SCIENCE CHINA
Technological Sciences

• Article •

February 2024 Vol. 67 No. 2: 380–394
https://doi.org/10.1007/s11431-023-2514-6

Stability and performance analysis of the compressed Kalman filter algorithm for sparse stochastic systems

LI RongJiang[1,2], GAN Die[3], XIE SiYu[4] & LÜ JinHu[3,5*]

[1] *Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China;*
[2] *School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China;*
[3] *Zhongguancun Laboratory, Beijing 100049, China;*
[4] *School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu 611731, China;*
[5] *School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China*

Received May 22, 2023; accepted September 20, 2023; published online December 18, 2023

This paper considers the problem of estimating unknown sparse time-varying signals for stochastic dynamic systems. To deal with the challenges of extensive sparsity, we resort to the compressed sensing method and propose a compressed Kalman filter (KF) algorithm. Our algorithm first compresses the original high-dimensional sparse regression vector via the sensing matrix and then obtains a KF estimate in the compressed low-dimensional space. Subsequently, the original high-dimensional sparse signals can be well recovered by a reconstruction technique. To ensure stability and establish upper bounds on the estimation errors, we introduce a compressed excitation condition without imposing independence or stationarity on the system signal, and therefore suitable for feedback systems. We further present the performance of the compressed KF algorithm. Specifically, we show that the mean square compressed tracking error matrix can be approximately calculated by a linear deterministic difference matrix equation, which can be readily evaluated, analyzed, and optimized. Finally, a numerical example demonstrates that our algorithm outperforms the standard uncompressed KF algorithm and other compressed algorithms for estimating high-dimensional sparse signals.

sparse signal, compressed sensing, Kalman filter algorithm, compressed excitation condition, stochastic stability, tracking performance

Citation: Li R J, Gan D, Xie S Y, et al. Stability and performance analysis of the compressed Kalman filter algorithm for sparse stochastic systems. Sci China Tech Sci, 2024, 67: 380–394, https://doi.org/10.1007/s11431-023-2514-6

1 Introduction

Parameter estimation or filtering problems have been extensively studied due to their broad applications, including signal processing, target tracking, navigation, and adaptive control [1–5]. Several efficient adaptive algorithms have been proposed to fulfill the desired estimation task, such as the least mean squares (LMS) algorithm, the recursive least squares with forgetting factor (FFLS) algorithm, and, in particular, the Kalman filter (KF) algorithm [2, 6, 7]. The last one triggers extensive research interest since the KF algorithm can produce an optimal state estimation in the minimum mean square error sense if noise processes obey Gaussian distribution.

Over the past few years, considerable progress has been made in estimating unknown parameters for a linear regression model or state space model [5, 8–16]; among them, refs. [14–16] investigated secure state estimation problems for linear systems with deterministic observation matrices. Most of the existing literature (e.g., refs. [10–16]) focuses on the case where regression vectors or observation matrices are deterministic while paying insufficient attention to the

*Corresponding author (email: jhlu@iss.ac.cn)

stochastic one. In fact, stochastic dynamic systems are ubiquitous and include many classical systems, such as the Hammerstein system, the autoregressive model with exogenous inputs (ARX), and the nonlinear ARX system [17, 18]. The corresponding theoretical investigation is also of particular significance.

In the context of stability and performance evaluation for stochastic dynamic systems, analyzing the product of random matrices poses a significant theoretical challenge. Most of the existing literature relies on assumptions of signal independence and stationarity within the system [19–21]. Such statistical assumptions are too stringent and simplistic to meet practical requirements under certain circumstances. For instance, the stochastic regression vectors are usually correlated in a general feedback control system since regression vectors contain input and output signals, and current inputs rely on previous inputs and outputs. Note that a stochastic excitation condition was proposed in ref. [22], relaxing independence and stationarity assumptions. Under this condition, the stability of three adaptive filtering algorithms (LMS, KF, FFLS) was eventually developed, which were widely utilized for the case of general stochastic signals. Nevertheless, challenges remain when practical systems are high-dimensional but sparse since the above excitation condition is difficult or even impossible to satisfy, and the traditional filtering algorithms may become invalid. This status quo motivates us to weaken the excitation and improve the estimation performance.

There have been attempts to consider sparsity as a priori to improve the tracking performance of unknown signals [3,23]. One technique for sparse system identification is to add a regularization term into the cost function, motivated by the fact that signal tracking can be recast as the optimization of model prediction error based on the input-output data. Ref. [24] proposed an adaptive filter based on the LMS algorithm and incorporated a ℓ_1-norm penalty of the coefficients into its cost function to accelerate convergence and reduce the mean squares error. Ref. [19] presented the ℓ_1-norm regularized versions of the recursive least squares algorithm, assuming that regression vectors are stationary and ergodic, and generated consistent estimates for linear stochastic systems with sparse parameters. Ref. [25] proposed an online alternating minimization (OAM) algorithm to track sparse signals and established the strong consistency of the proposed algorithm. In addition to the ℓ_1-norm penalty in refs. [19, 24, 25], there are several popular alternatives to induce sparse solutions, including ℓ_γ $(0 < \gamma < 1)$ [26] and smoothed shear absolute deviation [27]. It is noteworthy that the above literature takes the sparsity of unknown parameters into account, while the case where regression vectors are sparse also deserves attention.

The compressed sensing (CS) theory, as another technique for estimating sparse signals [28, 29], is beneficial to deal with possible degeneration of covariance matrices of regression vectors (i.e., insufficient excitation), especially when the regression vectors are high-dimensional but sparse. It guarantees the recovery of high-dimensional signals from fewer observations than the Nyquist/Shannon sampling principle considers necessary. We remark that refs. [30–32] incorporated the CS technique into the least squares, the LMS, and the FFLS algorithms, respectively, followed by relatively elegant theoretical results for the stability and upper bounds for the tracking error. When it comes to the KF algorithm, ref. [33] combined it with the Dantzig selector, which is an auxiliary CS optimization algorithm to estimate sparse signals. Ref. [10] utilized the pseudo-measurement technique to take the sparsity constraint into account while minimizing the tracking error based on the KF algorithm. Then numerical simulations demonstrated that the algorithms in refs. [10, 33] were viable for improving tracking performance. However, a rigorous stability analysis and performance results for tracking errors are still lacking.

Inspired by promising advances in CS, our paper proposes a compressed KF algorithm for identifying time-varying sparse stochastic regression models. We first perform a Kalman iteration based on the compressed regression data to generate a low-dimensional estimate for the compressed unknown signal. After that, an appropriate signal reconstruction algorithm is exploited to recover the original high-dimensional sparse signal. We establish stability and tracking performance bounds without independent or stationary signal assumptions, which makes our results applicable to feedback systems. The main challenge is to analyze the product of non-independent and non-stationary random matrices and deal with the inherent dependence of the signals. We resort to stochastic stability theory and CS theory to handle the above issues. The primary contributions are summarized in the following three aspects.

• A compressed KF algorithm on the basis of the CS theory is proposed to deal with high-dimensional but sparse signals in stochastic dynamic systems, while most of the literature revolves around the case where regression vectors or observation matrices are deterministic (see, e.g., refs. [10–16]).

• The stability analysis of the compressed KF algorithm is provided without resorting to the independence or stationarity of the system signal as commonly used in refs. [19, 34–36]. This paper further provides the approximate expressions for the covariance matrix of the compressed tracking error, which can be easily evaluated, analyzed, and even optimized.

• Compared with the excitation conditions imposed on the original high-dimensional regression vectors (cf., refs. [22, 37, 38]), a much weaker compressed excitation condition is introduced, which implies that even though the traditional uncompressed KF algorithm may fail in parameter estimation on account of insufficient excitation, the compressed KF algorithm can still fulfill the task of estimating high-dimensional sparse signals.

2 Preliminaries

In this section, we introduce some notations and preliminary knowledge.

2.1 Notations

$\mathbb{R}^n$ represents the set of n-dimensional real vectors. For a vector $x \in \mathbb{R}^n$, $\|x\|_0 = \#\{i \mid x_{(i)} \neq 0\}$ with $x_{(i)}$ being the ith element of the vector x and # being the cardinality of the set. That is, $\|x\|_0$ is the number of non-zero elements in x. $\mathbb{R}^{m\times n}$ stands for the set of $m\times n$ real matrices. I_m describes the m-dimensional square identity matrix. The notation $X > Y$ ($X \geq Y$) denotes that $X - Y$ is positive definite (semi-positive definite), where X and Y are symmetric matrices. Given a matrix $X \in \mathbb{R}^{m\times n}$, $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the minimum and maximum eigenvalues of the matrix. The spectral norm $\|X\|$ is defined by $\|X\| = \{\lambda_{\max}(XX^T)\}^{\frac{1}{2}}$. The notation tr$(X)$ denotes the trace of the corresponding matrix X. Given a random matrix Z, let $\|Z\|_{L_p} = \{\mathbb{E}[\|Z\|^p]\}^{\frac{1}{p}}$, $p \geqslant 1$ be its L_p-norm where $\mathbb{E}[\cdot]$ denotes the expectation operator. We use $\mathbb{E}[\cdot|\cdot]$ to represent the conditional expectation operator and use $P\{\cdot\}$ to denote the probability. For a matrix sequence $\{X_t\}$ and a positive scalar sequence $\{x_t\}$, $X_t = O(x_t)$ means that there exists a constant $C > 0$ independent of t such that $\|X_t\| \leqslant Cx_t$ holds for all $t \geqslant 0$.

2.2 Compressed sensing theory

Compressed sensing (CS) theory, a significant signal processing technique for efficient recovery and reconstruction of signals, has been widely used in various fields, including information theory, communication, and computer vision. It considers how to recover a signal $x \in \mathbb{R}^m$ from a noisy measurement:

$$z = Dx + \varepsilon, \tag{1}$$

where $D \in \mathbb{R}^{d\times m}$ is the sensing matrix ($d \ll m$)[1)] and $\varepsilon \in \mathbb{R}^d$ is the noise or measurement perturbation bounded by $\|\varepsilon\| \leqslant C$. Clearly, it is challenging or even impossible to solve this ill-posed problem. However, with the help of the sparse priors, the original high-dimensional signal x is likely to be recovered from the observation z. Hence, assume that x is s-sparse, i.e., $\|x\|_0 \leqslant s$ for some $s \leqslant d \ll m$.

As discussed below, the efficiency of the CS method will depend heavily on the construction of the sensing matrix and the design of the signal reconstruction algorithm.

Construction of the sensing matrix To guarantee that eq. (1) can discriminate approximately sparse unknown parameter vectors, ref. [28] introduced the restricted isometry property (RIP) as a condition on the sensing matrix D.

Definition 1 (RIP) For an integer s and the sensing matrix $D \in \mathbb{R}^{d\times m}$ ($1 \leqslant s \leqslant m$), we say that the matrix D satisfies the RIP of order s if there exists a constant $\delta_s \in [0, 1)$, which is the smallest quantity such that

$$(1-\delta_s)\|b\|^2 \leqslant \|D_L b\|^2 \leqslant (1+\delta_s)\|b\|^2, \quad \forall b \in \mathbb{R}^{\#L}, \tag{2}$$

holds for every submatrix D_L which is formed by columns of D corresponding to the indices in the set $L \subset \{1, \cdots, m\}$ with $\#L \leqslant s$.

Remark 1 The RIP, at least when applied to sparse vectors, characterizes matrices that are nearly orthonormal. From eq. (2), it is straightforward that $1-\delta_s \leqslant \lambda_{\min}(D_L^T D_L) \leqslant \lambda_{\max}(D_L^T D_L) \leqslant 1+\delta_s$. As can be seen, the s-restricted isometry constant δ_s grows as s does.

