Proceedings of the 14th Asian Control Conference
July 5-8, 2024, Dalian, China

# Tracking Bound of Compressed Distributed Recursive Least Squares with Forgetting Factor

Shuning Chen, Die Gan, Siyu Xie and Jinhu Lü

*Abstract*—**In this paper, we consider the problem of distributed time-varying sparse parameter estimation over sensor networks. By first compressing the regression signals to remove the sparsity, and then estimating the compressed parameters with the compressed signals, a compressed distributed recursive least squares algorithm with forgetting factor (FFLS) is proposed based on compressive sensing theory. Under the compressed cooperative stochastic excitation condition, we analyze the tracking bound of the estimation error and establish the stability of the compressed distributed FFLS algorithm. Our theoretical results do not rely on independency and stationarity of the regression signals, which makes it possible to be applied to the feedback system. Finally, some simulation results are presented to demonstrate the superiority of our proposed algorithm over the compressed distributed least mean squares (LMS) algorithm and the uncompressed distributed FFLS algorithm.**

*Index Terms*—**Sparse parameter identification, distributed recursive least squares algorithm, compressive sensing, stochastic dynamic system**

## I. Introduction

In recent years, distributed adaptive identification and filtering algorithms have attracted widespread attention and been applied to many fields such as signal processing, adaptive control and bio-medicine [1]. A hot issue of distributed adaptive filtering is to process and estimate the high-dimensional sparse signals. On the premise that the sparsity of the signal is known, the algorithm can be pertinently designed to improve the performance of parameter identification.

With the development of sparse estimation, researchers have also proposed many distributed sparse parameter estimation algorithms, among which a commonly used method is to add penalty terms to the cost function. In [2], a sparse distributed least mean squares (LMS) algorithm based on adapt-then-combine strategy was developed by adding penalty term to the cost function. Liu et al. in [3] adopted maximum likelihood framework with $l_0$ and $l_1$ norm penalties and designed distributed sparse recursive least squares (RLS) estimators. However, we note that these techniques took the full dimension of the data into account, which may result in excessive computational complexity and slow convergence speed.

Compressive sensing theory, as a burgeoning theory of restoring sparse signals, has promising advancements in processing and estimating sparse signals. Candès et al. in [4] introduced some important properties to ensure the compressive sensing performance, and this theory has been applied to adaptive filtering thereafter [5]–[7]. However, in the investigation of the theoretical analysis about distributed sparse identification algorithms, most existing literature assumes the independency, stationarity or ergodicity of input signals, which generally can not be satisfied in complex dynamic systems with various feedback loops. To weaken the assumptions on input signals and take the sparsity of regressors into consideration, some alternative cooperative excitation conditions on the stochastic signals are proposed, and the performance analysis of the compressed distributed LMS and RLS algorithms are provided in our previous works [8]–[10]. Compared with [9], [10], the current paper aims at dealing with a more general case where the parameters are time-varying.

In this paper, we investigate unknown sparse parameter estimation for a discrete-time stochastic regression model over sensor networks. We integrate compressive sensing methods to the recursive least squares algorithm with forgetting factor and propose a compressed distributed FFLS algorithm, which has faster convergence speed than the compressed distributed LMS algorithm in [8]. Compared with the traditional uncompressed FFLS algorithm, it is proved that the proposed algorithm can achieve more efficient parameter estimation in the case of sparse signals. We introduce a compressed excitation condition and theoretically establish the upper bound of the tracking error in the case of time-varying parameter. Moreover, our excitation condition also reveals the cooperative effect of multiple sensors in the sense that sensors can collaborate to achieve estimation, even if any sensor cannot do it alone. It is worth pointing out that our results are applicable to feedback systems, unlike the theoretical results in [5]–[7], which rely on the assumption of independency or stationarity of the regression signals.

The remainder of this paper is organized as follows. Section II gives the problem formulation including the design of the compressed distributed FFLS algorithm. The stability results of the proposed algorithm are presented in Section III. Some simulation results are provided in Section IV and the concluding remarks are given in Section V.

This work was supported by the National Natural Science Foundation of China under Grants 61621003, 62103015, 62141604, and China Postdoctoral Science Foundation (2022M722926) (Corresponding author: Die Gan).

S. Chen is with the Key Laboratory of Systems and Control at the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, as well as the School of Mathematical Sciences at the University of Chinese Academy of Sciences in Beijing 100190, China (email: chenshuning@amss.ac.cn).

D. Gan is with Zhongguancun Laboratory, located in Beijing 100094, China (email: gandie@amss.ac.cn).

S. Xie is with the School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: syxie@uestc.edu.cn).

J. Lü is affiliated with the School of Automation Science and Electrical Engineering at Beihang University, Beijing 100191, China. He is also associated with Zhongguancun Laboratory, Beijing 100094, China (email: jhlu@iss.ac.cn).

