

Received 8 June 2022, accepted 24 June 2022, date of publication 29 June 2022, date of current version 7 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3187112

RESEARCH ARTICLE

AG-LRTR: An Adaptive and Generic Low-Rank Tensor-Based Recovery for IIoT Network Traffic Factors Denoising

GANG YUE¹, ZHUO SUN¹, (Member, IEEE), AND JINPO FAN

Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Zhuo Sun (zhuosun@bupt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62171063 and in part by the Manufacturing High Quality Development Fund Project in China under Grant TC210H03D.

ABSTRACT The ubiquitous 5G-enabled industrial Internet of Things interconnects a great number of intelligent sensors and actuators. Network management becomes challenging due to massive traffic data generated by industrial equipment. However, the conventional single traffic factor is insufficient for the increasingly complicated network engineering tasks due to the poor representation capability. Besides, the insecure equipment with open communication access easily brings irregular network fluctuations to network traffic which interferes with the primary traffic factor. The simple and interfered traffic factor decreases the network management efficiency and misleads the operators. Motivated by that, we construct a comprehensive tensor model representing multi-dimension traffic factors to describe the network traffic beneficial characteristics. Meanwhile, an adaptive and generic low-rank tensor recovery (AG-LRTR) algorithm in the tensor singular value decomposition (t-SVD) framework is proposed for denoising. For effective tensor recovery, the alternating direction method of multipliers is employed to theoretically solve the partial augmented Lagrangian function of our objective with a closed-form solution. Numerical experiments on both synthetic data and real-world traffic data in IIoT validate that our proposed algorithm outperforms other state-of-the-art of tensor recovery algorithms.

INDEX TERMS Adaptive nuclear norm, generic noise, IIoT traffic factor, noise interference, low-rank tensor, tensor recovery.

I. INTRODUCTION

The Industrial Internet of Things (IIoT) is rapidly developing with the emerging 5G-enabled communication technology. The 5G communication technique provides coinstantaneous broadband access [1], and thus the IIoT network expands the scope of various intelligent equipment extensively. Effective network management becomes increasingly critical for operators. However, with the tremendous growth of traffic data, one challenge is to abstract the beneficial characteristics from the massive traffic data. Nowadays, the traffic volume is still the basic factor in the IIoT network management [2], [3] and only reflects the flow-based correlations in

successive slots. It loses the other intrinsic correlations in the traffic data, such as the packet-based correlation. For example, the packet interval arrival time is preferable to the routing program than traffic volume. As the IIoT network is more complicated, only a flow-based model is insufficient to represent network traffic characters completely. Besides, another challenge is the irregular network fluctuations due to the insecure equipment with open communication access. Parts of the network engineering tasks are susceptible to noise interference, so the irregular network fluctuations degenerate network management efficiency. For example, the network intrinsic resilient capacity should have avoided route reprogramming in instant network congestion, but this fluctuation is likely to cause once redundant reprogramming. To enhance the representation capability and improve the network

The associate editor coordinating the review of this manuscript and approving it for publication was Qilian Liang¹.

management efficiency, a new network traffic model is needed to represent the valuable factors, and precise denoising is essential to reduce unnecessary operations.

The supervisory control and data acquisition (SCADA) system is the central platform in the industry for aggregating and coordinating the network traffic transaction from edge equipment. The irregular network fluctuations are caused by network congestion, equipment failure or operator error as Fig. 1 shows. Many researchers have proposed network traffic modeling and denoising algorithms. The vector [4], [5], matrix [6], [7] and tensor [8] models are proposed respectively to recover the original traffic volume in which the tensor-based model outperformed the other models. As mentioned above, the volume only represents one single traffic factor, and the model representation capability is deficient. Inspired by the tensor recovery performance, we naturally choose the tensor to construct a novel traffic factor model.

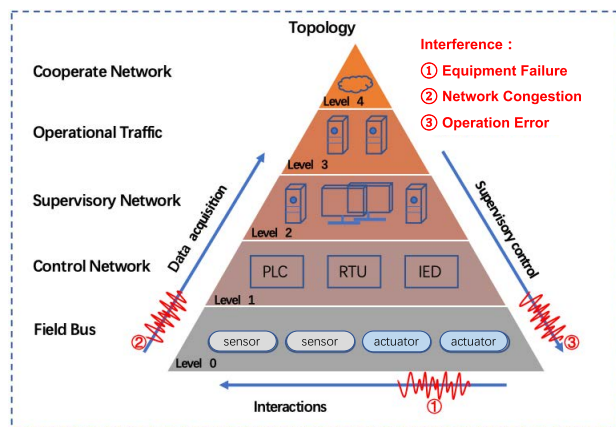


FIGURE 1. SCADA framework and the interference.

The essence of successful recovery is the low-rank property due to the multiple types of correlations in the traffic. Tensor recovery aims to realize low-rank tensor approximation from the noisy traffic factors. The regular optimization object of tensor recovery is to minimize the sum of the tensor nuclear norm and the reconstruction error. The same amplitude shrinkage for the tensor nuclear norm introduces additional bias and variance by the thresholding estimator in the recovery procedures. The reconstruction error only assumes the noise satisfies the zero-mean and ignores the non-zero mean influence. Both deficiencies result in the sub-optimal solution for the regular optimization object. To improve the optimization performance, our tensor recovery strategies in this paper are summarized as follows:

1) To enhance the tensor representation capability, we construct a compact and comprehensive traffic tensor model with ten factors: traffic volume, packet number, inter-arrival time (IAT), etc. Moreover, we further reveal that such a tensor satisfies the low-rank property.

2) We propose an adaptive and generic optimization object for precise denoising by minimizing the sum of the weighted

tensor nuclear norm and the noise variance. The weighted tensor nuclear and generic noise Frobenius norm respectively alleviate the influence of the estimator and the non-zero mean noise.

3) The adaptive and generic optimization object is solved by the alternating direction method of multipliers (ADMM) algorithm. Each optimization procedure has a closed-form solution in ADMM. We further perform numerical experiments on synthetic data and a real SCADA traffic dataset as an IIoT example to validate the effectiveness of our algorithm.

The rest of this paper is structured as follows. Section II introduces the preliminaries of an effective low-rank tensor recovery procedure and the tensor singular value decomposition (t-SVD) framework. Section III details the tensor model, adaptive nuclear norm formulation, and solver algorithm for ADMM. Numerical experiments conducted on synthetic and real-world data are presented in Section IV, and we conclude this work in Section V.

II. PRELIMINARY OF TENSOR RECOVERY

Tensor recovery is realized on the basis of the tensor low-rank property and decomposition approach. The low-rank property is embedded in the structure of entity arrangement, and the decomposition approach affects the recovery performance. This section summarizes an effective low-rank tensor recovery procedure and then introduces the tensor singular value decomposition framework involved in this paper.

A. TENSOR RECOVERY

Multi-dimensional data is becoming prevalent in many areas, such as computer vision [9], [10] and information science [11]. Tensor as a multi-dimensional extension of the matrix is a natural choice in these cases and has the capability of capturing these underlying multi-linear structures. Although often residing in extremely high-dimensional spaces, the tensor of interest is frequently of low rank, or approximately so [12]. Lying at the core of high-dimensional data analysis, tensor decomposition serves as a valuable tool for revealing when a tensor can be modeled as lying close to a low-dimensional subspace [13].

As for data analysis by tensor decomposition, the first step is to construct an appropriate tensor that could contain intrinsic correlations in the data. Except for some kinds of natural tensor data, such as the hyperspectral data, the seismic data, or the colorful picture, other data need to be rearranged as some regulation based on the intrinsic correlation. For example, in some applications for video background model in [9], [14], the original 4-dimensions video data was reshaped to a 3-dimension tensor through matricization of the colorful frame along the time mode. However it would degenerate the performance due to the information loss of frame matricization. Furthermore, the IT traffic tensor constructed in [8] gains periodic pattern in addition to temporal stability and spatial correlation only for traffic volume. Although the tensor entries can be substituted by other network traffic factors, this model intrinsically ignores the multiple correlations

between the factors and could not represent the complete traffic properties. Therefore, the constructed tensor model should contain the necessary correlations as much as possible.

The second step is to select a practical tensor decomposition approach to approximate the low-rank tensor. However, the best low-rank tensor approximation [15] can't be achieved by the convenient two most popular tensor decomposition approaches, namely CANDECOMP/PARAFAC(CP) [16], [17] and Tucker decomposition [18]. Compared with them, tensor singular value decomposition(t-SVD) [19] is based on the tensor-tensor product operator [20] and the calculation procedure needs to transform a tensor from the original domain to the Fourier domain along a fixed mode by discrete Fourier transform(DFT).As a newly emerged tensor decomposition paradigm, it has several properties similar to the traditional matrix SVD and decomposes any tensor with less prior information, so it is the optimality of the truncated t-SVD for data approximation. Therefore in this paper, the task of IIoT traffic denoising lies in the t-SVD framework.

