Atom-Refined Multiway Greedy Algorithm for Tensor-Based Compressive Sensing

RONGQIANG ZHAO[1,2], **(Student Member, IEEE), JUN FU**[1,2]**, LUQUAN REN**[1,2]**,
AND QIANG WANG**[3]**, (Member, IEEE)**
[1]Key Laboratory of Bionic Engineering, Ministry of Education, Jilin University, Changchun 130022, China
[2]College of Biological and Agricultural Engineering, Jilin University, Changchun 130022, China
[3]Department of Control Science and Engineering, Harbin Institute of Technology, Harbin 150001, China

Corresponding author: Jun Fu (fu_jun@jlu.edu.cn)

ABSTRACT In this paper, we develop a novel multiway greedy algorithm, named atom-refined multiway orthogonal matching pursuit, for tensor-based compressive sensing (TCS) reconstruction. The alternative supports of each dimension are selected using the respective inner product tensors and refined via a global least square coefficients tensor. For each inner product tensor, the Frobenius-norm (F-norm) of the tensor bands, instead of the largest magnitude entry, is employed to measure the correlation between the atoms and the residual. Theoretical analysis shows that the proposed algorithm could guarantee to exactly reconstruct an arbitrary multi-dimensional block-sparse signal in the absence of noise, provided that the sensing matrices for each dimension satisfy restricted isometry properties with constant parameters. The maximum required number of iterations for exact reconstruction shows an approximate logarithmic growth as the signal size increases. Furthermore, under the noise condition, it is presented that the F-norm of the reconstruction error can be upper-bounded by using the F-norm of noise and the restricted isometry constants of sensing matrices for each dimension. The simulation results demonstrate that the proposed algorithm exhibits obvious advantages as regards both reconstruction accuracy and speed compared with the existing multiway greedy algorithms. Besides TCS, the proposed algorithm also has the potential to be applied in diverse fields, such as hyperspectral image processing and tensor-based dictionary learning.

INDEX TERMS Compressive sensing, sparse representation, multi-dimensional block-sparsity, greedy algorithm.

I. INTRODUCTION

Since its proposal, compressive sensing (CS) has attracted considerable attention from researchers in signal processing and many other fields [1]–[6]. CS aims to reconstruct a signal from a set of measurements that are considerably smaller than the original signal, so that the data transmission and storage loads can be reduced significantly [1], [7]. Considering a standard CS problem, let $x \in \mathbb{R}^n$ denote a k-sparse signal. Here, $k < n$, meaning that only k coefficients of x are non-zero, with the others being zero or approximately zero. The signal x can be compressively sampled with respect to a sensing matrix $\Phi \in \mathbb{R}^{m \times n}$; this process is expressed as

$$y = \Phi x, \tag{1}$$

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

where $y \in \mathbb{R}^m$ represents the measurements and $k < m < n$. Recovery of the original x from y and Φ is referred to as reconstruction, and is expressed as the following NP-hard l_0 problem [8]:

$$\hat{x} = arg\, min\|x\|_0 \quad s.t.\, y = \Phi x, \tag{2}$$

where the norm $\|x\|_0$ denotes the number of non-zero entries of x. To date, a set of greedy algorithms has been developed to solve the l_0 problem, including the orthogonal matching pursuit (OMP) [9], subspace pursuit (SP) [10], generalized orthogonal matching pursuit (GOMP) [11], stage-wise OMP (StOMP) [12], compressive sampling matching pursuit (CoSAMP) [13], and regularized orthogonal matching pursuit (ROMP) algorithms [14].

The standard CS theory primarily focuses on one-dimensional (1D) signals. However, many applications in the CS field are based on multi-dimensional signals (tensors),

including wireless technology [15], [16], hyperspectral imaging [17], [18], video processing [19], [20], medical imaging [21], [22], and so on [23]–[26]. Based on the standard CS theory, one must multiplex the data of all dimensions using a global sampling device, corresponding to a dense and large sensing matrix. This process introduces a high physical complexity to the sampling hardware and is difficult to implement [27]–[29]. To relieve complexity of the sampling implementation, researchers have employed the Kronecker structure for the sensing matrix and sparsifying base to replace the distributed scheme for global operation [30]–[33]. Under this framework, conventional greedy algorithms are hindered by the considerable computational complexity of the data reconstruction [23], [30], [34]. To address this problem, Caiafa and Cichocki [35] have developed a well-known greedy algorithm, the N-way block orthogonal matching pursuit (NBOMP) algorithm, by exploiting the multi-dimensional block-sparsity of tensors. Instead of a single support for a global sensing matrix, NBOMP aims to find a set of supports corresponding to the sensing matrices of each dimension, based on the Tucker model. This strategy is called ''multiway'', meaning that an entry in the inner product tensor indicates the correlation of the measurement tensor and all sensing matrices [36], so that multiple supports are updated in each iteration; thus, the total number of iterations is reduced considerably. As an extension of NBOMP, another multiway greedy algorithm known as the multi-atom tensor orthogonal matching pursuit (MaTOMP) algorithm has been proposed [37], in which, more than one atom can be added to a support in each update. This approach accelerates the reconstruction but loses the accuracy.