## II. Problem Formulation

### A. Some Preliminaries

*1) Matrix Theory and Graph Theory:* For an $n \times m$-dimensional real matrix $A$, the Euclidean norm of $A$ is symbolically represented by $\|A\|$, i.e., $\|A\| \triangleq \{\lambda_{\max}(AA^T)\}^{\frac{1}{2}}$, where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of the matrix. Correspondingly, the smallest eigenvalue of the matrix is denoted by $\lambda_{\min}(\cdot)$. For an $m$-dimensional real vector $x$, its $p$-norm is defined as $\|x\|_p = (\sum_{i=1}^m |x_i|^p)^{1/p}$, where $x_i$ is the $i$th component of $x$ and $1 \leq p < \infty$. If there is no special indication, $\|x\|$ refers to the 2-norm (also the Euclidean norm). We also use $\|x\|_0$ to denote the number of non-zero elements in $x$. The $m$-dimensional identity matrix is denoted by $I_m$. Considering a matrix sequence $\{A_k, k \geq 0\}$ and a positive scalar sequence $\{a_k, k \geq 0\}$, if there exists a positive constant $C$, such that $\|A_k\| \leq Ca_k$ holds for all $k \geq 0$, then we say $A_k = O(a_k)$.

For a multi-sensor network, we can construct a corresponding topology $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{A}\}$ to show the information interaction between sensors. Take an $n$-sensor network for instance, let the node set $\mathcal{V} = \{1, 2, \cdots, n\}$. The elements in the adjacency matrix $\mathcal{A} = [a_{ij}]_{1 \leq i,j \leq n}$ represent the weight of information interaction between sensors. The diameter $D_{\mathcal{G}}$ of the graph $\mathcal{G}$ is defined as the maximum shortest length of paths between any two nodes. In this paper, we suppose the adjacency matrix is symmetric and stochastic.

*2) Compressive Sensing Theory:* A vector $x_0 \in \mathbb{R}^m$ is called $s$-sparse, if there are at most $s$ non-zero elements in it and $s$ is far smaller than $m$. The essence of compressive sensing is to restore high-dimensional sparse signals from a small number of linear measurements. The sensing matrix $M \in \mathbb{R}^{d \times m} (d < m)$, which can map a vector in a high-dimensional space $\mathbb{R}^m$ to a low-dimensional space $\mathbb{R}^d$, is the core of compressive sensing theory. Therefore, how to select the sensing matrix so that the high-dimensional signal can be well restored is an important issue. With regard to this point, Candès and Tao [4] put forward the concept of "restricted isometry property (RIP)" to measure the properties of sensing matrices, and gave the proof that the sensing matrix satisfying RIP can successfully restore the signal. The followings are RIP-related concepts and conclusions.

*Definition 2.1 (Restricted Isometry Property):* For a $d \times m$-dimensional matrix $M$, if there exists a minimum constant $\delta_s \in (0, 1)$ such that

$$(1 - \delta_s)\|x\|^2 \leq \|Mx\|^2 \leq (1 + \delta_s)\|x\|^2$$

holds for any $s$-sparse vector $x \in \mathbb{R}^m$, then we say that the sensing matrix $M$ satisfies the RIP of order $s$, and $\delta_s$ is the restricted isometry constant of order $s$, correspondingly.

*Lemma 2.1:* [4] Consider a sparse signal observation model with bounded disturbance $\|\varepsilon\| \leq C$:

$$y_0 = Mx_0 + \varepsilon,$$

where $x_0 \in \mathbb{R}^m$ is the $s$-sparse parameter vector to be recovered, $y_0 \in \mathbb{R}^d$ $(d < m)$ is a low-dimensional observation. Solve the following optimization problem to get the recovered signal:

$$x_0^* = \arg\min_x \{\|x\|_1, \text{ s.t. } \|y_0 - Mx\| \leq C\}.$$

If the sensing matrix $M$ satisfies $3s$- and $4s$-RIP, where $\delta_{3s} + 3\delta_{4s} < 2$, then the recovery error $\|x_0 - x_0^*\|$ can be controlled within a constant neighborhood of zero:

$$\|x_0 - x_0^*\| \leq C_s C,$$

where $C_s = \frac{4}{\sqrt{3(1-\delta_{4s})} - \sqrt{1+\delta_{3s}}}$.

### B. Discrete Regression Model

Let us consider the following discrete-time stochastic regression model of $n$ sensors (labeled as $1, 2, \cdots, n$):

$$y_{t+1,i} = \varphi_{t,i}^T \theta_t + w_{t+1,i}, \ t \geq 0, \tag{1}$$

where $y_{t,i}$ is the scalar observation of sensor $i$ at time $t$, $\theta_t \in \mathbb{R}^m$ is an $s$-sparse time-varying parameter vector to be estimated by $n$ sensors, $\varphi_{t,i} \in \mathbb{R}^m$ is the $3s$-sparse random regression vector of sensor $i$ at time $t$, and the scalar $w_{t,i}$ is the random measurement noise of sensor $i$ at time $t$.

We denote the variation of the parameter vector as:

$$\Delta\theta_{t+1} = \theta_{t+1} - \theta_t, t \geq 0.$$

In particular, when $\Delta\theta_t \equiv 0$, it degenerates to the time-invariant parameter case.

It is worth pointing out that the regression vector sequence $\{\varphi_{t+1,i}\}_{t=0}^{\infty}$ in model (1) is often random and not independent. For example, in the auto-regressive exogenous (ARX) model (cf., [11]), the regression vector can be specifically expressed as $\varphi_{t,i} = [y_{t,i}, \cdots, y_{t-p+1,i}, u_{t,i}, \cdots, u_{t-q+1,i}]^T$, where $u_{t,i}$ denotes the input of sensor $i$ at time $t$, and integers $p, q$ indicate that the regression vector contains the output information over the past $p$ time period as well as the input information over the past $q$ time period. Then, due to the existence of the system noise $w_{t,i}$, the regression vector is random rather than deterministic and can not satisfy the i.i.d. assumption when the control law $u_{t,i} = f(y_{k,i}, k \leq t)$ is designed based on past observations.