Considerable progress has been made to generate matrices satisfying the RIP; for example, the deterministic matrix methods, including Vandermond matrices [39] and the deterministic Fourier measurements [40], and stochastic matrix methods, including Gaussian and Bernoulli random sensing matrices [41], random Weibull matrices [42]. Also, regarding the random sensing matrices, we have the following result [41].

Lemma 1 For given d, m, and $0 < \delta < 1$, if the sensing matrix $D \in \mathbb{R}^{d\times m}$ is a Gaussian or Bernoulli random matrix, then it holds that for some constants c_1, c_2 depending only on δ, for a prescribed δ and for any $s \leqslant c_1 d/\log(m/s)$, the probability that RIP holds is no less than $1 - 2\mathrm{e}^{-c_2 d}$.

Remark 2 Ref. [41] established the connection between constants c_1 and c_2. It is clear that when c_1 is small enough, c_2 enables it to be larger than 0. Let $c_1 = \delta^3/120$, then for given $d \geqslant 120s\log(m/s)/\delta^3$, the sensing matrix D has probability not less than $1 - 2\mathrm{e}^{-c_2 d}$ of satisfying RIP (eq. (2)).

Design of the signal construction algorithm The recovery problem of the model eq. (1) can be formulated as a convex optimization problem:

$$\begin{aligned} &\min_{x\in\mathbb{R}^m} \|x\|_1, \\ &\text{s.t. } \|z - Dx\| \leqslant C, \end{aligned} \tag{3}$$

1) The relation $d \ll m$ means that d is much smaller than m.

where $\|x\|_1 \triangleq \sum_{i=1}^{m} \left|x_{(i)}\right|$.

The following lemma from ref. [28] analyzed the upper bound of the reconstruction error.

Lemma 2 Suppose that s satisfies $\delta_{3s} + 3\delta_{4s} < 2$ where δ_{3s} and δ_{4s} are defined in Definition 1. Let the signal x_0 be s-sparse and perturbation ε be subject to $\|\varepsilon\| \leqslant C$, the recovered signal x^* derived from solving (3) obeys

$$\|x - x^*\| \leqslant C_s C,$$

where $C_s \triangleq \dfrac{4}{\sqrt{3(1-\delta_{4s})} - \sqrt{1+\delta_{3s}}}$ is a positive constant.

Remark 3 This lemma shows that the recovery of the signal is stable. In other words, a small measurement perturbation ε can only lead to a small derivation of the recovered signal from the true signal. What's more, the signal can be accurately reconstructed when the measurement perturbation is zero.

3 Problem formulation

3.1 System model

Consider the following time-varying stochastic sparse regression model:

$$y_k = \varphi_k^T \theta_k + v_k, \ k \geqslant 0, \tag{4}$$

where y_k and v_k are scalar observation and noise, respectively, $\theta_k \in \mathbb{R}^m$ is an unknown s-sparse parameter vector of interest (i.e., $\|\theta_k\|_0 \leqslant s$), and $\varphi_k \in \mathbb{R}^m$ is the $3s$-sparse stochastic regression vector (i.e., $\|\varphi_k\|_0 \leqslant 3s$). The time-variation of θ_k is denoted as follows:

$$\tau\omega_k \triangleq \theta_k - \theta_{k-1}, \ k \geqslant 1, \tag{5}$$

where θ_k is time invariant when $\tau = 0$, otherwise, is time varying. Here we concentrate on the case where the regression vector and the unknown parameter vector are sparse. It is prevalent in practical scenarios such as industrial robots, field monitoring, and channel estimation [43–45].

In contrast to most of the literature that considers deterministic vectors or matrices [10–16], here, the sparse regression vector φ_k in eq. (4) is stochastic, which has practical significance in feedback systems. For example, if φ_k consists of current and past input-output data, i.e., $\varphi_k = [y_k, \ldots, y_{k-p}, x_k, \ldots, x_{k-q}]^T$ with x_k being the input signal at time k, then the model eq. (4) can be reduced to ARX model [18] with time-varying coefficients, and it is clear that φ_k is stochastic and can not satisfy stringent statistical conditions such as the independent and identically distributed condition when the control $x_k = f(y_j, j \leqslant k)$ is designed based on past observations.

In this paper, our aim is to track the time-varying sparse signal θ_k by using the observations $\{y_k, k \geqslant 1\}$ and the sparse stochastic regression vectors $\{\varphi_k, k \geqslant 1\}$.

3.2 Compressed Kalman filter algorithm

The KF algorithm is widely used in various realms including system control and signal processing due to the optimality in the posterior mean square sense when the noises have Gaussian distribution. Note that eqs. (4) and (5) can be recast as a state space model with state θ_k, then it is natural to consider the following KF algorithm [6, 7]:

$$\widehat{\theta}_{k+1} = \widehat{\theta}_k + \rho \frac{P_{k-1}\varphi_k}{R + \rho\varphi_k^T P_{k-1}\varphi_k}(y_k - \varphi_k^T \widehat{\theta}_k), \tag{6}$$

$$P_k = P_{k-1} - \frac{\rho P_{k-1}\varphi_k\varphi_k^T P_{k-1}}{R + \rho\varphi_k^T P_{k-1}\varphi_k} + \rho Q, \tag{7}$$

where $\rho \in (0, 1)$ is defined as step size, $P_0 > 0$, $R > 0$, $Q > 0$, and $\widehat{\theta}_0$ are deterministic and can be arbitrarily chosen.

With respect to theoretical analysis, refs. [22, 37, 38] proved the stability of the standard KF algorithm (i.e., eqs. (6) and (7)) under the following stochastic excitation condition imposed on the original regression vectors $\{\varphi_k\}$: There exists a constant $h > 0$ such that $\left\{\lambda_k^0, k \geqslant 0\right\} \in \mathcal{S}^0(\lambda^0)$ for some $\lambda^0 \in (0, 1)$, where $\mathcal{S}^0(\lambda^0)$ is defined by eq. (17), and λ_k^0 is defined as follows:

$$\lambda_k^0 \triangleq \lambda_{\min}\left(\mathbb{E}\left[\frac{1}{1+h}\sum_{i=kh+1}^{(k+1)h}\frac{\varphi_i\varphi_i^T}{1+\|\varphi_i\|^2}\middle|\mathcal{F}_{kh}\right]\right) \tag{8}$$

with $\mathcal{F}_k \triangleq \sigma\{\varphi_i, \omega_i, v_{i-1}, i \leqslant k\}$. However, when dealing with high-dimensional but sparse signals, the KF algorithm does not always perform well since the mentioned excitation condition eq. (8) is hard or even impossible to satisfy. To improve this situation, we resort to the CS theory to weaken the condition eq. (8) by reducing the dimension of measurements.

Now, in combination with the CS method, we propose the compressed Kalman filter algorithm (see Algorithm 1) to fulfill the tracking task of sparse signals.

To be specific, the sensor first utilizes the sensing matrix $D \in \mathbb{R}^{d\times m}$ ($s \leqslant d \ll m$) to compress the regression vector by $\psi_k = D\varphi_k$. Hence we rewrite the original model eq. (4) as

$$\begin{aligned} y_k &= \varphi_k^T\theta_k + v_k = \psi_k^T\zeta_k + \varphi_k^T\theta_k - \psi_k^T\zeta_k + v_k \\ &= \psi_k^T\zeta_k + \varphi_k^T[I_m - D^TD]\theta_k + v_k \triangleq \psi_k^T\zeta_k + \bar{v}_k, \end{aligned} \tag{9}$$

where $\zeta_k, \bar{v}_k$ are transformed according to $\zeta_k = D\theta_k$, $\bar{v}_k = \varphi_k^T[I_m - D^TD]\theta_k + v_k$. $\bar{v}_k$ can be regarded as the new "noise" term. Note also that for $k \geqslant 1$,

$$\zeta_k = \zeta_{k-1} + \tau\overline{\omega}_k, \tag{10}$$

Algorithm 1 Compressed Kalman filter algorithm

Input: $\{y_k, \varphi_k\}_{k\geqslant 1}$, step size $\rho \in (0,1)$, $Q > 0, R > 0$, sensing matrix $D \in \mathbb{R}^{d\times m}$.
Output: $\{\widehat{\theta}_{k+1}\}, k \geqslant 1$.
Intialize: Begin with an arbitrary initial vector $\widehat{\zeta}_1$ and matrix $P_0 > 0$,
for each time $k = 1, 2, \cdots$ **do**
Step 1. Compression: $\psi_k = D\varphi_k$.
Step 2. Estimation in a low-dimensional dimension.

$$L_k = \frac{P_{k-1}\psi_k}{R + \rho\psi_k^T P_{k-1}\psi_k}, \tag{11}$$

$$P_k = P_{k-1} - \frac{\rho P_{k-1}\psi_k\psi_k^T P_{k-1}}{R + \rho\psi_k^T P_{k-1}\psi_k} + \rho Q, \tag{12}$$

$$\widehat{\zeta}_{k+1} = \widehat{\zeta}_k + \rho L_k(y_k - \psi_k^T\widehat{\zeta}_k). \tag{13}$$

Step 3. Reconstruction:

$$\widehat{\theta}_{k+1} = \arg\min_{\theta\in\Theta} \|\theta\|_1, \tag{14}$$

where $\Theta = \{\theta \in \mathbb{R}^m \mid \|D\theta - \widehat{\zeta}_{k+1}\| \leqslant C\}$.

where $\overline{\omega}_k = D\omega_k$.

Next, based on the new model of data, we can view $\{\psi_k, y_k\}$ as new measurements and perform the KF algorithm to obtain a low-dimensional estimate $\widehat{\zeta}_k$ for the compressed unknown signal ζ_k. Finally, we recover a high-dimensional estimate $\widehat{\theta}_k$ for the original unknown sparse signal θ_k by solving the convex optimization problem (eq. (3)).

Remark 4 The reconstruction step (Step 3) in Algorithm 1, not affecting the execution of the remainder of the algorithm, need not be executed at every iteration. In practice, Step 3 can be performed every K ($K \gg 1$) iteration to recover the original high-dimensional sparse signal, i.e., $\theta_K, \theta_{2K}, \cdots$, thus lessening the computation burden. There are several algorithms applicable to Step 3, e.g., interior-point, the orthogonal matching pursuit (OMP), the basis pursuit, and the basis pursuit denoising algorithms [46–48]. In addition, we can choose the upper bound on the compressed estimation error $C(\tau, \delta_{4s})$ in Remark 13 to act as the constant C.

Subsequently, we analyze the stability and performance of Algorithm 1 regarding the upper bound of the estimation error and the approximate expression of the true mean square compressed tracking error matrix.

4 Definitions and assumptions

For further discussion, we need to introduce the following notations and definitions in ref. [38].

