Distributed State Estimation for Sparse Stochastic Systems Based on Compressed Sensing

Rongjiang Li, Die Gan, *Member, IEEE*, Haibo Gu, *Member, IEEE*, and Jinhu Lü, *Fellow, IEEE*

***Abstract*—This brief proposes a compressed distributed Kalman filter to cooperatively estimate the sparse state vector of a dynamic system with general stochastic coefficients. Based on the compressed sensing theory and the diffusion strategy, each sensor first compresses the original high-dimensional and sparse coefficient matrices via the sensing matrix. Then, each sensor diffuses the local innovation pairs with neighbors to obtain a distributed Kalman estimate in the compressed low-dimensional space. Subsequently, the original high-dimensional sparse state vector can be well recovered by the reconstruction technique. Under the compressed collective stochastic observability condition, the upper bound for the estimation error is established. Note that our theoretical results are established without such stringent conditions as independence or stationarity of the coefficient matrices and are thus applicable to feedback systems. Finally, a simulation example is given to illustrate our theoretical results.**

***Index Terms*—Sparse state estimation, distributed Kalman filter, compressed sensing, stochastic dynamic system.**

I. Introduction

DISTRIBUTED state estimation has attracted considerable research attention in many fields such as environmental monitoring and spacecraft navigation. In the absence of a fusion center, distributed estimation algorithms have advantages over centralized ones in terms of robustness and scalability. Hence, many distributed estimation algorithms have been proposed [1], [2], [3], [4], [5]; among them, the distributed Kalman filter (DKF) is one of the well-known estimators for cooperatively estimating the state of interest. Most studies (e.g., [3], [4], [5]) focus on deterministic coefficient matrices while paying insufficient attention to stochastic ones. In fact, the stochastic dynamic system is ubiquitous, and the corresponding theoretical investigation is also of particular significance.

In many scenarios, the unknown state vectors to be estimated can be sparse, such as speech signals, image signals, and solar waves [6], [7], [8]. Given the prevalence of sparsity, attempts have been made to consider sparsity as a prior to improve the estimation performance. One technique for sparse state estimation is to add a regularization term to the cost function. For instance, reference [8] presented a distributed adaptive filter and utilized the ℓ_0- or ℓ_1-norm as the sparsity penalties, assuming that measurement matrices are independent and identically distributed (i.i.d.). Similarly, reference [9] proposed a distributed sparse identification algorithm by incorporating a ℓ_1-regularization term into the ℓ_2-estimation error. It is worth noting that the above literature considers the sparsity of unknown state vectors, while the sparsity of coefficient matrices is also of interest.

The compressed sensing (CS) theory, as an alternative technique for estimating sparse signals [10], is beneficial to deal with possible degeneration of covariance matrices of measurement matrices (i.e., insufficient excitation), especially when the measurement matrices are high-dimensional and sparse. We remark that [11] and [12] incorporated the CS technique into the least squares and the normalized least mean squares algorithms, respectively, followed by the relatively elegant stability analysis. As for the Kalman filter, [13], [14] developed its variant based on the CS theory and provided simulation examples for the effectiveness of this CS-based Kalman filter. However, the rigorous stability analysis is still lacking.

Inspired by promising advances in CS, this brief investigates the distributed state estimation problem for sparse stochastic systems. The main contributions are summarized as follows:

- A novel compressed distributed Kalman filter (CDKF) is proposed to cooperatively estimate high-dimensional and sparse state vectors according to the CS theory and the diffusion strategy. Based on the compressed data, each sensor diffuses the local innovation pairs with neighbors to generate a low-dimensional estimate. Then, an appropriate signal reconstruction algorithm is adopted to reconstruct the state vector of interest.
- Without independent and stationary signal assumptions, as commonly used in [8], [15], we establish the stability of CDKF under the compressed collective stochastic observability condition, which makes the stability results applicable to stochastic feedback systems.
- Compared with the observability conditions imposed on the uncompressed coefficient matrices [16], [17], the introduced compressed collective stochastic observability condition in this brief is much weaker. It implies that even if the DKF in [16], [17] may fail in sparse state

Manuscript received 27 October 2023; revised 29 January 2024; accepted 27 February 2024. Date of publication 1 March 2024; date of current version 31 July 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFB3305600; in part by the National Natural Science Foundation of China under Grant 62103015 and Grant 62141604; in part by the China Postdoctoral Science Foundation under Grant 2022M722926; and in part by the Fundamental Research Funds for Central Universities of China under Grant YWF-23-03-QB-019. This brief was recommended by Associate Editor C. Wang. *(Corresponding author: Jinhu Lü.)*

Rongjiang Li is with the Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: lirongjiang@amss.ac.cn).

Die Gan is with Zhongguancun Laboratory, Beijing 100094, China (e-mail: gandie@amss.ac.cn).

Haibo Gu and Jinhu Lü are with the School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China, and also with Zhongguancun Laboratory, Beijing 100094, China (e-mail: guhaibo@amss.ac.cn; jhlu@iss.ac.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCSII.2024.3372020.

Digital Object Identifier 10.1109/TCSII.2024.3372020

estimation due to insufficient excitation, our compressed algorithm can accomplish the estimation task well.

The rest of this brief is organized as follows: Section II gives the problem formulation. In Section III, we present the stability analysis. A simulation example is given in Section IV, and the concluding remarks are made in Section V.

