - 1 \right\rceil, 0 \right\},$$

and consequently we have,

$$\|\widehat{Z}_k - Z_k\|_{L_p} \qquad \leq 3\sqrt{\gamma c_{1,2p}c_{2,2p}} + c_{3,2p}\delta_{4s}/(1-\delta_{4s})$$
$$\stackrel{\triangle}{=} c_{2p}(\gamma,\delta_{4s}), \quad k \geq K_{0,2p}, \quad (22)$$

where $\lceil\cdot\rceil$ denotes rounding up operator, $c_{1,2p}, c_{2,2p}, c_{3,2p}$ are defined in Theorem 2, and δ_{4s} is the RIP constant of the matrix M. Moreover, for $p = 1$, we have

$$K_{0,2} = \max\{\lceil \frac{\log(\gamma c_{1,2}c_{2,2}) - 2\log(c_{4,2})}{2\log(1-\mu\alpha_2)} - 1\rceil, 0\},$$

and

$$\mathbb{E}[\|\widehat{Z}_k - Z_k\|] \leq 3\sqrt{\gamma c_{1,2}c_{2,2}} + c_{3,2}\delta_{4s}/(1-\delta_{4s})$$
$$= \left(3N'_2\alpha_2^{-2}e^{\alpha_2}\sqrt{C^{LV}_2 C^{\Omega}_2}\right)\cdot\sqrt{\gamma} + \frac{2N'_2 M^{\Theta}_2\alpha_2^{-1}}{1-\delta_{4s}}\cdot\delta_{4s}$$
$$= c_2(\gamma,\delta_{4s}), \quad k \geq K_{0,2}. \quad (23)$$

where $c_{1,2}, c_{2,2}, c_{3,2}, c_{4,2}, \alpha_2, N'_2$ are defined in Theorem 2 with $p = 1$, and $C^{LV}_2, C^{\Omega}_2, M^{\Theta}_2$ are defined in Assumption 4 with $p = 1$.

When $\gamma = 0$, there exists an integer

$$K'_{0,2} = \max\{\lceil \frac{\log(c_{1,2}\sqrt{\mu}) - \log(c_{4,2})}{\log(1-\mu\alpha_2)} - 1\rceil, 0\},$$

such that

$$\mathbb{E}[\|\widehat{Z}_k - Z_k\|] \leq 2c_{1,2}\sqrt{\mu} + c_{3,2}\delta_{4s}/(1-\delta_{4s})$$
$$= 2N'_2 C^{LV}_2\alpha_2^{-2}e^{\alpha_2}\sqrt{\mu} + 2N'_2 M^{\Theta}_2\alpha_2^{-1}\delta_{4s}/(1-\delta_{4s})$$
$$= c'_2(C^{LV}_2\sqrt{\mu},\delta_{4s}), \quad k \geq K'_{0,2}. \quad (24)$$

By Remark 5, we can obtain the following theorem on the compressed estimation error bound in probability.

Theorem 3. *Consider the model* (5) *and the compressed estimation error Eq.* (8). *Under the same conditions as used in* Theorem 2, *we assume that $\gamma > 0$, $\mu = \gamma c_{2,2}(c_{1,2})^{-1}$, and $v \in (0,1)$ satisfying $\mu(1+2v) < 1$. Then there exists an integer $K_0 \geq 0$, such that for any integer $k \geq K_0$ and any constant $\xi \in (0,1)$, there exists a positive constant $\eta = \max\{1, 2c_2^{\xi}(\gamma,\delta_{4s})\}$ such that*

$$P\{\|\widehat{Z}_k - Z_k\| \leq \eta c_2^{1-\xi}(\gamma,\delta_{4s})\} \geq 1 - \frac{c_2^{\xi}(\gamma,\delta_{4s})}{\eta},$$

where $c_2(\gamma,\delta_{4s}) = 3\sqrt{\gamma c_{1,2}c_{2,2}} + c_{3,2}\delta_{4s}/(1-\delta_{4s})$, and $c_{1,2}, c_{2,2}, c_{3,2}$ are defined in Theorem 2 *with $p = 1$.*

Proof. By the analysis in Remark 5, we know that there exists an integer $K_0 = K_{0,2}$, such that for any $k \geq K_0$, we have $\mathbb{E}[\|\widehat{Z}_k - Z_k\|] \leq c_2(\gamma,\delta_{4s})$, where $\alpha_2, c_{1,2}, c_{2,2}, c_{4,2}$ are defined in Theorem 2 with $p = 1$. Since $\eta = \max\{1, 2c_2^{\xi}(\gamma,\delta_{4s})\}$, then by the Markov inequality, we have for any $k \geq K_0$,

$$P\{\|\widehat{Z}_k - Z_k\| \geq \eta c_2^{1-\xi}(\gamma,\delta_{4s})\}$$
$$\leq \frac{\mathbb{E}[\|\widehat{Z}_k - Z_k\|]}{\eta c_2^{1-\xi}(\gamma,\delta_{4s})} \leq \frac{c_2(\gamma,\delta_{4s})}{\eta c_2^{1-\xi}(\gamma,\delta_{4s})} = \frac{c_2^{\xi}(\gamma,\delta_{4s})}{\eta} \leq \frac{1}{2}.$$

This completes the proof.

