Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 56-2 (2023) 10240–10245

Stability of Compressed Recursive Least Squares with Forgetting Factor Algorithm ⋆

Shuning Chen * **Die Gan** ** **Kexin Liu** **,*** **Jinhu Lü** **,***

* *Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100190, China. (e-mail: chenshuning@amss.ac.cn)*
** *Zhongguancun Laboratory, Beijing 100094, China. (e-mails: gandie@amss.ac.cn, skxliu@163.com, jhlu@iss.ac.cn)*
*** *School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China.*

Abstract: To identify the unknown sparse time-varying parameters of the stochastic dynamic system, we integrate compressive sensing theory with the traditional recursive least squares with forgetting factor (FFLS) algorithm, and propose a compressed adaptive filtering algorithm. Our algorithm is designed to first compress the original high-dimensional sparse regression vector by using the sensing matrix, and then apply the FFLS algorithm to estimate the compressed parameters. Subsequently, the original high-dimensional sparse parameters can be well recovered by a reconstruction technique. We introduce an excitation condition on the compressed stochastic regressors, under which the stability of the proposed algorithm (i.e., the upper bound of the estimation error) is established without assuming independence, stationarity or ergodicity of the system signals. The effectiveness of our theoretical results is demonstrated by a numerical example, which also shows that our proposed algorithm has better performance than both the compressed least mean squares algorithm and the uncompressed FFLS algorithm for tracking high-dimensional sparse parameters.

Keywords: Stochastic dynamic system, compressive sensing, time-varying sparse parameters, recursive least squares, forgetting factor, excitation condition

1. INTRODUCTION

Parameter estimation or adaptive filtering is one of the important problems in many fields including robot dynamics (Khosla and Kanade, 1985), data analysis (Gutschker, 2008) and energy management (Yang et al., 2020). As one of the most popular adaptive filter algorithms, the stability and performance analysis of recursive least squares with forgetting factor (FFLS) algorithm has been widely investigated since it has better performance than many other algorithms (Cioffi and Kailath, 1984; Kamali et al., 2011).

However, most theoretical results are established either by requiring the regression vector to be deterministic, or by adding assumptions of independence, stationarity or ergodicity to the signals, which can not be guaranteed in many cases, such as signals in feedback systems. A key mathematical difficulty in the identification problem of stochastic signals is to analyze the properties of random matrix product in the estimation error equation. To deal with it, Guo (1994) proposed a unified excitation condition to ensure the stability of three algorithms (standard Kalman filter, least mean squares (LMS) and FFLS algorithms) without the conditions of independence and stationarity. Later, Gan and Liu (2020) developed the stability of the Kalman filter with random coefficients under a general system model, and our work is also partly motivated by the framework of their paper.

On the other hand, with the development of network technology, sparsity becomes one of the important characteristics of high-dimensional signals, such as wireless cognitive radios, crystal structure and civil images. Knowing the sparsity of signals in advance can help to design a suitable algorithm to improve the estimation performance. Many adaptive sparse algorithms have been proposed in the existing literature, see Angelosante et al. (2010); Li and Li (2020). The majority of previous findings performed sparse signal estimation by inserting penalty terms in the cost function, which may lead to high computational complexity and slow estimation speed. As we consider additional techniques to enhance the performance of the FFLS algorithm, we note a recently developed theory: compressive sensing (CS) theory.

Compressive sensing theory is a new sampling theory that emerged at the start of the twenty-first century. To estimate sparse signals, it offers a reliable framework and requires fewer measurements. In numerous research areas, including image processing (Romberg, 2008), geological exploration (Herrmann et al., 2012), and medical imaging

⋆ This work was supported by the National Natural Science Foundation of China under Grants 61621003, 62103015, 62141604. Corresponding author: Die Gan.

10.1016/j.ifacol.2023.10.905

(Li et al., 2011), CS theory has received a great deal of attention. Zhao et al. (2021) presented a compressive sensing technique based on FFLS to learn high-dimensional sparse signals. Xie and Guo (2020) presented the compressed consensus normalized LMS algorithm to estimate unknown high-dimensional sparse signals in the network based on the CS method. The performance of a compressed distributed least squares algorithm was then studied by Gan and Liu (2022).

Due to the benefits of the FFLS algorithm in convergence speed and estimation accuracy, we integrate compressive sensing with the FFLS algorithm. A compressed FFLS algorithm with noisy observations and sparse regression vectors is proposed to track unknown time-varying sparse parameters of a stochastic dynamic system. In contrast to the sparse optimization frameworks in Yazdanpanah and Diniz (2017); Li and Li (2020), we employ the FFLS algorithm to estimate the unknown time-varying signals in a low-dimensional space using compressed regression vectors. And the estimates of the original high-dimensional sparse parameters can be well obtained using the signal reconstruction algorithm. Then we introduce a compressed excitation condition on the compressed regressors to guarantee the stability of the proposed algorithm, and the upper bound of the estimation error is established. Due to the sparsity of signals, our compressed excitation condition is weaker than the excitation conditions in Guo (1994); Zhao et al. (2021). We remark that compared with Babadi et al. (2010); Qin et al. (2021), our theoretical results are established without relying on the assumptions of the independence and stationarity of regression signals, which makes it applicable to the stochastic feedback system.