Notations: For a vector $z \in \mathbb{R}^n$, $\|z\|_{\ell_0}$ denotes the number of non-zero elements in z. The ℓ_1-norm $\|z\|_{\ell_1}$ is defined by $\|z\|_{\ell_1} = \sum_{i=1}^{m} |z_{(i)}|$ with $z_{(i)}$ being the i-th element of the vector z. I_m describes the m-dimensional square identity matrix. 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 $\|X\| = \{\lambda_{\max}(XX^T)\}^{\frac{1}{2}}$. Given a stochastic matrix Z, let $\|Z\|_p = \{\mathbb{E}[\|Z\|^p]\}^{\frac{1}{p}}$, $p \geq 1$ be its L_p-norm where $\mathbb{E}[\,\cdot\,]$ denotes the expectation operator. We use $\mathbb{E}[\,\cdot \mid \cdot\,]$ to represent the conditional expectation operator and use $\mathbb{P}\{\cdot\}$ to denote the probability. For the set L, the notation $\#L$ denotes its cardinality.

II. Problem Formulation

A. System Model

This brief considers a sensor network consisting of n sensors. For any sensor i, we assume that its signal model is described by a discrete-time stochastic dynamic system as follows:

$$\begin{cases} x_{k+1} = F_k x_k + \omega_{k+1}, \\ y_{k,i} = H_{k,i}^T x_k + v_{k,i}, \end{cases} \tag{1}$$

where $x_k \in \mathbb{R}^s$ is the state vector and $F_k \in \mathbb{R}^{s\times s}$ is the stochastic state evolution matrix. $y_{k,i} \in \mathbb{R}^d$ and $H_{k,i} \in \mathbb{R}^{s\times d}$ represent the measurement vector and the stochastic measurement matrix for sensor i at time instant k, respectively. $\{\omega_k \in \mathbb{R}^s\}$ and $\{v_{k,i} \in \mathbb{R}^d\}$ are noise processes.

Note that in practical scenarios, including channel estimation and high-dimensional data classification [6], [7], sparsity appears not only in the state vector x_k but also in coefficient matrices F_k and $H_{k,i}$. Here, we concentrate on the case where $H_{k,i}$ and x_k are $3t$-sparse and t-sparse, respectively (i.e., $\|H_{k,i}\|_{\ell_0} \leq 3t$, $\|x_k\|_{\ell_0} \leq t$). Furthermore, we remark that the coefficient matrices here are stochastic, while most of the literature focuses on deterministic coefficient matrices [3], [4], [5].

B. Topology Structure

The communication links between sensors are modeled by a weighted undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, where $\mathcal{V} = \{1, \ldots, n\}$ is the set of nodes, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges, and $\mathcal{A} = (a_{ij})_{n\times n}$ is the weighted adjacency matrix. The edge $(i, j) \in \mathcal{E}$ if and only if nodes i and j can communicate with each other. All elements in $\mathcal{A}$ are nonnegative, and $a_{ij} = a_{ji} > 0$ if and only if $(i, j) \in \mathcal{E}$. $\mathcal{N}_i = \{j \in \mathcal{V} | (j, i) \in \mathcal{E}\}$ represents the set of neighbors of the node i and the node i is also included. A path between nodes i and j is defined as a sequence of edges $(i, i_1), (i_1, i_2), \ldots, (i_k, j) \in \mathcal{E}$ with distinct nodes $i_l \in \mathcal{V}, l = 1, \ldots, k$. The graph is connected if there exists a path between any pair of distinct nodes i and j. The diameter $D_{\mathcal{G}}$ of $\mathcal{G}$ is defined as the maximum shortest path length between any two nodes. Here, we assume that weighted adjacency matrix $\mathcal{A}$ is doubly stochastic, i.e., the row and column sums of matrix $\mathcal{A}$ are all equal to one.

Algorithm 1 Compressed Distributed Kalman Filter

Input: $\{y_{k,i}, H_{k,i}\}_{i\in\mathcal{V}}$, $\{F_k\}$, $k = 0, 1, 2, \ldots$
Output: $\{\widehat{x}_{k+1,i}\}_{i\in\mathcal{V}}$, $k = 0, 1, 2, \ldots$
for every sensor $i = 1, 2, \ldots, n$ **do**
Intialize: Begin with an arbitrary initial estimate $\widehat{\zeta}'_{0,i}$ and an initial covariance matrix $P'_{0,i} > 0$.
for each time $k = 0, 1, 2, \ldots$ **do**
Step 1. Compression: $\varphi_{k,i} = DH_{k,i}$, $F'_k = DF_kD^T$.
Step 2. Estimation in a low-dimensional dimension.
(1) Initialization process of local innovation pair

$$\xi^0_{k,i} = \varphi_{k,i}R^{-1}_{k,i}\varphi^T_{k,i},\ \eta^0_{k,i} = \varphi_{k,i}R^{-1}_{k,i}y_{k,i};$$

(2) Diffusion process for $l = 0, 1, \ldots, L$ with $L \geq D_{\mathcal{G}}$

$$\xi^{l+1}_{k,i} = \textstyle\sum_{j\in\mathcal{N}_i} a_{ij}\xi^l_{k,j},\ \ \eta^{l+1}_{k,i} = \textstyle\sum_{j\in\mathcal{N}_i} a_{ij}\eta^l_{k,j};$$