The goals of this paper are to design a compressed distributed adaptive estimation algorithm where all sensors cooperatively track the unknown time-varying sparse parameter $\theta_t$, and establish the stability of the proposed algorithm without imposing stringent independency or stationarity conditions on the random regression vectors $\{\varphi_{t,i}\}$.

### C. Compressed Distributed FFLS algorithm

In this section, we devise the compressed distributed FFLS algorithm aiming at estimating the unknown sparse parameter vector.

Consider the sparsity of regression vector $\varphi_{t,i}$, in order to save the calculation cost and improve the tracking performance, we utilize the sensing matrix $M \in \mathbb{R}^{d \times m}$ to compress the original high-dimensional sparse regression vector $\varphi_{t,i}$ into the low-dimensional regressors $\phi_{t,i} = M\varphi_{t,i}$. The corresponding compressed unknown parameter vector is denoted as

$\zeta_t = M\theta_t$. Thus, the regression model (1) can be transformed into

$$
\begin{aligned}
y_{t+1,i} &= \varphi_{t,i}^T\theta_t + \phi_{t,i}^T\zeta_t - \phi_{t,i}^T\zeta_t + w_{t+1,i} \\
&= \phi_{t,i}^T\zeta_t + \varphi_{t,i}^T(I_m - M^TM)\theta_t + w_{t+1,i} \\
&= \phi_{t,i}^T\zeta_t + \bar{w}_{t+1,i},
\end{aligned} \tag{2}
$$

where $\bar{w}_{t+1,i} = \varphi_{t,i}^T(I_m - M^TM)\theta_t + w_{t+1,i}$ can be seen as a new extended error, including the original measurement noise and the error caused by compressive sensing.

Next, the compressed system parameter estimates can be obtained by applying the distributed FFLS algorithm [12] (i.e., (3)-(7)) to the compressed model (2), and finally, using the restoration technique, we can get the corresponding sparse original parameter values. Details are shown in Algorithm 1.

---

**Algorithm 1:** Compressed Distributed FFLS algorithm

**Data:** $\{y_{t,i}, \varphi_{t,i}\}_{i=1}^n$, $t = 0, 1, 2, \cdots$

**Result:** $\{\hat{\theta}_{t+1,i}\}_{i=1}^n$, $t = 0, 1, 2, \cdots$

For every sensor $i \in \{1, 2, \cdots, n\}$, give an initial estimation $\hat{\zeta}_{0,i} \in \mathbb{R}^d$, a positive definite matrix $P_{0,i} \in \mathbb{R}^{d\times d}$, and a common sensing matrix $M$;

**for** *each time* $t = 0, 1, 2, \cdots$ **do**

  **for** *each sensor* $i \in \{1, 2, \cdots, n\}$ **do**

    Compress the regressors: $\phi_{t,i} = M\varphi_{t,i}$;

    Adapt sensor's own estimates:

$$
L_{t,i} = \frac{P_{t,i}\phi_{t,i}}{\lambda + \phi_{t,i}^T P_{t,i}\phi_{t,i}}, \tag{3}
$$

$$
\bar{\zeta}_{t+1,i} = \hat{\zeta}_{t,i} + L_{t,i}(y_{t+1,i} - \phi_{t,i}^T\hat{\zeta}_{t,i}), \tag{4}
$$

$$
\bar{P}_{t+1,i} = \frac{1}{\lambda}\big(P_{t,i} - L_{t,i}\phi_{t,i}^T P_{t,i}\big); \tag{5}
$$

    Combine with neighbors' estimates:

$$
P_{t+1,i}^{-1} = \sum_{j\in\mathcal{N}_i} a_{ij}\bar{P}_{t+1,j}^{-1}, \tag{6}
$$

$$
\hat{\zeta}_{t+1,i} = P_{t+1,i}\sum_{j\in\mathcal{N}_i} a_{ij}\bar{P}_{t+1,j}^{-1}\bar{\zeta}_{t+1,j}; \tag{7}
$$

    Restore the original signal:

$$
\hat{\theta}_{t+1,i} = \arg\min_{\theta\in\mathbb{B}} \|\theta\|_1, \tag{8}
$$

    where $\mathbb{B} = \{\theta \| \|\hat{\zeta}_{t+1,i} - M\theta\| \le C\}$ and $C$ is a given constant.

  **end**

**end**

---

## III. Tracking Performance of the Algorithm

Due to the lag of estimation, the error of time-varying parameter estimation generally cannot converge to zero. What we can do is to track the unknown time-varying parameters and keep the tracking error within a small range, which is also the meaning of algorithmic stability in this paper.

In this section, we will investigate the tracking performance of our algorithm.