The remainder of this paper is organized as follows. We first introduce some notations and basic properties of compressive sensing theory in Section 2. The compressed FFLS algorithm is presented in Section 3, followed by its stability analysis in Section 4. A simulation example is given in Section 5 and some concluding remarks are provided in Section 6.

2. PRELIMINARIES

2.1 Notations

For an m-dimensional vector x, the p-norm of x is defined as $\|x\|_p = (\sum_{i=1}^m |x_i|^p)^{1/p}$ $(1 \le p < \infty)$ with x_i being the ith element of x. In particular, when $p = 2$, $\|x\|_2$ represents the Euclidean norm. If no subscript of the norm is specified in the following text, the Euclidean norm is used by default. We use $\|x\|_0$ to denote the number of non-zero elements in x. A vector $x \in \mathbb{R}^m$ is said to be s-sparse if it has at most s non-zero elements, where $s \ll m$. For an $m \times n$-dimensional real matrix A, $\|A\|$ denotes the matrix norm induced by the vector Euclidean norm, and can be equally calculated by $(\lambda_{\max}(A^T A))^{\frac{1}{2}}$, where $(\cdot)^T$ denotes the transpose operator and $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a matrix. Correspondingly, we denote the smallest eigenvalue of a matrix as $\lambda_{\min}(\cdot)$. For two positive scalar sequences $\{a_k, k \ge 0\}$ and $\{b_k, k \ge 0\}$, $a_k = O(b_k)$ means that there exists a constant C such that $a_k \le C b_k$ holds for every $k \ge 0$.

For matrices A, B, C and D with appropriate dimensions, if the relevant matrices are invertible, the following matrix inversion formula can be obtained (see e.g., Zielke (1968))

$$(A + BDC)^{-1} = A^{-1} - A^{-1}B(D^{-1} + CA^{-1}B)^{-1}CA^{-1}. \tag{1}$$

We introduce some definitions for the random matrix given by Guo (1994).

Definition 1. A random matrix sequence $\{A_t, t \ge 0\}$ defined on the basic probability space $(\Omega, \mathcal{F}, P)$ is called L_p-stable $(p > 0)$ if $\sup_{t \ge 0} \mathbb{E}\|A_t\|^p < \infty$. We define $\|A_t\|_{L_p} \triangleq (\mathbb{E}\|A_t\|^p)^{\frac{1}{p}}$ as the L_p-norm of the random matrix A_t.

Definition 2. A sequence of $n \times n$ random matrices $A = \{A_t, t \ge 0\}$ is called L_p-exponentially stable $(p \ge 0)$ with parameter $\lambda \in [0, 1)$, if it belongs to the following set

$$S_p(\lambda) = \Big\{A : \Big\| \prod_{j=k+1}^{t} (I_n - A_j) \Big\|_{L_p} \le M\lambda^{t-k},$$
$$\forall k \ge 0, \forall t \ge k, \text{ for some constant } M > 0\Big\}.$$

For convenience, we introduce the following subclass of $S_1(\lambda)$ for a scalar sequence $a = \{a_t, t \ge 0\}$.

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

2.2 Compressive sensing theory

The observation of an m-dimensional signal x_0 is defined as a d-dimensional vector y_0 with

$$y_0 = Mx_0 + \epsilon, \tag{2}$$

where $M \in \mathbb{R}^{d \times m}$ is the sensing matrix and $\epsilon \in \mathbb{R}^d$ is the measurement perturbation with a constant bound C, i.e., $\|\epsilon\| \le C$.

Generally, only down-sampled observation y_0 is transmitted to the sensor in order to save space and measurements. The goal of compressive sensing is to accurately recover the original signal x_0 from the observation y_0, which is challenging or perhaps impossible to achieve. However, if the signal x_0 is sparse, then its restoration is achievable. To deal with the challenge of reconstructing sparse signals, Candès and Tao (2005) established the concept of the restricted isometry property (RIP) of the sensing matrix in CS theory. They demonstrated that if M has the RIP and the noise is small, the sparse signal can be reconstructed with high accuracy.

Let $M \in \mathbb{R}^{d \times m}$ be the sensing matrix, and denote $Q \subseteq \{1, ..., m\}$ as a set of column indices. $\#(Q)$ is the number of the elements in set Q, and M_Q is a $d \times \#(Q)$-dimensional matrix, whose columns are the same as those in M that correspond to indexes in Q. Similarly, for a vector $x \in \mathbb{R}^m$, $x_Q \in \mathbb{R}^{\#(Q)}$ is defined as the column vector obtained after reserving the elements corresponding to the indices in Q.