(3) Measurement update process

$$P^{-1}_{k,i} = P'^{-1}_{k,i} + \xi^L_{k,i},\ \ P^{-1}_{k,i}\widehat{\zeta}_{k,i} = P'^{-1}_{k,i}\widehat{\zeta}`_{k,i} + \eta^L_{k,j};$$

(4) State prediction process

$$\widehat{\zeta}`_{k+1,i} = F`_k\widehat{\zeta}_{k,i},\ \ P'_{k+1,i} = F`_kP_{k,i}F^{`T}_k + Q_k. \tag{7}$$

Step 3. Reconstruction:

$$\widehat{x}_{k+1,i} = \arg\min_{x\in\mathcal{X}} \|x\|_{\ell_1}, \tag{8}$$

where $\mathcal{X} = \{x \in \mathbb{R}^s | \|Dx - \widehat{\zeta}'_{k+1,i}\| \leq C\}$.

C. Compressed Distributed Kalman filter

In sensor networks, the DKF algorithm is designed to cooperatively estimate the unknown state vector via local communications and interactions among sensors. Note that [17] proposed a DKF based on the diffusion strategy and established its stability under the collective stochastic observability condition. However, this algorithm suffers from several weaknesses when state vectors and coefficient matrices are high-dimensional and sparse. On the one hand, the high-dimensional local information is diffused over the sensor network, which may yield high computational costs. On the other hand, the collective stochastic observability condition proposed in [17] is hard to satisfy due to sparsity.

To improve this situation, we propose the CDKF, which relies on the CS theory to reduce the computational cost and improve the estimation performance in sparse and high-dimensional scenarios (see Algorithm 1). Specifically, by virtue of a sensing matrix $D \in \mathbb{R}^{l\times s}$ ($t \leq l \ll s$), the compressed data $\varphi_{k,i} = DH_{k,i} \in \mathbb{R}^{l\times d}$ and $F'_k = DF_kD^T \in \mathbb{R}^{l\times l}$ are obtained at k. Then in Step 2, we adopt the combine-then-adapt strategy to obtain a low-dimensional estimate $\widehat{\zeta}'_{k+1,i}$ for the compressed unknown parameter $\zeta_k = Dx_k \in \mathbb{R}^l$. Note that the original model (1) can be rewritten as follows:

$$\begin{aligned} \zeta_{k+1} &= F'_k\zeta_k + DF_k(I_s - D^TD)x_k + \overline{\omega}_{k+1} \\ &= F'_k\zeta_k + \widehat{\omega}_{k+1}, \\ y_{k,i} &= \varphi^T_{k,i}\zeta_k + H^T_{k,i}(I_s - D^TD)x_k + v_{k,i} \\ &= \varphi^T_{k,i}\zeta_k + \overline{v}_{k,i}, \end{aligned} \tag{2}$$

where $\widehat{\omega}_{k+1}$ and $\overline{v}_{k,i}$ are transformed according to $\widehat{\omega}_{k+1} = DF_k(I_s - D^TD)x_k + \overline{\omega}_{k+1}$ with $\overline{\omega}_k = D\omega_k$, and $\overline{v}_{k,i} = H^T_{k,i}(I_s - D^TD)x_k + v_{k,i}$. In this case, we regard $\widehat{\omega}_{k+1}$ and $\overline{v}_{k,i}$ as the new "noise" terms. Hence, the sensor i first diffuses local innovation pairs $\{\varphi_{k,i}R^{-1}_{k,i}\varphi^T_{k,i}, \varphi_{k,i}R^{-1}_{k,i}y_{k,i}\}$ with its neighbors and then performs the Kalman iteration based on the compressed data with $Q_k \geq Q > 0$ and $R_{k,i} \geq \alpha_i I_d > 0$ being arbitrarily chosen. Finally, in Step 3, we tackle the convex optimization problem (8) to recover a high-dimensional estimate $\{\widehat{x}_{k+1,i}\}$ for the original state vector x_k.

D. Error Equation

To analyze the CDKF, we first need to derive the compressed estimation error equation. For every sensor, we define $\widetilde{\zeta}'_{k,i} = \zeta_k - \widehat{\zeta}'_{k,i}$. Then by (2) and (7), we have $\widetilde{\zeta}'_{k+1,i} = F'_k(\zeta_k - \widehat{\zeta}_{k,i}) + \widehat{\omega}_{k+1}$. From Step 2, we obtain that

$$P_{k,i}^{-1} = P_{k,i}'^{-1} + \sum_{j=1}^{n} a_{ij}^{(L)} \varphi_{k,j} R_{k,j}^{-1} \varphi_{k,j}^T, \tag{3}$$

$$P_{k,i}^{-1} \widehat{\zeta}_{k,i} = P_{k,i}'^{-1} \widehat{\zeta}'_{k,i} + \sum_{j=1}^{n} a_{ij}^{(L)} \varphi_{k,j} R_{k,j}^{-1} y_{k,j}, \tag{4}$$

where $a_{ij}^{(L)}$ is the element in the i-th row and j-th column of $\mathcal{A}^L$ and $\mathcal{A}^L$ is the L-th power of $\mathcal{A}$. Hence by (3), we have

$$\zeta_k = \left(P_{k,i} P_{k,i}'^{-1} + P_{k,i} \sum_{j=1}^{n} a_{ij}^{(L)} \varphi_{k,j} R_{k,j}^{-1} \varphi_{k,j}^T \right) \zeta_k. \tag{5}$$