4.1 Definitions

Definition 2 For any random matrix or vector sequence $x = \{x_k(\rho), k \geqslant 1\}$ and real numbers $p \geqslant 1$, $\rho^* \in (0,1)$, the L_p-stable family is defined by

$$\mathcal{L}_p(\rho^*) = \left\{x : \sup_{\rho\in(0,\rho^*]} \sup_{k\geqslant 1} \|x_k(\rho)\|_{L_p} < \infty\right\}.$$

Definition 3 For any random square matrix sequence $Z = \{Z_k(\rho)\}$ and real numbers $p \geqslant 1, \rho^* \in (0,1)$, the stochastic exponentially stable family is defined by

$$\mathcal{S}_p(\rho^*) = \Bigg\{Z : \left\|\prod_{t=s+1}^{k}(I - \rho Z_k(\rho))\right\|_{L_p} \leqslant M(1-\alpha\rho)^{k-s}, \quad \forall\rho \in (0, \rho^*],\ \forall k \geqslant s \geqslant 0, \quad \text{for some } M > 0, \text{ and } \alpha \in (0,1)\Bigg\}. \tag{15}$$

Correspondingly, the deterministic exponential stability family is defined by

$$\mathcal{S}(\rho^*) = \Bigg\{Z : \left\|\prod_{t=s+1}^{k}(I - \rho\mathbb{E}[Z_k(\rho)])\right\| \leqslant M(1-\alpha\rho)^{k-s}, \quad \forall\rho \in (0, \rho^*],\ \forall k \geqslant s \geqslant 0, \quad \text{for some } M > 0, \text{ and } \alpha \in (0,1)\Bigg\}. \tag{16}$$

For convenience, the following sets are introduced.

$$\mathcal{S}_p \triangleq \bigcup_{\rho\in(0,\rho^*]} \mathcal{S}_p(\rho^*),\ \mathcal{S} \triangleq \bigcup_{\rho\in(0,\rho^*]} \mathcal{S}(\rho^*).$$

Next, for a scalar sequence $x = \{x_k, k \geqslant 0\}$, we define

$$\mathcal{S}^0(\lambda) = \Bigg\{x : x_k \in [0,1], \mathbb{E}\left[\prod_{t=s+1}^{k}\left(1 - x_j\right)\right] \leqslant M\lambda^{k-s}, \quad \forall k \geqslant s \geqslant 0, \text{ for some } M > 0\Bigg\}, \tag{17}$$

where $\lambda \in [0,1)$ is a parameter reflecting the stability margin. Also,

$$\mathcal{S}^0 \triangleq \bigcup_{\lambda\in(0,1)} \mathcal{S}^0(\lambda).$$

Definition 4 A random process $x = \{x_k, k \geqslant 0\}$ is said to belong to the weakly dependent class $\mathcal{M}_p$, $p \geqslant 1$, if there exists a constant C_p^x depending only on p and the distribution of $\{x_k\}$ such that for $k \geqslant 0$ and $h \geqslant 1$,

$$\left\|\sum_{i=k+1}^{k+h} x_i\right\|_{L_p} \leqslant C_p^x h^{1/2}.$$

Remark 5 The set $\mathcal{M}_p$ includes the martingale difference sequences (m.d.s.), zero mean ϕ- and α-mixing sequences, and linear processes (a process generated from a white noise source via a linear filter with absolutely summable impulse response) as special cases (cf., [38]).

4.2 Assumptions

For the stability and performance analysis, we make the following assumptions about the sensing matrix, the regression vectors, and the observation noise.

Assumption 1 The sensing matrix $D \in \mathbb{R}^{d\times m}$ satisfies the RIP with order $4s$ where the $4s$-restricted isometry constant is denoted as δ_{4s} (see Definition 1).

Assumption 2 There exists a constant $h > 0$ such that $\{\lambda_k, k \geqslant 0\} \in \mathcal{S}^0(\lambda)$ for some $\lambda \in (0,1)$, where $\mathcal{S}^0(\lambda)$ is defined by eq. (17), and λ_k is defined as follows:

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

with $\mathcal{F}_k \triangleq \sigma\{\varphi_i, \omega_i, v_{i-1}, i \leqslant k\}$ and $\sup_{k\geqslant 1}\|\psi_k\| \triangleq C_\psi < \infty$.

Remark 6 Intuitively, Assumption 2 implies that the smallest eigenvalue λ_k is not "too small". Consider an extreme case where $\varphi_k = 0$. It is straightforward that $\lambda_k = 0$ and Assumption 2 is not satisfied. At this point, the measurement $\{\varphi_k, y_k\}$ contains no information about the unknown parameter, and thus the system is non-identifiable. Inspired by this, some "not too small" or "nonzero" excitation conditions are imposed on the regression vectors φ_k to fulfill the task of parameter estimation or signal tracking.

Remark 7 We remark that the main difference between the traditional excitation condition in refs. [22, 37, 38] and our assumption (see Assumption 2) is that Assumption 2 is assumed for the compressed regression vectors $\{\psi_k\}$ instead of the original high-dimensional ones $\{\varphi_k\}$. Since the dimension of $\{\psi_k\}$ is much smaller than that of $\{\varphi_k\}$, for the high-dimensional and sparse regression vectors, Assumption 2 is much easier to satisfy than the traditional excitation condition in refs. [22, 37, 38]. In other words, even in the case where the non-compressed KF algorithm cannot fulfill the estimation tasks, our proposed compressed KF algorithm may still get the compressed estimation results stably, see the simulation results in Section 7.

Assumption 3 For some $z \geqslant 2$, the unknown parameter θ is L_z-bounded, i.e., $\sup_k\|\theta_k\|_{L_z} = C^\theta < \infty$. Furthermore, $\{L_k v_k\} \in \mathcal{M}_r$, $\{\omega_k\} \in \mathcal{M}_r, r > 2$ with the parameters C_r^n and C_r^ω, respectively.

Remark 8 By Assumption 3, noises and parameter variations are assumed to be weakly dependent with certain bounded moments to build a more refined bound in Theorem 2 than that in Theorem 1. In contrast, weaker assumptions (i.e., moment conditions) are imposed on v_k and ω_k in Theorem 1.

Assumption 4 Let $\mathcal{F}_k \triangleq \sigma\{\varphi_i, \omega_i, v_{i-1}, i \leqslant k\}$, for each $k > 1$ we have

$$\begin{aligned}
&\mathbb{E}[v_k|\mathcal{F}_k] = 0,\ \mathbb{E}[\omega_{k+1}|\mathcal{F}_k] = \mathbb{E}[\omega_{k+1}v_k|\mathcal{F}_k] = 0,\\
&\mathbb{E}[v_k^2|\mathcal{F}_k] = R_k^v,\ \mathbb{E}[\omega_{k+1}\omega_{k+1}^T] = Q_{k+1}^\omega,\\
&\sup_{k\geqslant 1}\{\mathbb{E}[|v_k|^r|\mathcal{F}_k] + \mathbb{E}[\|\omega_{k+1}|^r\|]\} < \infty,\ r \geqslant 1.
\end{aligned}$$

Remark 9 Assumption 4 implies that the noise and the parameter variation are endowed with white noise characters, which are used to analyze the approximate tracking performance of Algorithm 1 (see Theorem 4), although it is stronger than Assumption 3.

Assumption 5 There exists a number $t \geqslant 7$ together with a function $\phi(m) \to 0$ (as $m \to \infty$) such that

$$\left\|\mathbb{E}\left[\psi_k\psi_k^T|\mathcal{F}_{k-m}\right] - \mathbb{E}\left[\psi_k\psi_k^T\right]\right\|_{L_t} \leqslant \phi(m)\ \ \forall k, m. \tag{18}$$

Remark 10 Eq. (18) describes the decaying correlation between $\psi_k\psi_k^T$ and $\mathcal{F}_{k-m}$. In other words, the larger the time distance m, the smaller the correlation. This assumption can be readily satisfied by introducing a certain weak dependence requirement on the compressed regression vector $\{\psi_k\}$, e.g., ϕ-mixing property.

5 Main results

Now, we are in the position to describe the main technical results about the proposed algorithm.

5.1 Stability results

In this section, in order to perform the stability analysis for Algorithm 1, we first obtain the compressed estimation error equation. By eqs. (13), (9), and (10), we have

$$\begin{aligned}
&\widetilde{\zeta}_{k+1} = (I - \rho F_k)\widetilde{\zeta}_k - \rho L_k\bar{v}_k + \tau\overline{\omega}_{k+1},\\
&F_k = L_k\psi_k^T,
\end{aligned} \tag{19}$$

where $\widetilde{\zeta}_k = \zeta_k - \widehat{\zeta}_k$ denotes the compressed estimation error. Then, we provide the exponential stability of the homogeneous part of the compressed error equation (19) and a preliminary upper bound for the compressed tracking error of the proposed compressed KF algorithm.

Theorem 1 Consider the time-varying model (4) and the compressed error equation (19). Under Assumption 1 and Assumption 2, we have $\{\rho F_k\} \in \mathcal{S}_p$, $\forall p \geqslant 1$. Furthermore, if for some $\beta > 2$, $p \geqslant 1$, and

$$\pi_p \triangleq \sup_k\left\|\Xi_k\log^\beta(e+\Xi_k)\right\|_{L_p} < \infty$$

hold, where $\Xi_k = \frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}} \|\theta_k\| + \|v_k\| + \tau\sqrt{1+\delta_{4s}} \|\omega_{k+1}\|$, then the compressed tracking error $\{\widetilde{\zeta}_k, k \geqslant 0\}$ is L_p-stable and

$$\limsup_{k\to\infty} \left\|\widetilde{\zeta}_k\right\|_{L_p} \leqslant c\left[\pi_p \log^{\beta}\left(e+\pi_p^{-1}\right)\right], \tag{20}$$

where c is a positive constant.

Remark 11 See the proof of Theorem 1 in Section 6.1. From eq. (19) and the definition of Ξ_k, we know that the upper bound of tracking error consists of three parts: The first part is caused by the compression error which gets smaller as δ_{4s} gets smaller; the second part is related to the system noise v_k while the third part is relevant to the parameter variation ω_k. Note that by Remark 2, when dimensions of the sensing matrix D satisfy the inequality $d \geqslant 480s\log(m/4s)/\delta_{4s}^3$, the RIP parameter δ_{4s} can be arbitrarily small. On this basis, if the magnitudes of the system noise v_k and the parameter variation ω_k are also small, then the upper bound of the estimation error will be small.

With more conditions about the system noise v_k and the parameter variation ω_k, we can further construct a more refined upper bound on the compressed tracking error $\widetilde{\zeta}_{k+1}$.

Theorem 2 Under Assumptions 1, 2, and 3, we have for any $k \geqslant 0$, for some constant v,

$$\begin{aligned}\|\widetilde{\zeta}_{k+1}\|_{L_v} \leqslant & B_{1,v}\sqrt{\rho} + B_{2,v}\frac{\tau}{\sqrt{\rho}} + B_{3,v}\frac{\delta_{4s}}{\sqrt{1-\delta_{4s}}} \\ & + B_{4,v}(1-\alpha\rho)^{k+1},\end{aligned}$$

where $B_{1,v}, B_{2,v}, B_{3,v}, B_{4,v}, \alpha \in (0,1)$ are positive constants defined in the proof, which is irrelevant to ρ, τ, δ_{4s}.

Remark 12 The proof of Theorem 2 is given in Section 6.2. Theorem 2 roughly indicates the trade-off between tracking ability and noise sensitivity.

Remark 13 Consider the special case that $v = 1$. When $\tau > 0$ (i.e., the unknown parameter is time-varying), we know $B_{1,1}\sqrt{\rho} + B_{2,1}\tau/\sqrt{\rho}$ reaches its minimum $2\sqrt{\tau B_{1,1}B_{2,1}}$ at the optimal step-size $\rho^* = \tau B_{2,1}(B_{1,1})^{-1}$. Then there exists an integer $K > 0$, such that we have for $k \geqslant K$, $B_{4,1}(1-\alpha\rho)^{k+1} \leqslant \sqrt{\tau B_{1,1}B_{2,1}}$. Here,

$$K = \max\left\{\left\lceil \frac{\log(\tau B_{1,1}B_{2,1} - 2\log(B_{4,1}))}{2\log(1-\alpha\rho)}\right\rceil, 0\right\} \tag{21}$$

and then we have for $k \geqslant K$,

$$\left\|\widetilde{\zeta}_k\right\|_{L_1} \leqslant 3\sqrt{\tau B_{1,1}B_{2,1}} + B_{3,1}\frac{\delta_{4s}}{\sqrt{1-\delta_{4s}}} \triangleq C(\tau, \delta_{4s}),$$

where $\lceil\cdot\rceil$ denotes rounding up operator.