Remark 6. Firstly, we know that $c_2(\gamma,\delta_{4s})$ tends to zero as γ and δ_{4s} tends to zero. Moreover, when $c_2(\gamma,\delta_{4s})$ is small enough, we have $\eta = 1$. Thus, when γ and δ_{4s} tend to zero, the compressed estimation error $\|\widehat{Z}_k - Z_k\|$ goes to zero with a large probability, which tends to 1.

Remark 7. Moreover, when $\gamma = 0$, under the same conditions as used in Theorem 2, we assume that $\mu \in (0,1)$ and $v \in (0,1)$ satisfying $\mu(1+2v) < 1$. Then there exists an integer $K_0 \geq 0$, such that for any integer $k \geq K_0$ and any constant $\xi \in (0,1)$, there exists a positive constant $\eta = \max\{1, 2(c'_2(C^{LV}_2\sqrt{\mu},\delta_{4s}))^{\xi}\}$ such that

$$P\{\|\widehat{Z}_k - Z_k\| \leq \eta(c'_2(C^{LV}_2\sqrt{\mu},\delta_{4s}))^{1-\xi}\}$$
$$\geq 1 - \frac{(c'_2(C^{LV}_2\sqrt{\mu},\delta_{4s}))^{\xi}}{\eta}, \quad (25)$$

where $c'_2(C^{LV}_2\sqrt{\mu},\delta_{4s}) = 2c_{1,2}\sqrt{\mu} + c_{3,2}\delta_{4s}/(1-\delta_{4s})$, $c_{1,2}, c_{3,2}$ are defined in Theorem 2 with $p = 1$ and C^{LV}_2 is defined in Assumption 4 with $p = 1$. Thus, when $C^{LV}_2\sqrt{\mu}$ and δ_{4s} tends to

zero, the compressed estimation error $\|\widehat{Z}_k - Z_k\|$ goes to zero with a large probability, which tends to 1.

By Theorem 3 and Algorithm 2, we can finally obtain the estimation error bound for the original high-dimensional parameter vector in the following theorem.

Theorem 4. *Consider the model* (5) *and the compressed estimation error Eq.* (8). *Suppose that* $M \in \mathbb{R}^{d \times m}$ *satisfies 4sth RIP with s satisfying* $\delta_{3s} + 3\delta_{4s} < 2$. *Under the same conditions and the same step-size* μ *and* ν *as used in Theorem 3, there exists an integer* $K_0 \geq 0$, *such that for any integer* $k \geq K_0$ *and any constant* $\xi \in (0, 1)$, *there exists a positive constant* $\eta = \max\{1, 2c_2^{\xi}(\gamma, \delta_{4s})\}$ *such that for any* $i(i = 1, \ldots, n)$,

$$P\left\{\|\widehat{\boldsymbol{\theta}}_k^i - \boldsymbol{\theta}_k\| \leq C_s \eta c_2^{1-\xi}(\gamma, \delta_{4s})\right\} \geq 1 - \frac{c_2^{\xi}(\gamma, \delta_{4s})}{\eta},$$

where $c_2(\gamma, \delta_{4s}) = 3\sqrt{\gamma c_{1,2} c_{2,2}} + c_{3,2}\delta_{4s}/(1 - \delta_{4s})$, *and* $c_{1,2}, c_{2,2}, c_{3,2}$ *are defined in Theorem 2 with* $p = 1$, *the constant* C_s *may only depend on* δ_{4s}, *and* C_s *gets smaller as* δ_{4s} *gets smaller.*