### A. Estimation Error Equation

To obtain the compressed distributed estimation error equation, we first denote the compressed estimation error of single sensor $i$ as $\tilde{\zeta}_{t,i} \triangleq \zeta_t - \hat{\zeta}_{t,i}$ with $\zeta_t = M\theta_t$. Then, from (6) and (7), we have

$$
\begin{aligned}
&\tilde{\zeta}_{t+1,i} = \zeta_{t+1} - P_{t+1,i}\sum_{j\in\mathcal{N}_i} a_{ij}\bar{P}_{t+1,j}^{-1}\bar{\zeta}_{t+1,j} \\
&= P_{t+1,i}\sum_{j\in\mathcal{N}_i} a_{ij}\bar{P}_{t+1,j}^{-1}\zeta_{t+1} - P_{t+1,i}\sum_{j\in\mathcal{N}_i} a_{ij}\bar{P}_{t+1,j}^{-1}\bar{\zeta}_{t+1,j} \\
&= P_{t+1,i}\sum_{j\in\mathcal{N}_i} a_{ij}\bar{P}_{t+1,j}^{-1}(\zeta_{t+1} - \bar{\zeta}_{t+1,j}).
\end{aligned} \tag{9}
$$

Let $\Delta\zeta_t := M\Delta\theta_t$, then by equations (1)-(5), we can further obtain that

$$
\begin{aligned}
&\zeta_{t+1} - \bar{\zeta}_{t+1,j} \\
&= \zeta_t + \Delta\zeta_{t+1} - \hat{\zeta}_{t,j} - L_{t,j}(y_{t+1,j} - \phi_{t,j}^T\hat{\zeta}_{t,j}) \\
&= \big(I_d - L_{t,j}\phi_{t,j}^T\big)\tilde{\zeta}_{t,j} - L_{t,j}\bar{w}_{t+1,j} + \Delta\zeta_{t+1} \\
&= \lambda\bar{P}_{t+1,j}P_{t,j}^{-1}\tilde{\zeta}_{t,j} - L_{t,j}\bar{w}_{t+1,j} + \Delta\zeta_{t+1}.
\end{aligned} \tag{10}
$$

In order to centrally measure the estimation error of each sensor, we introduce the following series of notations.

$$
\begin{aligned}
&Y_t = \mathrm{col}\{y_{t,1}, \cdots, y_{t,n}\}, && W_t = \mathrm{col}\{w_{t,1}, \cdots, w_{t,n}\}, \\
&\Phi_t = \mathrm{diag}\{\phi_{t,1}, \cdots, \phi_{t,n}\}, && \overline{W}_t = \mathrm{col}\{\bar{w}_{t,1}, \cdots, \bar{w}_{t,n}\}, \\
&P_t = \mathrm{diag}\{P_{t,1}, \cdots, P_{t,n}\}, && \bar{P}_t = \mathrm{diag}\{\bar{P}_{t,1}, \cdots, \bar{P}_{t,n}\}, \\
&Z_t = \mathrm{col}\{\underbrace{\zeta_t, \cdots, \zeta_t}_{n}\}, && \Delta Z_t = \mathrm{col}\{\underbrace{\Delta\zeta_t, \cdots, \Delta\zeta_t}_{n}\}, \\
&\widetilde{Z}_t = \mathrm{col}\{\tilde{\zeta}_{t,1}, \cdots, \tilde{\zeta}_{t,n}\}, && L_t = \mathrm{diag}\{L_{t,1}, \cdots, L_{t,n}\},
\end{aligned}
$$

Hence, combining the equations (9) and (10), we can obtain the distributed augmented error formula:

$$
\widetilde{Z}_{t+1} = \lambda P_{t+1}\mathscr{A}P_t^{-1}\widetilde{Z}_t - P_{t+1}\mathscr{A}\bar{P}_{t+1}^{-1}(L_t\overline{W}_{t+1} + \Delta Z_t), \tag{11}
$$

where $\mathscr{A} = \mathcal{A}\otimes I_d$.

### B. Definitions and Assumptions

For the randomness of regression vector $\{\varphi_{t,i}, t \ge 0\}_{i=1}^n$, we first give some necessary definitions and assumptions on random matrix before going straight to the theorem discussion.

*Definition 3.1:* [8] A sequence of random matrices $\{A_t, t \ge 0\}$ defined on the basic probability space $(\Omega, \mathcal{F}, P)$ is deemed $L_p$-stable $(p > 0)$ if $\sup_{t\ge 0}\mathbb{E}(\|A_t\|^p) < \infty$. To quantify the stability of random variable $A_t$, its $L_p$-norm is defined as $\|A_t\|_{L_p} \triangleq (\mathbb{E}(\|A_t\|^p))^{\frac{1}{p}}$, where $\mathbb{E}(\cdot)$ denotes the expectation operator.

For convenience, we introduce the following class [8] for a scalar sequence $a = \{a_t, t \ge 0\}$:

$$
\begin{aligned}
S^0(\alpha) = \Big\{&a : a_t \in [0,1), \mathbb{E}\left(\prod_{j=k+1}^{t}(1 - a_j)\right) \le M\alpha^{t-k}, \\
&\forall t \ge k, \forall k \ge 0, \text{for some } M > 0\Big\}.
\end{aligned}
$$

*Assumption 3.1:* (Compressed Cooperative Excitation Condition) For the adapted sequences $\{\phi_{t,i}, \mathcal{F}_t, t \geq 0\}$, with $\mathcal{F}_t$ denoting a non-decreasing sequence of $\sigma$-algebras, there exists an integer $l > 0$ such that $\{\alpha_t\} \in S^0(\alpha)$ for some $\alpha \in (0,1)$, where $\alpha_t$ is defined by

$$\alpha_t \triangleq \lambda_{\min}\Big[\mathbb{E}\Big(\frac{1}{n(1+l)}\sum_{i=1}^{n}\sum_{k=tl+1}^{(t+1)l}\frac{\phi_{k,i}\phi_{k,i}^T}{1+\|\phi_{k,i}\|^2}\Big|\mathcal{F}_{tl}\Big)\Big]. \quad (12)$$

with $\mathbb{E}(\cdot|\cdot)$ being the conditional expectation operator.