In this paper, we propose a novel multiway greedy algorithm called the atom-refined multiway orthogonal matching pursuit (ArMOMP) algorithm. Its convergence, low-complexity, and robustness are verified through theoretical analysis and numerical simulations. Our contributions can be summarized as follows:

- We develop a novel strategy for multiway greedy recovery, which is reflected in two aspects. First, we employ the respective inner product tensors to select the alternative supports for each dimension while refining them via a global least square (LS) coefficients tensor. Second, we use the F-norm of the tensor bands instead of the largest magnitude entry to measure the correlation between the atoms and the residual.
- We prove that the proposed algorithm could guarantee to exactly reconstruct an arbitrary multi-dimensional block-sparse signal in the absence of noise, provided that the sensing matrices for each dimension satisfy restricted isometry properties with constant parameters, i.e., sufficient condition for convergence. Additionally, we derive the maximum required number of iterations for exact reconstruction when the sufficient condition for convergence is satisfied.
- For the noise condition, we provided the theoretical upper-bound for the F-norm of reconstruction error,

which is determined by the noise level and the restricted isometry constants of sensing matrices for each dimension. We also analyze the advantages of the proposed algorithm as regards both reconstruction accuracy and speed compared to existing multiway greedy algorithms via numerical simulations.

This paper is organized as follows. Section II introduces notation used in this paper and reviews related works. Section III provides a detailed description of the ArMOMP algorithm, including analysis of the algorithm convergence and complexity. In Section IV, we study the results obtained in Section III under noise conditions. Section V concludes the paper. Finally, proofs of certain theorems are provided in the Appendix.

II. PRELIMINARIES
A. NOTATION AND DEFINITIONS
The scalars, vectors, matrices, and tensors are denoted by italic lowercase letters (e.g., x), bold italic lowercase letters (e.g., $\boldsymbol{x}$), bold italic capital letters (e.g., $\boldsymbol{X}$), and bold underlined capital letters (e.g., $\underline{\boldsymbol{X}}$), respectively. The transpose and pseudo inverse of a matrix $\boldsymbol{X}$ are denoted by $\boldsymbol{X}^*$ and $\boldsymbol{X}^\dagger$, respectively. For a set comprised of the matrices $\{\boldsymbol{\Phi}_j\}$ and the respective supports $\{\boldsymbol{\gamma}_j\}$, we denote $\boldsymbol{\Phi}_{\boldsymbol{\gamma},j} = \boldsymbol{\Phi}_j(:, \boldsymbol{\gamma}_j)$ in this paper.

Definition 1: The operation that unfolds a tensor $\underline{\boldsymbol{X}} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_d}$ to its $j-$mode matrix $\boldsymbol{X}_{(j)} = unfold_{(j)}(\underline{\boldsymbol{X}})$ is defined as

$$X_{(j)}(i_j, p) = \underline{\boldsymbol{X}}(i_1, \cdots, i_j, \cdots, i_d),$$
$$p = 1 + \sum_{k=1}^{d-1}(i_{v_k} - 1)w_k, \quad w_k = \prod_{l=1}^{k-1} n_{v_l}, \quad (3)$$

where $\boldsymbol{v} = [j+1, j+2, \cdots, d, 1, \cdots, j-1]^*$, $\boldsymbol{X}_{(j)} \in \mathbb{R}^{n_j \times N_j}$, *and* $N_j = \prod_{p \neq j} n_p$.

Definition 2: We state that $\underline{\boldsymbol{Y}} = \underline{\boldsymbol{X}} \times_j \boldsymbol{\Phi}_j$, when $\underline{\boldsymbol{Y}}_{(j)} = \boldsymbol{\Phi}_j \underline{\boldsymbol{X}}_{(j)}$, where $\underline{\boldsymbol{X}}_{(j)} = unfold_{(j)}(\underline{\boldsymbol{X}})$, $\underline{\boldsymbol{Y}}_{(j)} = unfold_{(j)}(\underline{\boldsymbol{Y}})$, $\underline{\boldsymbol{X}} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, $\underline{\boldsymbol{Y}} \in \mathbb{R}^{n_1 \times \cdots \times m_j \times \cdots \times n_d}$, *and* $\boldsymbol{\Phi}_j \in \mathbb{R}^{m_j \times n_j}$.