Proof. Since $\widehat{\widetilde{\zeta}}_k^i = \zeta_k + \widetilde{\zeta}_k^i = M\boldsymbol{\theta}_k + \widetilde{\zeta}_k^i$, then by Theorem 3, and Lemma A.2 in the Appendix, we have for any $k \geq K_0$, and $i = 1, \ldots, n$,

$$P\{\|\widehat{\boldsymbol{\theta}}_k^i - \boldsymbol{\theta}_k\| \leq C_s \eta c_2^{1-\xi}(\gamma, \delta_{4s})\}$$
$$\geq P\{\widetilde{\zeta}_k^i = \widehat{\zeta}_k^i - M\boldsymbol{\theta}_k, \|\widetilde{\zeta}_k^i\| \leq \eta c_2^{1-\xi}(\gamma, \delta_{4s})\}$$
$$= P\{\|\widetilde{\zeta}_k^i\| \leq \eta c_2^{1-\xi}(\gamma, \delta_{4s})\}$$
$$\geq P\{\|\widetilde{Z}_k\| \leq \eta c_2^{1-\xi}(\gamma, \delta_{4s})\} \geq 1 - \frac{c_2^{\xi}(\gamma, \delta_{4s})}{\eta}.$$

Moreover, by the relationship between C_s and δ_{4s} in Remark 9 in the Appendix, we know that C_s gets smaller as δ_{4s} gets smaller. This completes the proof.

By Remark 6, we know that when γ and δ_{4s} tend to zero, the estimation error $\|\widehat{\boldsymbol{\theta}}_k^i - \boldsymbol{\theta}_k\|$ also goes to zero with a large probability which tends to 1. When $\gamma = 0$, we can obtain a similar result. Here we omit it.

4. Simulation results

Let us consider a network with 20 sensors, which is the same connected graph as used in Xie and Guo (2018b) with the same adjacency matrix $\mathcal{A}$. Thus, $\mathcal{A}$ is symmetrical. Here we compare two different kinds of algorithms to illustrate the advantage of the compressed consensus NLMS algorithm, i.e., the consensus NLMS algorithm (Sayed, 2014; Xie & Guo, 2018b), and the compressed consensus NLMS algorithm.

Example 1. All the sensors will estimate or track an unknown 50-dimensional 3-sparse signal $\boldsymbol{\theta}_k$, and all the coefficients of $\boldsymbol{\theta}_k$ are zero except the first 3 coefficients. Here we only consider the time-invariant case, i.e., $\gamma = 0$. When $\gamma \neq 0$, we can also obtain some similar results, which are omitted here for space reason. We assume that the observation noises $\{v_k^i, k \geq 1, i = 1, \ldots, 20\}$ are temporally and spatially independently
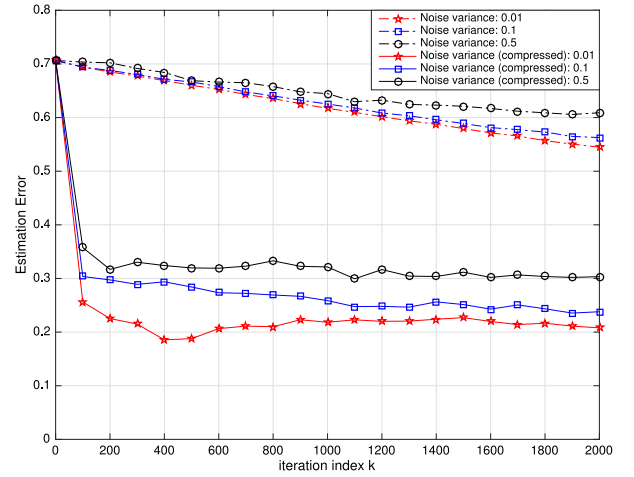

Fig. 1. Estimation errors in Example 1.

distributed with $v_k^i \sim N(0, \sigma^2, 1, 1)$ (Gaussian distribution with zero mean, σ^2 variance and 1 dimension) in (3), where σ^2 is chosen to be 0.01, 0.1 and 0.5, respectively. Then the corresponding averaged signal-to-noise ratios (SNRs) are given in the following table.

Noise variance	0.01	0.1	0.5
SNR	−8 dB	−18 dB	−25 dB

Let the regressors $\boldsymbol{\varphi}_k^i \in \mathbb{R}^{50}(i = 1, \ldots, 20)$ be generated by the equations $\boldsymbol{x}_k^i = A_i \boldsymbol{x}_{k-1}^i + B_i \boldsymbol{\xi}_k^i$ and $\boldsymbol{\varphi}_k^i = C_i \boldsymbol{x}_k^i$, where $\boldsymbol{x}_k^i \in \mathbb{R}^{50}, A_i \in \mathbb{R}^{50 \times 50}, B_i \in \mathbb{R}^{50 \times 3}, C_i \in \mathbb{R}^{50 \times 50}$ and $\boldsymbol{\xi}_k^i \in \mathbb{R}^3$. Let $\boldsymbol{\xi}_k^i \sim N(0, 0.3^2, 3, 1)$, and let $A_i(i = 1, \ldots, 20)$ be all diagonal matrices with the diagonal elements equal to $1/3$ and let B_i, C_i be sparse enough to ensure the sparsity 12 of regressors. It can be verified that the information condition in Xie and Guo (2018b) is not satisfied. However, Assumption 2 in this paper can be satisfied for compressed regressors $\boldsymbol{\psi}_k^i = M\boldsymbol{\varphi}_k^i \in \mathbb{R}^{10}$, where $M \in \mathbb{R}^{10 \times 50}$ is a Gaussian random matrix which is defined in Appendix. Since the sparse value is 3, $d = 10$ is suitable for compression and decompression algorithm.