*Remark 3.1:* The excitation conditions imposed on the regression vectors are widely used to guarantee the effective estimation of the identification algorithm. Since the regression vector $\{\varphi_{t,i}\}$ and the estimated signals are high-dimensional and sparse, excitation conditions directly on the original regression vector $\varphi_{t,i}$ may not be satisfied. So it is more reasonable to consider the excitation condition on compressed signals $\phi_{t,i}$, which is weaker than on the uncompressed one.

Assumption 3.1 serves as a fundamental underpinning for ensuring the stability of the distributed LMS algorithm (e.g., [13]). In fact, the assumption indicates that the sequence $\{\alpha_t\}$ has a "lower bound" that may change over time. For the special case that $\inf_t \alpha_t \geq \alpha_0$ with $\alpha_0 \in (0,1)$, it is clear that Assumption 3.1 can be satisfied.

*Assumption 3.2:* The graph $\mathcal{G}$ of the sensor network is undirected and connected.

*Remark 3.2:* The connectivity of the network topology is a very reasonable and common assumption. Under Assumption 3.2, we have $a_d \triangleq \min_{i,j\in\{1,2,\cdots,n\}}[\mathcal{A}^{D_\mathcal{G}}]_{ij} > 0$.

*Assumption 3.3:* The sensing matrix $M \in \mathbb{R}^{d\times m}$ satisfies the RIP with order $4s$. The $3s$- and $4s$-restricted isometry constants are designated as $\delta_{3s}$ and $\delta_{4s}$, and fulfill $\delta_{3s}+3\delta_{4s} < 2$.

*Remark 3.3:* Assumption 3.3 aligns with the condition in Lemma 2.1 to ensure a tightly controlled upper bound on the reconstruction error when recovering the original signal from its compressed estimate. There are many construction methods to make the sensing matrix $M$ satisfy Assumption 3.3, such as random matrices with i.i.d. entries, Fourier transform-based ensembles and general orthogonal measurement ensembles [4].

## C. Main Results

*Lemma 3.1:* [12] For $P_t$ generated by (5) and (6), under Assumptions 3.1 and 3.2, if the forgetting factor $\lambda$ satisfies $\alpha^{\frac{a_d^2}{32pml(4l+D_\mathcal{G}-1)}} < \lambda < 1$, then for any $p \geq 1$, $P_t$ is $L_p$-stable, where $\alpha, l$ are defined in Assumption 3.1.

Based on Lemma 3.1, we can get the upper bound of tracking error under some conditions on the estimated signal and its variation. The key step is to analyze the impact of sensing error $\varphi_{t,i}^T(I_m - M^TM)\theta_t$ brought by compressive sensing on stability. We use the properties of the sensing matrix $M$ to deal with the cumulative effect of sensing error.

*Theorem 3.1:* Consider model (2) and the error equation (11), under the conditions in Lemma 3.1, if for some $p \geq 1$, $\sigma_p(\delta_{4s}) := \sup_t \|\mu_t\|_{L_p} < \infty$, where $\mu_t = \frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}}\|\theta_t\|_{L_{6p}} + \|W_t\|_{L_{3p}} + \sqrt{1+\delta_{4s}}\|\Delta\theta_{t+1}\|_{L_{3p}}$. And for any $i \in \{1,\cdots,n\}$, $\sup_t \|\phi_{t,i}\|_{L_{6p}} < \infty$, then the compressed tracking error $\{\tilde{Z}_t\}$ is $L_p$-stable, i.e., there exists a constant $C_1$ such that

$$\limsup_{t\to\infty} \|\tilde{Z}_t\|_{L_p} \leq C_1\sigma_p(\delta_{4s}).$$

*Proof 3.1:* To simplify the expression, let the state transition matrix $\Psi(t,k)$ be recursively defined by

$$\Psi(t+1,k) = \lambda P_{t+1}\mathscr{A}P_t^{-1}\Psi(t,k), \Psi(k,k) = I_{dn}.$$

From equation (5) and the matrix inverse formula [14], it can be easily conducted that $\bar{P}_{t+1,i}^{-1} = \lambda P_{t,i}^{-1} + \phi_{t,i}\phi_{t,i}^T$. Combine the fact $\bar{P}_{t+1}^{-1} = \lambda P_t^{-1} + \Phi_t\Phi_t^T$ and the definition of $L_t$, it holds $\bar{P}_{t+1}^{-1}L_t = \Phi_t$. Then substituting it into (11), we have

$$\widetilde{Z}_{t+1} = \lambda P_{t+1}\mathscr{A}P_t^{-1}\widetilde{Z}_t - P_{t+1}\mathscr{A}(\Phi_t W_{t+1} + \bar{P}_{t+1}^{-1}\Delta Z_t).$$