Combining (4) and (5), we can obtain the compressed estimation error equation as follows:

$$\begin{aligned}\widetilde{\zeta}'_{k+1,i} &= F'_k P_{k,i} P_{k,i}'^{-1} \widetilde{\zeta}'_{k,i} \\ &\quad - F'_k P_{k,i} \sum_{j=1}^{n} a_{ij}^{(L)} \varphi_{k,j} R_{k,j}^{-1} \overline{v}_{k,j} + \widehat{\omega}_{k+1}.\end{aligned} \tag{6}$$

For convenience, we introduce the following notations:

$$\begin{aligned}
&P_k \triangleq \mathrm{diag}\{P_{k,1}, \ldots, P_{k,n}\},\ P'_k \triangleq \mathrm{diag}\{P'_{k,1}, \ldots, P'_{k,n}\},\\
&\overline{F}_k \triangleq \mathrm{diag}\{F'_k, \ldots, F'_k\},\ \overline{Q}_k \triangleq \mathrm{diag}\{Q_k, \ldots, Q_k\}\\
&\widetilde{Z}_k \triangleq \mathrm{col}\{\widetilde{\zeta}'_{k,1}, \ldots, \widetilde{\zeta}'_{k,n}\},\ \Pi_k \triangleq \mathrm{diag}\{\varphi_{k,1}, \ldots, \varphi_{k,n}\},\\
&R_k \triangleq \mathrm{diag}\{R_{k,1}, \ldots, R_{k,n}\},\ \widehat{\Omega}_k \triangleq \mathrm{col}\{\widehat{\omega}_k, \ldots, \widehat{\omega}_k\},\\
&\overline{V}_k \triangleq \mathrm{col}\{\overline{v}_{k,1}, \ldots, \overline{v}_{k,n}\},\ \mathscr{A} \triangleq \mathcal{A} \otimes I_l \in \mathbb{R}^{nl \times nl},
\end{aligned}$$

where $\mathrm{col}\{\cdots\}$ denotes a vector by stacking the specified vectors, $\mathrm{diag}\{\cdots\}$ denotes a block matrix formed in a diagonal manner of the corresponding vectors or matrices. $\otimes$ represents the Kronecker product.

By the above notations, we rewrite the compressed estimation error equation (6) in the following compact form:

$$\widetilde{Z}_{k+1} = \overline{F}_k P_k P_k'^{-1} \widetilde{Z}_k - \overline{F}_k P_k \mathscr{A}^L \Pi_k R_k^{-1} \overline{V}_k + \widehat{\Omega}_{k+1}. \tag{9}$$

III. Stability of the Compressed Distributed Kalman Filter

A. Assumptions

Definition 1 [10]: For an integer t and the sensing matrix $D \in \mathbb{R}^{l \times s} (1 \leq t \leq l)$. We say that the matrix D satisfies the restricted isometry property (RIP) of order t if there exists a constant $\delta_t \in [0, 1)$, which is the smallest quantity such that

$$(1 - \delta_t)\|b\|^2 \leq \|D_L b\|^2 \leq (1 + \delta_t)\|b\|^2,\ \ \forall b \in \mathbb{R}^{\#L} \tag{10}$$

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

Remark 1: The RIP characterizes matrices which are nearly orthonormal, at least when operating on sparse vectors. By (10), it is straightforward that $1 - \delta_t \leq \lambda_{\min}(D_L^T D_L) \leq \lambda_{\max}(D_L^T D_L) \leq 1 + \delta_t$.

Assumption 1: The sensing matrix $D \in \mathbb{R}^{l \times s}$ satisfies the RIP of order $4t$ with δ_{4t} being $4t$-th RIP constant.

Assumption 2: The graph $\mathcal{G}$ is connected.

Define the state transition matrix $\Psi(k, j)$ as follows: $\Psi(k, j) = F'_{k-1}, \ldots F'_j,\ \forall k \geq j+1;\ \ \Psi(j, j) = I_l$.

Assumption 3 (compressed collective stochastic observability condition): For any $\epsilon > 0$, there exists $\delta > 0$ such that for any k, $\mathbb{P}\{\lambda_{\min}(G(k+h, k)) > \delta | \mathcal{F}_{k-1}\} > 1 - \epsilon$, where $h > 0$ is an integer, and $\mathcal{F}_k = \sigma\{F'_j, \varphi_{j,i}, j \leq k, 1 \leq i \leq n\}$, $G(k+h, h)$ is the collective observability matrix, i.e., $G(k+h, h) = \sum_{i=1}^{n} \sum_{j=k+1}^{k+h} \Psi^T(j, k) \varphi_{j,i} \varphi_{j,i}^T \Psi(j, k)$.

Remark 2: Assumption 3 is expressed in probability form due to the stochasticity of our system model. Then, we give some intuitive explanations in terms of "collective" and "compressed", and these characteristics are illustrated in the simulation part. (i) "Collective". Compared to the observability condition for the single case in [18], our assumption contains not only temporal union information but also spatial union information of all the sensors, which implies that multiple sensors can cooperate to accomplish the estimation task even if any individual sensor cannot. (ii) "Compressed". Compared to collective observability conditions in [16], [17], Assumption 3 is assumed for the compressed coefficient matrices instead of the non-compressed ones. Thus, Assumption 3 is much weaker than that in [16], [17], which implies that the CDKF may still get the compressed estimation results stably even if the non-compressed DKFs cannot fulfill the estimation tasks.