Let $\boldsymbol{x}_0^i = \text{col}\{\underbrace{1, \ldots, 1}_{50}\}(i = 1, \ldots, 20)$ and the first 3 coordinates of variation $\boldsymbol{\theta}_0 \sim 0.1 * N(0, 1, 3, 1)$, while the rest of its coordinates are all zero. We choose $\hat{\boldsymbol{\theta}}_0^i = \text{col}\{\underbrace{0.1, \ldots, 0.1}_{50}\}, \hat{\boldsymbol{\zeta}}_0^i = \text{col}\{\underbrace{0.1, \ldots, 0.1}_{10}\}(i = 1, \ldots, 20), \mu = 0.3, \nu = 0.2$. For the decompression part, we apply the OMP algorithm to solve the optimization problem in Algorithm 2 to generate the recovery signal every 100 iterations. For each algorithm, we repeat the simulation for $t = 500$ times with the same initial states. Then for sensor $i(i = 1, \ldots, 20)$, we can get t sequences $\{\|\hat{\boldsymbol{\theta}}_k^{i,j} - \boldsymbol{\theta}_k\|, k = 1, 100, 200, \ldots, 2000\}(j = 1, \ldots, t)$, where the superscript j denotes the jth simulation result. Here we use

$$\frac{1}{20t} \sum_{i=1}^{20} \sum_{j=1}^{t} \|\hat{\boldsymbol{\theta}}_k^{i,j} - \boldsymbol{\theta}_k\|, \quad k = 100, 200, \ldots, 2000,$$

to approximate the mean estimation errors for every 100 iterations in Fig. 1.

Note that the three upper curves represent the estimation errors of the consensus NLMS algorithm, which keep large because the whole sensor network cannot satisfy the information condition in Xie and Guo (2018b). However, the estimation errors

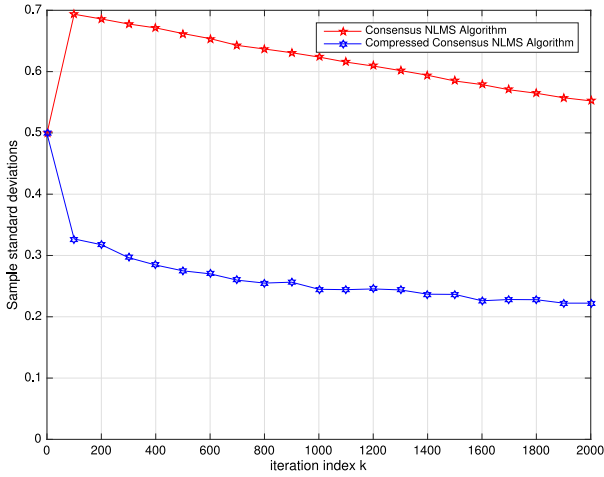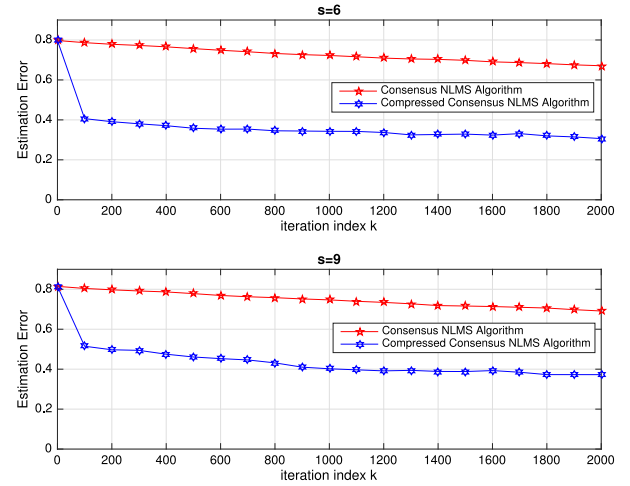**Fig. 2.** Sample standard deviations in Example 1.