Therefore, by Hölder inequality and the assumption that $\mathcal{A}$ is stochastic, we have

$$\begin{aligned}
&\|\widetilde{Z}_{t+1}\|_{L_p} = \|\Psi(t+1,0)\widetilde{Z}_0 \\
&\quad -\sum_{k=0}^{t}\Psi(t+1,k+1)(P_{k+1}\mathscr{A}(\Phi_k\overline{W}_{k+1}+\bar{P}_{k+1}^{-1}\Delta Z_k))\|_{L_p} \\
&\leq \|\lambda^{t+1}P_{t+1}\mathscr{A}^{t+1}P_0^{-1}\widetilde{Z}_0\|_{L_p} \\
&\quad + \Big\|\sum_{k=0}^{t}\lambda^{t-k}P_{t+1}\mathscr{A}^{t-k+1}(\Phi_k\overline{W}_{k+1} + \bar{P}_{k+1}^{-1}\Delta Z_k)\Big\|_{L_p} \\
&\leq \lambda^{t+1}\|P_{t+1}\|_{L_{3p}}\|P_0^{-1}\widetilde{Z}_0\|_{L_{3p}} \\
&\quad + \sum_{k=0}^{t}\lambda^{t-k}\|P_{t+1}\|_{L_{3p}}\|\Phi_k\|_{L_{3p}}\|\overline{W}_{t+1}\|_{L_{3p}} \\
&\quad + \sum_{k=0}^{t}\lambda^{t-k}\|P_{t+1}\|_{L_{3p}}\|\bar{P}_{k+1}^{-1}\|_{L_{3p}}\|\Delta Z_k\|_{L_{3p}}.
\end{aligned}$$

We should examine each of the three terms on the right-hand side of the above formula to determine their respective upper bounds. By Lemma 3.1, we have $\sup_{t\geq 0}\mathbb{E}(\|P_t\|)^p < \infty, \forall p$. Thus our task reduces to figure out the boundedness of $\{\overline{W}_{t+1}\}$, $\{\Delta Z_t\}$ and $\{P_t^{-1}\}$.

Since $\varphi_{t,i}$ is $3s$-sparse and $\theta_t$ is $s$-sparse, their corresponding non-zero entries can be indexed by $i_1, ..., i_{3s}$ and $j_1, ..., j_s$, respectively. We can construct reduced versions of these vectors by keeping only the components at these specific positions and denote the truncated forms of $\varphi_{t,i}$ and $\theta_t$ as $\varphi_{t,i}^{(4s)}$ and $\theta_{t,4s}$, which now contain the combined total of $4s$ non-zero elements from both vectors. In parallel, select and retain the column vectors corresponding to the $4s$ positions specified for $\varphi_{t,i}$ and $\theta_t$, discarding the rest. The resulting matrix after this dimensionality reduction is a $d \times 4s$-dimensional matrix, which is denoted as $M_{4s}$.

Under Assumption 3.3, it is posited that all eigenvalues of $M_{4s}^TM_{4s}$ lie within the interval $[1-\delta_{4s}, 1+\delta_{4s}]$, and it can be obtained that

$$\begin{aligned}&\|\varphi_{t,i}^T(I_m - M^TM)\theta_t\| = \|(\varphi_{t,i}^{(4s)})^T(I_{4s} - M_{4s}^TM_{4s})\theta_{t,4s}\|\\ \le&\|(\varphi_{t,i}^{(4s)})^T\|\|(1+\delta_{4s})I_{4s} - M_{4s}^TM_{4s}\|\|\theta_{t,4s}\|\\ &+\delta_{4s}\|(\varphi_{t,i}^{(4s)})^T\|\|\theta_{t,4s}\|\\ \le&2\delta_{4s}\|\varphi_{t,i}^{(4s)}\|\|\theta_{t,4s}\| + \delta_{4s}\|\varphi_{t,i}^{(4s)}\|\|\theta_{t,4s}\|\\ =&3\delta_{4s}\|\varphi_{t,i}\|\|\theta_t\| \le \frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}}\|M\varphi_{t,i}\|\|\theta_t\|\\ =&\frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}}\|\phi_{t,i}\|\|\theta_t\|. \end{aligned}\tag{13}$$

By the definition of $\bar{w}_{t+1,i}$ in (2) and inequality (13), it can be derived that

$$\begin{aligned}&\|\bar{w}_{t+1,i}\|_{L_{3p}} = \|\varphi_{t,i}^T(I_m - M^TM)\theta_t + w_{t+1,i}\|_{L_{3p}}\\ \le&\|\varphi_{t,i}^T(I_m - M^TM)\theta_t\|_{L_{3p}} + \|w_{t+1,i}\|_{L_{3p}}\\ \le&\frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}}\|\phi_{t,i}\|_{L_{6p}}\|\theta_t\|_{L_{6p}} + \|w_{t+1,i}\|_{L_{3p}}. \end{aligned}\tag{14}$$

Meanwhile, note that $\Delta\theta_{k+1}$ is $2s$-sparse, by the RIP of the matrix $M$, it can be obtained that

$$\|\Delta\zeta_{t+1}\|_{L_{3p}} = \|M\Delta\theta_{t+1}\|_{L_{3p}} \le \sqrt{1+\delta_{4s}}\|\Delta\theta_{t+1}\|_{L_{3p}}.\tag{15}$$

As for $P_{t+1}^{-1}$, based on the condition $\sup_t\|\phi_{t,i}\|_{L_{6p}} < \infty$, it follows that

$$\begin{aligned}&\|P_{t+1}^{-1}\|_{L_{3p}} \le \lambda\|P_t^{-1}\|_{L_{3p}} + \|\phi_{t,i}\|_{L_{6p}}^2 \le \cdots\\ \le&\lambda^{t+1}\|P_0^{-1}\|_{L_{3p}} + \sum_{k=0}^{t}\lambda^{t-k}\|\phi_{k,i}\|_{L_{6p}}^2 < \infty. \end{aligned}\tag{16}$$