The last and most important step is defining the tensor rank based on the tensor decomposition approach and solving the optimization object by corresponding rank relaxation as the tensor nuclear norm. In the t-SVD framework, the basic tensor rank is called tubal rank [20] which is defined as the number of nonzero singular tubes in the original domain and relaxed by the sum of all singular values in the Fourier domain. Due to the lack of considering the relations of the singular values in the original and Fourier domain, the basic algebra calculation is more complex and computational. For simplicity and elegance, the average tubal rank is defined as the mean rank of the block circulant matrix in [14]. Its rank relaxation is to sum the singular values of the first frontal slice in the original domain. It is rigorously deduced theoretically as a new tensor nuclear norm and has similar theorems with the matrix SVD. The other approaches are either extension [10], [21] or combination [22] underlain by the rank definition and relaxation in [20] or [14]. In this paper, we have the same rank definition and relaxation with [14], because the adaptive coefficients to shrink the nuclear norm needs to be calculated by the unique singular values.

B. TENSOR SINGULAR VALUE DECOMPOSITION

T-SVD operation as an extension to matrix SVD is based on the tensor-tensor product(t-product). For simplicity, we mainly introduce the correlation between the original domain and the Fourier domain caused by DFT. The basic definitions related to the t-SVD framework are given in Appendix.

Let $\mathcal{A} \in \mathbb{R}^{n_1 \times l \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{l \times n_2 \times n_3}$. The Fourier transformations along the third mode are denoted as $\overline{\mathcal{A}} = \text{fft}(\mathcal{A}, [], 3)$ and $\overline{\mathcal{B}} = \text{fft}(\mathcal{B}, [], 3)$. Then the t-product $\mathcal{C} = \mathcal{A} * \mathcal{B}$ is defined as a tensor of size $n_1 \times n_2 \times n_3$ and represented in the original and Fourier domain as follow

$$\begin{aligned} \text{Original domain} : \mathcal{C} &= \mathcal{A} * \mathcal{B} = \text{fold}(\text{bcirc}(\mathcal{A}) \cdot \text{unfold}(\mathcal{B})) \\ \text{Fourier domain} : \overline{\mathcal{C}} &= \overline{\mathcal{A}} \cdot \overline{\mathcal{B}} \end{aligned} \quad (1)$$

where $\text{fold}(\cdot)$, $\text{bcirc}(\cdot)$, $\text{unfold}(\cdot)$ denote the fold, block circulant and unfold operation for a tensor respectively and $\overline{\mathcal{A}}, \overline{\mathcal{B}}, \overline{\mathcal{C}}$ is block diagonal matrix of $\overline{\mathcal{A}}, \overline{\mathcal{B}}, \overline{\mathcal{C}}$. In the original domain, a 3-mode tensor can be regarded as a matrix, with each entry being a tube that lies in the third mode. Thus, the t-product is analogous to the matrix multiplication, except that the circular convolution replaces the multiplication operation between the entries. In the Fourier domain, the t-product is equivalent to the matrix multiplication. The t-product enjoys many similar properties to the matrix-matrix product.

Then for any 3-mode tensors, $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ the t-SVD is defined in the original and Fourier domain as follows:

$$\begin{aligned} \text{Original domain} : \mathcal{A} &= \mathcal{U} * \mathcal{S} * \mathcal{V}^* \\ \text{Fourier domain} : \overline{\mathcal{A}} &= \overline{\mathcal{U}} \cdot \overline{\mathcal{S}} \cdot \overline{\mathcal{V}}^* \end{aligned} \quad (2)$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal, and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor. See Fig. 2 for an intuitive illustration of the t-SVD operation.

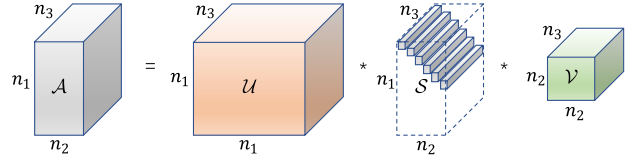


FIGURE 2. An illustration of tensor singular value decomposition.

In this paper, we refer to the rank definition and relaxation in [14]. Let $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$ be the t-SVD of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. The entries on the diagonal of the first frontal slice $\mathcal{S}(:, :, 1)$ have the decreasing property as follow

$$\mathcal{S}(1, 1, 1) \geq \mathcal{S}(2, 2, 1) \geq \dots \geq \mathcal{S}(n', n', 1) \quad (3)$$

where $n' = \min(n_1, n_2)$. It holds since the inverse DFT gives

$$\mathcal{S}(i, i, 1) = \frac{1}{n_3} \sum_{j=1}^{n_3} \overline{\mathcal{S}}(i, i, j) \quad (4)$$

and the entries on the diagonal of $\overline{\mathcal{S}}(:, :, j)$ are the singular values of $\overline{\mathcal{A}}(:, :, j)$, so the tensor tubal rank is determined by the first frontal slice $\mathcal{S}(:, :, 1)$ and equivalent to the number of non-zero singular values of \mathcal{A} . Based on the above properties, tensor average rank defined in [14] is the slice mean of the total rank in Fourier domain as $\text{rank}_a(\mathcal{A}) = \frac{1}{n_3} \text{rank}(\overline{\mathcal{A}})$ and proved that the low average rank assumption is weaker than the low Tucker rank and low CP rank assumption, so it is more convenient to decompose a low rank tensor. Then the relaxation of tensor average rank can be rigorously deduced as summation of the singular values in the first frontal slice $\mathcal{S}(:, :, 1)$ and denoted as $\|\mathcal{A}\|_* = \sum_{i=1}^r \mathcal{S}(i, i, 1)$ where r is tensor tubal rank.

The discrimination between the tubal and average rank is the coefficient $\frac{1}{n_3}$ that is crucial to guarantee the convex envelope of the average tensor rank in a specific scope. Therefore, adopting an adaptive strategy for IIoT traffic denoising is possible based on the above-defined tensor nuclear norm.

III. TENSOR-BASED MODELING AND DENOISING FOR IIoT NETWORK TRAFFIC

To improve the IIoT network management efficiency, we conduct a novel tensor model with ten traffic factors based on the periodical transaction mechanism to enhance the representation capability firstly. We further validate that such a representation model has a low-rank property that underlies the effective denoising. Then an adaptive and generic optimization object is reformulated to improve the denoising performance, and finally a closed-form ADMM algorithm is proposed to solve the object.

A. TENSOR MODELLING AND LOW-RANK ANALYSIS

It is impossible to analyze the per packet due to the massive amount of traffic data in the industrial internet of things. Effective network management relies on multiple statistical factors which could represent the complete data exchange. In IIoT, SCADA systems are the central platform and automatically coordinate and manage the equipment actions to ensure that the infrastructure operates correctly and safely. The primary transaction mechanism [23] is polling field information and sending corresponding control commands periodically, as Fig. 3 shows. The acquisition and control data are transmitted at the determined time of one period. In addition, the same equipment category has stable operation logic and generates similar traffic data. Therefore, the total network traffic data can be characterized by periodic throughput patterns, clear statistics of packet size, predictable flow direction, and expected connection lifetime [24].

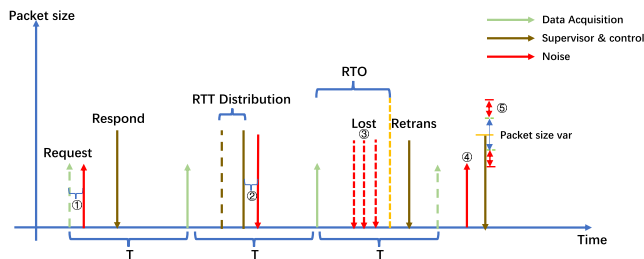


FIGURE 3. SCADA network traffic transaction interfered by noise.

However, irregular network fluctuations will be caused by package loss, data delay or retransmission and payload changes as depicted in Fig. 3. These fluctuations finally reflect the variances in traffic volume, packet number, and packet interval arrival time (IAT). To enhance the representation capability, ten statistical factors listed in Table 1 are calculated periodically by the way provided in [25]. These factors are sufficient to represent the data exchange at the flow-based and packet-based levels and can be applied in most network engineering tasks. In all factors, the traffic volume and packet IAT contain four factors: maximum value, minimum value, mean value, and variance.

Therefore, a novel low-rank tensor model can be constructed based on the factors to represent the beneficial characteristics of IIoT network traffic. In this paper, a testbed

TABLE 1. The comprehensive factors of SCADA traffic.

No.	factor	Explanation
1	Traffic volume	Total volume of traffic in a time slot
2	Packet number	Total number of packets in a time slot
3	Volume mean	Mean value of volume in a time slot
4	Volume variance	Variance value of volume in a time slot
5	Volume max	Maximum volume in a time slot
6	Volume min	Minimum volume in a time slot
7	IAT mean	Average inter-arrival time in a time slot
8	IAT variance	Various inter-arrival time in a time slot
9	IAT max	Maximum inter-arrival time in a time slot
10	IAT min	Minimum inter-arrival time in a time slot

SCADA traffic data named Electrical Power and Intelligent Control (EPIC) [26] is used as the real IIoT network traffic dataset, and it mimics a real-world power system in small scale smart-grid. This SCADA system interacts with six categories of equipment which are access point(AP), programmable logic controller(PLC), intelligent electronic equipment(IED), switch(SW), history database(HIST), firewall(FW), and the others. Obviously, the most prominent traffic volumes are periodic as Fig. 4(a) shows, and the least common multiple periods can be set to 30s with the 1s time slot. For each traffic factor, we calculate each statistical value per second as a row vector and continuously repeat 30 times to form a factor matrix as the frontal slice, then stack the ten factor frontal slices along the third mode to construct a traffic tensor model with the size of $30 \times 30 \times 10$ as Fig. 4(b) shows.