Fig. 3. Estimation errors in Example 2.

of the compressed one are smaller than the other situation. Moreover, we can see that when the step size parameters are fixed, the estimation performance improves with the increase of SNR.

For each algorithm, we can also consider the standard deviations of the estimation errors over the sensor network, which is usually defined as the square root of the variances. Here, the sample standard deviations are defined as

$$\sqrt{\frac{1}{20t}\sum_{i=1}^{20}\sum_{j=1}^{t}\mathrm{Tr}((\hat{\boldsymbol{\theta}}_k^{i,j}-\boldsymbol{\theta}_k)(\hat{\boldsymbol{\theta}}_k^{i,j}-\boldsymbol{\theta}_k)^T)},\quad k=100,200,\dots,2000,$$

which is plotted in Fig. 2 with $\sigma^2 = 0.1$ for every 100 iterations. We can see that the sample standard deviations of the compressed algorithm are smaller than the other situation.

Example 2. We consider higher s-sparse value, and we choose that the noise variance equals 0.1 and $\gamma = 0$ to simplify the experiment. We assume that all the sensors will estimate an unknown 50-dimensional s-sparse signal $\boldsymbol{\theta}_k$, and all the coefficients of $\boldsymbol{\theta}_k$ are zero except the first s coefficients, where s cannot be larger than $50/4$ according to Theorem 1 and the first s coordinates of variation $\boldsymbol{\theta}_0 \sim 0.1 * N(0, 1, s, 1)$. Then s is chosen to be 6 and 9 in this experiment, and d is chosen to be 12 and 15, respectively. We select the same regressors, initial states and step size parameters as in the first example and apply the consensus NLMS and compressed consensus NLMS algorithms for all sensors; then plot the estimation errors of the sensor network with $s = 6$ and 9 in Fig. 3, which shows that the estimation errors of the compressed one are smaller than the other situation for both cases.

5. Concluding remarks

In this work, we have introduced a class of compressed distributed adaptive filtering algorithms based on consensus strategies and CS theory, aiming at estimating an unknown high-dimensional sparse signal. We remark that the traditional non-compressed distributed algorithm may not be able to estimate the unknown signal due to the sparsity of the regressors. However, the current compressed consensus NLMS algorithm can fulfill the estimation task under a compressed information condition, which is much weaker than the original information condition (Xie & Guo, 2018b) as illustrated by the numerical example in the paper. Furthermore, we have not only provided an upper bound for the compressed estimation error, but also given an upper bound for the desired high-dimensional estimation error with large probability. Of course, there are still a number of interesting problems for further research, for example, to consider real applications of the compressed filtering algorithms, and to study other adaptive filtering algorithms based on this compressed method.

Appendix. The CS theory and signal reconstruction algorithm

A vector $\boldsymbol{x}_0 \in \mathbb{R}^m$ is called s-sparse ($s \ll m$) if at most s of its coordinates are non-zero. The noisy observation of $\boldsymbol{x}_0$ is a $d(s \le d \ll m)$-dimensional vector $\boldsymbol{y}$, which is defined as

$$\boldsymbol{y} = M\boldsymbol{x}_0 + \boldsymbol{e}, \tag{26}$$

where $M \in \mathbb{R}^{d \times m}$ is a measurement matrix, and $\boldsymbol{e} \in \mathbb{R}^d$ is some unknown perturbation bounded by a known amount $\|\boldsymbol{e}\| \le C$. The main goal of CS is to recovery $\boldsymbol{x}_0$ from $\boldsymbol{y}$ by properly choosing M and solving the signal reconstruction problem.

Candès et al. (2006) introduced the restricted isometry property (RIP) constraint and showed that the RIP on the measurement matrix M can guarantee the high quality of $\boldsymbol{y}$. Given any set Q of column indices, we denote by $\#(Q)$ the number of elements in the set Q, and M_Q the $d \times \#(Q)$ matrix composed of these columns. Similarly, for a vector $\boldsymbol{x} \in \mathbb{R}^m$, we denote by $\boldsymbol{x}_Q$ the vector obtained by retaining only the entries in $\boldsymbol{x}$ corresponding to the column indices Q. We say that the matrix M satisfies the RIP of order s if there exists a constant $\delta_s \in [0, 1)$, which is the smallest quantity such that

$$(1 - \delta_s)\|\boldsymbol{x}_Q\|^2 \le \|M_Q\boldsymbol{x}_Q\|^2 \le (1 + \delta_s)\|\boldsymbol{x}_Q\|^2, \tag{27}$$

for all sets Q with $\#(Q) \le s$. Then δ_s is called the sth RIP constant. The condition (27) is equivalent to requiring that the Grammian matrix $M_Q^T M_Q$ has all of its eigenvalues in $[1 - \delta_s, 1 + \delta_s]$, where $\#(Q) \le s$.