Assumption 4: For some $r \geq 1$, there exist a positive constant β such that for any $k \geq 0$,

(i) $\sup_k \mathbb{E}[\|\varphi_{k,i}\|^{32r}] < \infty,\ i \in \mathcal{V}$;

(ii) $\sup_{k \leq j \leq m \leq k+h} \mathbb{E}[\|\Psi(m, j)\|^{32r+\beta}] < \infty$;

(iii) $\sup_k \mathbb{E}[\|\Psi(k+h, k)\|^{16r} | \mathcal{F}_{k-1}] < \infty$,

where $h > 0$ is defined in Assumption 3.

Assumption 5: For any k, we assume that

(i) $\sup_{k \leq j \leq k+h} \mathbb{E}[\|\varphi_{j,i}\|^8 | \mathcal{F}_{k-1}] < \infty,\ \ i \in \mathcal{V}$;

(ii) $\sup_{k \leq j \leq m \leq k+h} \mathbb{E}[\|\Psi(m, j)\|^{8+\beta} | \mathcal{F}_{k-1}] < \infty$,

where constants $h, \beta > 0$ are defined in Assumption 4.

Assumption 6: The initial value and the noises satisfy: $\mathbb{E}[\|\zeta_0\|^{2r}] < \infty$, $\sup_k \mathbb{E}[\|\omega_{k+1}\|^{2r} + \|v_{k,i}\|^{4r}] < \infty,\ \ i \in \mathcal{V}$.

B. The Main Results

In this subsection, we present the stability of our proposed CDKF algorithm in the following theorem.

Theorem 1: Assume that Assumptions 1-6 hold. If the unknown state vector is L_{8r}-bounded, i.e., $\sup_k \|x_k\|_{8r} < \infty$, where $r > 0$ is defined in Assumption 4, the compressed estimation error $\widetilde{Z}_k$ defined in (9) is L_r-bounded, i.e., $\sup_k \|\widetilde{Z}_k\|_r \triangleq C(\delta_{4t})$, where $C(\delta_{4t})$ is a constant irrespective to k and is positively related to the RIP constant δ_{4t}.

Proof: Define $\Delta_{k+1} = \widehat{\Omega}_{k+1} - \overline{F}_k P_k \mathscr{A}^L \Pi_k R_k^{-1} \overline{V}_k$. Then from (9), we get $\widetilde{Z}_{k+1} = \prod_{j=0}^{k} (\overline{F}_j P_j P_j'^{-1}) \widetilde{Z}_0 + \sum_{j=1}^{k+1} \prod_{m=j}^{k} (\overline{F}_m P_m P_m'^{-1}) \Delta_j$. By [17, Th. 2], we know that there exist constants $M > 0$ and $\lambda \in [0, 1)$ such that for any $k \geq j \geq 0$, $\|\prod_{m=j}^{k} (\overline{F}_m P_m P_m'^{-1})\|_{2r} \leq M \lambda^{k-j+1}$, where r is defined in Assumption 4.

Then by Assumption 6, and Hölder inequality, we have

$$\begin{aligned}\|\widetilde{Z}_{k+1}\|_r &\le \left\|\prod_{j=0}^{k}\left(\overline{F}_j P_j P_j'^{-1}\right)\right\|_{2r}\|\widetilde{Z}_0\|_{2r}\\&\quad+\sum_{j=1}^{k+1}\left\|\prod_{m=j}^{k}\left(\overline{F}_m P_m P_m'^{-1}\right)\right\|_{2r}\left\|\Delta_j\right\|_{2r}\\&\le O(\lambda^{k+1})+M\sum_{j=1}^{k+1}\lambda^{k-j+1}\left\|\Delta_j\right\|_{2r}\\&=O(1)+M\sum_{j=0}^{k}\lambda^{j}\left\|\Delta_{k-j+1}\right\|_{2r}.\end{aligned}\tag{11}$$

So our problem reduces to estimate $\|\Delta_k\|_{2r}$. By Hölder inequality and Minkowski inequality, we have, for all $k \ge 1$,

$$\begin{aligned}\|\Delta_k\|_{2r} &= \|\widehat{\Omega}_k - \overline{F}_{k-1}P_{k-1}\mathscr{A}^L\Pi_{k-1}R_{k-1}^{-1}\overline{V}_{k-1}\|_{2r}\\&\le \|\widehat{\Omega}_k\|_{2r} + \|\overline{F}_{k-1}P_{k-1}\mathscr{A}^L\Pi_{k-1}R_{k-1}^{-1}\|_{4r}\left\|\overline{V}_{k-1}\right\|_{4r}.\end{aligned}\tag{12}$$