From Remark 13, the probability that the compressed tracking error stays within a specific range is analyzed in the following corollary.

Corollary 1 Under the same conditions in Theorem 2, we suppose that $\tau > 0$, $\rho = \tau B_{2,1}(B_{1,1})^{-1}$. Then there exists an integer $K \geqslant 0$, such that for any integer $k \geqslant K$ and any constant $\xi \in (0,1)$, there exists a positive constant $\eta = \max\left\{1, 2C^{\xi}(\tau, \delta_{4s})\right\}$ such that

$$P\left\{\left\|\widetilde{\zeta}_k\right\| \leqslant \eta C^{1-\xi}(\tau, \delta_{4s})\right\} \geqslant 1 - \frac{C^{\xi}(\tau, \delta_{4s})}{\eta},$$

where $C(\tau, \delta_{4s}) = 3\sqrt{\tau B_{1,1}B_{2,1}} + B_{3,1}\frac{\delta_{4s}}{\sqrt{1-\delta_{4s}}}$ with constants $B_{1,1}, B_{2,1}, B_{3,1}$ being defined in Theorem 2.

Remark 14 The proof of Corollary 1 can be found in Section 6.3. $C(\tau, \delta_{4s})$ goes to zero as τ and δ_{4s} goes to zero. Then, $\eta = 1$. Finally, $\left\|\widetilde{\zeta}_k\right\|$ tends to zero with a probability of nearly one.

From Corollary 1 and the step of the reconstruction of the estimates in Algorithm 1, we ultimately establish an upper bound on the tracking error of the uncompressed sparse signal

Theorem 3 Assume that the sensing matrix $D \in \mathbb{R}^{d\times m}$ satisfies 4sth RIP with s satisfying $\delta_{3s} + 3\delta_{4s} < 2$. Under the same conditions as used in Corollary 1, there exists an integer $K \geqslant 0$, such that for any integer $k \geqslant K$ and any $\xi \in (0,1)$,

$$P\left\{\left\|\widehat{\theta}_k - \theta_k\right\| \leqslant C_s\eta C^{1-\xi}(\tau, \delta_{4s})\right\} \geqslant 1 - \frac{C^{\xi}(\tau, \delta_{4s})}{\eta}$$

holds for some positive constant $\eta = \max\left\{1, 2C^{\xi}(\tau, \delta_{4s})\right\}$ with $C(\tau, \delta_{4s})$ being defined in Corollay 1.

Remark 15 Detailed proof of Theorem 3 is given in Section 6.4. By the expression for the constant C_s, it is straightforward that C_s may be only relevant to δ_{4s}, and it becomes smaller when δ_{4s} becomes smaller.

From Theorems 1, 2, and 3, we can see that the stability results of the compressed KF algorithm make no appeal to the independence or stationarity conditions of the regression vectors $\{\varphi_k\}$ and thus can be applied to the stochastic feedback systems.

5.2 Performance results

In Section 5.1, under the compressed excitation condition and RIP of the sensing matrix, the upper bound of the L_p-norm of the estimation error is established, indicating how the estimation error is influenced by RIP constant δ_{4s}, the noise v_k and parameter drift $\tau\omega_k$. In this subsection, we provide the performance results to accurately evaluate or optimize the performance of the compressed KF algorithm. It is then necessary to derive an explicit expression for the mean square compressed tracking error, defined as follows:

$$\Pi_k = \mathbb{E}\left[\widetilde{\zeta}_k\widetilde{\zeta}_k^T\right].$$

Nevertheless, one of the leading technical difficulties for calculating Π_k lies in the dependence among $\{\psi_k\}$. The cross-term originating from this dependence is generally negligible but quite complicated. We therefore introduce the following simple linear deterministic difference equation for $\widehat{\Pi}_k$:

$$\widehat{\Pi}_{k+1} =(I-\rho\mathbb{E}[F_k])\widehat{\Pi}_k(I-\rho\mathbb{E}[F_k])^T+\rho^2\mathbb{E}[R_k^vL_kL_k^T] + \tau^2DQ_{k+1}^{\omega}D^T \tag{22}$$

where $\widehat{\Pi}_0=\mathbb{E}[\widetilde{\zeta}_0\widetilde{\zeta}_0^T]$, and R_k^v and Q_{k+1}^{ω} are defined in Assumption 4.

Also, we demonstrate that this simple expression $\widehat{\Pi}_k$ can arbitrarily well approximate Π_k with the help of the assumption on the weak (or decaying) dependence. The following theorem centers on this idea.

Theorem 4 Suppose that Assumptions 1, 2, 4, and 5, and for any $z\geqslant 1$, $\{\theta_k\}\in\mathcal{L}_z$ hold. Considering the tracking error $\widetilde{\zeta}_k$ defined by eq. (19) and the approximate expression $\widehat{\Pi}_k$ defined by eq. (22), we have, for any $\rho\in(0,\rho^*]$ and any $k\geqslant 1$,

$$\|\Pi_{k+1}-\widehat{\Pi}_{k+1}\|\leqslant c\sigma(\rho)\left[\rho+\frac{\tau^2}{\rho}+\frac{\delta_{4s}^2}{1-\delta_{4s}}+\left(1-\frac{\alpha\rho}{2}\right)^k\right],$$

where $\sigma(\rho)$ is a function that tends to zero as ρ tends to zero. It is defined by $\sigma(\rho)=\min\limits_{m\geqslant 1}\{\sqrt{\rho}m+\phi(m,\rho)\}$. $c>0$ and $\alpha\in(0,1)$ are constants.

The proof of Theorem 4 is presented in Section 6.5. Theorem 4 presents an approximate expression for the covariance matrix of the compressed estimation error and explicitly calculates the approximation error. The approximation performs well in the case where adaption gain ρ is appropriately small. Moreover, when δ_{4s} gets smaller, the quality of approximation gets better.

To further simplify the expression for $\widehat{\Pi}_k$, the strengthened assumption is introduced and gain a clearer view of the effect of ρ on Π_k. Assuming that $\mathbb{E}[F_k]$ and $\mathbb{E}\left[L_kL_k^T\right]$ do not vary with k, the following corollary is able to be derived by Theorem 4 directly.

Corollary 2 Assume that $R_k^v\equiv R_v$, $Q_k^{\omega}\equiv Q_{\omega}$, $\mathbb{E}[F_k]=F$, and $\mathbb{E}\left[L_kL_k^T\right]=G$. Then under the same conditions of Theorem 4, we have for all $\rho\in(0,\rho^*]$, $k\geqslant 1$,

$$\Pi_k=\rho\overline{R}_v+\frac{\tau^2}{\rho}\overline{Q}_{\omega}+O\left(\sigma(\rho)\cdot\left[\rho+\frac{\tau^2}{\rho}+\frac{\delta_{4s}^2}{1-\delta_{4s}}\right]\right)+o(1),$$

where the term $o(1)$ converges to 0 at an exponential rate as $k\to\infty$, $\sigma(\rho)$ is defined in Theorem 4, and

$$\overline{R}_v=R_v\int_0^{\infty}\mathrm{e}^{-Ft}G\mathrm{e}^{-F^Tt}\mathrm{d}t,\ \overline{Q}_{\omega}=\int_0^{\infty}\mathrm{e}^{-Ft}DQ_{\omega}D^T\mathrm{e}^{-F^Tt}\mathrm{d}t.$$

Remark 16 A trivial proof is listed in Section 6.6. Since $\sigma(\rho)$ goes to zero as ρ goes to zero, it follows that

$$\Pi_k\sim\rho\overline{R}_v+\frac{\tau^2}{\rho}\overline{R}_{\omega},\ k\to\infty,\ \rho\to 0. \tag{23}$$

Clearly, the error term caused by noise positive correlates with ρ while the error term caused by parameter drift negatively correlates with ρ. Hence, for the sake of minimizing the tracking error, the choice of ρ concerns the trade-off between noise and parameter drift. Therefore, considering that the right hand of eq. (23) is in the form of the matrix, we take "trace" on both sides and obtain

$$\mathbb{E}\left[\left\|\widetilde{\zeta}_k\right\|^2\right]\sim\rho\mathrm{tr}(\overline{R}_v)+\frac{\tau^2}{\rho}\mathrm{tr}(\overline{R}_{\omega}).$$

The right-hand side reaches its minimum $2\tau\sqrt{\mathrm{tr}(\overline{R}_v)\cdot\mathrm{tr}(\overline{R}_{\omega})}$ at the optimal step size $\tau\sqrt{\mathrm{tr}(\overline{R}_v)/\mathrm{tr}(\overline{R}_{\omega})}$, indicating the practical consequences of the Theorem 4. An argument similar to the one in Corollary 1 and Theorem 3 shows that the tracking error bound for the original high-dimensional signal gets quite small with a large probability.

6 Proofs of the main results

6.1 Proof of Theorem 1

One of the main difficulties of stability analysis lies in the analysis of the errors arising from the compression step. To this end, we resort to the RIP property of the sensing matrix D and then present the following key lemma.

Lemma 3 Under Assumption 1, we have

$$\|L_k\varphi_k^T[I_m-D^TD]\theta_k\| \leqslant\frac{3\delta_{4s}}{2\sqrt{\rho R}(1-\delta_{4s})}\|P_{k-1}\|^{1/2}\cdot\|\psi_k\|\cdot\|\theta_k\|.$$

Proof. Since $\|\varphi_k\|_0\leqslant 3s$ and $\|\theta_k\|_0\leqslant s$, we first define index sets of nonzero elements as $L_1=\{m_1,m_2,\cdots,m_{3s}\}$ and $L_2=\{n_1,n_2,\cdots,n_s\}$, respectively. Next, we retain all elements indexed by $L=L_1\cup L_2$ from φ_k, θ_k, and the corresponding columns of the matrix D. Then we remove the remaining elements and denote the resulting vectors and matrix as $\varphi_{k,4s}$, $\theta_{k,4s}$, and D_{4s}, respectively (In cases where the elements of φ_k and θ_k are non-zero with the same indexes, we just keep $\#L$ ($\#L<4s$) elements of φ_k, θ_k, and D. Obviously, the subsequent analysis is almost identical, the major change being the substitution of $4s$ for $\#L$, so it is omitted).