There are many existing constructive methods to make M satisfying the RIP condition (DeVore, 2007; Xu & Xu, 2015). For example, the deterministic matrix methods including the deterministic Fourier measurements and the incoherent measurements. Also, random matrix methods, e.g., Gaussian measurements, Binary measurements, random Fourier measurements, may be used to get probabilistic results. In this work, we employ Gaussian

measurements for our simulations, i.e., the entries of the measurement matrix M are independently sampled from a Gauss distribution with zero mean and variance $1/d$ (Baraniuk, Davenport, DeVore, & Wakin, 2008). About random matrices, we have the following result.

Lemma A.1 (*Baraniuk et al., 2008*). *Suppose that d, m and $0 < \delta < 1$ are given. If the measurement matrix $M \in \mathbb{R}^{d \times m}$ is a Gaussian or Bernoulli random matrix, then there exist constants C_1 and C_2 depending only on δ such that the RIP condition* (27) *holds for the measurement matrix M with the prescribed δ and any $s \leq C_1 d / \log(m/s)$ with probability not less than $1 - 2exp(-C_2 d)$.*

Remark 8. In fact, Baraniuk et al. (2008) give a simple proof for Lemma A.1 and show the connection between C_1 and C_2. Here we should guarantee $C_2 > 0$ and $C_2 \leq \frac{\delta^2}{16} - \frac{\delta^3}{48} - C_1\left[1 + \frac{1+\log(12/\delta)}{\log(m/s)}\right]$. The constant C_1 should be small enough to ensure $C_2 > 0$, see Theorem 5.2 in Baraniuk et al. (2008) for details. Here we can choose $C_1 = \delta^3/120$. Then we can obtain that if $d \geq 120 s \log(m/s)/\delta^3$, then the measurement matrix M satisfies condition (27) with probability not less than $1 - 2exp(-C_2 d)$.

In order to recover the signal $\boldsymbol{x}_0$, we consider the following reconstruction problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^m} \quad \|\boldsymbol{x}\|_{\ell_1}, \quad s.t. \quad \|M\boldsymbol{x} - \boldsymbol{y}\| \leq C. \tag{28}$$

We denote the decoding result as $\boldsymbol{x}^*$, and from Candès et al. (2006), we know that the recovery procedure above is stable, which means that small changes in the observations should result in small changes in the recovery.

Lemma A.2 (*Candès et al., 2006*). *Let s be such that $\delta_{3s} + 3\delta_{4s} < 2$. Then for any signal $\boldsymbol{x}_0$ which has not more than s non-zero coordinates, and any perturbation $\boldsymbol{e}$ with $\|\boldsymbol{e}\| \leq C$, the solution $\boldsymbol{x}^*$ obeys $\|\boldsymbol{x}^* - \boldsymbol{x}_0\| \leq C_s C$, where the constant C_s may only depend on δ_{4s}.*

Remark 9. The condition for stable recovery involves the restricted isometry constants. For s obeying the condition of the theorem, the reconstruction from noiseless data is exact. Moreover, from the proof of Theorem 1 in Candès et al. (2006), we know that $C_s = 4/(\sqrt{3(1 - \delta_{4s})} - \sqrt{1 + \delta_{3s}})$, which shows that as δ_{4s} gets smaller, C_s will also get smaller. Then for reasonable values of δ_{4s}, C_s is well behaved; e.g. $C_s \approx 8.82$ for $\delta_{4s} = 1/5$ and $C_s \approx 10.47$ for $\delta_{4s} = 1/4$.

References

Agaev, R., & Chebotarev, P. (2006). The matrix of maximum out forests of a digraph and its applications. *Automation & Remote Control*, 61(9), 1424–1450.

Angelosante, D., Bazerque, J. A., & Giannakis, G. B. (2010). Online adaptive estimation of sparse signals: Where RLS meets the ℓ_1-norm. *IEEE Transactions on Signal Processing*, 58(7), 3436–3447.

Babadi, B., Kalouptsidis, N., & Tarokh, V. (2010). SPARLS: The sparse RLS algorithm. *IEEE Transactions on Signal Processing*, 58(8), 4013–4025.

Bajwa, W., Haupt, J., Sayeed, A., & Nowak, R. (2006). Compressive wireless sensing. In *International conference on information processing in sensor networks* (pp. 134–142). Nashville, TN, USA.

Baraniuk, R. G., Davenport, M., DeVore, R., & Wakin, M. (2008). A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3), 253–263.