Definition 3. (RIP). It is called that a matrix M satisfies the RIP with order s, if there exists a minimum constant $\delta_s \in [0, 1)$ such that the following inequality

$$(1 - \delta_s)\|x_Q\|^2 \le \|M_Q x_Q\|^2 \le (1 + \delta_s)\|x_Q\|^2 \tag{3}$$

holds for every column index set Q with $\#(Q) \leq s$ and every $x_Q \in \mathbb{R}^{\#(Q)}$. The constant δ_s is called the s-restricted isometry constant.

Remark 1. Condition (3) characterizes the ability of the matrix M to retain vector norms, and is actually equivalent to the property that all eigenvalues of the matrix $M_Q^T M_Q$ lie in $[1-\delta_s, 1+\delta_s]$. There are many available methods including deterministic and random methods to make M satisfy the RIP condition. For example, Baraniuk et al. (2008) proved that if the sensing matrix M is a Gaussian random matrix (i.e., the entries of the matrix are independently sampled from a Gauss distribution with zero mean and variance $1/d$) , the RIP holds with a high probability.

The following result shows the upper bound of the reconstruction error under an RIP condition:

Lemma 1. (Candès et al., 2006) Consider the recovery problem of the model (2). For any s-sparse signal x_0 and any perturbation ϵ with $\|\epsilon\| \leq C$, the recovered signal can be obtained from

$$x_0^* = \arg\min_x \{\|x\|_1 \quad \text{s.t. } \|y_0 - Mx\| \leq C\}. \tag{4}$$

If s satisfy $\delta_{3s} + 3\delta_{4s} < 2$, then the recovered signal x_0^* obeys $\|x_0 - x_0^*\| \leq C_s C$, where the positive constant C_s can be taken as $C_s \triangleq \frac{4}{\sqrt{3(1-\delta_{4s})} - \sqrt{1+\delta_{3s}}}$.

3. PROBLEM FORMULATION

3.1 System Model

Consider the identification problem of time-varying sparse parameters. At instant t, the scalar output y_t is subject to the following discrete-time stochastic regression model with the noise w_t,

$$y_t = \varphi_t^T \theta_t + w_t, \tag{5}$$

where $\varphi_t \in \mathbb{R}^m$ is a $3s$-sparse stochastic regression vector, and $\theta_t \in \mathbb{R}^m$ is a time-varying s-sparse parameter vector to be estimated ($s \ll m$). We denote the variation of the parameter process at time t as

$$\Delta\theta_t = \theta_t - \theta_{t-1}, t \geq 1. \tag{6}$$

It is clear that when $\Delta\theta_t \equiv 0$, (5) degenerates into the time-invariant case.

3.2 Compressed FFLS algorithm

We know that for the general stochastic regression vector φ_t without sparsity, the following recursive least squares algorithm with a constant forgetting factor α is commonly used to estimate the unknown parameter vector:

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \frac{P_t\varphi_t}{\alpha + \varphi_t^T P_t \varphi_t}(y_t - \varphi_t^T \hat{\theta}_t), \tag{7}$$

$$P_{t+1} = \frac{1}{\alpha}\Big(P_t - \frac{P_t\varphi_t\varphi_t^T P_t}{\alpha + \varphi_t^T P_t \varphi_t}\Big), t \geq 0. \tag{8}$$

Guo (1994) introduced the following excitation condition to analyze the stability of the FFLS algorithm for a large class of stochastic processes $\{\varphi_t\}$, i.e. $\{\lambda_t\} \in S^0(\lambda)$ where

$$\lambda_t \triangleq \lambda_{\min}\Big[\mathbb{E}\Big(\frac{1}{1+h}\sum_{k=th+1}^{(t+1)h} \frac{\varphi_k\varphi_k^T}{1+\|\varphi_k\|^2}\big|\mathcal{F}_{th}\Big)\Big], \tag{9}$$

with $\mathcal{F}_t$ being a sequence of non-decreasing σ-algebras and h being a constant integer. However, for the high-dimensional sparse regression vector φ_t, the excitation condition (9) is hard to be satisfied. Hence, the classic FFLS algorithm is not able to accurately track the high-dimensional sparse time-varying signal θ_t. Now, we propose the compressed FFLS algorithm by using the CS method to improve the tracking performance.

At instant t, we first utilize the sensing matrix $M \in \mathbb{R}^{d\times m}(s \leq d \ll m)$ to obtain the compressed d-dimensional regression vector $\phi_t = M\varphi_t$. Hence model (5) is converted to the following equation:

$$\begin{aligned} y_t &= \varphi_t^T\theta_t + \phi_t^T\zeta_t - \phi_t^T\zeta_t + w_t \\ &= \phi_t^T\zeta_t + \varphi_t^T(I_m - M^T M)\theta_t + w_t \\ &= \phi_t^T\zeta_t + \bar{w}_t, \end{aligned} \tag{10}$$

where $\bar{w}_t = \varphi_t^T(I_m - M^T M)\theta_t + w_t$ can be regarded as the extended "noise" term, including the compression error and the measurement noise. After that, the FFLS algorithm is used to estimate the d-dimensional parameters $\zeta_t = M\theta_t$ in the compressed space based on the data $\{\phi_t, y_t\}$. Finally, we recover the original estimates $\hat{\theta}_t$ for the unknown parameter vector θ_t by solving the convex optimization problem (4). The details of the compressed distributed FFLS algorithm are stated in Algorithm 1.