We decompose the traffic tensor of each category of equipment by t-SVD and illustrate the tensor singular values of the first frontal slice as Fig. 5 shows. If the traffic has the same period, such as AP, IED, SW and FW, most of the singular values of the traffic tensor are relatively small, which means the optimum low rank. However, when the traffic is random as HIST or with various periods as PLC, the low-rank property is relatively weak due to the uniform dispersion in the Fourier domain caused by random. As for sparse traffic such as the others, it has the worst low rank due to the independent and poor intrinsic correlation.

Furthermore, we decompose the sum of all traffic, and the singular values are depicted as Fig. 6. Although the summation consists of different kinds of traffic, it still has the property of low rank, which means the periodical traffics data is overwhelming. Our proposed tensor model is capable to capture the intrinsic correlations in various aspects. In the following paper, the IIoT traffic factor tensor is deemed to contain all traffic by default.

B. ADAPTIVE AND GENERAL OPTIMIZATION OBJECT

We briefly introduce the related denotations before description of the proposed algorithm. \mathcal{L} denotes the actual low-rank traffic tensor that can be decomposed as $\mathcal{L} = \mathcal{U} * \mathcal{S} * \mathcal{V}$ by t-SVD. \mathcal{N} denotes the noise with identical distribution and \mathcal{M} denotes the noisy measurement. The denoising purpose is to recover the actual low-rank tensor from the noisy measurement.

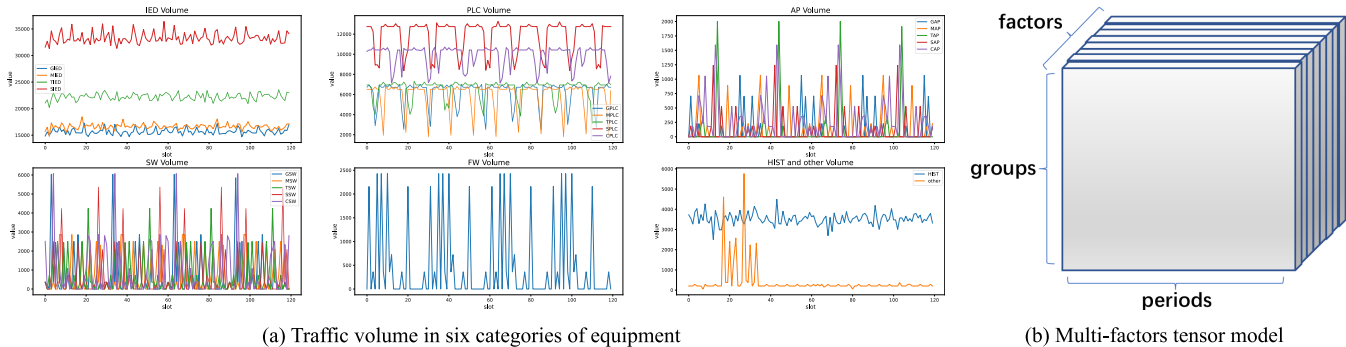


FIGURE 4. SCADA traffic volume and tensor model.

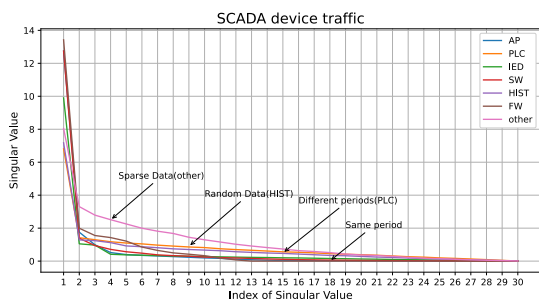


FIGURE 5. SCADA equipment traffic tensor model singular values.

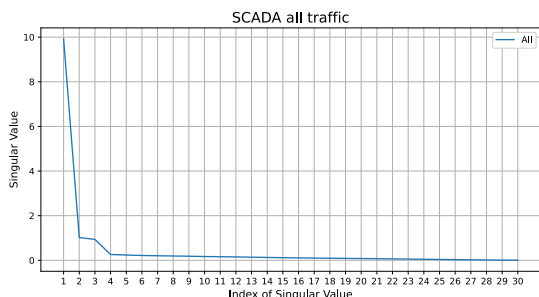


FIGURE 6. SCADA all traffic tensor model singular values.

The convenient optimization object of the low-rank tensor recovery is to minimize the sum of tensor rank and reconstruction error as follows:

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{N}} \quad & \text{rank}(\mathcal{L}) + \frac{\lambda}{2} \|\mathcal{N}\|_F^2 \\ \text{s.t.} \quad & \mathcal{M} = \mathcal{L} + \mathcal{N} \end{aligned} \quad (5)$$

where λ denotes the penalty coefficient. Because the rank operator is non-convex, the nuclear norm $\|\cdot\|_*$ as the tightest relaxation of rank replaces the first term in (5). In the paper, we define the tensor rank as the average rank and the optimization object can be reformulated as follow:

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{N}} \quad & \sum_{i=1}^r \mathcal{S}(i, i, 1) + \frac{\lambda}{2} \|\mathcal{N}\|_F^2 \\ \text{s.t.} \quad & \mathcal{M} = \mathcal{L} + \mathcal{N}, \mathcal{L} = \mathcal{U} * \mathcal{S} * \mathcal{V} \text{ and } \|\mathcal{L}\|_* = \sum_{i=1}^r \mathcal{S}(i, i, 1) \end{aligned} \quad (6)$$

Two terms of the above optimization object have a deficiency in the optimization process. As for the first term $\sum_{i=1}^r \mathcal{S}(i, i, 1)$, the solver involving soft-thresholding shrink would cause some unavoidable biases [27], so the variance of the estimated tensor would be smaller than the original tensor when equally shrinking every singular value. Conspired by the adaptive nuclear norm for low-rank matrix approximation in [28], we can extend it to the tensor model by assigning weights to the singular values of a tensor.

Another term $\|\mathcal{N}\|_F^2$ assumes that the noise has zero mean and ignores the common non-zero mean situation, which leads to the sub-optimal solution. Assuming the actual mean $\hat{\mu}$ and variance $\hat{\delta}^2$ of noise are unknown and μ is a variable representing noise's means distribution, the variance of $\mathcal{N} - \mu \mathcal{I}_\mu$ can be expressed as $\text{var}(\mathcal{N} - \mathcal{I}_\mu) = \text{var}(\mathcal{N}) + \text{var}(\mathcal{I}_\mu)$ where \mathcal{I}_μ is the tensor with the same size of \mathcal{N} and its all entries are μ . The variance of tensor can be calculated by the Frobenius norm $\|\cdot\|_F^2$. If $\mu \neq \hat{\mu}$, the unbiased variance will be larger than $\hat{\delta}^2$ as the simple proof below where $N = n_1 \times n_2 \times n_3 - 1$ denotes the element number of tensor:

$$\begin{aligned} \|\mathcal{N} - \mathcal{I}_\mu\|_F^2 &= \|\mathcal{N} - \mathcal{I}_{\hat{\mu}} + \mathcal{I}_{\hat{\mu}} - \mathcal{I}_\mu\|_F^2 \\ &= \|\mathcal{N} - \mathcal{I}_{\hat{\mu}}\|_F^2 + \|\mathcal{I}_{\hat{\mu}} - \mathcal{I}_\mu\|_F^2 \\ &\geq \|\mathcal{N} - \mathcal{I}_{\hat{\mu}}\|_F^2 \\ &= N \times \hat{\delta}^2 \end{aligned} \quad (7)$$

Then the adaptive and generic optimization object is formulated as follow:

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{N}, \mu} \quad & \sum_{i=1}^r \alpha_i \mathcal{S}(i, i, 1) + \frac{\lambda}{2} \|\mathcal{N} - \mathcal{I}_\mu\|_F^2 \\ \text{s.t.} \quad & \mathcal{M} = \mathcal{L} + \mathcal{N}, \mathcal{L} = \mathcal{U} * \mathcal{S} * \mathcal{V} \\ \text{and } \|\mathcal{L}\|_* &= \sum_{i=1}^r \mathcal{S}(i, i, 1) \end{aligned} \quad (8)$$

where α_i denotes the i -th adaptive coefficient. The optimization object (8) can be explained to minimize the sum of the tensor adaptive nuclear norm and the noise variance.

Then an adaptive tensor soft-thresholding (ATSVT) operation as a closed-form solution could solve the optimization object with the adaptive nuclear norm as follows.