Definition 3: The projection of a tensor $\underline{\boldsymbol{Y}} \in \mathbb{R}^{m_1 \times m_2 \cdots \times m_d}$ onto the space spanned by matrices $\{\boldsymbol{\Phi}_j\}$ is given by

$$pro(\underline{\boldsymbol{Y}}, \{\boldsymbol{\Phi}_j\}) = \underline{\boldsymbol{Y}} \times_1 \boldsymbol{\Phi}_1 \boldsymbol{\Phi}_1^\dagger \times_2 \boldsymbol{\Phi}_2 \boldsymbol{\Phi}_2^\dagger \cdots \times_d \boldsymbol{\Phi}_d \boldsymbol{\Phi}_d^\dagger, \quad (4)$$

where $\boldsymbol{\Phi}_j \in \mathbb{R}^{m_j \times k_j}$, $k_j \leq m_j$. *The residual between* $\underline{\boldsymbol{Y}}$ *and* $pro(\underline{\boldsymbol{Y}}, \{\boldsymbol{\Phi}_j\})$ *is given by*

$$res(\underline{\boldsymbol{Y}}, \{\boldsymbol{\Phi}_j\}) = \underline{\boldsymbol{Y}} - pro(\underline{\boldsymbol{Y}}, \{\boldsymbol{\Phi}_j\}), \quad (5)$$

Definition 4: The $k-$RIP constant of a sensing matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n}$ is denoted by δ_k, the minimum value that satisfies $0 \leq \delta_k \leq 1$, and

$$(1 - \delta_k)\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{\Phi}\boldsymbol{x}\|_2^2 \leq (1 + \delta_k)\|\boldsymbol{x}\|_2^2, \quad (6)$$

for all vectors $\boldsymbol{x} \in \mathbb{R}^n$ *with* $\|\boldsymbol{x}\|_0 \leq k$. *The relationship* $\delta_{k_1} \leq \delta_{k_2}$ *holds if* $k_1 \leq k_2$ *for integers* k_1 *and* k_2.

B. PROBLEM STATEMENT

A multi-dimensional signal $\underline{X} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_d}$ is said to be $(k_1, k_2, \cdots, k_d)-$block-sparse with regard to a set of supports $\{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_d\}$ when the following equation holds:

$$\underline{X}(i_1, i_2, \cdots, i_d) = 0$$
$$\forall \underline{X}(i_1, i_2, \cdots, i_d) \notin \underline{X}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_d), \quad (7)$$

with $|\boldsymbol{\gamma}_j| = k_j, j = 1, 2, \cdots, d$. The term "multi-dimensional block-sparse" is used here because the non-zero entries of $\underline{X}$ are positioned in a block (core sub-tensor). This concept is also important for compressive multi-dimensional signals, because they exhibit multi-dimensional block-sparse behavior when sparsely represented under the TCS framework [30], [35]. Let $\boldsymbol{\Phi}_j \in \mathbb{R}^{m_j \times n_j}, j = 1, 2, \cdots, d$ be the sensing matrices for each dimension of $\underline{X}$. The compressive sampling can be expressed as

$$\underline{Y} = \underline{X} \times_1 \boldsymbol{\Phi}_1 \times_2 \boldsymbol{\Phi}_2 \cdots \times_d \boldsymbol{\Phi}_d, \quad (8)$$

where $\underline{Y} \in \mathbb{R}^{m_1 \times m_2 \cdots \times m_d}$ denotes the measurements. Hence, the reconstruction problem becomes the problem of solving $\underline{X}$ from $\underline{Y}$ and $\boldsymbol{\Phi}_j, j = 1, 2, \cdots, d$. This is expressed as

$$\hat{\underline{X}} = arg\, min \|\underline{X}\|_0 \quad s.t.\ \underline{Y} = \underline{X} \times_1 \boldsymbol{\Phi}_1 \times_2 \boldsymbol{\Phi}_2 \cdots \times_d \boldsymbol{\Phi}_d, \quad (9)$$

where $\underline{X}$ is under the constraint given in (7). Eq. (9) is the main focus of this paper, and constitutes the objective problem addressed by the proposed algorithm. Note that subsequent derivations require the below lemma, the proof of which is provided in the Appendix (part V-A).