Clearly, $\mathscr{A}^L$ is still a doubly stochastic matrix. By Lemma 4.2 in [20] and (3), we have $\mathscr{A}^L\Pi_k R_k^{-1}\Pi_k^T\mathscr{A}^L \le P_k^{-1}$. By (7), we have $P'_{k+1} = \overline{F}_k P_k \overline{F}_k^T + \overline{Q}_k$. Then it can be seen that

$$\begin{aligned}&\|\overline{F}_k P_k\mathscr{A}^L\Pi_k R_k^{-1}\|^2\\&= \lambda_{\max}(\overline{F}_k P_k\mathscr{A}^L\Pi_k R_k^{-2}\Pi_k^T\mathscr{A}^L P_k\overline{F}_k^T)\\&\le \lambda_{\max}(P_k^{\frac{1}{2}}\overline{F}_k^T\overline{F}_k P_k^{\frac{1}{2}})\lambda_{\max}(P_k^{\frac{1}{2}}\mathscr{A}^L\Pi_k R_k^{-2}\Pi_k^T\mathscr{A}^L P_k^{\frac{1}{2}})\\&\le \lambda_{\max}(P'_{k+1})\lambda_{\max}(R_k^{-\frac{1}{2}}\Pi_k^T\mathscr{A}^L P_k^{\frac{1}{2}}P_k^{\frac{1}{2}}\mathscr{A}^L\Pi_k R_k^{-\frac{1}{2}}R_k^{-1})\\&\le \lambda_{\max}(P'_{k+1})\lambda_{\max}(\mathscr{A}^L\Pi_k R_k^{-1}\Pi_k^T\mathscr{A}^L P_k)\lambda_{\max}(R_k^{-1})\\&\le \|P'_{k+1}\|\mathrm{tr}(R_k^{-1}) \le d\|P'_{k+1}\|\sum_{i=1}^{n}\alpha_i^{-1},\end{aligned}$$

where d is the dimension of $v_{k,i}$. According to [17, Th. 1], we know for $k \ge 0$, $\|P'_k\|_{2r} = O(1)$. Then we have $\|\overline{F}_k P_k\mathscr{A}^L\Pi_k R_k^{-1}\|_{4r} \le \sqrt{d\sum_{i=1}^{n}\frac{1}{\alpha_i}}\|P'_{k+1}\|_{2r}^{\frac{1}{2}} = O(1)$.

Now, it remains to estimate $\|\overline{V}_k\|_{4r}$ and $\|\widehat{\Omega}_k\|_{2r}$ in (12). Since $\|x_k\|_{\ell_0} \le t$ and $\|H_{k,i}\|_{\ell_0} \le 3t$, we first define index sets of nonzero elements in x_k and $H_{k,i}(j)$ as L_t and L_{3t} where $H_{k,i}(j)$ denotes the j-th column of the matrix $H_{k,i}$. Then, we extract elements indexed by $L = L_t \cup L_{3t}$ from x_k, $H_{k,i}(j)$, and the corresponding columns of the sensing matrix D. Without loss of generality, assume that $\#L = 4t$. Then, denote the resulting vectors and matrix as $x_{k,4t}$, $H_{k,i,4t}(j)$, and D_{4t}, respectively. Then by Assumption 1, C_r-inequality, and some properties of L_p-norm, we have for some $L > 0$,

$$\begin{aligned}&\|H_{k,i}^T(I_s - D^T D)x_k\|_{4r}\\&= \|\mathrm{col}\{H_{k,i}^T(1)(I_s - D^T D)x_k, \ldots, H_{k,i}^T(d)(I_s - D^T D)x_k\}\|_{4r}\\&\le d^{\frac{2r-1}{4r}}\sum_{j=1}^{d}\|H_{k,i}^T(j)(I_s - D^T D)x_k\|_{4r}\\&= d^{\frac{2r-1}{4r}}\sum_{j=1}^{d}\|H_{k,i,4t}^T(j)(I_{4t} - D_{4t}^T D_{4t})x_{k,4t}\|_{4r}\\&\le d^{\frac{2r-1}{4r}}\sum_{j=1}^{d}\|H_{k,i,4t}(j)\|_{8r}\|(I_{4t} - D_{4t}^T D_{4t})\|\|x_{k,4t}\|_{8r}\\&\le 3\delta_{4t}\cdot d^{\frac{2r-1}{4r}}\sum_{j=1}^{d}\|H_{k,i,4t}(j)\|_{8r}\|x_{k,4t}\|_{8r}\\&\le 3\delta_{4t}\cdot d^{\frac{2r-1}{4r}}\cdot L\|H_{k,i}\|_{8r}\|x_k\|_{8r}\\&\le \frac{3\delta_{4t}}{\sqrt{1-\delta_{4t}}}\cdot d^{\frac{2r-1}{4r}}\cdot L\|\varphi_{k,i}\|_{8r}\|x_k\|_{8r} \triangleq C_0(\delta_{4t}).\end{aligned}\tag{13}$$

Then, by Assumption 4, and the L_{8r}-bounded condition for x_k, it is yielded that $C_0(\delta_{4t})$ is the constant irrespective to k and is positively related to the RIP constant δ_{4t}.