Algorithm 1 Compressed FFLS algorithm

Input:$\{\varphi_t, y_t\}, t = 0, 1, 2, \cdots$

Output:$\{\hat{\theta}_{t+1}\}, t = 0, 1, 2, \cdots$

Initialization: Start with an initial estimation $\hat{\zeta}_0 \in \mathbb{R}^d$ and an initial positive definite matrix $P_0 \in \mathbb{R}^{d\times d}$

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

Step 1. Compression: $\phi_t = M\varphi_t$.

Step 2. Estimation in a lower dimension:

$$\hat{\zeta}_{t+1} = \hat{\zeta}_t + \frac{P_t\phi_t}{\alpha + \phi_t^T P_t\phi_t}(y_t - \phi_t^T\hat{\zeta}_t), \tag{11}$$

$$P_{t+1} = \frac{1}{\alpha}\big(P_t - \frac{P_t\phi_t\phi_t^T P_t}{\alpha + \phi_t^T P_t\phi_t}\big). \tag{12}$$

Step 3. Reconstruction:

$$\hat{\theta}_{t+1} = \arg\min_{\theta\in\mathbb{B}} \|\theta\|_1, \tag{13}$$

where $\mathbb{B} = \Big\{\theta \in \mathbb{R}^m \Big| \|M\theta - \hat{\zeta}_{t+1}\| \leq C\Big\}$.

Remark 2. With regard to the selection of the optimization target for the sparse signal reconstruction problem (13), the most sparse solution can be obtained by optimizing $\|\theta\|_0$, but it is nonconvex and NP-hard in general. Therefore, we consider the 1-norm $\|\theta\|_1$ which is easy to be solved by using orthogonal matching pursuit (OMP) and compressive sampling matching pursuit algorithms (Tropp et al., 2007). Additionally, the constant C in Step 3 of Algorithm 1 should be chosen appropriately. In fact, it can be taken as the upper bound of the compressed estimation error (see $c\sigma_{3p}$ in Theorem 1).

4. STABILITY ANALYSIS

Before investigating the tracking performance of our compressed algorithm, we first introduce some necessary assumptions.

Assumption 1. (Compressed Excitation Condition): For the adapted sequences $\{\phi_t, \mathcal{F}_t, t \geq 0\}$, where $\mathcal{F}_t$ is a sequence of non-decreasing σ-algebras, there exists an integer $h > 0$ such that $\{\lambda'_t\} \in S^0(\lambda)$ for some $\lambda \in (0,1)$, where λ'_t is defined by

$$\lambda'_t \triangleq \lambda_{\min}\Big[\mathbb{E}\Big(\frac{1}{1+h}\sum_{k=th+1}^{(t+1)h}\frac{\phi_k\phi_k^T}{1+\|\phi_k\|^2}\big|\mathcal{F}_{th}\Big)\Big]. \quad (14)$$

Remark 3. Consider a special scenario: the sparse parameter vector θ_t and the sparse regression vector φ_t do not have the same nonzero indexes, then the output y_t is equal to the noise w_t, which does not contain any useful information about θ_t. As a result, the estimated value is manifestly impossible to approach the true value of system parameters. In the meantime, condition (9) illustrated in the last section may not be satisfied for the sparse regression vector. Assumption 1 is introduced by replacing the original high-dimensional φ_t with the compressed signal ϕ_t, which is weaker than the condition (9).

Assumption 2. The sensing matrix $M \in \mathbb{R}^{d\times m}$ satisfies the RIP with order $4s$ where the $3s$- and $4s$-restricted isometry constants denoted as δ_{3s} and δ_{4s} satisfy $\delta_{3s} + 3\delta_{4s} < 2$.

Remark 4. Assumption 2 is consistent with the condition in Lemma 1 to guarantee a small error upper bound upon the recovery of the compressed estimate to the original signal.

Lemma 2. (Guo, 1994) Let P_k be generated by (12) with forgetting factor $\alpha \in (0,1)$. If Assumption 1 holds, then for any $p \geq 0$,

$$\sup_{k\geq 0}\mathbb{E}\|P_k\|^p < \infty,$$

provided that α satisfies $\lambda^{[16hd(2h-1)p]^{-1}} < \alpha < 1$, where λ and h are given by Assumption 1, and d is the dimension of $\{\phi_t\}$.