Theorem 1: For any $\lambda \geq 0$, $\mathcal{V} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $0 \leq \alpha_1 \leq \dots \leq \alpha_r$ ($r = \min(n_1, n_2)$), a global optimal solution

to the optimization problem

$$\begin{aligned} \min_{\mathcal{X}} \quad & \sum_{i=1}^r \alpha_i \sigma_i(\mathcal{X}) + \frac{\lambda}{2} \|\mathcal{X} - \mathcal{Y}\|_F^2 \\ \text{s.t.} \quad & \sigma_i(\mathcal{X}) = \mathcal{S}_{\mathcal{X}}(i, i, 1) \end{aligned} \quad (9)$$

is given by the ATSVT as $\hat{\mathcal{X}} := D_{\frac{\alpha}{\lambda}}(\mathcal{Y})$,

$$\begin{aligned} D_{\frac{\alpha}{\lambda}}(\mathcal{Y}) &= \mathcal{U} * \mathcal{S}_{\frac{\alpha}{\lambda}} * \mathcal{V} \\ \mathcal{S}_{\frac{\alpha}{\lambda}} &= \text{if } \text{ft}(\text{Diag}\{\bar{\mathcal{S}}(i, i, :) - \frac{\alpha_i}{\lambda}\}_+, i = 1, 2, \dots, r) \end{aligned} \quad (10)$$

where $\mathcal{S}_{\tau}(\cdot)$ denotes the adaptive soft-thresholding operation. Further, if \mathcal{Y} has a unique t-SVD, $\hat{\mathcal{X}}$ is the unique optimal solution. The soft-thresholding operation for tensors shrinks the singular values in the Fourier domain. Following [29], the weights can be set as some power of the singular values of the tensor, i.e., $\alpha_i = \frac{1}{\sigma_i(\mathcal{X})^\gamma}$, where $\gamma \geq 0$ is a predefined constant. In this way, the order constraint in the Theorem is automatically satisfied. The fact that a closed-form global minimizer for the optimization object (9) would be proved as follows based on von Neumann's trace inequality [30] and the properties of t-SVD.

Proof: We first prove that $\hat{\mathcal{X}}$ is indeed a global optimal solution to (9). Since the weighted coefficients only depend on the singular values of \mathcal{X} , by letting $\mathbf{g} = \{g_i\}_{i=1}^r = \sigma(\mathcal{X})$ (which implies the entries of \mathbf{g} are in non-increasing order), (9) can be written as:

$$\min_{\mathbf{g}: g_1 \geq \dots \geq g_r \geq 0} \left\{ \min_{\sigma(\mathcal{X})} \left\{ \frac{1}{2} \|\mathcal{X} - \mathcal{Y}\|_F^2 \right\} + \frac{1}{\lambda} \sum_{i=1}^r \alpha_i g_i \right\} \quad (11)$$

For the inner minimization, we have the inequality

$$\begin{aligned} \|\mathcal{X} - \mathcal{Y}\|_F^2 &= \frac{1}{\sqrt{n_3}} \|\bar{\mathbf{X}} - \bar{\mathbf{Y}}\|_F^2 \\ &= \frac{1}{\sqrt{n_3}} \text{Tr}(\bar{\mathbf{X}} - \bar{\mathbf{Y}})(\bar{\mathbf{X}} - \bar{\mathbf{Y}})^* \\ &= \frac{1}{\sqrt{n_3}} \text{Tr}(\bar{\mathbf{X}}\bar{\mathbf{X}}^*) - \frac{2}{\sqrt{n_3}} \text{Tr}(\bar{\mathbf{X}}\bar{\mathbf{Y}}^*) \\ &\quad + \frac{1}{\sqrt{n_3}} \text{Tr}(\bar{\mathbf{Y}}\bar{\mathbf{Y}}^*) \\ &= \sum_{i=1}^r g_i^2 - \frac{2}{\sqrt{n_3}} \text{Tr}(\bar{\mathbf{X}}\bar{\mathbf{Y}}^*) + \sum_{i=1}^r \sigma_i^2(\mathcal{Y}) \\ &\geq \sum_{i=1}^r g_i^2 - 2\mathbf{g}\sigma(\mathcal{Y})^* + \sum_{i=1}^r \sigma_i^2(\mathcal{Y}) \end{aligned} \quad (12)$$

The last inequality is due to von Neumann's trace inequality. The equality holds when \mathcal{X} admits the singular value decomposition $\mathcal{X} = \mathcal{U} * \text{Diag}(\mathbf{g}) * \mathcal{V}^*$ where \mathcal{U} and \mathcal{V} are the left and right orthogonal tensor in the t-SVD of \mathcal{Y} . Then the optimization can be reformulated as follow

$$\min_{\mathbf{g}: g_1 \geq \dots \geq g_r \geq 0} \left\{ \sum_{i=1}^r \left(\frac{1}{2} g_i^2 - [\sigma_i(\mathcal{Y}) - \frac{\alpha_i}{\lambda}] g_i + \frac{1}{2} \sigma_i^2(\mathcal{Y}) \right) \right\} \quad (13)$$

The optimization object is completely separable and takes minimum when $g_i = (\sigma_i(\mathcal{Y}) - \frac{\alpha_i}{\lambda})_+$. This is a feasible solution because $\{\sigma_i(\mathcal{Y})\}$ is in non-increasing order, while $\{\alpha_i\}$ is in non-decreasing order. Therefore, equation (14) is a global optimal solution to the objection function (10). The uniqueness follows by the equality condition for von Neumann's trace inequality when \mathcal{Y} has a unique t-SVD, and the uniqueness of the strictly convex optimization. \square

C. CLOSED-FORM ADMM SOLVER ALGORITHM

ADMM algorithm is very efficient for some convex or non-convex programming problems. The closed-form solution to each optimization procedure guarantees recovery performance. To solve the optimization object (8), the problem can be reformulated by the partial augmented Lagrangian function as follows, and we deduce the closed-form solutions to all formulations.

$$\begin{aligned} L(\mathcal{L}, \mathcal{N}, \mu, \mathcal{P}, \beta) &= \sum_{i=1}^r \alpha_i \sigma_i(\mathcal{L}) \frac{\lambda}{2} \|\mathcal{N} - \mathcal{I}_\mu\|_F^2 \\ &\quad + \langle \mathcal{P}, \mathcal{M} - \mathcal{L} - \mathcal{N} \rangle + \frac{\beta}{2} \|\mathcal{M} - \mathcal{L} - \mathcal{N}\|_F^2 \end{aligned} \quad (14)$$

where $\mathcal{P} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the tensor of Lagrange multipliers and $\beta > 0$ is a penalty parameter. So the variables are updated sequentially in each iteration as follows

$$\begin{aligned} \min_{\mathcal{L}_{k+1}} \quad & \sum_{i=1}^r \alpha_i \sigma_i(\mathcal{L}) + \langle \mathcal{P}_k, \mathcal{M} - \mathcal{L} - \mathcal{N}_k \rangle \\ & + \frac{\beta_k}{2} \|\mathcal{M} - \mathcal{L} - \mathcal{N}_k\|_F^2 \end{aligned} \quad (15)$$

$$\begin{aligned} \min_{\mathcal{N}_{k+1}} \quad & \frac{\lambda}{2} \|\mathcal{N} - \mathcal{I}_\mu\|_F^2 + \langle \mathcal{P}_k, \mathcal{M} - \mathcal{L}_{k+1} - \mathcal{N} \rangle \\ & + \frac{\beta_k}{2} \|\mathcal{M} - \mathcal{L}_{k+1} - \mathcal{N}\|_F^2 \end{aligned} \quad (16)$$

$$\min_{\mu_{k+1}} \quad \frac{\lambda}{2} \|\mathcal{N}_{k+1} - \mathcal{I}_\mu\|_F^2 \quad (17)$$

$$\mathcal{P}_{k+1} = \mathcal{P}_k + \beta_k (\mathcal{M} - \mathcal{L}_{k+1} - \mathcal{N}_{k+1}) \quad (18)$$

$$\beta_{k+1} = \min(\rho \beta_k, \beta_{max}) \quad (19)$$

where $\rho \in (1.0, 1.1]$ denotes the adjustment coefficient to accelerate the convergence speed and β_0 is a small constant. And there exists a closed-form solution for each component in (14)-(18).

The term $\langle \mathcal{P}, \mathcal{M} - \mathcal{L} - \mathcal{N} \rangle + \frac{\beta}{2} \|\mathcal{M} - \mathcal{L} - \mathcal{N}\|_F^2$ can be merged as $\frac{\beta}{2} \|\mathcal{M} - \mathcal{L} - \mathcal{N} + \frac{\mathcal{P}}{\beta}\|_F^2$. For optimization problem (15) and (16), they can be rewritten as follow

$$\min_{\mathcal{L}_{k+1}} \quad \sum_{i=1}^r \alpha_i \sigma_i(\mathcal{L}) + \frac{\beta_k}{2} \|\mathcal{M} - \mathcal{L} - \mathcal{N}_k + \frac{\mathcal{P}_k}{\beta_k}\|_F^2 \quad (20)$$

$$\min_{\mathcal{N}_{k+1}} \quad \frac{\lambda}{2} \|\mathcal{N} - \mathcal{I}_\mu\|_F^2 + \frac{\beta_k}{2} \|\mathcal{M} - \mathcal{L}_{k+1} - \mathcal{N} + \frac{\mathcal{P}_k}{\beta_k}\|_F^2 \quad (21)$$

where the problem (20) has the same format as problem (9) and can be solved by the closed-form solution (10). For optimization problem (21), the closed-form solution needs to be calculated in Fourier domain as follow