Lemma 1: Given a $(k_1, k_2, \cdots, k_d)-$block-sparse signal $\underline{X} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_d}$ with non-zero supports $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_d$ and sensing matrices $\boldsymbol{\Phi}_j \in \mathbb{R}^{m_j \times n_j}$, suppose that the sensing matrices satisfy the k_j-RIP with constants δ_{k_j}. Then, the following inequalities hold.

$$\prod_{j=1}^{d} (1 - \delta_{k_j})^{\frac{1}{2}} \|\underline{X_B}\|_F \leq \|\underline{X} \times_1 \boldsymbol{\Phi}_1 \cdots \times_d \boldsymbol{\Phi}_d\|_F,$$

$$= \|\underline{X_B} \times_1 \boldsymbol{\Phi}_{\boldsymbol{\gamma},1} \cdots \times_d \boldsymbol{\Phi}_{\boldsymbol{\gamma},d}\|_F$$

$$\leq \prod_{j=1}^{d} (1 + \delta_{k_j})^{\frac{1}{2}} \|\underline{X_B}\|_F, \quad (10)$$

$$\prod_{j=1}^{d} (1 - \delta_{k_j}) \|\underline{X_B}\|_F \leq \|\underline{X_B} \times_1 \boldsymbol{\Phi}_{\boldsymbol{\gamma},1}^* \boldsymbol{\Phi}_{\boldsymbol{\gamma},1} \cdots$$

$$\times_d \boldsymbol{\Phi}_{\boldsymbol{\gamma},d}^* \boldsymbol{\Phi}_{\boldsymbol{\gamma},d}\|_F,$$

$$\leq \prod_{j=1}^{d} (1 + \delta_{k_j}) \|\underline{X_B}\|_F, \quad (11)$$

$$\|\underline{X_B} \times_1 \boldsymbol{\Phi}_{\boldsymbol{\nu},1}^* \boldsymbol{\Phi}_{\boldsymbol{\gamma},1} \cdots \times_d \boldsymbol{\Phi}_{\boldsymbol{\nu},d}^* \boldsymbol{\Phi}_{\boldsymbol{\gamma},d}\|_F$$

$$\leq \prod_{j=1}^{d} \delta_{k_j+l_j} \|\underline{X_B}\|_F, \quad (12)$$

where $\underline{X_B} = \underline{X}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_d)$, $\boldsymbol{\Phi}_{\boldsymbol{\gamma},j} \in \mathbb{R}^{m_j \times k_j}$, $\boldsymbol{\Phi}_{\boldsymbol{\nu},j} \in \mathbb{R}^{m_j \times l_j}$, and $\boldsymbol{\Phi}_{\boldsymbol{\gamma},j} \bigcap \boldsymbol{\Phi}_{\boldsymbol{\nu},j} = \emptyset$.

III. PROPOSED ALGORITHM

In this section, we first detailedly introduce the ArMOMP algorithm. Then we discuss the convergence and computational complexity of ArMOMP. Following the numerical simulations are provided to verify the algorithm performance.

A. INTRODUCTION OF ARMOMP

1) OVERVIEW

This subsection provides a detailed introduction of the ArMOMP algorithm, which is designed to solve Eq. (9). For the ArMOMP algorithm, all indexes of support for each sensing matrix are selected simultaneously; then, refinement is conducted until all supports are correct. The ArMOMP algorithm is divided it into two stages and presented as Algorithm 1. In the first stage, the used parameters, the supports for the sensing matrices, and the corresponding coefficients are initialized. The second stage consists of two processes, mixture and refinement, which cycle alternately until the termination condition is reached.

Algorithm 1 The ArMOMP Algorithm

Require: Sensing matrices $\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \cdots, \boldsymbol{\Phi}_d$ with $\boldsymbol{\Phi}_j \in \mathbb{R}^{m_j \times n_j}$, measurements $\underline{Y} \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_d}$, sparsity level $\{k_j\}$, maximum number of iterations l_{max}, and tolerance ϵ.

    Initialization stage:
1: Execute Algorithm 2
    Iteration stage (begin with $l = 1$):
2: **while** $l \leq l_{max}$ and the termination condition is not reached **do**
3:     Execute Algorithm 3.
4:     $l = l + 1$.
5: **end while**
6: $\hat{\underline{X}}(T_1^l, T_2^l, \cdots, T_d^l) = \underline{X}_{C,T^l}$.

Ensure: Reconstruction $\hat{\underline{X}}$.

---

2) INITIALIZATION

The initialization stage is summarized in Algorithm 2. Steps 1–5 aim to determine the number of mixed indexes $\{k_1', k_2', \cdots, k_d'\}$. In general, we set $k_j' = k_j$. However, a special condition should be considered, in which $k_j > m_j/2$. In this case, the LS problem has infinite solutions. To avoid this scenario, k_j' is set to $m_j - k_j$ so that $k_j + k_j' \leq m_j$. For convenience, we consider $k_j' = k_j$ only. One can obtain similar conclusions under the condition $k_j' = m_j - k_j$.