In order to analyze the stability of the compressed FFLS algorithm (Algorithm 1), we denote the compressed estimation error as $\tilde{\zeta}_t := \zeta_t - \hat{\zeta}_t$, and the variation of the compressed signal as $\Delta\zeta_t = M\Delta\theta_t$, then from (10) and (11), we have

$$\begin{aligned}&\tilde{\zeta}_{t+1} = \zeta_{t+1} - \hat{\zeta}_{t+1}\\ =&\zeta_t + \Delta\zeta_{t+1} - \Big(\hat{\zeta}_t + \frac{P_t\phi_t}{\alpha+\phi_t^TP_t\phi_t}(y_t - \phi_t^T\hat{\zeta}_t)\Big)\\ =&(I_d - \frac{P_t\phi_t\phi_t^T}{\alpha+\phi_t^TP_t\phi_t})\tilde{\zeta}_t - \frac{P_t\phi_t}{\alpha+\phi_t^TP_t\phi_t}\bar{w}_t + \Delta\zeta_{t+1}. \quad (15)\end{aligned}$$

Theorem 1. Consider the time-varying model (5) with Algorithm 1. Under Assumptions 1 and 2 , if the following conditions hold:

(1) $\sigma_{3p} := \sup_t \mu_t < \infty$, where $\mu_t = \frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}}\|\theta_t\|_{L_{6p}} + \|w_t\|_{L_{3p}} + \sqrt{1+\delta_{4s}}\|\Delta\theta_{t+1}\|_{L_{3p}}$;
(2) $\sup_t \|\phi_t\|_{L_{6p}} < \infty$ for some $p \geq 1$;
(3) the forgetting factor α satisfies $\lambda^{[48hd(2h-1)p]^{-1}} < \alpha < 1$, where d is the dimension of $\{\phi_t\}$, λ and h are given by Assumption 1.

Then the compressed tracking error $\{\tilde{\zeta}_t, t \geq 1\}$ is L_p-stable, i.e., there exists a constant c such that

$$\limsup_t \|\tilde{\zeta}_t\|_{L_p} \leq c\sigma_{3p}.$$

Proof. For the simplicity of the proof, denote $L_t = \frac{P_t\phi_t}{\alpha+\phi_t^TP_t\phi_t}$ and it can be easily obtained from (12) that

$$(I_d - L_t\phi_t^T) = \alpha P_{t+1}P_t^{-1}.$$

Then by (15), we have

$$\begin{aligned}\tilde{\zeta}_{t+1} &= (I_d - L_t\phi_t^T)\tilde{\zeta}_t - L_t\bar{w}_t + \Delta\zeta_{t+1}\\ &= \alpha P_{t+1}P_t^{-1}\tilde{\zeta}_t - L_t\bar{w}_t + \Delta\zeta_{t+1}\\ &= \alpha^2 P_{t+1}P_{t-1}^{-1}\tilde{\zeta}_{t-1} - \alpha P_{t+1}P_t^{-1}L_{t-1}\bar{w}_{t-1}\\ &\quad + \alpha P_{t+1}P_t^{-1}\Delta\zeta_t - L_t\bar{w}_t + \Delta\zeta_{t+1} = \cdots\\ &= \alpha^{t+1}P_{t+1}P_0^{-1}\tilde{\zeta}_0 - \sum_{k=0}^t \alpha^{t-k}P_{t+1}P_{k+1}^{-1}L_k\bar{w}_k\\ &\quad + \sum_{k=0}^t \alpha^{t-k}P_{t+1}P_{k+1}^{-1}\Delta\zeta_{k+1}. \quad (16)\end{aligned}$$

On the other hand, multiply both sides of (12) by ϕ_t, and we can see that $P_{t+1}^{-1}L_t$ can be equally replaced by ϕ_t. Therefore, (16) can be further transformed into

$$\begin{aligned}\tilde{\zeta}_{t+1} =&\alpha^{t+1}P_{t+1}P_0^{-1}\tilde{\zeta}_0 - \sum_{k=0}^t \alpha^{t-k}P_{t+1}\phi_k\bar{w}_k\\ &+ \sum_{k=0}^t \alpha^{t-k}P_{t+1}P_{k+1}^{-1}\Delta\zeta_{k+1}.\end{aligned}$$

Thus for the L_p-norm of the tracking error, we have

$$\begin{aligned}\|\tilde{\zeta}_{t+1}\|_{L_p} \leq&\alpha^{t+1}\|P_{t+1}P_0^{-1}\tilde{\zeta}_0\|_{L_p} + \sum_{k=0}^t \alpha^{t-k}\|P_{t+1}\phi_k\bar{w}_k\|_{L_p}\\ &+ \sum_{k=0}^t \alpha^{t-k}\|P_{t+1}P_{k+1}^{-1}\Delta\zeta_{k+1}\|_{L_p}. \quad (17)\end{aligned}$$

The upper bounds of the three terms on the right-hand side of (17) need to be analyzed one by one. By Lemma 2, we have $\sup_{k\geq 0}\mathbb{E}\|P_k\|^p < \infty$. Hence we just need to think about the L_p-stability of $\bar{w}_k$, $\Delta\zeta_{k+1}$ and P_k^{-1}.