Fourier domain :

$$\begin{aligned} f(\bar{N}) &= \arg \min_{\bar{N}} \left\{ \frac{\lambda}{2} \|\bar{N} - \bar{I}_\mu\|_F^2 \right. \\ &\quad \left. + \frac{\beta_k}{2} \|\bar{M} - \bar{L}_{k+1} - \bar{N} + \frac{\bar{P}_k}{\beta_k}\|_F^2 \right\} \\ \frac{\partial f(\bar{N})}{\partial \bar{N}} &= \lambda(\bar{N} - \bar{I}_\mu) + \beta_k(\bar{N} + \bar{L}_{k+1} - \bar{M} - \frac{\bar{P}_k}{\beta_k}) \\ &= (\lambda + \beta)\bar{N} - \lambda_k \bar{I}_\mu + \beta_k \bar{L}_{k+1} - \beta_k \bar{M} - \bar{P}_k \\ &= 0 \\ \bar{N} &= \frac{\lambda_k \bar{I}_\mu - \beta_k \bar{L}_{k+1} + \beta_k \bar{M} + \bar{P}_k}{\lambda + \beta} \end{aligned}$$

Original domain :

$$\mathcal{N} = \frac{\lambda_k \mathcal{I}_\mu - \beta_k \mathcal{L}_{k+1} + \beta_k \mathcal{M} + \mathcal{P}_k}{\lambda + \beta} \quad (22)$$

Then the problem (17) can be solved in original domain through calculating the mean of all entries as follow

$$\mu = \frac{\sum_{i,j,k} \mathcal{N}_{ijk}}{n_1 \cdot n_2 \cdot n_3} \quad (23)$$

Finally, the pseudocode for low-rank tensor recovery from noisy measurement by adaptive nuclear norm is described in Algorithm 1

Algorithm 1 Low-Rank Tensor Recovery From Noisy Measurement by Adaptive Nuclear Norm

Input: \mathcal{M} , γ , λ , $Iter_max$ and ϵ

Output: \mathcal{L} , μ

Initialize $\mathcal{L}_0 = \mathcal{N}_0 = \mathcal{P}_0 = 0$, $\beta_0 = 10e - 6$, $\rho = 1.1$, $iter = 0$

while not converged **do**

update \mathcal{L}_{k+1} :

 1. decompose the tensor $(\mathcal{M} - \mathcal{N}_k + \frac{\mathcal{P}_k}{\beta_k})$

 2. calculate the adaptive coefficient $\alpha_i =$

$$\frac{1}{\sigma_i(\mathcal{M} - \mathcal{N}_k + \frac{\mathcal{P}_k}{\beta_k})^\gamma}$$

 3. update the tensor \mathcal{L}_{k+1} by (10)

update \mathcal{N}_{k+1} by (22)

update μ_{k+1} by (23)

update \mathcal{P}_{k+1} by (18)

update β_{k+1} by (19)

$iter += 1$

check if convergence is meet:

$\|\mathcal{L}_{k+1} - \mathcal{L}_k\|_\infty < \epsilon$ and

$\|\mathcal{N}_{k+1} - \mathcal{N}_k\|_\infty < \epsilon$ and

$\|\mu_{k+1} - \mu_k\|_\infty < \epsilon$ and

$\|\mathcal{L} + \mathcal{N} - \mathcal{M}\|_\infty < \epsilon$

$oriter > Iter_{max}$

end while

Output: \mathcal{L} , μ

IV. PERFORMANCE EVALUATION

We conduct numerical experiments to validate the efficiency of our proposed AG-LRTR in zero mean and non-zero mean noise interference scenarios. Three experiments on the synthetic data are about to analyze the noise influence, evaluate the recovery performance and discuss the parameters variation respectively. One more experiment on the EPIC SCADA traffic data is conducted for practical application. The Peak Signal-to-Noise Ratio (PSNR) is used as the metric to measure the quality of the recovery performance and it is defined as equation (24). All the experiments are conducted on a PC with a 2.9 GHz CPU and 8 GB RAM.

$$PSNR = 10 \log_{10} \left(\frac{\|\mathcal{L}\|_\infty^2}{\frac{1}{n_1 n_2 n_3} \|\mathcal{L} - \bar{\mathcal{L}}\|_F^2} \right) \quad (24)$$

where \mathcal{L} denotes true ground tensor, and $\bar{\mathcal{L}}$ denotes the recovered low-rank tensor. The larger value of PSNR corresponds to the higher quality of the results.

Moreover, we compare the AG-LRTR algorithm with the other four algorithms for tensor recovery. They are abbreviated as LRMR, LRTR, G-LRTR, and TRPCA. LRMR represents the low-rank matrix recovery which needs to flatten the original tensor to a matrix and then recover the low-rank component by RPCA. The LRTR represents the low-rank tensor recovery in [31] to restore the hyperspectral image, and G-LRTR extends the LRTR only by generalizing the noise formulation. The TRPCA represents the tensor robust principal components analysis used in [14] to recover from the sparse noisy tensor. The results illustrate that our AG-LRTR algorithm outperforms the other algorithms in tensor recovery.

A. SYNTHETIC DATA

The synthetic tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with $rank(\mathcal{A}) = r$ can be generated by the tensor-tensor product directly as follows.

$$\mathcal{A} = \mathcal{Q} * \mathcal{R} \quad (25)$$

where $\mathcal{Q} \in \mathbb{R}^{n_1 \times r \times n_3}$, $\mathcal{R} \in \mathbb{R}^{r \times n_2 \times n_3}$ and all entries sampled in independent and identical uniform distribution of $U(0, 1)$. Firstly, we summarize the influence of different kinds of noise on the tensor singular values, then compare the recovery performance of the five algorithms and finally discuss the parameters set of the proposed algorithm.

1) NOISE INFLUENCE

As Fig. 6 depicted, the size of SCADA network traffic tensor is $30 \times 30 \times 10$ and the rank is four. We generate the synthetic tensor in the experiment with the same tensor size and rank for simplicity. It is noted that the proposed algorithm can be generalized to arbitrary size tensors with the low-rank property. To exploit noise influence on tensor rank, two major categories are specified: the zero-mean noise and the non-zero mean noise. For zero mean noise, the noise with various variances ranging from 1 to 10 is added to the original tensor as Fig. 7 (a) shows. Because the non-zero mean noise

can be divided as the sum of a zero-mean noise and a constant value, only the influence of the constant value is depicted in the right picture of Fig. 7 (b).

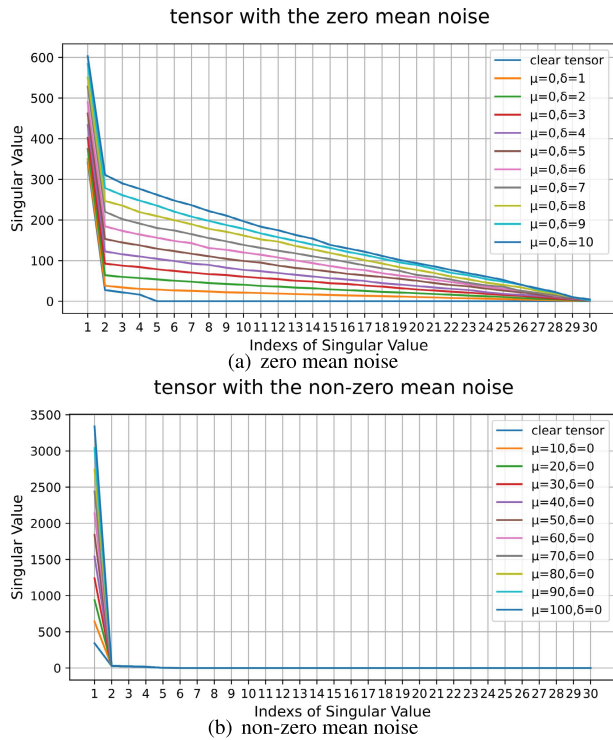


FIGURE 7. Tensor singular values interfered with various kinds of noise.

As we can see, no matter whether the mean of noise is zero or non-zero, the randomness property of noise increases all the singular values of the tensor in a similar tendency and the increment is positively correlated with the variance of the noise. As for the non-zero mean noise, the constant mean only affects the largest singular value, as proved as follows. Moreover, the discrepancy between the original tensor and noisy tensor is diverse from each singular value as Fig. 8 shows. The influence of noise with different variances on each singular value of the original tensor differs, so the adaptive nuclear norm could shrink each singular value to a different extent, thus improves recovery performance.

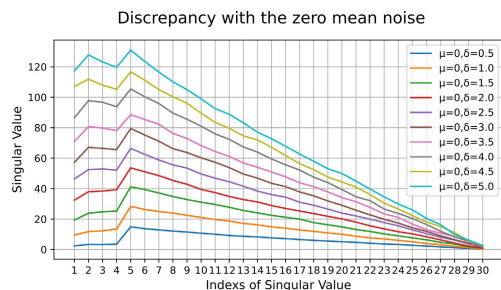


FIGURE 8. Discrepancy between original tensor and noise tensor.