In ArMOMP, the initial supports of each sensing matrix are found by using the respective inner product tensor, and the correlation between the atoms and the residual are measured by the F-norm of the bands of the inner product tensor. For example, e.g., for j^{th} dimension, the problem can be expressed as

$$T_j^0 = arg\, max_{T_j^0} \|\underline{P}(:, \cdots, T_j^0, \cdots, :)\|_F,$$

$$s.t.\ |T_j^0| = k_j, \quad \underline{P} = \underline{Y} \times_j \boldsymbol{\Phi}_j^*, \quad (13)$$

Algorithm 2 Initialization Stage

Require: Sensing matrices $\mathbf{\Phi}_1, \mathbf{\Phi}_2, \cdots, \mathbf{\Phi}_d$ with $\mathbf{\Phi}_j \in \mathbb{R}^{m_j \times n_j}$, measurements $\underline{\mathbf{Y}} \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_d}$, and sparsity level $\{k_j\}$.

1: **if** $k_j < m_j/2$, **then**
2:     $k'_j = k_j, \ \forall j = 1, 2, \cdots, d,$
3: **else**
4:     $k'_j = m_j - k_j.$
5: **end if**
6: **for** $j = 1$ to d **do**
7:     $\underline{\mathbf{P}}^0 = \underline{\mathbf{Y}} \times_j \mathbf{\Phi}_j^* \in \mathbb{R}^{m_1 \times m_2 \cdots \times n_j \cdots \times m_d}.$
8:     $\mathbf{P}^0_{(j)} = unfold_{(j)}(\underline{\mathbf{P}}^0).$
9:     $\boldsymbol{e}_j(i) = \|\mathbf{P}^0_{(j)}(i, :)\|_F, i = 1, 2, \cdots, n_j.$
10:    Find the k_j maximum entries of $\boldsymbol{e}_j$, the indexes of which are denoted T_j^0, supporting $\mathbf{\Phi}_{T^0,j}$.
11: **end for**
12: $\underline{\mathbf{Y}}_{R^0} = res(\underline{\mathbf{Y}}, \mathbf{\Phi}_{T^0,1}, \mathbf{\Phi}_{T^0,2}, \cdots, \mathbf{\Phi}_{T^0,d}).$
Ensure: $\underline{\mathbf{Y}}_{R^0}, \{k'_j\}, \{T_j^0\}, j = 1, 2, \cdots, d.$

which is solved by steps 7–10 in the initialization stage. The convergence is analyzed in the following subsection.

3) MIXTURE AND REFINEMENT

The mixture and refinement stage is summarized in Algorithm 3. The mixture process is realized by steps 1–6, with the aim of searching for the most likely atoms from the remaining options. We compute the inner product between the residual tensor and sensing matrix for each dimension, and then compare the relevance between the residual and each atom, which is expressed as

$$\Delta_j = arg\,max_{\Delta_j} \|\mathbf{P}_j^l(:, \cdots, \Delta_j, \cdots, :)\|_F,$$
$$s.t. \ \Delta_j \cap T_j^{l-1} = \emptyset, \ |\Delta_j| = k'_j, \ \mathbf{P}_j^l = \underline{\mathbf{Y}}_{R^{l-1}} \times_j \mathbf{\Phi}_j^*. \tag{14}$$

The indexes of the k'_j most relevant atoms Δ_j are selected to merge with the current indexes T_j^{l-1}, denoted by $\widetilde{T}_j = T_j^{l-1} \cup \Delta_j$. It should be noted that, for simplicity, we omit the superscripts of the intermediate variables $\widetilde{T}$, Δ_j, and so on.

Next, it is necessary to refine the supports $\{\widetilde{T}_j\}$ and abandon the k'_j indexes for each support. The specific method is to first compute the LS coefficient for $\underline{\mathbf{Y}}$ using $\{\mathbf{\Phi}_{\widetilde{T},j}\}$, $j = 1, 2, \cdots, d$, such that

$$\underline{\mathbf{X}}_{C,\widetilde{T}} = argmin_{\underline{\mathbf{C}}} \|(\underline{\mathbf{C}} \times_1 \mathbf{\Phi}_{\widetilde{T},1} \times_2 \mathbf{\Phi}_{\widetilde{T},2} \cdots \times_d \mathbf{\Phi}_{\widetilde{T},d} - \underline{\mathbf{Y}}\|_F. \tag{15}$$