Based on this, by C_r-inequality and Assumption 6, it is yielded that

$$\begin{aligned}\|\overline{V}_k\|_{4r} &\le n^{\frac{2r-1}{4r}}\sum_{i=1}^{n}\|\overline{v}_{k,i}\|_{4r}\\&\le n^{\frac{2r-1}{4r}}\sum_{i=1}^{n}\left(C_0(\delta_{4t}) + \|v_{k,i}\|_{4r}\right) \triangleq C_1(\delta_{4t}).\end{aligned}\tag{14}$$

Similarly, we obtain that

$$\begin{aligned}\|\widehat{\omega}_{k+1}\|_{2r} &\le \|DF_k(I_s - D^T D)x_k\|_{2r} + \|\overline{\omega}_{k+1}\|_{2r}\\&\le \|DF_k\|_{4r}\|(I_s - D^T D)x_k\|_{4r} + \|D\omega_{k+1}\|_{2r}\\&\le 3\delta_{4t}\|D\|\|F_k\|_{4r}\|x_k\|_{4r} + \sqrt{1+\delta_{4t}}\|\omega_k\|_{2r}.\end{aligned}$$

and $\|\widehat{\Omega}_k\|_{2r} = n\|\widehat{\omega}_k\|_{2r} \triangleq C_2(\delta_{4t})$. Substituting this equation and (14) into (12), we have for all $k \ge 1$, $\|\Delta_k\|_{2r} = C_2(\delta_{4t}) + O(1)\cdot C_1(\delta_{4t}) = O(1)$. Finally, by equation (11), the proof is completed. ■

Remark 3: For instance, let the sensing matrix D be a Gaussian or Bernoulli random matrix, and then when dimensions of the sensing matrix D satisfy the inequality $l > 480t\log(s/4t)/\delta_{4t}^3$, the RIP constant δ_{4t} can be arbitrarily small [19]. On this basis, if the magnitudes of the noise and parameter variation are also small then the upper bound of the compressed estimation error will be small.

By Theorem 1 and the Step 3 in Algorithm 1, we obtain the upper bound for the estimation error of the original state vector in Corollary 1. The corollary will be based on the following result on the reconstruction error in [10].

Lemma 1: Assume that the t-sparse signal $x \in \mathbb{R}^s$ obeys $z = Dx + \varepsilon$ where $D \in \mathbb{R}^{l\times s}$ satisfies $4t$-th RIP with t being subject to $\delta_{3t} + 3\delta_{4t} < 2$ and $\|\varepsilon\| \le C$. Then the recovered signal x^* derived from solving the optimization problem $x^* = \arg\min_{x\in\mathcal{X}}\|x\|_{\ell_1}$ where $\mathcal{X} = \{x \in \mathbb{R}^s | \|Dx - z\| \le C\}$ obeys $\|x - x^*\| \le C_t C$, with $C_t \triangleq \frac{4}{\sqrt{3(1-\delta_{4t})}-\sqrt{1+\delta_{3t}}} > 0$.

Corollary 1: Suppose that $D \in \mathbb{R}^{l\times s}$ satisfies $4t$-th RIP with t satisfying $\delta_{3t} + 3\delta_{4t} < 2$. Under the same conditions as used in Theorem 1, for any given constant $\gamma \in (0, 1)$, $\mathbb{P}(\|x_k - \widehat{x}_{k,i}\| \le C_t\Gamma C^{1-\gamma}(\delta_{4t})) \ge 1 - \frac{C^\gamma(\delta_{4t})}{\Gamma}$ holds where $\Gamma = \max\{1, 2C^\gamma(\delta_{4t})\}$, $C(\delta_{4t})$ is defined in Theorem 1, and C_t depends only on δ_{4t}.

Remark 4: The proof of Corollary 1 can be readily accomplished according to Markov inequality, so it is omitted. By Theorem 1, we know that $C(\delta_{4t})$ gets small as the RIP constant δ_{4t} gets small. Moreover, if $C(\delta_{4t})$ is close to zero, we have $\Gamma = 1$. Thus, the estimation error will be stable in a small neighborhood of zero with a large probability.

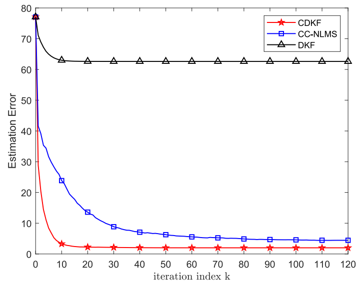
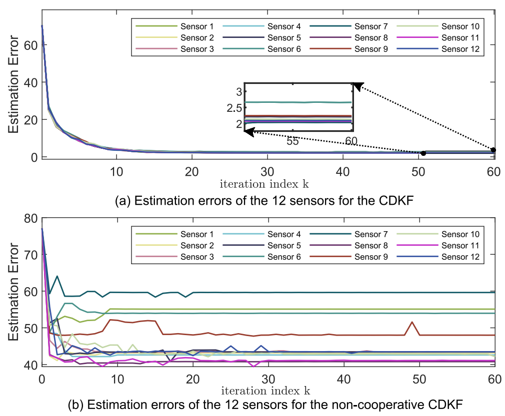


Fig. 1. Estimation errors for different algorithms.



(a) Estimation errors of the 12 sensors for the CDKF

(b) Estimation errors of the 12 sensors for the non-cooperative CDKF

Fig. 2. Estimation errors of the 12 sensors for different algorithms.

IV. A Simulation Example

In a sensor network whose weights are generated by the Metropolis rule [17], 12 sensors cooperate to estimate a 2-sparse state vector $x_k \in \mathbb{R}^{80}$ with only the last two elements being non-zero. Here, we assume $F_k = I_{80}$ and the last two elements of ω_k in (1) follow $1/k^2 \cdot \mathcal{N}(0, 0.1^2)$. The noise sequence $\{v_{k,i} \in \mathbb{R}\}$ is supposed to be i.i.d. with Gaussian distribution $\mathcal{N}(0, 0.2^2/i)$. All elements of the stochastic measurement matrix $H_{k,i} \in \mathbb{R}^{80}$ are assumed to be zero except for the i-th element which is generated by $1.1^k + \sum_{p=0}^{k-1} 1.1^p \tau_{k-p,i}$ with $\tau_{p,i} \sim \mathcal{N}(0, 0.1^2/i)$. Obviously, the compressed measurement matrix $\varphi_{k,i} = DH_{k,i} \in \mathbb{R}^5$ satisfies Assumption 3 while the uncompressed high-dimensional measurement matrix $H_{k,i}$ fails to meet the observability condition in [17].