Proof: Assume a low rank tensor $\mathcal{L} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and a constant tensor $\mathcal{I}_\mu \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then representation of the

tensor $\mathcal{M} = \mathcal{L} + \mathcal{I}_\mu$ in Fourier domain can be calculated as follow

$$\begin{aligned} \bar{\mathbf{M}} &= (\mathbf{F}_{n_3} \otimes \mathbf{I}_{n_1}) \cdot \text{bcirc}(\mathcal{L} + \mathcal{I}_\mu) \cdot (\mathbf{F}_{n_3}^{-1} \otimes \mathbf{I}_{n_2}) \\ &= \bar{\mathbf{L}} + \mu(\mathbf{F}_{n_3} \otimes \mathbf{I}_{n_1}) \cdot \text{bcirc}(\mathcal{I}_1)(\mathbf{F}_{n_3}^{-1} \otimes \mathbf{I}_{n_2}) \end{aligned} \quad (26)$$

As for the second term in equation (26), all entries of the tensor \mathcal{I}_1 are 1 and its block circulant matrix is an ones-matrix in which all entries are equal to 1. Then any frontal slices in Fourier domain only have the same value as follow

$$\bar{\mathbf{I}}_1^{(i)} = (\mathbf{F}_{n_3}(i, :) \cdot \mathbf{1}) \times \mathbf{I}_1, i = 1, 2, \dots, n_3 \quad (27)$$

where $\mathbf{F}_{n_3}(i, :)$ denotes the i -th row vector, $\mathbf{1}$ is an ones-vector and \mathbf{I}_1 denotes an ones-matrix with size $n_1 \times n_2$. So each frontal slice is a rank one matrix and has only one singular value and its left and right singular vectors are ones-vector with size n_1 and n_2 . Because the first row and column in \mathbf{F}_{n_3} is ones-vector, both the $\bar{\mathbf{L}}^{(i)}$ and $\bar{\mathbf{I}}_1^{(i)}$ have the same singular vectors and $\bar{\mathbf{I}}_1^{(i)}$ is in the subspace of the $\bar{\mathbf{L}}^{(i)}$. Then the constant tensor only affects the largest singular value of the original tensor. \square

2) RECOVERY COMPARISON

A low-rank tensor with size $30 \times 30 \times 10$ is generated as the original and interfered with different noise as the measurement to comprehensively compare the recovery performance. The results of recovery performance are listed in Table 2. We can see that our proposed AG-LRTR outperforms other algorithms in all cases. Especially when the noise has low variance, the AG-LRTR could improve recovery performance effectively because low variance on small singular values is much greater than on large singular values, as Fig. 8 shows. The adaptive nuclear norm could retain the information of the original tensor in main singular values and alleviate the noise influence on others. Besides, the G-LRTR and the LRTR have similar performance, which is better than the LRMR that loses the structure information by matricization. The TRPCA has the denoising capability when the mean of noise is zero. However, it fails to recover the original tensor from non-zero noise that dissatisfies sparsity.

3) PARAMETER DISCUSSION

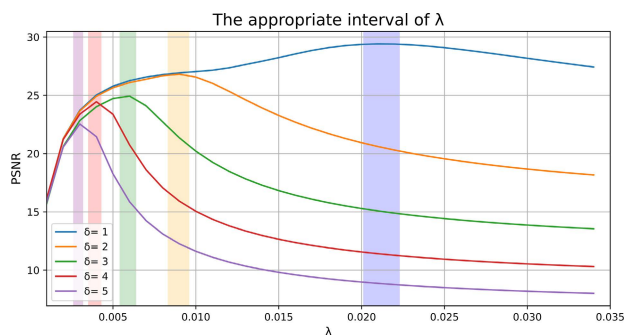
The recovery performance of our proposed AG-LRTR algorithm mainly depends on the parameters of the problem (8) in which the adaptive coefficients $\alpha_i, i = 1, 2, \dots, r$ are determined by the parameter γ , so we will discuss the influence of two core parameters λ and γ on recovery performance.

Before that, we make the connection from our algorithm to other algorithms. If the parameter γ is set to 0, all adaptive coefficients are equal to 1 and the sum of weighted singular values can be regarded as the standard nuclear norm, so our algorithm degenerates to the G-LRTR. Moreover, if the mean variable of noise μ is fixed to 0, the optimization object would be the same with LRTR, then if all tensors only have one frontal slice, the tensor degenerates to a matrix and the LRTR is the same with LRMR. So based on the connection,

TABLE 2. The comparison of recovery performance.

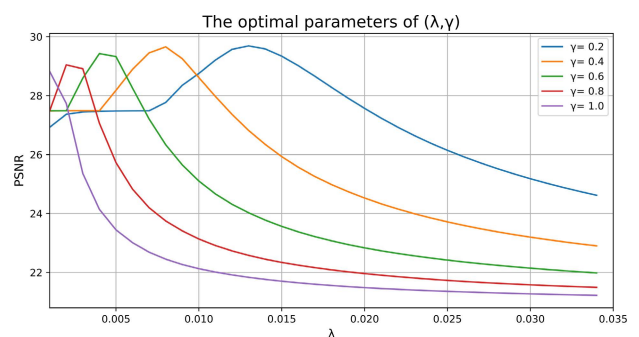
Rank	Var	observed	AG-LRTR	G-LRTR	LRTR	LRMR	TRPCA
Rank = 3	$\mu = 0$						
	$\delta = 1$	22.69	31.12	29.66	29.67	27.07	26.52
	$\delta = 3$	16.12	29.06	26.75	24.68	19.70	17.55
	$\delta = 5$	15.11	27.66	26.71	22.29	16.52	13.28
	$\mu = 3$						
	$\delta = 1$	14.26	30.13	28.23	27.98	26.14	12.54
	$\delta = 3$	14.90	25.96	24.92	24.66	23.39	11.72
	$\delta = 5$	15.06	24.88	24.66	23.38	20.83	10.21
	$\mu = 5$						
$\delta = 1$	11.28	27.64	24.67	24.90	24.06	8.13	
$\delta = 3$	12.84	24.50	23.49	23.40	22.01	7.79	
$\delta = 5$	13.05	27.89	23.09	23.14	19.90	7.06	
Rank = 5	$\mu = 0$						
	$\delta = 1$	26.11	34.64	32.32	31.82	26.71	29.20
	$\delta = 3$	18.75	29.77	28.33	27.19	23.26	20.64
	$\delta = 5$	16.08	27.46	27.17	26.31	19.55	16.34
	$\mu = 3$						
	$\delta = 1$	16.93	30.86	28.65	28.21	27.72	15.37
	$\delta = 3$	16.50	29.73	27.96	27.70	25.65	14.70
	$\delta = 5$	15.46	27.98	27.23	27.20	23.19	13.31
	$\mu = 5$						
$\delta = 1$	13.52	29.96	27.04	26.94	26.44	11.01	
$\delta = 3$	14.44	28.21	26.98	26.26	25.14	10.86	
$\delta = 5$	14.20	27.04	26.51	25.70	23.00	10.17	
Rank = 7	$\mu = 0$						
	$\delta = 1$	27.66	36.73	33.84	33.05	30.64	31.38
	$\delta = 3$	20.53	32.25	29.61	29.64	27.96	23.01
	$\delta = 5$	17.19	30.54	29.59	28.07	23.87	18.91
	$\mu = 3$						
	$\delta = 1$	18.64	33.68	30.96	30.81	30.31	17.79
	$\delta = 3$	17.88	31.01	29.94	29.45	27.66	17.12
	$\delta = 5$	16.38	30.86	29.74	29.60	24.98	15.56
	$\mu = 5$						
$\delta = 1$	15.09	33.26	29.88	29.60	28.63	13.46	
$\delta = 3$	15.37	30.01	28.27	27.82	27.16	13.26	
$\delta = 5$	15.31	27.65	27.18	27.08	24.97	12.77	

we firstly set the parameter $\gamma = 0$ to discuss the influence of penalty coefficient λ as the same as the LRTR and then explore the optimal parameters (γ, λ) of AG-LRTR by the grid search strategy.

**FIGURE 9.** The optimal interval of penalty coefficient when $\gamma = 0$.

In the experiments, the size of the original tensor is $30 \times 30 \times 10$ and the rank is 3. When $\gamma = 0$, the influence of λ on recovery performance is depicted in Fig. 9 which is consistent with the intuition. The penalty coefficient is to balance the nuclear norm of the tensor and the noise variance, so as the

variance increases, the penalty coefficient needs to decrease to keep the nuclear norm and the variance in the same order.

**FIGURE 10.** The optimal parameter of AG-LRTR when the variance of noise is 1.

As for the optimal parameters (λ, γ) of AG-LRTR, the λ depends on the γ due to the adaptive nuclear norm as Fig. 10 shows where the variance of the noise is 1. We can see that the optimal λ corresponding to the peak PSNR decreases as the γ increases because more information of the larger singular values is retained so that there needs a lower λ to reduce the noise variance and keep the balance. The other variances of the noise have the same regulation.

B. TENSOR RECOVERY FOR THE REAL-WORLD DATA

We have introduced the SCADA systems and constructed a tensor model for the traffic of a testbed called EPIC in Section III-A, so the recovery performances of all algorithms are compared based on the low-rank SCADA traffic tensor as Fig. 11 shows. Obviously, the rank of the SCADA traffic tensor is 3 in Fig. 6, and the largest singular value is 10 that is much smaller than the synthetic data, so the variance of noise needs to be chosen cautiously.