Baron, D., Duarte, M. F., Wakin, M. B., Sarvotham, S., & Baraniuk, R. G. (2009). *Distributed compressive sensing*: Technical report, Houston, TX, USA: ECE Department, Rice University.

Berg, E., & Friedlander, M. (2008). Probing the Pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2), 890–912.

Candès, E. J., Romberg, J. K., & Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8), 1207–1223.

Candès, E. J., Wakin, M. B., & Boyd, S. P. (2008). Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier Analysis and Applications*, 14(5–6), 877–905.

Carli, R., Chiuso, A., Schenato, L., & Zampieri, S. (2008). Distributed Kalman filtering based on consensus strategies. *IEEE Journal on Selected Areas in Communications*, 26(4), 622–633.

Chartrand, R. (2007). Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Processing Letters*, 14(10), 707–710.

Chen, Y., Gu, Y., & Hero, A. O. (2009). Sparse LMS for system identification. In *IEEE international conference on acoustics, speech and signal processing* (pp. 3125–3128). Taipei.

Chen, J. S., & Sayed, A. H. (2015a). On the learning behavior of adaptive networks part I: Transient analysis. *IEEE Transactions on Information Theory*, 61(6), 3487–3517.

Chouvardas, S., Slavakis, K., Kopsinis, Y., & Theodoridis, S. (2011). A sparsity promoting adaptive algorithm for distributed learning. *IEEE Transactions on Signal Processing*, 60(10), 5412–5425.

Chow, Y. S., & Teicher, H. (1978). *Probability theory*. New York: Springer.

DeVore, R. A. (2007). Deterministic constructions of compressed sensing matrices. *Journal of Complexity*, 23, 918–925.

Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, 52(4), 1289–1306.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456), 1348–1360.

Figueiredo, M., Nowak, R., & Wright, S. (2007). Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal on Selected Topics in Signal Processing*, 1(4), 586–597.

Guo, L. (1994). Stability of recursive stochastic tracking algorithms. *SIAM Journal on Control and Optimization*, 32(5), 1195–1225.

Guo, L., & Ljung, L. (1995). Performance analysis of general tracking algorithms. *IEEE Transactions on Automatic Control*, 40(8), 1388–1402.

Herrmann, F. J., Friedlander, M. P., & Yilmaz, O. (2012). Fighting the curse of dimensionality: Compressive sensing in exploration seismology. *IEEE Signal Processing Magazine*, 29(3), 88–100.

Hosseini, S. H., & Shayesteh, M. G. (2012). Compressed sensing for denoising in adaptive system identification. In *20th Iranian conference on electrical engineering* (pp. 1238–1242). Tehran, Iran.

Khalifa, M. O., Abdelhafiz, A. H., & Zerguine, A. (2013). Sparse channel estimation using adaptive filtering and compressed sampling. In *1st international conference on computing, electrical and electronics engineering* (pp. 144–147). Khartoum, Sudan.

Kim, S. J., Koh, K., Lustig, M., Boyd, S., & Gorinevsky, D. (2007). An interior-point method for large-scale ℓ_1-regularized least squares. *IEEE Journal on Selected Topics in Signal Processing*, 1(4), 606–617.

Kopsinis, Y., Slavakis, K., & Theodoridis, S. (2011). Online sparse system identification and signal reconstruction using projections onto weighted ℓ_1 balls. *IEEE Transactions on Signal Processing*, 59(3), 936–952.

Li, D., Zhang, Q., Shen, H., Shi, J., & Shen, M. (2011). The application of compressive sensing based on wavelet in the reconstruction of ultrasonic medical image. In *2011 second international conference on mechanic automation and control engineering (MACE)* (pp. 5350–5353). Hohhot, China.

Liu, Y., Li, C., & Zhang, Z. (2012). Diffusion sparse least-mean squares over networks. *IEEE Transactions on Signal Processing*, 60(8), 4480–4485.

Liu, Z., Liu, Y., & Li, C. (2014). Distributed sparse recursive least-squares over networks. *IEEE Transactions on Signal Processing*, 62(6), 1386–1395.

Lorenzo, P. D., & Sayed, A. H. (2013). Sparse distributed learning based on diffusion adaptation. *IEEE Transactions on Signal Processing*, 61(6), 1419–1433.

Paredes, J. L., Arce, G. R., & Wang, Z. (2007). Ultra-wideband compressed sensing: Channel estimation. *IEEE Journal of Selected Topics in Signal Processing*, 1(3), 383–395.

Piggott, M. J., & Solo, V. (2016). Diffusion LMS with correlated regressors I: Realization-wise stability. *IEEE Transactions on Signal Processing*, 64(21), 5473–5484.