This can be solved via

$$\underline{\mathbf{X}}_{C,\widetilde{T}} = \underline{\mathbf{Y}} \times_1 \mathbf{\Phi}_{\widetilde{T},1}^{\dagger} \times_2 \mathbf{\Phi}_{\widetilde{T},2}^{\dagger} \cdots \times_d \mathbf{\Phi}_{\widetilde{T},d}^{\dagger}. \tag{16}$$

Then, the refined indexes that make the greatest contributions to the LS representation are determined. For each dimension j,

Algorithm 3 Mixture and Refinement Stage

Require: Sensing matrices $\mathbf{\Phi}_1, \mathbf{\Phi}_2, \cdots, \mathbf{\Phi}_d$ with $\mathbf{\Phi}_j \in \mathbb{R}^{m_j \times n_j}$, measurements $\underline{\mathbf{Y}} \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_d}, \underline{\mathbf{Y}}_{R^0}$, $l, \{k_j\}, \{k'_j\}, \{T_j^0\}$, and $j = 1, 2, \cdots, d$

1: **for** $j = 1$ to d **do**
2:     $\mathbf{P}_j^l = \underline{\mathbf{Y}}_{R^{l-1}} \times_j \mathbf{\Phi}_j^* \in \mathbb{R}^{m_1 \times m_2 \cdots \times n_j \cdots \times m_d}.$
3:     $\overline{\mathbf{P}_{(j)}^l} = unfold_{(j)}(\mathbf{P}_j^l).$
4:     $\boldsymbol{e}_j(i) = \|\mathbf{P}_{(j)}^l(i, :)\|_F, i = 1, 2, \cdots, n_j.$
5:     Find the k'_j maximum entries of $\boldsymbol{e}_j$, the indexes of which are denoted as Δ_j, s.t $\Delta_j \cap T_j^{l-1} = \emptyset.$
6:     $\widetilde{T}_j = T_j^{l-1} \cup \Delta_j$, supporting $\mathbf{\Phi}_{\widetilde{T},j}.$
7: **end for**
8: $\underline{\mathbf{X}}_{C,\widetilde{T}} = argmin_{\underline{\mathbf{C}}} \|(\underline{\mathbf{C}} \times_1 \mathbf{\Phi}_{\widetilde{T},1} \times_2 \mathbf{\Phi}_{\widetilde{T},2} \cdots \times_d \mathbf{\Phi}_{\widetilde{T},d} - \underline{\mathbf{Y}}\|_F.$
9: **for** $j = 1$ to d **do**
10:    $\mathbf{X}_{C,\widetilde{T}(j)} = unfold_{(j)}(\underline{\mathbf{X}}_{C,\widetilde{T}}).$
11:    $\boldsymbol{f}_j(i) = \|\mathbf{X}_{C,\widetilde{T}(j)}(i, :)\|_F, i = 1, 2, \cdots, k_j + k'_j.$
12:    Find the k_j maximum entries of $\boldsymbol{f}_j$, the indexes of which are denoted T_j^l, supporting $\mathbf{\Phi}_{T^l,j}.$
13: **end for**
14: $\underline{\mathbf{X}}_{C,T^l} = \underline{\mathbf{Y}} \times_1 \mathbf{\Phi}_{T^l,1}^{\dagger} \times_2 \mathbf{\Phi}_{T^l,2}^{\dagger} \cdots \times_d \mathbf{\Phi}_{T^l,d}^{\dagger}.$
15: $\underline{\mathbf{Y}}_{R^l} = \underline{\mathbf{Y}} - \underline{\mathbf{X}}_{C,T^l} \times_1 \mathbf{\Phi}_{T^l,1} \times_2 \mathbf{\Phi}_{T^l,2} \cdots \times_d \mathbf{\Phi}_{T^l,d}.$
Ensure: $\underline{\mathbf{Y}}_{R^l}, \underline{\mathbf{X}}_{C,T^l}, \{T_j^l\}, j = 1, 2 \cdots d.$

we select the supports T_j^l that satisfies

$$T_j^l = arg\,max_{T_j^l} \|\underline{\mathbf{X}}_{C,\widetilde{T}}(:, \cdots, T_j^l, \cdots, :)\|_F. \tag{17}$$

Here, we employ the F−norm of the coefficients supported by T_j^l to measure the contribution. The convergence is proven in later subsections. The mixture process and refinement process are executed alternately until the maximum number of iterations or the termination condition is reached. There are two different termination conditions, which are introduced as follows.