Since φ_k is $3s$-sparse and θ_k is s-sparse, the positions of their non-zero elements can be corresponding denoted as $i_1, ..., i_{3s}$ and $j_1, ..., j_s$, respectively. Retain all the aforementioned $4s$ non-zero elements in φ_k and θ_k, remove the other positions, and denote the new low-dimension one as $\varphi_{k,4s}$ and $\theta_{k,4s}$. Similarly, retain the corresponding $4s$ column vectors of matrix M, remove the column vectors in other positions, and denote the $d \times 4s$-dimensional matrix as M_{4s}. When φ_k and θ_k share the same indexes of nonzero elements, the analysis for such case is similar. Due to the similarity of the analysis process, only $4s$ non-zero elements are taken into consideration.

According to Assumption 2, all eigenvalues of $M_{4s}^TM_{4s}$ are within the interval $[1-\delta_{4s}, 1+\delta_{4s}]$, and it can be obtained that

$$\begin{aligned}&\|\varphi_k^T[I_m - M^TM]\theta_k\|\\ =&\|\varphi_{k,4s}^T[I_{4s} - M_{4s}^TM_{4s}]\theta_{k,4s}\|\\ \leq&\|\varphi_{k,4s}^T[(1+\delta_{4s})I_{4s} - M_{4s}^TM_{4s}]\theta_{k,4s}\| + \delta_{4s}\|\varphi_{k,4s}^T\theta_{k,4s}\|\\ \leq&2\delta_{4s}\|\varphi_{k,4s}\|\|\theta_{k,4s}\| + \delta_{4s}\|\varphi_{k,4s}\|\|\theta_{k,4s}\|\\ \leq&\frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}}\|M\varphi_k\|\|\theta_k\| = \frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}}\|\phi_k\|\|\theta_k\|. \quad (18)\end{aligned}$$

Combined the definition of $\bar{w}_k$ with (18), it can be derived that

$$\begin{aligned}\|\bar{w}_k\|_{L_{3p}} &= \|\varphi_k^T(I_m - M^T M)\theta_k + w_k\|_{L_{3p}} \\ &\leq \|\varphi_k^T(I_m - M^T M)\theta_k\|_{L_{3p}} + \|w_k\|_{L_{3p}} \\ &\leq \frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}}\|\phi_k\|_{L_{6p}}\|\theta_k\|_{L_{6p}} + \|w_k\|_{L_{3p}}. \qquad (19)\end{aligned}$$

Meanwhile, note that $\Delta\theta_{k+1}$ is $2s$-sparse, with the RIP of the matrix M, we can obtain

$$\|\Delta\zeta_{k+1}\|_{L_{3p}} \leq \sqrt{1+\delta_{4s}}\|\Delta\theta_{k+1}\|_{L_{3p}}. \qquad (20)$$

As for P_{t+1}^{-1}, by the matrix inversion formula (1), it can be derived from (12) that $P_{t+1}^{-1} = \alpha P_t^{-1} + \phi_t\phi_t^T$.

According to the condition $\sup_t \|\phi_t\|_{L_{6p}} < \infty$, it can be known that

$$\begin{aligned}\|P_{t+1}^{-1}\|_{L_{3p}} &\leq \alpha\|P_t^{-1}\|_{L_{3p}} + \|\phi_t\|_{L_{6p}}^2 \leq \cdots \\ &\leq \alpha^{t+1}\|P_0^{-1}\|_{L_{3p}} + \sum_{k=0}^{t}\alpha^{t-k}\|\phi_k\|_{L_{6p}}^2 < \infty. \qquad (21)\end{aligned}$$

Finally, combining the conditions $\sup_t \|\phi_t\|_{L_{6p}} < \infty$, $\sigma_{3p} < \infty$ with inequalities (19),(20) and (21), the tracking error (17) ultimately comes to the conclusion:

$$\begin{aligned}\|\tilde{\zeta}_{t+1}\|_{L_p} &\leq O(\alpha^{t+1}) + \sum_{k=0}^{t}\alpha^{t-k}\|P_{t+1}\|_{L_{3p}}\|\phi_k\|_{L_{3p}}\|w_k\|_{L_{3p}} \\ &+ \sum_{k=0}^{t}\alpha^{t-k}\|P_{t+1}\|_{L_{3p}}\|\phi_k\|_{L_{3p}}\|\phi_k\|_{L_{6p}}\frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}}\|\theta_k\|_{L_{6p}} \\ &+ \sum_{k=0}^{t}\alpha^{t-k}\|P_{t+1}\|_{L_{3p}}\|P_{k+1}^{-1}\|_{L_{3p}}\sqrt{1+\delta_{4s}}\|\Delta\theta_{k+1}\|_{L_{3p}} \\ &\leq O(\alpha^{t+1}) + c\sigma_{3p},\end{aligned}$$

where c is a positive constant related to the supremum of L_p-norm of $\{P_k\}, \{\phi_k\}$ and $\{P_k^{-1}\}$. This completes the proof of the theorem. ■

We finally establish the upper bound of the estimation error for the uncompressed signal by using Lemma 1.