By Assumption 1, D satisfies $4s$th RIP which implies that $1-\delta_{4s}\leqslant\lambda_i\leqslant 1+\delta_{4s}$ with λ_i being any eigenvalue of the matrix $D_{4s}^TD_{4s}$. Then we have

$$\left\|L_k\varphi_k^T\left[I_m-D^TD\right]\theta_k\right\|$$

$$
\begin{aligned}
&= \left\| L_k \varphi_{k,4s}^T \left[I_{4s} - D_{4s}^T D_{4s} \right] \theta_{k,4s} \right\| \\
&\leqslant \left\| L_k \varphi_{k,4s}^T \left[(1+\delta_{4s}) I_{4s} - D_{4s}^T D_{4s} \right] \theta_{k,4s} \right\| \\
&\quad + \delta_{4s} \left\| L_k \varphi_{k,4s}^T \theta_{k,4s} \right\| \\
&\leqslant \left\| L_k \varphi_{k,4s}^T \right\| \cdot \left\| \left[(1+\delta_{4s}) I_{4s} - D_{4s}^T D_{4s} \right] \right\| \cdot \left\| \theta_{k,4s} \right\| \\
&\quad + \delta_{4s} \left\| L_k \varphi_{k,4s}^T \right\| \cdot \left\| \theta_{k,4s} \right\| \\
&\leqslant 2\delta_{4s} \left\| L_k \varphi_{k,4s}^T \right\| \cdot \left\| \theta_{k,4s} \right\| + \delta_{4s} \left\| L_k \varphi_{k,4s}^T \right\| \cdot \left\| \theta_{k,4s} \right\| \\
&= 3\delta_{4s} \left\| L_k \varphi_k^T \right\| \cdot \|\theta_k\| \\
&\leqslant 3\delta_{4s} \|L_k\| \cdot \|\varphi_k\| \cdot \|\theta_k\| \\
&\leqslant \frac{3\delta_{4s}}{2\sqrt{\rho R}(1-\delta_{4s})} \|P_{k-1}\|^{1/2} \cdot \|\psi_k\| \cdot \|\theta_k\|. \qquad (24)
\end{aligned}
$$

The last inequality holds by noting that $\|L_k\| \leqslant \frac{\|P_{k-1}\|^{1/2}}{2\sqrt{\rho R}}$ and eq. (2), which is the desired conclusion. ■

Proof of Theorem 1. By error equation (19), we get

$$
\begin{aligned}
\widetilde{\zeta}_{k+1} &= \prod_{i=0}^{k} (I - \rho F_i) \widetilde{\zeta}_0 \\
&\quad + \sum_{i=0}^{k} \prod_{j=i+1}^{k} (I - \rho F_j)(-\rho L_i \overline{v}_i + \tau \overline{\omega}_{i+1}). \qquad (25)
\end{aligned}
$$

From Theorem 3.5 in ref. [37], we know that for any $p \geqslant 1$, $\{F_k\} \in \mathcal{S}_p$, i.e., for any $\rho \in [0, \rho^*]$, $\rho^* \in (0,1)$, there exists $M > 0$, $\alpha \in (0,1)$,

$$
\left\| \prod_{j=i+1}^{k} \left(I_d - \rho F_j \right) \right\|_{L_p} \leqslant M (1 - \alpha\rho)^{k-i}, \ k \geqslant i \geqslant 0.
$$

Following the similar procedure as eq. (24) in Lemma 3, we have

$$
\begin{aligned}
&\| -\rho L_k \overline{v}_k + \tau \overline{\omega}_{k+1} \| \\
&\leqslant \left(1 + \frac{\sqrt{\rho} \max\left\{ 1, C_\psi \right\} \|P_{k-1}\|^{1/2}}{2\sqrt{R}} \right) \Xi_k. \qquad (26)
\end{aligned}
$$

The remainder of the argument is analogous to that in Theorem 4.1 in ref. [22] and so is omitted. Combining eqs. (25) and (26), this theorem is proved. ■

6.2 Proof of Theorem 2

Now we define

$$
\Lambda(k+1, s) = (I_d - \rho F_k) \Lambda(k, s).
$$
$$
\Lambda(s, s) = I_d. \ \forall k \geqslant s \geqslant 0.
$$

Note that by eq. (25), we have

$$
\widetilde{\zeta}_{k+1} = \Lambda(k+1, 0) \widetilde{\zeta}_0 + \sum_{i=0}^{k} \Lambda(k+1, i+1) \cdot [-\rho L_i \overline{v}_i + \tau \overline{\omega}_{i+1}].
$$

According to Lemma 7.2 in ref. [37], for any $s \geqslant 1$, $\{P_k\} \in \mathcal{L}_s(\rho^*)$. By Lemma 3 and Hölder inequality, we have for any $e < \min\{s, z\}$,

$$
\begin{aligned}
&\left\| L_k \varphi_k^T [I_m - D^T D] \theta_k \right\|_{L_e} \\
&\leqslant \frac{3\delta_{4s}}{2\sqrt{\rho R}(1-\delta_{4s})} \left\| C^\psi \right\|_{L_t} \cdot \|P_{k-1}\|_{L_s}^{1/2} \cdot \|\theta_k\|_{L_z} \\
&\leqslant \frac{C^\theta \cdot \sqrt{C^P} \left\| C^\psi \right\|_{L_t}}{2\sqrt{\rho R}} \frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}}, \qquad (27)
\end{aligned}
$$

where C^P is the L_s-bound of $\{P_k\}$ and $t = \left(\frac{1}{e} - \frac{1}{s} - \frac{1}{z} \right)^{-1}$.

By eqs. (11) and (12), we have

$$
F_k = (P_k - \rho Q) R^{-1} \psi_k \psi_k^T.
$$

An obvious induction gives $\{F_k\} \in \mathcal{L}_w(\rho^*)$ for any $w < s$. Combining with $\{F_k\} \in \mathcal{S}_p(\rho^*)$, by Lemma A.2 in ref. [22] and Hölder inequality, we have for some positive constant v,

$$
\begin{aligned}
&\left\| \widetilde{\zeta}_{k+1} \right\|_{L_v} \\
&\leqslant \|\Lambda(k+1, 0)\|_{L_p} \cdot \left\| \widetilde{\zeta}_0 \right\|_{L_r} + \rho \left\| \sum_{i=0}^{k} \Lambda(k+1, i+1) L_i v_i \right\|_{L_v} \\
&\quad + \rho \left\| \sum_{i=0}^{k} \Lambda(k+1, i+1) \cdot L_k \varphi_k^T [I_m - D^T D] \theta_k \right\|_{L_v} \\
&\quad + \tau \left\| \sum_{i=0}^{k} \Lambda(k+1, i+1) \cdot \overline{\omega}_{i+1}] \right\|_{L_v} \\
&\leqslant M (1 - \alpha\rho)^{k+1} \cdot \left\| \widetilde{\zeta}_0 \right\|_{L_r} + B_{1,v} \sqrt{\rho} + B_{2,v} \frac{\tau}{\sqrt{\rho}} \\
&\quad + \rho \sum_{i=0}^{k} (1 - \alpha\rho)^{k-i} \frac{M \cdot C^\theta \cdot \sqrt{C^P} \cdot \left\| C^\psi \right\|_{L_t}}{2\sqrt{\rho R}} \frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}} \\
&\leqslant B_{4,v} (1 - \alpha\rho)^{k+1} + B_{1,v} \sqrt{\rho} + B_{2,v} \frac{\tau}{\sqrt{\rho}} \\
&\quad + M \frac{C^\theta \cdot \sqrt{C^P} \cdot \left\| C^\psi \right\|_{L_t}}{2\alpha\sqrt{R}} \frac{3\delta_{4s}}{\sqrt{1-\delta_{4s}}} \\
&= B_{1,v} \sqrt{\rho} + B_{2,v} \frac{\tau}{\sqrt{\rho}} + B_{3,v} \frac{\delta_{4s}}{\sqrt{1-\delta_{4s}}} + \\
&\quad + B_{4,v} (1 - \alpha\rho)^{k+1}, \qquad (28)
\end{aligned}
$$

where $B_{1,v} = M \cdot C_r^n$, $B_{2,v} = M \cdot C_r^\omega$, $B_{3,v} = M \cdot \frac{3C^\theta \cdot \sqrt{C^P} \cdot \|C^\psi\|_{L_t}}{2\alpha\sqrt{R}}$, $B_{4,v} = M \cdot \|\widetilde{\zeta}_0\|_{L_r}$. This proof is completed. ■

6.3 Proof of Corollary 1

Following Remark 13, there exists an integer K defined as eq. (21), such that for any $k \geqslant K$, we have $\|\widetilde{\zeta}_k\|_{L_1} \leqslant C(\tau, \delta_{4s})$. Since $\eta = \max\left\{1, 2C^\xi(\tau, \delta_{4s})\right\}$, then by Markov inequality, we have for $\forall k \geqslant K$,

$$
P\left\{ \left\| \widetilde{\zeta}_k \right\| \geqslant \eta C^{1-\xi}(\tau, \delta_{4s}) \right\} \leqslant \frac{\mathbb{E}\left[\left\| \widetilde{\zeta}_k \right\| \right]}{\eta C^{1-\xi}(\tau, \delta_{4s})} \leqslant \frac{C(\tau, \delta_{4s})}{\eta C^{1-\xi}(\tau, \delta_{4s})}
$$

$$= \frac{C^{\xi}(\tau, \delta_{4s})}{\eta} \leqslant \frac{1}{2},$$

which proofs the corollary. ■

6.4 Proof of Theorem 3

Since $\widehat{\zeta}_k = \zeta_k - \widetilde{\zeta}_k = D\theta_k - \widetilde{\zeta}_k$, then by Corollary 1 and Lemma 2, we have for $\forall k \geqslant K$,

$$P\left\{\left\|\widehat{\theta}_k - \theta_k\right\| \leqslant C_s \eta C^{1-\xi}(\tau, \delta_{4s})\right\}$$
$$\geqslant P\left\{\widehat{\zeta}_k = D\theta_k - \widetilde{\zeta}_k, \left\|\widetilde{\zeta}_k\right\| \leqslant \eta C^{1-\xi}(\tau, \delta_{4s})\right\}$$
$$= P\left\{\left\|\widetilde{\zeta}_k\right\| \leqslant \eta C^{1-\xi}(\tau, \delta_{4s})\right\}$$
$$\geqslant 1 - \frac{C(\tau, \delta_{4s})}{\eta}.$$

Furthermore, C_s becomes smaller as δ_{4s} becomes smaller, in view of the relationship between C_s and δ_{4s} presented in Remark 3. This proof is completed. ■

6.5 Proof of Theorem 4

Before proving Theorem 4, we first introduce Lemma 4 and Lemma 5 from ref. [22].

Lemma 4 Assume that Assumptions 2 and 5 hold, then for any $p < 2t$ and $q < 4t/7$, we have,
(i) There are $\rho^* \in (0, 1)$, and $p \geqslant 2$ such that

$$\{F_k\} \in \mathcal{S}_p(\rho^*) \cap \mathcal{S}(\rho^*). \tag{29}$$

(ii) There is a real number $q \geqslant 3$ together with a bounded function $\phi(m, \rho) \geqslant 0$, with

$$\lim_{m\to\infty, \rho\to 0} \phi(m, \rho) = 0$$

(taking first m to infinity and then ρ to zero) such that $\forall m$, $\forall k$, $\forall \rho \in (0, \rho^*]$

$$\|\mathbb{E}[F_k|\mathcal{F}_{k-m}] - \mathbb{E}[F_k]\|_{L_q} \leqslant \phi(m, \rho). \tag{30}$$

(iii) $L_i \in \mathcal{F}_i$, $\forall i \geqslant 1$, and there exists a $\rho^* \in (0, 1)$ such that

$$\{L_i\} \in \mathcal{L}_r(\rho^*), \quad \{F_i\} \in \mathcal{L}_{2q}(\rho^*) \tag{31}$$

with $r = (1/2 - 1/p - 3/2q)^{-1}$, and with p and q defined as in (i) and (ii).

Lemma 5 Let $\alpha \in (0, 1)$ be a constant. Then for any $\rho \in (0, 1)$, we have

$$\sup_{k\geqslant 0}(1 - \alpha\rho)^k \sqrt{k} = O\left(\rho^{-(1/2)}\right), \tag{32}$$

$$\sum_{k=0}^{\infty}(1 - \alpha\rho)^k k = O\left(\rho^{-2}\right), \tag{33}$$

$$\sum_{k=0}^{\infty}(1 - \alpha\rho)^k \sqrt{k} = O\left(\rho^{-(3/2)}\right), \tag{34}$$

where the " O" constant is dependent on α.