First, $\|\underline{\mathbf{Y}}_{R^l}\|_F < \epsilon$. This condition means the residual is sufficiently small that the reconstruction has succeeded. For multi-dimensional block-sparse signals, there theoretically exists $\underline{\mathbf{Y}}_{R^l} = 0$ when all supports are found correctly. However, for approximate multi-dimensional block-sparse signals or compressible signals under multi-dimensional block-sparse representation, the reconstruction residual cannot be completely eliminated. Generally, we do not know the degree of approximation; thus, it is difficult to determine ϵ. In that case, we prefer to use the below termination condition.

Second, $\underline{\mathbf{Y}}_{R^l} \geq \underline{\mathbf{Y}}_{R^{l-1}}$. The residual after the latter iteration is larger than that after the former iteration, demonstrating that the iteration has reached a point where the reconstruction may be moving towards non-convergence. In this case, we should break the iteration and revert the outputs $\underline{\mathbf{X}}_{C,T^l}$ and $\{T_j^l\}$ to $\underline{\mathbf{X}}_{C,T^{l-1}}$ and $\{T_j^{l-1}\}$, respectively.

B. CONVERGENCE

In this subsection, we provide a series of theorems regarding the algorithm convergence, which explain why the ArMOMP algorithm can recover a multi-dimensional block-sparse signal with a few iterations.

Theorem 1: Let $\underline{X}_T \in \mathbb{R}^{k_1 \times k_2 \times \cdots \times k_d}$ represent the core sub-tensor of a $(k_1, k_2, \cdots, k_d)$-block-sparse tensor $\underline{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with regard to the supports $\{T_1, T_2, \cdots, T_j\}$, and $\underline{X}_{T-T^{l-1}}$ and $\underline{X}_{T-\widetilde{T}}$ denote the unrecovered parts in $\underline{X}_T$ after the $(l-1)$th iteration and the mixture process of the lth iteration, respectively. Then, the following inequality holds.

$$\|\underline{X}_{T-\widetilde{T}}\|_F \leq \Omega_1 \|\underline{X}_{T-T^{l-1}}\|_F, \tag{18}$$

$$\Omega_1 = \frac{2d\delta_{3k}(1+\delta_{3k})^{\frac{d-1}{2}}\left[(1-\delta_{3k})^d + (2^d-1)^{\frac{1}{2}}\delta_{3k}\right]}{(1-\delta_{3k})^{\frac{3d+1}{2}}}, \tag{19}$$

where δ_{3k} is defined as the maximum of the $3k-$RIP constants of all sensing matrices, i.e., $\delta_{3k} = max\{\delta_{3k_j}\}$, $j = 1, 2, \cdots, d$.
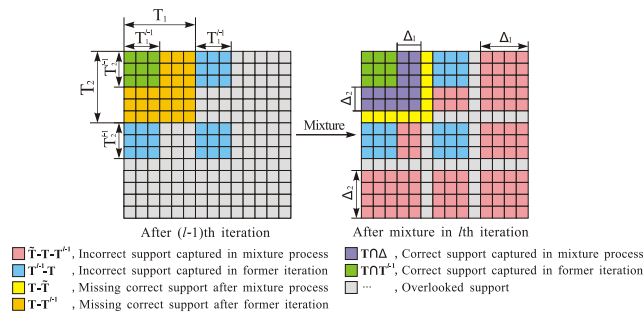
FIGURE 1. Intuitive representation of supports $T - T^{l-1}$ and $T - \widetilde{T}$, with $d = 2$, $n_1 = n_2 = 12$, and $k_1 = k_2 = 6$.

An intuitive representation of $\underline{X}_{T-T^{l-1}}$ and $\underline{X}_{T-\widetilde{T}}$ is displayed in Fig. 1. A fundamental difference is that, compared to the 1-D scenario, a position is said to be on a support only when all its coordinate indexes are on the support. This is an embodiment of coupling under multi-dimensional conditions.

This theorem is derived from the below two lemmas, indicating the relationship between the unrecovered signal after the preceding iteration and that after the mixture process in the current iteration.

Lemma 2: Let $\underline{X}_{R^{l-1}}$ denote the residual after the $(l-1)$th iteration. Then, there exists

$$\|\underline{X}_{R^{l-1}}\|_F \leq \left[1 + \frac{(2^d-1)^{\frac{1}{2}}\delta_{2k}}{(1-\delta_k)^d}\right]\|\underline{X}_{T-T^{l-1}}\|_F, \tag{20}$$

where δ_k and δ_{2k} are defined as the maximum of the $k-$ and $2k-$RIP constants of all sensing matrices, i.e., $\delta_k = max\{\delta_{k_j}\}$, $\delta_{2k} = max\{\delta_{2k_j}\}$, and $j = 1, 2, \cdots, d$.