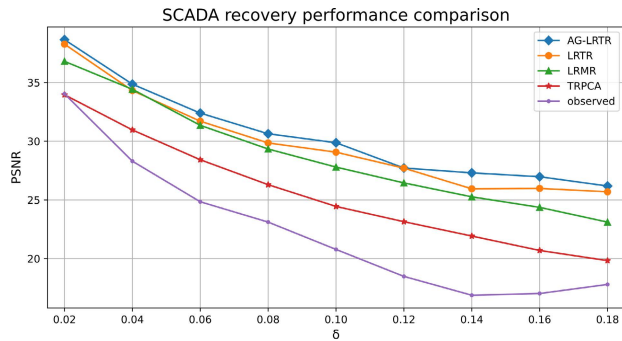


FIGURE 11. The recovery performance for SCADA traffic tensor.

The recovery performances for SCADA traffic are depicted when the mean of noise is 0, and the noise variance varies from 0.02 to 0.18. Our proposed AG-LRTR algorithm still outperforms other algorithms and improves much better than the observed PSNR. The performances of the LRTR and LRMR are similar, which means that the correlations between the frontal slices are weaker than the synthetic data. The TRPCA performance is the worst even if it has the ability to denoising. When the mean of noise is non-zero, the recovery performances of all algorithms have the same tendency except the TRPCA, which fails to denoising.

V. CONCLUSION

Based on the recently developed t-SVD, an adaptive and generic low-rank tensor recovery algorithm is proposed to recover the original traffic factor tensor from the irregular network fluctuations in IIoT scenario. We construct a novel tensor model to abstract multiple correlations from the traffic data and retain as many traffic factors as possible by the adaptive and general optimization object. Numerical experiments on the synthetic data and real-world SCADA traffic verify our algorithm is effective in denoising for network management.

There are some interesting future works left. Because the increment of each singular value depends on the noise variance, the adaptive nuclear norm criterion needs to combine the information of noise with the singular values, which is ignored in this paper. Besides, the optimal parameters are short of a theoretical guarantee and found by grid search, so there needs to explore an effective approach for parameter selection and save the computational cost. Moreover, deep learning can be employed if we do not have much prior information about the original tensor and the noise.

APPENDICES

PRELIMINARY DEFINITION AND PROPERTIES OF T-SVD

Firstly we briefly introduce the basics of the tensor notion and related operation. Scalars are denoted by lower-case letters such as i, j, k and vectors by bold lower-case letters such as $\mathbf{a}, \mathbf{b}, \mathbf{c}$. Matrices are denoted by upper-case letters, e.g., \mathbf{X} . Tensors are denoted by a calligraphic letter, e.g., \mathcal{X} , and its entry are denoted by x_{i_1, \dots, i_n} for a N-mode tensor. Identity matrix with size $n \times n$ is denoted as \mathbf{I}_n . The fields of real numbers and complex numbers are denoted as \mathbb{R} and \mathbb{C} .

For a 3-mode tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, its i -th horizontal, lateral and frontal slice are denoted as $\mathcal{A}(i, :, :)$, $\mathcal{A}(:, i, :)$ and $\mathcal{A}(:, :, i)$ respectively and especially for the frontal slice, it can be abbreviated by $\mathbf{A}^{(i)}$. The tube along the third mode is denoted as $\mathcal{A}(i, j, :)$. In the field of complex number, the complex conjugate of \mathcal{A} is denoted as $\text{conj}(\mathcal{A})$ which takes the complex conjugate of each entry of \mathcal{A} .

The inner product between \mathbf{A} and \mathbf{B} in $\mathbb{C}^{n_1 \times n_2}$ is defined as $\langle \mathbf{A}, \mathbf{B} \rangle = \text{Tr}(\mathbf{A}^* \mathbf{B})$ where \mathbf{A}^* denotes the conjugate transpose of \mathbf{A} and $\text{Tr}(\cdot)$ denotes the matrix trace. The inner product of two same-sized tensor $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ can be represented as the sum of the inner matrix product of each frontal slice and equals to the sum of numerical product of each entry.

$$\langle \mathcal{A}, \mathcal{B} \rangle := \sum_{i=1}^{n_3} \langle \mathbf{A}^{(i)}, \mathbf{B}^{(i)} \rangle = \sum_{i_1, \dots, i_n} a_{i_1, \dots, i_n} b_{i_1, \dots, i_n} \tag{28}$$

Due to the need for rank relaxation, some norms of matrix and tensor are used. For a matrix \mathbf{A} , the i -th singular value is denoted as $\sigma_i(\mathbf{A})$, then the spectral and nuclear norm are defined as follows

$$\begin{aligned} \|\mathbf{A}\| &:= \max_i \sigma_i(\mathbf{A}) \\ \|\mathbf{A}\|_* &:= \sum_i \sigma_i(\mathbf{A}) \end{aligned} \tag{29}$$

For a tensor \mathcal{A} , the l_1 , infinity and Frobenius norm are defined as follow

$$\begin{aligned} \|\mathcal{A}\|_1 &:= \sum_{i,j,k} |a_{ijk}| \\ \|\mathcal{A}\|_\infty &:= \max_{ijk} |a_{ijk}| \\ \|\mathcal{A}\|_F &:= \sqrt{\sum_{i,j,k} a_{ijk}^2} \end{aligned} \tag{30}$$

Definition 1.1 (DFT): For a vector $\mathbf{v} \in \mathbb{R}^n$, the Discrete Fourier Transformation of \mathbf{v} is denoted as $\bar{\mathbf{v}}$ and given by

$$\bar{\mathbf{v}} = \mathbf{F}_n \mathbf{v} \in \mathbb{C}^n \tag{31}$$

where \mathbf{F}_n is the DFT matrix defined as

$$\mathbf{F}_n = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \dots & \omega^{(n-1)(n-1)} \end{bmatrix} \in \mathbb{C}^{n \times n}$$

where $\omega = e^{-\frac{2\pi i}{n}}$ is a primitive n -th root of unity in which $i = \sqrt{-1}$. Note that \mathbf{F}_n/\sqrt{n} is unitary matrix.

Lemma 1.1: Given any real vector $\mathbf{v} \in \mathbb{R}^n$, the associated $\bar{\mathbf{v}}$ satisfied

$$\bar{v}_1 \text{ is real and } \text{conj}(\bar{v}_i) = \bar{v}_{n-i+2}, i = 2, \dots, \lfloor \frac{n+1}{2} \rfloor \quad (32)$$

Conversely, for any given complex $\bar{\mathbf{v}} \in \mathbb{C}^n$ satisfying (32), there exists a real block circulant matrix $\text{circ}(\mathbf{v})$ holds

$$\mathbf{F}_n \cdot \text{circ}(\mathbf{v}) \cdot \mathbf{F}_n^{-1} = \text{Diag}(\bar{\mathbf{v}})$$

$$\text{where } \text{circ}(\mathbf{v}) = \begin{bmatrix} v_1 & v_n & \cdots & v_2 \\ v_2 & v_1 & \cdots & v_3 \\ \vdots & \vdots & \ddots & \vdots \\ v_n & v_{n-1} & \cdots & v_1 \end{bmatrix} \in \mathbb{C}^{n \times n} \text{ and}$$

$$\text{Diag}(\bar{\mathbf{v}}) = \begin{bmatrix} \bar{v}_1 & 0 & \cdots & 0 \\ 0 & \bar{v}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{v}_n \end{bmatrix} \quad (33)$$

To extend similar properties for tensor, some operation especially for tensor are needed as follow

Operation 1.1 (Bdiag): For any $\bar{\mathbf{A}} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, the block diagonal matrix is denoted as $\bar{\mathbf{A}} \in \mathbb{C}^{n_1 n_3 \times n_2 n_3}$ with its i -th block on the diagonal as the i -th frontal slice $\bar{\mathbf{A}}^{(i)}$ of $\bar{\mathbf{A}}$.

$$\bar{\mathbf{A}} = \text{bdiag}(\bar{\mathbf{A}}) = \begin{bmatrix} \bar{\mathbf{A}}^{(1)} & & & \\ & \bar{\mathbf{A}}^{(2)} & & \\ & & \ddots & \\ & & & \bar{\mathbf{A}}^{(n_3)} \end{bmatrix}$$

Operation 1.2 (Bcirc): For any $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the block circulant matrix of \mathcal{A} is denoted as $\text{bcirc}(\mathcal{A}) \in \mathbb{R}^{n_1 n_3 \times n_2 n_3}$

$$\text{bcirc}(\mathcal{A}) = \begin{bmatrix} \mathbf{A}^{(1)} & \mathbf{A}^{(n_3)} & \cdots & \mathbf{A}^{(2)} \\ \mathbf{A}^{(2)} & \mathbf{A}^{(1)} & \cdots & \mathbf{A}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}^{(n_3)} & \mathbf{A}^{(n_3-1)} & \cdots & \mathbf{A}^{(1)} \end{bmatrix}$$

Operation 1.2 (Unfold): For any $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the unfold of \mathcal{A} is denoted as follow

$$\text{unfold}(\mathcal{A}) = \begin{bmatrix} \mathbf{A}^{(1)} \\ \mathbf{A}^{(2)} \\ \vdots \\ \mathbf{A}^{(n_3)} \end{bmatrix}, \text{fold}(\text{unfold}(\mathcal{A})) = \mathcal{A}$$

Then based on the DFT of vector, the **Lemma 1.1** can be extended to the tensor as follow