Theorem 2. Under the conditions in Theorem 1, the upper bound for the original signal estimation error is as follows:

$$\limsup_t \|\theta_t - \hat{\theta}_t\|_{L_p} \leq C_s c\sigma_{3p},$$

where C_s is defined in Lemma 1.

Proof. From Theorem 1, we can know that the upper bound of compressed estimation error

$$\begin{aligned}\sup_t \|\tilde{\zeta}_{t+1}\|_{L_p} &= \sup_t \|\zeta_{t+1} - \hat{\zeta}_{t+1}\|_{L_p} \\ &= \sup_t \|M\theta_{t+1} - \hat{\zeta}_{t+1}\|_{L_p} \leq c\sigma_{3p}.\end{aligned}$$

Let $C = c\sigma_{3p}$ in equation (13) of Algorithm 1, then from Lemma 1, we can obtain that the recovered signal $\hat{\theta}_{t+1}$ obeys:

$$\limsup_t \|\theta_{t+1} - \hat{\theta}_{t+1}\|_{L_p} \leq C_s c\sigma_{3p},$$

which completes the proof. ■

The following corollary depicts the probability of the estimated error of the original high dimensional signal falling within a certain range.

Corollary 1. Consider the model (10) and the estimation error (16), under the conditions in Theorem 1. Then for any given constant $\varepsilon > 0$ and $\gamma \in (0,1)$, there exists time instant T_ε such that for any $t \geq T_\varepsilon$,

$$P\{\|\theta_t - \hat{\theta}_t\| \leq \eta(C_s c\sigma_{3p} + \varepsilon)^{1-\gamma}\} \geq 1 - \frac{(C_s c\sigma_{3p} + \varepsilon)^\gamma}{\eta}$$

holds with $\eta = \max\{1, 3(C_s c\sigma_{3p} + \varepsilon)^\gamma\}$.

Proof. Consider the special case of $p = 1$ in Theorem 2, and we have $\limsup_t \mathbb{E}\|\theta_t - \hat{\theta}_t\| \leq C_s c\sigma_{3p}$. That is to say, $\forall\varepsilon > 0, \exists T_\varepsilon \in \mathbb{R}$ s.t. $\forall t \geq T_\varepsilon, \mathbb{E}\|\theta_t - \hat{\theta}_t\| \leq C_s c\sigma_{3p} + \varepsilon$. Then by the Markov inequality, we have for any $t \geq T_\varepsilon$,

$$\begin{aligned}&P\{\|\theta_t - \hat{\theta}_t\| \geq \eta(C_s c\sigma_{3p} + \varepsilon)^{1-\gamma}\} \\ \leq &\frac{\mathbb{E}[\|\theta_t - \hat{\theta}_t\|]}{\eta(C_s c\sigma_{3p} + \varepsilon)^{1-\gamma}} \leq \frac{(C_s c\sigma_{3p} + \varepsilon)^\gamma}{\eta} \leq \frac{1}{3}.\end{aligned}$$

This completes the proof. ■

Remark 5. From Theorems 1 and 2, we know that when the restricted isometry constant δ_{4s} is small, σ_{3p} becomes small. Moreover, if σ_{3p} is close to zero, we have $\eta = 1$. Thus, the tracking error $\|\theta_t - \hat{\theta}_t\|_{L_p}$ will be stable in a small neighborhood of zero with a large probability. We remark that compared with Babadi et al. (2010); Qin et al. (2021), our theoretical results for the stability analysis of the compressed FFLS algorithm in this paper are derived without using any independence or stationarity assumptions on the regression vectors, which makes our results more suitable for practical feedback systems.

5. SIMULATION

Consider a 2-sparse parameter identification problem with a total dimension $m = 50$, and assume that only the first two indexes of unknown time-varying parameter vector θ_t are non-zero. To simplify the statement, denote $N(\mu, \sigma^2, m, n)$ to represent an $m \times n$-dimensional matrix in which every element follows the normal distribution with mean value μ and standard deviation σ. Let the first two indexes of $\Delta\theta_t \sim \frac{1}{t^2}N(0, 0.01^2, 2, 1)$. Assume the observation noises are independently distributed with $w_t \sim N(0, 0.5^2)$. Let the regressors be 6-sparse and the last six elements in $\varphi_t \in \mathbb{R}^{50}$ are generated by $x_{t+1} = Ax_t + \xi_t, x_0 \sim N(0, 1, 6, 1)$, where the matrix $A \in \mathbb{R}^{6\times 6}$ is a diagonal matrix with the diagonal elements equal to 4/5 and $\xi_t \sim N(0, 1, 6, 1)$. In the compressed FFLS algorithm, the sensing matrix M is selected as a 5×50-dimensional random matrix and $M \sim N(0, 1/5, 5, 50)$.

From the setting of signals θ_t and regressor φ_t, we can see that their non-zero elements have no common position, and thus it makes the uncompressed FFLS algorithm invalid. Moreover, we can verify that condition (9) is not satisfied while the compressed excitation condition (14) (i.e., Assumption 1) holds.