To demonstrate the estimation performance of our algorithm, we repeat our algorithm, the DKF in [17], and compressed consensus normalized least mean squares algorithm (CC-NLMS) in [12] for ten times with the same initial values. As for the CDKF and CC-NLMS, we set the sensing matrix as the Gaussian matrix $D \sim \mathcal{N}(0, 1/5, 5, 80)$ and utilize the OMP algorithm [21] to perform the reconstruction step. The Fig. 1. shows that the estimation error of CDKF is apparently smaller than that of CC-NLMS while the estimation error of the DKF stays large. Also, we compare the CDKF with the non-cooperative CDKF (i.e., the CDKF with $\mathcal{A} = I_{12}$) in Fig. 2. to demonstrate the cooperative effect of sensors. We can see that the estimation error for the CDKF falls into a small neighborhood of 0, while for the non-cooperative case, the estimation error of every sensor is large.

V. Concluding Remarks

This brief proposes the CDKF based on the compression-estimation-reconstruction scheme to estimate the sparse state vectors. Under a compressed collective stochastic observability condition, we establish the estimation error bound. It is shown that our algorithm performs well in the estimation task even if the traditional non-compressed DKF algorithm may not accurately estimate the unknown sparse state vector due to inadequate excitation. Many interesting problems deserve further investigation, for instance, incorporating an error feedback scheme to reduce the compression error.

References

[1] F. S. Cattivelli and A. H. Sayed, "Analysis of spatial and incremental LMS processing for distributed estimation," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1465–1480, Apr. 2011.

[2] Q. Wang, X. Dong, J. Lü, and Z. Ren, "Fully distributed adaptive formation tracking for directed heterogeneous multi-agent systems with unknown input," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 70, no. 2, pp. 700–704, Feb. 2023.

[3] Y. Luo, J. Zhou, and W. Yang, "Distributed state estimation with colored noises," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 6, pp. 2807–2811, Jun. 2022.

[4] D.-J. Xin, L.-F. Shi, and X. Yu, "Distributed Kalman filter with faulty/reliable sensors based on wasserstein average consensus," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 4, pp. 2371–2375, Apr. 2022.

[5] Z. Peng, Y. Zhang, G. Wen, S. Yang, and T. Huang, "Optimal distributed filter for state and unknown input estimations in multi-sensor networks," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 70, no. 11, pp. 4128–4132, Nov. 2023.

[6] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[7] S. Vinga, "Structured sparsity regularization for analyzing high-dimensional omics data," *Brief. Bioinf.*, vol. 22, no. 1, pp. 77–87, Jan. 2021.

[8] Z. Liu, Y. Liu, and C. Li, "Distributed sparse recursive least-squares over networks," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1386–1395, Mar. 2014.

[9] D. Gan and Z. Liu, "Distributed sparse identification for stochastic dynamic systems under cooperative non-persistent excitation condition," *Automatica*, vol. 151, May 2023, Art. no. 110958.

[10] E. J. Candès, J. K. Romberg, and T. Tao "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[11] D. Gan and Z. Liu, "Performance analysis of the compressed distributed least squares algorithm," *Syst. Control Lett.*, vol. 164, Jun. 2022, Art. no. 105228.

[12] S. Xie and L. Guo, "Analysis of compressed distributed adaptive filters," *Automatica*, vol. 112, Feb. 2020, Art. no. 108707.

[13] A. Carmi, P. Gurfil, and D. Kanevsky, "Methods for sparse signal recovery using Kalman filtering with embedded pseudo-measurement norms and quasi-norms," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2405–2409, Apr. 2010.

[14] N. Vaswani, "Kalman filtered compressed sensing," in *Proc. Int. Conf. Image Process.*, San Diego, CA, United states, 2008, pp. 893–896.

[15] L. Li and D. Li, "A distributed estimation method over network based on compressed sensing," *Int. J. Distrib. Sens. Netw.*, vol. 15, no. 4, pp. 1–10, Apr. 2019.

[16] Q. Liu, Z. Wang, X. He, and D. H. Zhou, "On Kalman-consensus filtering with random link failures over sensor networks," *IEEE Trans. Autom. Control*, vol. 63, no. 8, pp. 2701–2708, Aug. 2018.

[17] D. Gan, S. Xie, and Z. Liu, "Stability of the distributed Kalman filter using general random coefficients," *Sci. China Inf. Sci.*, vol. 64, May 2021, Art. no. 172204.

[18] R. Li, D. Gan, S. Xie, and J. Lu, "Stability and performance analysis of the compressed Kalman filter algorithm for sparse stochastic systems," *Sci. China Technol. Sci.*, vol. 67, pp. 380–394, Feb. 2024.

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

[20] S. Xie, Y. Zhang, and L. Guo, "Convergence of a distributed least squares," *IEEE Trans. Autom. Control*, vol. 66, no. 10, pp. 4952–4959, Oct. 2021.

[21] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.