Proof of Theorem 4. Introduce a new sequence $\{\overline{\zeta}_k\}$,

$$\overline{\zeta}_{k+1} = (I - \rho\mathbb{E}[F_k])\overline{\zeta}_k - \rho L_k v_k + \tau\overline{\omega}_{k+1}, \tag{35}$$

where $\overline{\zeta}_0 = \widetilde{\zeta}_0$. Recalling eqs. (22) and (35), by Assumption 4, it follows that

$$\widehat{\Pi}_{k+1} = \mathbb{E}\left[\overline{\zeta}_{k+1}\overline{\zeta}_{k+1}^T\right].$$

Then, by Schwarz inequality,

$$\left\|\Pi_{k+1} - \widehat{\Pi}_{k+1}\right\| = \left\|\mathbb{E}\left[\widetilde{\zeta}_{k+1}\widetilde{\zeta}_{k+1}^T - \overline{\zeta}_{k+1}\overline{\zeta}_{k+1}^T\right]\right\|$$
$$= \left\|\mathbb{E}\left[\left(\widetilde{\zeta}_{k+1} - \overline{\zeta}_{k+1}\right) \cdot \widetilde{\zeta}_{k+1}^T + \overline{\zeta}_{k+1} \cdot \left(\widetilde{\zeta}_{k+1}^T - \overline{\zeta}_{k+1}^T\right)\right]\right\|$$
$$\leqslant \left\|\widetilde{\zeta}_{k+1} - \overline{\zeta}_{k+1}\right\|_{L_2} \cdot \left[\left\|\widetilde{\zeta}_{k+1}\right\|_{L_2} + \left\|\overline{\zeta}_{k+1}\right\|_{L_2}\right]. \tag{36}$$

By Assumption 4, $\{\overline{\omega}_k\}$ and $\{L_k v_k\}$ are m.d.s., and then $\{\overline{\omega}_k\} \in \mathcal{M}_r$ and $\{L_k v_k\} \in \mathcal{M}_r$. We now proceed as in the proof of Theorem 2 and by Lemma 4, then we have

$$\|\overline{\zeta}_{k+1}\|_{L_2} = O\left(\sqrt{\rho} + \frac{\tau}{\sqrt{\rho}} + (1 - \alpha\rho)^{k+1}\right), \tag{37}$$

where $\rho \in (0, 1)$ is a constant. For simplicity, we may take ρ as the same as that in Theorem 2.

Denote

$$\varepsilon_k(\alpha) = \sqrt{\rho} + \frac{\tau}{\sqrt{\rho}} + (1 - \alpha\rho)^{k+1} + \frac{\delta_{4s}}{\sqrt{1 - \delta_{4s}}}. \tag{38}$$

By Assumption 4, Lemma 4, and Theorem 2, we have

$$\|\widetilde{\zeta}_{k+1}\|_{L_2} = O(\varepsilon_k(\alpha)). \tag{39}$$

Collecting eqs. (37) and (39) gives the result

$$\left\|\widetilde{\zeta}_{k+1}\right\|_{L_2} + \left\|\overline{\zeta}_{k+1}\right\|_{L_2} = O\left(\varepsilon_k(\alpha)\right). \tag{40}$$

So our task now is to consider the term $\left\|\widetilde{\zeta}_{k+1} - \overline{\zeta}_{k+1}\right\|_{L_2}$ in eq. (36). By eqs. (19) and (35), we have

$$\widetilde{\zeta}_{k+1} - \overline{\zeta}_{k+1}$$
$$= (I - \rho\mathbb{E}[F_k])\left(\widetilde{\zeta}_k - \overline{\zeta}_k\right) + \rho\left(\mathbb{E}[F_k] - F_k\right)\widetilde{\zeta}_k - \rho L_k u_k,$$

where $u_k = \varphi_k^T[I_m - D^T D]\theta_k$.

Define

$$\Omega(k+1, s) = (I - \rho\mathbb{E}[F_k])\,\Omega(k, s),$$
$$\Omega(s, s) = I, \ \forall k \geqslant s.$$

Then for any $k \geqslant 0$, we have

$$\left\|\widetilde{\zeta}_{k+1} - \overline{\zeta}_{k+1}\right\|_{L_2}$$

$$
\begin{aligned}
&\leqslant \rho \left\| \sum_{i=0}^{k} \Omega(k+1,i+1)\left(\mathbb{E}[F_i]-F_i\right)\widetilde{\zeta}_i \right\|_{L_2} \\
&\quad + \rho \left\| \sum_{i=0}^{k} \Omega(k+1,i+1)L_i u_i \right\|_{L_2} \\
&\leqslant \rho \left\| \sum_{i=0}^{m-1} \Omega(k+1,i+1)\left(\mathbb{E}[F_i]-F_i\right)\widetilde{\zeta}_i \right\|_{L_2} \\
&\quad + \rho \left\| \sum_{i=m}^{k} \Omega(k+1,i+1)\left(\mathbb{E}[F_i]-F_i\right)\left(\widetilde{\zeta}_i-\widetilde{\zeta}_{i-m}\right) \right\|_{L_2} \\
&\quad + \rho \left\| \sum_{i=m}^{k} \Omega(k+1,i+1)\left(\mathbb{E}[F_i]-F_i\right)\widetilde{\zeta}_{i-m} \right\|_{L_2} \\
&\quad + \rho \left\| \sum_{i=0}^{k} \Omega(k+1,i+1)L_i u_i \right\|_{L_2} \\
&\triangleq U_1+U_2+U_3+U_4,
\end{aligned} \tag{41}
$$

where $m = m(\rho) \triangleq \arg\min_{m\geqslant 1}\left[\sqrt{\rho}m + \phi(m,\rho)\right]$.

Notice that $\sqrt{\rho}m(\rho) \leqslant \sqrt{\rho}m(\rho)+\phi(m(\rho),\rho) \leqslant \sqrt{\rho}+\phi(1,\rho)$, which implies that for some constant $c > 0$,

$$
m(\rho) \leqslant 1 + \frac{\phi(1,\rho)}{\sqrt{\rho}} \leqslant \frac{c}{\sqrt{\rho}}, \ \forall \rho \in (0,1).
$$

Hence, $(1-\alpha\rho)^{-m(\rho)}$, $\rho \in (0,1)$ is a bounded function for any $\alpha \in (0,1)$ since $(1-\alpha\rho)^{-m(\rho)} \leqslant (1-\alpha\rho)^{-\frac{c}{\sqrt{\rho}}} \to 1, \ as\ \rho \to 0$. Henceforth, we will directly use this fact and abbreviate $m(\rho)$ to ρ.

The next thing to do in the proof is to consider U_1 in eq. (41) for $k < m$. Denote

$$
s = \left(\frac{1}{r}+\frac{1}{p}+\frac{1}{2q}\right)^{-1}. \tag{42}
$$

From eq. (38), Lemma 4, and Theorem 2, we have $\forall \rho \in (0,\rho^*]$

$$
\left\|\widetilde{\zeta}_{k+1}\right\|_{L_s} = O\left(\varepsilon_k(\alpha)\right). \tag{43}
$$

Due to the fact that s defined by eq. (42) satisfies $\left[s^{-1}+(2q)^{-1}\right]^{-1} > 2$, by eq. (43), the following inequality is obtained

$$
\begin{aligned}
U_1 &\leqslant \rho \sum_{i=0}^{m-1} \|\Omega(k+1,i+1)\| \cdot \|(\mathbb{E}[F_i]-F_i)\|_{L_{2q}} \cdot \left\|\widetilde{\zeta}_i\right\|_{L_s} \\
&= O\left(\rho m\left(\varepsilon_k(\alpha)\right)\right).
\end{aligned} \tag{44}
$$

Another step in the proof is to estimate U_2 in eq. (41). Denote

$$
u = \left(\frac{1}{s}+\frac{1}{2q}\right)^{-1}. \tag{45}
$$

Clearly, $\left\{-\rho L_j v_j + \tau\overline{\omega}_{j+1}\right\} \in \mathcal{M}_r \subset \mathcal{M}_u$. Then by Hölder inequality, it follows that for $\rho \in (0,\rho^*]$ and for any $i \geqslant m$

$$
\begin{aligned}
&\left\|\widetilde{\zeta}_i-\widetilde{\zeta}_{i-m}\right\|_{L_u} \\
&= \left\| \sum_{j=i-m}^{i-1} \left[-\rho F_j\widetilde{\zeta}_j - \rho L_j u_j - \rho L_j v_j + \tau\overline{\omega}_{j+1}\right] \right\|_{L_u} \\
&\leqslant \rho \sum_{j=i-m}^{i-1} \left\|F_j\right\|_{L_{2q}} \left\|\widetilde{\zeta}_j\right\|_{L_s} + \rho \sum_{j=i-m}^{i-1} \left\|L_j u_j\right\|_{L_u} \\
&\quad + \left\| \sum_{j=i-m}^{i-1} -\rho L_j v_j + \tau\overline{\omega}_{j+1} \right\|_{L_u} \\
&= O\left(\rho \sum_{j=i-m}^{i-1} \varepsilon_j(\alpha)\right) + O\left(\frac{m\delta_{4s}}{\sqrt{1-\delta_{4s}}}\right) + O(\sqrt{m}[\rho+\tau]) \\
&= O\left(m(\rho+\tau)\right) + O\left(\frac{m\delta_{4s}}{\sqrt{1-\delta_{4s}}}\right) + O\left(\rho m(1-\alpha\rho)^{i-m}\right).
\end{aligned} \tag{46}
$$

Eqs. (42) and (45), and the definition of r in eq. (31) make it obvious that $1/2 = 1/2q + 1/u$. Eventually, from Lemma 4, eqs. (46) and (29) yield that for any $k \geqslant m$,

$$
\begin{aligned}
U_2 &\leqslant \rho \sum_{i=m}^{k} \|\Omega(k+1,i+1)\| \cdot \|\mathbb{E}\left[F_i\right]-F_i\|_{L_{2q}} \cdot \left\|\widetilde{\zeta}_i-\widetilde{\zeta}_{i-m}\right\|_{L_u} \\
&= O\left(m(\rho+\tau)\right) + O\left(\frac{m\delta_{4s}}{\sqrt{1-\delta_{4s}}}\right) \\
&\quad + O\left(\rho^2 m(k-m)(1-\alpha\rho)^{k-m}\right) \\
&= O\left(m\left(\rho+\tau+\frac{\delta_{4s}}{\sqrt{1-\delta_{4s}}}\right)\right) + O\left(\rho^2 m \sup_{k\geqslant 0}\left\{k(1-\alpha\rho)^k\right\}\right) \\
&= O\left(m\left(\rho+\tau+\frac{\delta_{4s}}{\sqrt{1-\delta_{4s}}}\right)\right).
\end{aligned} \tag{47}
$$

It remains to estimate U_3 in eq. (41). To this end, for any $i \geqslant m$,

$$
F_i - \mathbb{E}\left[F_i\right] = \sum_{j=0}^{m-1} \delta_j(i) + \mathbb{E}\left[F_i|\mathcal{F}_{i-m}\right] - \mathbb{E}\left[F_i\right], \tag{48}
$$

where $\delta_j(i) \triangleq \mathbb{E}\left[F_i|\mathcal{F}_{i-j}\right] - \mathbb{E}\left[F_i|\mathcal{F}_{i-j-1}\right]$ with $j \geqslant 0$ and $i \geqslant m$.