Ultimately, the tracking error $\|\widetilde{Z}_{t+1}\|_{L_p}$ comes to the following conclusion by combining the conditions $\sup_t\|\phi_{t,i}\|_{L_{6p}} < \infty$, $\sigma_p(\delta_{4s}) < \infty$ with inequalities (14), (15) and (16):

$$\begin{aligned}&\|\tilde{Z}_{t+1,i}\|_{L_p}\\ \le&O(\lambda^{t+1}) + \sum_{k=1}^{t}\lambda^{t-k}\|P_{t+1}\|_{L_{3p}}\|\Phi_k\|_{L_{3p}}\|W_t\|_{L_{3p}}\\ &+\sum_{k=1}^{t}\lambda^{t-k}\|P_{t+1}\|_{L_{3p}}\|\Phi_k\|_{L_{3p}}\|\Phi_k\|_{L_{6p}}\frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}}\|\theta_k\|_{L_{6p}}\\ &+\sum_{k=1}^{t}\lambda^{t-k}\|P_{t+1}\|_{L_{3p}}\|P_{k+1}^{-1}\|_{L_{3p}}\sqrt{1+\delta_{4s}}\|\Delta\theta_{k+1}\|_{L_{3p}}\\ \le&O(\lambda^{t+1}) + C_1\sigma_p(\delta_{4s}), \end{aligned}$$

where $C_1$ is a positive constant depending on the upper bounds of $\{P_k\}, \{\Phi_k\}$ and $\{P_k^{-1}\}$. This completes the proof of the theorem.

Using Lemma 2.1, we finally determine the upper bound of the estimation error for the original high-dimensional signal.

*Theorem 3.2:* Under the same conditions in Theorem 3.1, the upper bound for the original estimation error is characterized thus:

$$\limsup_t\|\theta_t - \hat{\theta}_t\|_{L_p} \le C_sC_1\sigma_p(\delta_{4s}),$$

where $C_s$ is defined in Lemma 2.1.

*Proof 3.2:* Theorem 3.1 provides the following upper bound of compressed estimation error

$$\begin{aligned}\sup_t\|\tilde{\zeta}_{t+1,i}\|_{L_p} &= \sup_t\|\zeta_{t+1} - \hat{\zeta}_{t+1,i}\|_{L_p}\\ &= \sup_t\|M\theta_{t+1} - \hat{\zeta}_{t+1,i}\|_{L_p} \le C_1\sigma_p(\delta_{4s}). \end{aligned}$$

Based on Lemma 2.1 and $C = C_1\sigma_p(\delta_{4s})$ in equation (8) of Algorithm 1, it can be deduced that the recovered signal $\hat{\theta}_{t+1}$ obeys $\limsup_t\|\theta_{t+1} - \hat{\theta}_{t+1,i}\|_{L_p} \le C_sC_1\sigma_p(\delta_{4s})$, which completes the proof.

*Remark 3.4:* From Theorems 3.1 and 3.2, it is clear that the upper bound of the tracking error gets lower as the restricted isometry constant $\delta_{4s}$ decreases. In the special case of time-invariant parameters and noiseless systems, the estimation error can be very small as long as the restricted isometry constant $\delta_{4s}$ is small enough. Compared with [3], [15], we can see that our theoretical findings are derived without necessitating the independence or stationarity of the regression signal, which makes our conclusions more applicable to feedback systems.

## IV. Simulation Results

In this section, we present some simulation results to verify the efficacy of the compressed distributed FFLS algorithm based on high-dimensional sparse data.

We consider a parameter identification problem with a 2-sparse parameter vector $\theta_t$ having a total dimension $m = 80$ over a 12-sensor network. Only the first two components of the time-varying parameter vector $\theta_t$ are nonzero, whose variation at instant $t$ follows the normal distribution $\frac{1}{t^2}\mathscr{N}(0, 1^2)$. Assume the observation noises in (1) are independent Gaussian random variables with a distribution of $w_{t,i} \sim N(0, 0.2^2)$. Next, we generate the regressor vectors $\{\varphi_{t,i} \in \mathbb{R}^{80}, i = 1, \cdots, 12, t \ge 0\}$ by $\varphi_{t,i} = [0, \cdots, 0, \underbrace{1.1^t + \sum_{k=0}^{t-1}1.1^k\varepsilon_{t-k,i}}_{i^{th}}, 0, \cdots, 0]^T$, where the sequence $\{\varepsilon_{t,i}, t \ge 1, i = 1, \cdots, 12\}$ is independent and identically distributed in $N(0, 0.1^2)$. It was proved in [16] that when the sensing matrix is a Gaussian random matrix with zero mean and variance $1/d$, it satisfies the RIP with high probability under appropriate choices of dimensions. Thus, we configure the sensing matrix $M$ as a 10×80-dimensional Gaussian random matrix, whose every element $[M]_{ij} \sim N(0, 1/10)$.

From the setting of the regressor $\varphi_{t,i}$, we can verify that the compressed cooperative excitation condition (12), which corresponds to Assumption 3.1, is indeed satisfied. Besides, the connected sensor network is constructed by the Metropolis rule [17], thus Assumptions 3.1-3.3 can all be guaranteed.

We compare our algorithm with the compressed non-cooperative FFLS algorithm (i.e., the adjacency matrix $\mathcal{A} = I$) by using identical initial values. Besides, for solving the optimization problem represented by equation (8) in the compressed algorithms, we employ the OMP method described in

[18]. We run the simulation 200 times to ensure reliable and robust results.