Lemma 1.2: For any tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its DFT associated as $\bar{\mathcal{A}} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ satisfied

$$(\mathbf{F}_{n_3} \otimes \mathbf{I}_{n_1}) \cdot \text{bric}(\mathcal{A}) \cdot (\mathbf{F}_{n_3}^{-1} \otimes \mathbf{I}_{n_2}) = \bar{\mathcal{A}} \quad (34)$$

where \otimes denotes the Kronecker product and $(\mathbf{F}_{n_3} \otimes \mathbf{I}_{n_1})/\sqrt{n_3}$ is unitary. Then, we have

$$\begin{cases} \bar{\mathbf{A}}^{(1)} \in \mathbb{R}^{n_1 \times n_2} \\ \text{conj}(\bar{\mathbf{A}}^{(i)}) = \bar{\mathbf{A}}^{(n_3-i+2)}, i = 2, \dots, \lfloor \frac{n_3+1}{2} \rfloor \end{cases} \quad (35)$$

Conversely, for any given $\bar{\mathcal{A}} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ satisfying (35), there exists a real tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ such that (34) holds. Moreover, the following properties of tensor DFT are used frequently:

$$\|\mathcal{A}\|_F = \frac{1}{\sqrt{n_3}} \|\bar{\mathcal{A}}\|_F \quad (36)$$

$$\langle \mathcal{A}, \mathcal{B} \rangle = \frac{1}{\sqrt{n_3}} \langle \bar{\mathcal{A}}, \bar{\mathcal{B}} \rangle \quad (37)$$

Definition 1.2 (T-Product): Let $\mathcal{A} \in \mathbb{R}^{n_1 \times l \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{l \times n_2 \times n_3}$. Then the t-product $\mathcal{A} * \mathcal{B}$ is defined to be a tensor of size $n_1 \times n_2 \times n_3$

$$\mathcal{A} * \mathcal{B} = \text{fold}(\text{bcirc}(\mathcal{A})) \cdot \text{unfold}(\mathcal{B}) \quad (38)$$

There are some other concepts on tensor extended from the matrix as follow:

Definition 1.3 (Conjugate Transpose): The conjugate transpose of a tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is the tensor $\mathcal{A}^* \in \mathbb{C}^{n_2 \times n_1 \times n_3}$ obtained by conjugate transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through n_3 .

Definition 1.3 (Identity Tensor): The identity tensor $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ is the tensor with its first frontal slice being the $n \times n$ identity matrix, and other frontal slices being all zeros.

Definition 1.3 (Orthogonal Tensor): A tensor $\mathcal{Q} \in \mathbb{R}^{n \times n \times n_3}$ is orthogonal if it satisfies $\mathcal{Q}^* \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^* = \mathcal{I}$

Definition 1.3 (F-Diagonal Tensor): A tensor is called f-diagonal Tensor if each of its frontal slices is a diagonal matrix.

ACKNOWLEDGMENT

The authors would like to thank the iTrust, Centre for Research in Cyber Security, Singapore University of Technology and Design, for providing the open EPIC dataset.

REFERENCES

- [1] L. Nie, X. Wang, S. Wang, Z. Ning, M. S. Obaidat, B. Sadoun, and S. Li, "Network traffic prediction in industrial Internet of Things backbone networks: A multitask learning mechanism," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 7123–7132, Oct. 2021.
- [2] R. Wang, Y. Zhang, L. Peng, G. Fortino, and P.-H. Ho, "Time-varying-aware network traffic prediction via deep learning in IIoT," *IEEE Trans. Ind. Informat.*, early access, Mar. 30, 2022, doi: 10.1109/THI.2022.3163558.
- [3] M. Abdel-Basset, V. Chang, H. Hawash, R. K. Chakraborty, and M. Ryan, "Deep-IFS: Intrusion detection approach for industrial Internet of Things traffic in fog environment," *IEEE Trans. Ind. Informat.*, vol. 17, no. 11, pp. 7704–7715, Nov. 2021.
- [4] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. D. Kolaczyk, and N. Taft, "Structural analysis of network traffic flows," in *Proc. Int. Conf. Meas. Model. Comput. Syst.*, 2004, pp. 61–72.
- [5] R. Castro, M. Coates, G. Liang, R. Nowak, and B. Yu, "Network tomography: Recent developments," *Statist. Sci.*, vol. 19, no. 3, pp. 499–517, 2004.
- [6] M. Mardani and G. B. Giannakis, "Robust network traffic estimation via sparsity and low rank," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 4529–4533.
- [7] K. Xie, L. Wang, X. Wang, G. Xie, G. Zhang, D. Xie, and J. Wen, "Sequential and adaptive sampling for matrix completion in network monitoring systems," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2015, pp. 2443–2451.

- [8] K. Xie, L. Wang, X. Wang, G. Xie, J. Wen, and G. Zhang, "Accurate recovery of internet traffic data: A tensor completion approach," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun.*, Apr. 2016, pp. 1–9.
- [9] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5249–5257.
- [10] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, and L.-J. Deng, "Multi-dimensional imaging data recovery via minimizing the partial sum of tubal nuclear norm," *J. Comput. Appl. Math.*, vol. 372, Jul. 2020, Art. no. 112680.
- [11] T. Franz, A. Schultz, S. Sizov, and S. Staab, "TripleRank: Ranking semantic web data by tensor decomposition," in *Proc. Int. Semantic Web Conf. Cham, Switzerland: Springer*, 2009, pp. 213–228.
- [12] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Sep. 2009.
- [13] B. Huang, C. Mu, D. Goldfarb, and J. Wright, "Provable models for robust low-rank tensor completion," *Pacific J. Optim.*, vol. 11, no. 2, pp. 339–364, 2015.
- [14] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 925–938, Jan. 2020.
- [15] Z. Zhang and S. Aeron, "Exact tensor completion using t-SVD," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1511–1526, Mar. 2016.
- [16] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.
- [17] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multimodal factor analysis," *UCLA Work. Papers Phonetics*, vol. 16, pp. 1–84, 1970.
- [18] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, Sep. 1996.
- [19] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra Appl.*, vol. 435, no. 3, pp. 641–658, 2011.
- [20] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, 2013.
- [21] T.-H. Oh, Y.-W. Tai, J.-C. Bazin, H. Kim, and I. S. Kweon, "Partial sum minimization of singular values in robust PCA: Algorithm and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 4, pp. 744–758, Apr. 2015.
- [22] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, T.-H. Ma, and T.-Y. Ji, "Mixed noise removal in hyperspectral image via low-fibered-rank regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 734–749, Jan. 2020.
- [23] R. R. R. Barbosa, R. Sadre, and A. Pras, "A first look into SCADA network traffic," in *Proc. IEEE Netw. Oper. Manage. Symp.*, Apr. 2012, pp. 518–521.
- [24] B. Galloway and G. P. Hancke, "Introduction to industrial control networks," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 860–880, 2nd Quart., 2013.
- [25] G. Yue, Z. sun, J. Tian, H. Zhu, and B. Zhang, "Power grid industrial control system traffic classification based on two-dimensional convolutional neural network," in *Communications, Signal Processing, and Systems*. Singapore: Springer, 2022, pp. 41–48.
- [26] S. Adepun, N. K. Kandasamy, and A. Mathur, "EPIC: An electric power testbed for research and training in cyber physical systems security," in *Computer Security*. Cham, Switzerland: Springer, 2019, pp. 37–52.
- [27] T.-H. Oh, H. Kim, Y.-W. Tai, J.-C. Bazin, and I. S. Kweon, "Partial sum minimization of singular values in RPCA for low-level vision," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 145–152.
- [28] K. Chen, H. Dong, and K.-S. Chan, "Reduced rank regression via adaptive nuclear norm penalization," *Biometrika*, vol. 100, no. 4, pp. 901–920, 2012.
- [29] H. Zou, "The adaptive lasso and its oracle properties," *J. Amer. Stat. Assoc.*, vol. 101, no. 476, pp. 1418–1429, Jan. 2006.
- [30] L. Mirsky, "A trace inequality of John von Neumann," *Monatshefte für Math.*, vol. 79, no. 4, pp. 303–306, 1975.
- [31] H. Fan, Y. Chen, Y. Guo, H. Zhang, and G. Kuang, "Hyperspectral image restoration using low-rank tensor recovery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4589–4604, Oct. 2017.



GANG YUE received the B.Eng. degree in information engineering and the M.Eng. degree in instructional technology from the Beijing University of Posts and Telecommunications, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree with the WSPN Laboratory. His research interests include machine learning, signal processing, and anomaly detection.



ZHUO SUN (Member, IEEE) received the Ph.D. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2007. He is currently an Associate Professor with the Wireless Signal Processing and Networks Laboratory, Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications. His research interests include wireless communication, signal processing in communication, and machine learning in communication.



JINPO FAN received the B.S. degree in communication engineering and the M.S. degree in electronic and communication engineering from Beijing Electronic Science and Technology Institute, Beijing, China, in 2015 and 2020, respectively. He is currently pursuing the Ph.D. degree in information and communication engineering with the Beijing University of Posts and Telecommunications, Beijing. His research interests include machine learning and signal recognition, multidimensional data processing, and wireless sensor network security.

• • •