For any fixed $0 \leqslant j \leqslant m-1$, define $e_i = \delta_j(i)\widetilde{\zeta}_{i-m}$, it can be readily verified that $\left\{e_i, \mathcal{F}_{i-j}, i \leqslant m\right\}$ is a m.d.s.. By eq. (43) and the fact that $1/2 \geqslant 1/2q + 1/s$,

$$
\|e_i\|_{L_2} \leqslant 2\|F_i\|_{L_{2q}} \cdot \left\|\widetilde{\zeta}_{i-m}\right\|_{L_s} = O\left(\varepsilon_{i-m}(\alpha)\right),\ i \geqslant m.
$$

Hence, define $S(k,i) \triangleq \sum_{j=i}^{k} e_j$. We have for any $k \geqslant i \geqslant m$,

$$
\begin{aligned}
\|S(k,i)\|_{L_2} &= \sum_{j=i}^{k} \left\|e_j\right\|_{L_2} \\
&= O\left(\left\{\sum_{j=i}^{k}\left(\sqrt{\rho}+\frac{\tau}{\sqrt{\rho}}+(1-\alpha\rho)^{j-m}+\frac{\delta_{4s}}{\sqrt{1-\delta_{4s}}}\right)^2\right\}^{1/2}\right)
\end{aligned}
$$

$$= O\left(\left\{\sum_{j=i}^{k}\left(\sqrt{\rho}+\frac{\tau}{\sqrt{\rho}}+\frac{\delta_{4s}}{\sqrt{1-\delta_{4s}}}\right)^2+(1-\alpha\rho)^{2(j-m)}\right\}^{1/2}\right)$$
$$= O\Bigg(\sqrt{k-i+1}\left(\sqrt{\rho}+\frac{\tau}{\sqrt{\rho}}+\frac{\delta_{4s}}{\sqrt{1-\delta_{4s}}}\right) + (1-\alpha\rho)^{i-m}\rho^{-1/2}\Bigg).$$

Proceeding as in the proof of Lemma A.2 in ref. [22], and combining eq. (38) and Lemma 5, we have for any $k \geqslant m$,

$$\rho\left\|\sum_{i=m}^{k}\Omega(k+1,i+1)\sum_{j=0}^{m-1}\delta_j(i)\widetilde{\zeta}_{i-m}\right\|_{L_2}$$
$$\leqslant \rho\sum_{j=0}^{m-1}\left\|\sum_{i=m}^{k}\Omega(k+1,i+1)e_i\right\|_{L_2}$$
$$\leqslant \rho\sum_{j=0}^{m-1}\|\Omega(k+1,m+1)\|\cdot\|S(k,m)\|_{L_2} + \rho^2\sum_{j=0}^{m-1}\sum_{i=m+1}^{k}\|\Omega(k+1,i+1)\|_{L_2}\cdot\|\mathbb{E}[F_i]\|\cdot\|S(k,i)\|_{L_2}$$
$$= O\left(\sqrt{\rho}m\varepsilon_k\left(\frac{\alpha}{2}\right)\right). \tag{49}$$

Note that by eq. (42) and the definition of r in eq. (31), it is seen that $1/2 = 1/q + 1/s$. Hence from eqs. (30), (43), and Lemma 4, we have for any $\rho \in (0, \rho^*]$,

$$\rho\left\|\sum_{i=m}^{k}\Omega(k+1,i+1)\{\mathbb{E}[F_i|\mathcal{F}_{i-m}]-\mathbb{E}[F_i]\}\widetilde{\zeta}_{i-m}\right\|_{L_2}$$
$$\leqslant \rho\sum_{i=m}^{k}\|\Omega(k+1,i+1)\|\cdot\|\mathbb{E}[F_i|\mathcal{F}_{i-m}]-\mathbb{E}[F_i]\|_{L_q}\cdot\left\|\widetilde{\zeta}_{i-m}\right\|_{L_s}$$
$$= O\left(\phi(m,\rho)\varepsilon_k\left(\frac{\alpha}{2}\right)\right). \tag{50}$$

Combining with eqs. (49) and (50) and noting eq. (48), we see that

$$U_3 = O\left([\sqrt{\rho}m+\phi(m,\rho)]\varepsilon_k\left(\frac{\alpha}{2}\right)\right). \tag{51}$$

An argument similar to the one used in eq. (28) shows that

$$U_4 = O\left(\frac{\delta_{4s}}{\sqrt{1-\delta_{4s}}}\right),$$

this in conjunction with eqs. (44), (47), and (51) yields

$$\left\|\tilde{\zeta}_{k+1}-\overline{\zeta}_{k+1}\right\|$$
$$= O(\rho m\varepsilon_k(\alpha)) + O\left(m\left(\rho+\tau+\frac{\delta_{4s}}{\sqrt{1-\delta_{4s}}}\right)\right) + O\left([\sqrt{\rho}m+\phi(m,\rho)]\,\varepsilon_k\left(\frac{\alpha}{2}\right)\right)+O\left(\frac{\delta_{4s}}{\sqrt{1-\delta_{4s}}}\right)$$
$$= O\left([\sqrt{\rho}m+\phi(m,\rho)]\,\varepsilon_k\left(\frac{\alpha}{2}\right)\right).$$

Then, by substituting this and eq. (40) into eq. (36), we obtain $\forall\rho \in (0, \rho^*]$,

$$\left\|\mathbb{E}\left[\widetilde{\zeta}_{k+1}\widetilde{\zeta}_{k+1}^T\right]\right\| = O\left([\sqrt{\rho}m+\phi(m,\rho)]\left[\varepsilon_k\left(\frac{\alpha}{2}\right)\right]^2\right). \tag{52}$$

Finally, by substituting eq. (38) into eq. (52) we see that Theorem 4 is true. ■

6.6 Proof of Corollary 2

In view of $\{F_k\} \in \mathcal{S}_p(\rho^*)$, the stability of $I-\rho F$ is obtained. Then, it is straightforward that $\widehat{\Pi}_k$ defined in eq. (22) converges at an exponential rate to Π which satisfies that

$$\Pi = (I-\rho F)\Pi(I-\rho F)^T + \rho^2 R_v G + \tau^2 DQ_\omega D^T$$

or

$$F\Pi + \Pi F = \rho F\Pi F + \rho R_v G + \frac{\tau^2}{\rho}DQ_\omega D^T. \tag{53}$$

According to Theorem 2,

$$\Pi_k = O\left(\rho+\frac{\tau^2}{\rho}+\frac{\delta_{4s}^2}{1-\delta_{4s}}\right)+o(1).$$

By Theorem 4,

$$\widehat{\Pi}_k = O\left(\rho+\frac{\tau^2}{\rho}+\frac{\delta_{4s}^2}{1-\delta_{4s}}\right)+o(1).$$

Then,

$$\Pi = O\left(\rho+\frac{\tau^2}{\rho}+\frac{\delta_{4s}^2}{1-\delta_{4s}}\right).$$

Hence, eq. (53) can be further simplified as

$$F\Pi + \Pi F = \rho^2 R_v G + \tau^2 DQ_\omega D^T + O\left(\rho^2+\tau^2+\frac{\rho\delta_{4s}^2}{1-\delta_{4s}}\right).$$

According to the definition of $\overline{R}_v$ and $\overline{R}_\omega$, the solution of Lyapuonv equation can be transformed into

$$\Pi = \rho\overline{R}_v G + \frac{\tau^2}{\rho}\overline{R}_\omega + O\left(\rho^2+\tau^2+\frac{\rho\delta_{4s}^2}{1-\delta_{4s}}\right).$$

Then we have

$$\widehat{\Pi}_k = \rho\overline{R}_v G + \frac{\tau^2}{\rho}\overline{R}_\omega + O\left(\rho^2+\tau^2+\frac{\rho\delta_{4s}^2}{1-\delta_{4s}}\right)+o(1).$$

By substituting this equation into Theorem 4, the proof is completed. ■

7 Simulation results

An example is presented to illustrate the efficacy of the proposed compressed KF algorithm for tracking of high-dimensional sparse signals.

We will estimate 2-sparse signal $\theta_k \in \mathbb{R}^{50}$ with only the first two elements being non-zero. Here, we focus on the time-varying case where $\tau = 1$ and the first two elements of parameter variation in eq. (5) follow $1/k^2 \cdot \mathcal{N}(0, 0.1^2, 2, 1)$. Recall that the notation $\mathcal{N}(\rho, \sigma^2, m, n)$ represents an $m \times n$-dimensional matrix in which every element obeys the Gaussian distribution with mean value ρ and standard deviation σ. Also, the noise sequence v_k is supposed to be independent identically distributed with $\mathcal{N}(0, 0.5^2, 1, 1)$. To generate the regression vectors $\{\varphi_k\}$, assuming that they are 6-sparse, we first generate the last six elements in $\varphi_k \in \mathbb{R}^{50}$ by $x_{k+1} = Ax_k + \xi_k$, $x_0 \sim \mathcal{N}(0, 1, 6, 1)$, where the matrix $A \in \mathbb{R}^{6\times6}$ is a diagonal matrix with diagonal element being 4/5 and $\xi_k \sim \mathcal{N}(0, 1^2, 6, 1)$. It is obvious that the compressed regression vector $\psi_k = D\varphi_k$ satisfies the compressed excitation condition (i.e., Assumption 2) while the original high-dimensional regression vector φ_k fails to meet the excitation condition in (i.e., eq. (8)).

To demonstrate the tracking performance of our algorithm, we set $R = 0.25$ and $Q = 6.7 \times I_5$. Then we set the sensing matrix as the Gaussian matrix $D \sim \mathcal{N}(0, 1/5, 5, 50)$ and resort to the OMP algorithm [46] to yield the high-dimensional sparse estimate by tackling the optimization problem eq. (14) in the reconstruction step. To avoid accidents, we repeat 200 times for our numerical simulations. Figure 1 shows the impact of step size ρ on the performance of our proposed compressed KF algorithm, from which we see that with different step sizes, the tracking errors fall into a small neighborhood of zero and the large step size can promote the performance of the algorithm in a certain sense.

Using the same initial values, we compare our algorithm with the standard KF algorithm (i.e., eqs. (6) and (7)) with $R = 0.25$ and $Q = 6.7 \times I_{50}$ in Figure 2. Figure 2 shows both the tracking error and $\mathrm{tr}(P_k)$ for the standard KF algorithm are apparently larger than that for our compressed KF algorithm. It is straight that even when the standard KF algorithm fails in estimating high-dimensional and sparse parameters, our compressed algorithm can still fulfill the estimation task. This phenomenon is consistent with the theoretical results because the original high-dimensional regression vector φ_k in the standard KF algorithm lacks sufficient excitation condition, while the compressed regression vector $\psi_k = D\varphi_k$ in our compressed KF algorithm satisfies the compressed excitation condition (i.e., Assumption 2). We also compare our algorithm with the compressed LMS algorithm [31] (step size $\rho = 0.2$) and the compressed FFLS algorithm [32] (forgetting factor $\alpha = 0.8$) in Figure 3. From Figure 3, we can see that

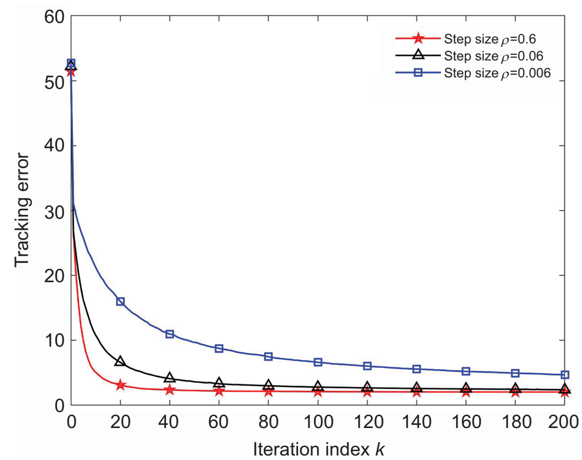

Figure 1 (Color online) Tracking errors for the compressed KF algorithm with different step sizes.