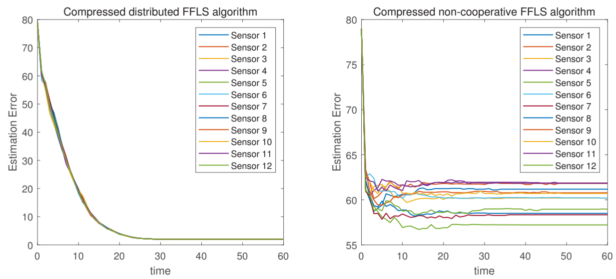


Fig. 1. Estimation errors of 12 sensors under compressed distributed FFLS algorithm (left) and the compressed non-cooperative FFLS algorithm (right).

The left subgraph of Fig. 1 shows the estimation performance of the proposed distributed algorithm, whose estimation errors of all sensors quickly decrease to around 0. The subgraph on the right shows the estimation performance of the sensors without information interaction, and none of the sensors can estimate the signal successfully, which verifies the importance of cooperation among sensors.

Next, we compare our algorithm with the compressed LMS algorithm and the uncompressed distributed FFLS algorithm. The estimation errors (computed over 100 times) of these three scenarios are depicted in Fig. 2, which demonstrates that our proposed algorithm has a much faster rate of decreasing in estimation error than the compressed distributed LMS algorithm, while the uncompressed distributed FFLS fails to achieve the estimation target.

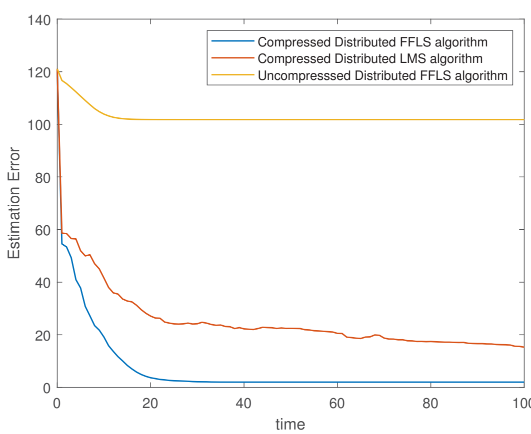


Fig. 2. Estimation errors of the compressed distributed FFLS algorithm, the compressed distributed LMS algorithm in [13], and the uncompressed distributed FFLS algorithm.

## V. Concluding Remarks

This paper developed a compressed distributed FFLS algorithm to estimate the time-varying sparse parameter. We established the stability of the proposed algorithm by combining compressive sensing theory with stochastic stability theory. The introduced compressed cooperative excitation condition guaranteed effective estimation without assuming the independency and stationarity of the regression signal. It is shown that the compressed distributed FFLS algorithm can realize accurate estimation of high-dimensional sparse signals, while the uncompressed one cannot complete the tracking task due to the lack of adequate excitation condition. Some interesting problems deserve to be further investigated, e.g., extending our main results to the time-delay scenario and utilizing the idea of feedback to overcome the influence of sensing error.

## References


[1] I. D. Schizas, G. Mateos, and G. B. Giannakis, "Distributed lms for consensus-based in-network adaptive processing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2365–2382, 2009.

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

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

[4] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[5] Y. Tang, V. Ramanathan, J. Zhang, and N. Li, "Communication-efficient distributed SGD with compressed sensing," *IEEE Control Systems Letters*, vol. 6, pp. 2054–2059, 2022.

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

[7] X. He, R. Song, and W.-P. Zhu, "Pilot allocation for distributed-compressed-sensing-based sparse channel estimation in MIMO-OFDM systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 2990–3004, 2016.

[8] S. Xie and L. Guo, "Analysis of compressed distributed adaptive filters," *Automatica*, vol. 112, p. 108707, 2020.

[9] D. Gan and Z. Liu, "Performance analysis of the compressed distributed least squares algorithm," *Systems & Control Letters*, vol. 164, p. 105228, 2022.

[10] ——, "Distributed sparse identification for stochastic dynamic systems under cooperative non-persistent excitation condition," *Automatica*, vol. 151, p. 110958, 2023.

[11] R. Johansson, "Identification of continuous-time models," *IEEE Transactions on Signal Processing*, vol. 42, no. 4, pp. 887–897, 1994.

[12] D. Gan, S. Xie, Z. Liu, and J. Lü, "Stability of FFLS-based diffusion adaptive filter under cooperative excitation condition," *IEEE Transactions on Automatic Control*, doi: 10.1109/TAC.2024.3388128, 2024.

[13] S. Xie and L. Guo, "Analysis of normalized least mean squares-based consensus adaptive filters under a general information condition," *SIAM Journal on Control and Optimization*, vol. 56, no. 5, pp. 3404–3431, 2018.

[14] G. Zielke, "Inversion of modified symmetric matrices," *Journal of the ACM*, vol. 15, no. 3, pp. 402–408, 1968.

[15] L. Li and D. Li, "A distributed estimation method over network based on compressed sensing," *International Journal of Distributed Sensor Networks*, vol. 15, no. 4, p. 1550147719841496, 2019.

[16] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, pp. 253–263, 2008.

[17] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *the Fourth International Symposium on Information Processing in Sensor Networks*, pp. 63–70, 2005.

[18] J. Tropp, A. C. Gilbert *et al.*, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.