Rastegarnia, A., Tinati, M. A., & Khalili, A. (2010). A distributed incremental LMS algorithm with reliability of observation consideration. In *IEEE international conference on communication systems* (pp. 67–70). Singapore.

Romberg, J. (2008). Imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2), 14–20.

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

Sayin, M. O., & Kozat, S. S. (2014). Compressive diffusion strategies over distributed networks for reduced communication load. *IEEE Transactions on Signal Processing*, 62(20), 5308–5323.

Schizas, I. D., Mateos, G., & Giannakis, G. B. (2009). Distributed LMS for consensus-based in-network adaptive processing. *IEEE Transactions on Signal Processing*, 57(6), 2365–2382.

Shi, K., & Shi, P. (2010). Convergence analysis of sparse LMS algorithms with l_1-norm penalty based on white input signal. *Signal Processing*, 90(12), 3289–3293.

Tropp, J. A., & Gilbert, A. C. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12), 4655–4666.

Widrow, B., & Stearns, S. D. (1985). *Adaptive signal processing*. NJ: Prentice-Hall Englewood Cliffs.

Wright, S., Nowak, R., & Figueiredo, M. (2009). Sparse reconstruction by separable approximation. *IEEE Transactions on Signal Processing, 57*(7), 2479–2493.

Xie, S. Y., & Guo, L. (2016). Compressive distributed adaptive filtering. In *Proc. 35th Chinese control conference (CCC)* (pp. 5229–5234). Chengdu, China.

Xie, S. Y., & Guo, L. (2018a). A necessary and sufficient condition for stability of LMS-based consensus adaptive filters. *Automatica, 93*, 12–19.

Xie, S. Y., & Guo, L. (2018b). Analysis of NLMS-based consensus adaptive filters under a general information condition. *SIAM Journal on Control and Optimization, 56*(5), 3404–3431.

Xie, S. Y., & Guo, L. (2018c). Analysis of distributed adaptive filters based on diffusion strategies over sensor networks. *IEEE Transactions on Automatic Control, 63*(11), 3643–3658.

Xu, Z. Q. (2010). A remark about orthogonal matching pursuit algorithm. ArXiv: 1005.3093.

Xu, S., de Lamare, R. C., & Poor, H. V. (2015). Distributed compressed estimation based on compressive sensing. *IEEE Signal Processing Letters, 22*(9), 1311–1315.

Xu, G. W., & Xu, Z. Q. (2015). Compressed sensing matrices from Fourier matrices. *IEEE Transactions on Information Theory, 61*(1), 469–478.

Yang, J., & Zhang, Y. (2011). Alternating direction algorithms for ℓ_1-problems in compressive sensing. *SIAM Journal on Scientific Computing, 33*(1), 250–278.

Zeng, J., Lin, S., & Xu, Z. (2016). Sparse regularization: Convergence of iterative jumping thresholding algorithm. *IEEE Transactions on Signal Processing, 64*(19), 5106–5118.

Siyu Xie received her B.S. degree in information and computing science (systems control) from Beijing University of Aeronautics and Astronautics in 2013, and Ph.D. degree in control theory from Academy of Mathematics and Systems Science, Chinese Academy of Sciences in 2018. She is currently a postdoctoral fellow at the Department of Electrical and Computer Engineering, Wayne State University, USA.

She received the IEEE CSS Beijing Chapter Young Author Prize in 2015, China National Scholarship for graduate students in 2017, and the President Scholarship of Chinese Academy of Sciences in 2018. Her research interests include networked systems, distributed adaptive filters, machine learning, compressive sensing and distributed control.

Lei Guo received his B.S. degree in mathematics from Shandong University in 1982, and Ph.D. degree in control theory from the Chinese Academy of Sciences (CAS) in 1987. He is currently a professor of the Academy of Mathematics and Systems Science (AMSS), CAS. His research interests include estimation and control of uncertain systems, stochastic systems, multi-agent systems, and game-based control systems.

He is a Fellow of IEEE, Member of Chinese Academy of Sciences, Fellow of Academy of Sciences for the Developing World (TWAS), Foreign Member of Royal Swedish Academy of Engineering Sciences, and Fellow of International Federation of Automatic Control (IFAC). He was awarded the Hendrik W. Bode Lecture Prize by the IEEE Control System Society in 2019, an honorary doctorate by Royal Institute of Technology (KTH), Sweden, in 2014, Outstanding Service Award of IFAC in 2011, National Natural Science Prizes of China in 1987 and 1997. He has served as a Council Member of IFAC, General Co-Chair of 48th IEEE-CDC, President of China Society for Industrial and Applied Mathematics, and President of AMSS, CAS, among others.