We compare our algorithm (Algorithm 1) with the uncompressed FFLS algorithm (i.e., (7)-(8)) and the compressed LMS algorithm, using the same initial values. Besides, we utilize the OMP algorithm to solve the optimization problem (13). To avoid accidents, we repeat the simulation 200 times. Then the tracking errors of these three algorithms are shown in Fig. 1, from which we can see that the compressed FFLS algorithm has better performance than the compressed LMS algorithm, while the uncompressed FFLS algorithm fails to track the sparse unknown parameters.

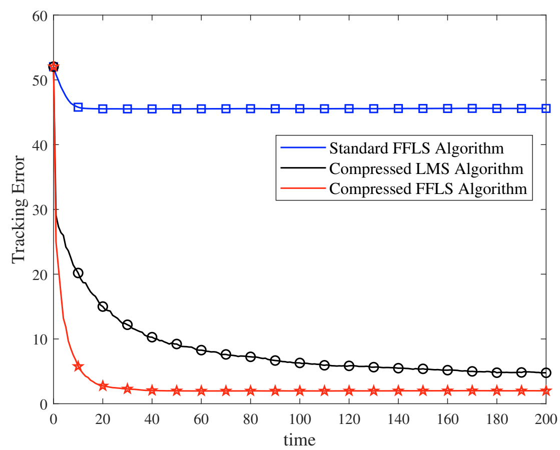

Fig. 1. Tracking errors of three different algorithms

6. CONCLUDING REMARKS

In this paper, a compressed recursive least squares with forgetting factor algorithm is proposed to track high-dimensional time-varying sparse signals by using compressive sensing methods. A compressed excitation condition is proposed to ensure that the compressed regressors can effectively achieve the estimation task, and the upper bound of the tracking error is established. Our algorithm can estimate the unknown high-dimensional sparse signal, while the traditional FFLS algorithm (cf. Guo (1994)) cannot due to the sparsity of the regressors. Many interesting problems deserve to be further studied, such as developing distributed compressed algorithms and optimizing the sensing matrix.

REFERENCES

Angelosante, D., Bazerque, J.A., and 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., and Tarokh, V. (2010). SPARLS: The sparse RLS algorithm. *IEEE Transactions on Signal Processing*, 58(8), 4013–4025.

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

Candès, E.J., Romberg, J.K., and Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8), 1207–1223.

Candès, E.J. and Tao, T. (2005). Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12), 4203–4215.

Cioffi, J. and Kailath, T. (1984). Fast, recursive-least-squares transversal filters for adaptive filtering. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(2), 304–337.

Gan, D. and Liu, Z. (2020). On the stability of Kalman filter with random coefficients. *IFAC-PapersOnLine*, 53(2), 2397–2402.

Gan, D. and Liu, Z. (2022). Performance analysis of the compressed distributed least squares algorithm. *Systems & Control Letters*, 164, 105228.

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

Gutschker, O. (2008). Parameter identification with the software package LORD. *Building and Environment*, 43(2), 163–169.

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

Kamali, C., Pashilkar, A., and Raol, J. (2011). Evaluation of recursive least squares algorithm for parameter estimation in aircraft real time applications. *Aerospace Science and Technology*, 15(3), 165–174.

Khosla, P.K. and Kanade, T. (1985). Parameter identification of robot dynamics. In *1985 24th IEEE Conference on Decision and Control*, 1754–1760. Fort Lauderdale, FL, USA.

Li, D., Zhang, Q., Shen, H., Shi, J., and 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*, 5350–5353. Hohhot, China.

Li, J. and Li, X. (2020). Online sparse identification for regression models. *Systems & Control Letters*, 141, 104710.

Qin, Z., Tao, J., and Xia, Y. (2021). A proportionate recursive least squares algorithm and its performance analysis. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 68(1), 506–510.

Romberg, J. (2008). Imaging via compressive sampling: Introduction to compressive sampling and recovery via convex programming. *IEEE Signal Processing Magazine*, 25(2), 14–20.

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

Xie, S. and Guo, L. (2020). Analysis of compressed distributed adaptive filters. *Automatica*, 112, 108707.

Yang, B., Wang, J., Zhang, X., Yu, T., Yao, W., Shu, H., Zeng, F., and Sun, L. (2020). Comprehensive overview of meta-heuristic algorithm applications on PV cell parameter identification. *Energy Conversion and Management*, 208, 112595.

Yazdanpanah, H. and Diniz, P.S.R. (2017). Recursive Least-Squares algorithms for sparse system modeling. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing*, 3879–3883. New Orleans, LA, USA.

Zhao, P., Wang, X., Xie, S., Guo, L., and Zhou, Z.H. (2021). Distribution-free one-pass learning. *IEEE Transactions on Knowledge and Data Engineering*, 33(3), 951–963.

Zielke, G. (1968). Inversion of modified symmetric matrices. *Journal of the ACM*, 15(3), 402–408.