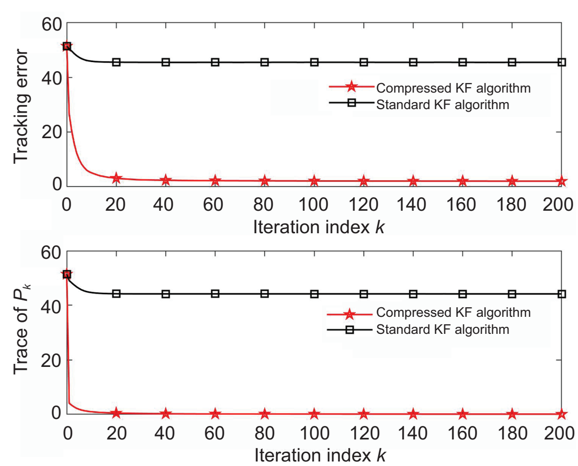

Figure 2 (Color online) The trace of P_k for the standard KF algorithm and the compressed KF algorithm.

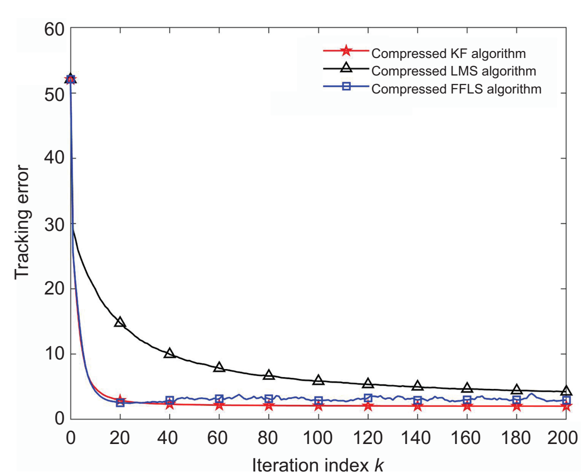

Figure 3 (Color online) Tracking errors for different algorithms.

the tracking error of our compressed KF algorithm is smaller than other compressed algorithms. Hence, our algorithm outperforms the compressed LMS algorithm and the compressed FFLS algorithm in tracking sparse signals.

8 Concluding remarks

In this paper, we propose the compressed Kalman filter algorithm based on the compression-estimation-reconstruction scheme to track unknown high-dimensional sparse signals. Our proposed algorithm performs well in the estimation task even if the traditional KF algorithm may not accurately track unknown sparse signals due to inadequate excitation. Under a compressed excitation condition, we provide stability analysis to establish estimation error bounds for the compressed parameter vector and the original high-dimensional parameter vector with a large probability. The stability results reveal that the estimation error is positively related to the restricted isometry constant. Furthermore, we present the tracking performance analysis of the compressed KF algorithm in terms of the covariance matrix of the compressed tracking error, which shows that the mean square compressed tracking error matrix can be approximately calculated by a linear deterministic difference matrix equation that can be easily evaluated and analyzed.

Our algorithm can be utilized to estimate sparse signals in diverse applications, including channel estimation and field monitoring. Moreover, many interesting problems deserve to be further investigated, for example, incorporating an error feedback scheme to reduce the compression error, optimizing the sensing matrix online, considering the compressed distributed KF algorithm, applying the CS method to estimate unknown sparse attacks, and so on.

This work was supported by the National Key Research and Development Program of China (Grant No. 2022YFB3305600), the National Natural Science Foundation of China (Grant Nos. 61621003, 62141604), the China Postdoctoral Science Foundation (Grant No. 2022M722926), and the Major Key Project of Peng Cheng Laboratory (Grant No. PCL2023AS1-2).

1 Liu Y W, Shi Q, Wei Y J, et al. State of charge estimation by square root cubature particle filter approach with fractional order model of lithium-ion battery. Sci China Tech Sci, 2022, 65: 1760–1771
2 Zhang S Z, Zhang X W. A comparative study of different online model parameters identification methods for lithium-ion battery. Sci China Tech Sci, 2021, 64: 2312–2327
3 Wu J H, Li W, Xiong Z H. Identification of robot dynamic model and joint frictions using a baseplate force sensor. Sci China Tech Sci, 2022, 65: 30–40
4 Liang C X, Xue W C, Fang H T, et al. On distributed Kalman filter based state estimation algorithm over a bearings-only sensor network. Sci China Tech Sci, 2023, 66: 3174–3185
5 Pei Y Q, Gu H B, Liu K X, et al. An overview on the designs of distributed observers in LTI multi-agent systems. Sci China Tech Sci, 2021, 64: 2337–2346
6 Kailath T, Sayed A H, Hassibi B. Linear Estimation. Upper Saddle River: Prentice Hall, 2000
7 Kalman R E. A new approach to linear filtering and prediction problems. J Basic Eng, 1960, 82: 35–45
8 Gan D, Liu Z. Distributed order estimation of ARX model under cooperative excitation condition. SIAM J Control Optim, 2022, 60: 1519–1545
9 Xie S, Guo L. Analysis of normalized least mean squares-based consensus adaptive filters under a general information condition. SIAM J Control Optim, 2018, 56: 3404–3431
10 Carmi A, Gurfil P, Kanevsky D. Methods for sparse signal recovery using Kalman filtering with embedded pseudo-measurement norms and quasi-norms. IEEE Trans Signal Process, 2010, 58: 2405–2409
11 Nourian M, Leong A S, Dey S, et al. An optimal transmission strategy for Kalman filtering over packet dropping links with imperfect acknowledgements. IEEE Trans Control Netw Syst, 2014, 1: 259–271
12 Rohr E R, Marelli D, Fu M. Kalman filtering with intermittent observations: On the boundedness of the expected error covariance. IEEE Trans Automat Contr, 2014, 59: 2724–2738
13 Wang J, Ren X, Dey S, et al. Remote state estimation with usage-dependent Markovian packet losses. Automatica, 2021, 123: 109342
14 Lu A Y, Yang G H. Distributed secure state estimation for linear systems against malicious agents through sorting and filtering. Automatica, 2023, 151: 110927
15 Lu A Y, Yang G H. Secure state estimation under sparse sensor attacks via saturating adaptive technique. IEEE Trans Control Netw Syst, 2023
16 Zhao Z, Xu Y, Li Y, et al. Sparse actuator attack detection and identification: A data-driven approach. IEEE Trans Cybern, 2023, 53: 4054–4064
17 Zhao W X. Parametric identification of hammerstein systems with consistency results using stochastic inputs. IEEE Trans Automat Contr, 2010, 55: 474–480
18 Sjöberg J, Zhang Q, Ljung L, et al. Nonlinear black-box modeling in system identification: A unified overview. Automatica, 1995, 31: 1691–1724
19 Angelosante D, Bazerque J A, Giannakis G B. Online adaptive estimation of sparse signals: Where RLS meets the ℓ_1-norm. IEEE Trans Signal Process, 2010, 58: 3436–3447
20 Babadi B, Kalouptsidis N, Tarokh V. SPARLS: The sparse RLS algorithm. IEEE Trans Signal Process, 2010, 58: 4013–4025
21 Qin Z, Tao J, Xia Y. A proportionate recursive least squares algorithm and its performance analysis. IEEE Trans Circuits Syst II, 2021, 68: 506–510
22 Guo L. Stability of recursive stochastic tracking algorithms. SIAM J Control Optim, 1994, 32: 1195–1225
23 Zhao W, Yin G, Bai E W. Sparse system identification for stochastic systems with general observation sequences. Automatica, 2020, 121: 109162
24 Chen Y, Gu Y, Hero A O. Sparse LMS for system identification. In Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. Taipei, 2009
25 Li J, Li X. Online sparse identification for regression models. Syst Control Lett, 2020, 141: 104710
26 Xu Z, Chang X, Xu F, et al. $L_{1/2}$ regularization: A thresholding representation theory and a fast solver. IEEE Trans Neural Netw Learn Syst, 2012, 23: 1013–1027
27 Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc, 2001, 96: 1348–1360
28 Candés E J, Romberg J K, Tao T. Stable signal recovery from incomplete and inaccurate measurements. Comm Pure Appl Math, 2006, 59: 1207–1223
29 Donoho D L. Compressed sensing. IEEE Trans Inform Theor, 2006, 52: 1289–1306
30 Gan D, Liu Z. Performance analysis of the compressed distributed least squares algorithm. Syst Control Lett, 2022, 164: 105228

31 Xie S, Guo L. Analysis of compressed distributed adaptive filters. Automatica, 2020, 112: 108707

32 Zhao P, Wang X, Xie S, et al. Distribution-free one-pass learning. IEEE Trans Knowl Data Eng, 2021, 33: 951–963

33 Vaswani N. Kalman filtered compressed sensing. In: 2008 15th IEEE International Conference on Image Processing. San Diego, 2008. 893–896

34 Di Lorenzo P, Sayed A H. Sparse distributed learning based on diffusion adaptation. IEEE Trans Signal Process, 2013, 61: 1419–1433

35 Liu Z, Liu Y, Li C. Distributed sparse recursive least-squares over networks. IEEE Trans Signal Process, 2014, 62: 1386–1395

36 Yazdanpanah H, Diniz P S. Recursive least-squares algorithms for sparse system modeling. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, 2017. 3879–3883

37 Guo L, Ljung L. Exponential stability of general tracking algorithms. IEEE Trans Automat Contr, 1995, 40: 1376–1387

38 Guo L, Ljung L. Performance analysis of general tracking algorithms. IEEE Trans Automat Contr, 1995, 40: 1388–1402

39 DeVore R A. Deterministic constructions of compressed sensing matrices. J Complex, 2007, 23: 918–925

40 Xu G, Xu Z. Compressed sensing matrices from fourier matrices. IEEE Trans Inform Theor, 2015, 61: 469–478

41 Baraniuk R, Davenport M, DeVore R, et al. A simple proof of the restricted isometry property for random matrices. Constr Approx, 2008, 28: 253–263

42 Zhang X, Li S. Compressed sensing via dual frame based ℓ_1-analysis with Weibull matrices. IEEE Signal Process Lett, 2013, 20: 265–268

43 Shen T, Dong Y L, He D X, et al. Online identification of time-varying dynamical systems for industrial robots based on sparse Bayesian learning. Sci China Tech Sci, 2022, 65: 386–395

44 Khalifa M O, Abdelhafiz A H, Zerguine A. Sparse channel estimation using adaptive filtering and compressed sampling. In: 2013 International Conference on Computing, Electrical and Electronic Engineering (ICCEEE). Khartoum, 2013. 144–147

45 Bajwa W U, Haupt J, Sayeed A M, et al. Compressed channel sensing: A new approach to estimating sparse multipath channels. Proc IEEE, 2010, 98: 1058–1076

46 Tropp J A, Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans Inform Theor, 2007, 53: 4655–4666

47 van den Berg E, Friedlander M P. Probing the Pareto frontier for basis pursuit solutions. SIAM J Sci Comput, 2009, 31: 890–912

48 Kim S J, Koh K, Lustig M, et al. An interior-Point method for large-scale-regularized least squares. IEEE J Sel Top Signal Process, 2007, 1: 606–617