Lemma 3: It holds that

$$\|\underline{X}_{T-\widetilde{T}}\|_F \leq \frac{2d\delta_{3k}(1+\delta_{2k})^{\frac{d-1}{2}}}{(1-\delta_k)^{\frac{d+1}{2}}}\|\underline{X}_{R^{l-1}}\|_F. \tag{21}$$
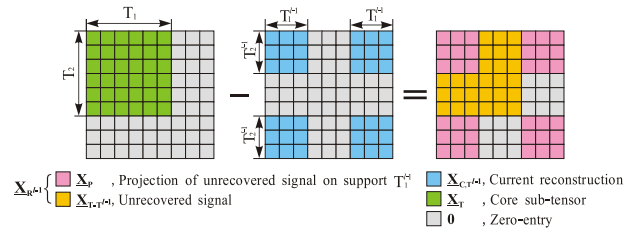

FIGURE 2. Intuitive representation of $\underline{X}_T, \underline{X}_{T^{l-1}}, \underline{X}_{R^{l-1}}$, with $d = 2$ and $k_1 = k_2 = 6$.

The proofs of Lemmas 2 and 3 are presented in the Appendix (parts V-B and V-C, respectively). The unrecovered signal $\underline{X}_{T-T^{l-1}}$ is derived by removing the recovered component $\underline{X}_{T \cap T^{l-1}}$ from $\underline{X}_T$, where $\underline{X}_{T \cap T^{l-1}} = \underline{X}_T(T_1 \cap T_1^{l-1}, \cdots, T_d \cap T_d^{l-1})$. It should be noted that the residual $\underline{X}_R^{l-1}$ between the original signal and the current reconstruction is not equal to the unrecovered signal $\underline{X}_{T-T^{l-1}}$. Fig. 2 illustrates their relationship, where $\underline{X}_R^{l-1}$ is composed of $\underline{X}_{T-T^{l-1}}$ and $\underline{X}_P$. The latter essentially corresponds to the coefficients with respect to the projection of the unrecovered signal onto the selected support, which is derived as follows.

$$\begin{aligned}
\underline{Y}_R^{l-1} &= res(\underline{Y}, \{\mathbf{\Phi}_{T^{l-1},j}\}), \\
&= res(\underline{Y}_{T \cap T^{l-1}}, \{\mathbf{\Phi}_{T^{l-1},j}\}) + res(\underline{Y}_{T-T^{l-1}}, \{\mathbf{\Phi}_{T^{l-1},j}\}), \\
&\overset{(a)}{=} \underline{0} + \underline{Y}_{T-T^{l-1}} - pro(\underline{Y}_{T-T^{l-1}}, \{\mathbf{\Phi}_{T^{l-1},j}\}), \\
&\overset{(b)}{=} \underline{Y}_{T-T^{l-1}} - \underline{X}_P \times_1 \mathbf{\Phi}_{T^{l-1},1} \cdots \times_d \mathbf{\Phi}_{T^{l-1},d},
\end{aligned} \tag{22}$$

where $\underline{Y}_{T-T^{l-1}}$ represents the component of $\underline{Y}$ from the unrecovered signal $\underline{X}_{T-T^{l-1}}$. Equation (a) is valid because $\underline{Y}_{T \cap T^{l-1}}$ can be obtained through linear operations of $\{\mathbf{\Phi}_{T^{l-1},j}\}$, and (b) holds as a result of the definition

$$\underline{X}_P = -\underline{Y}_{T-T^{l-1}} \times_1 \mathbf{\Phi}_{T^{l-1},1}^\dagger \times_2 \mathbf{\Phi}_{T^{l-1},2}^\dagger \cdots \times_d \mathbf{\Phi}_{T^{l-1},d}^\dagger. \tag{23}$$

Combining Lemmas 2 and 3, we obtain

$$\begin{aligned}
\|\underline{X}_{T-\widetilde{T}}\|_F &\leq \frac{2d\delta_{3k}(1+\delta_{2k})^{\frac{d-1}{2}}}{(1-\delta_k)^{\frac{d+1}{2}}}\left[1 + \frac{(2^d-1)^{\frac{1}{2}}\delta_{2k}}{(1-\delta_k)^d}\right] \\
&\quad \times \|\underline{X}_{T-T^{l-1}}\|_F, \\
&\leq \frac{2d\delta_{3k}(1+\delta_{3k})^{\frac{d-1}{2}}\left[(1-\delta_{3k})^d + (2^d-1)^{\frac{1}{2}}\delta_{3k}\right]}{(1-\delta_{3k})^{\frac{3d+1}{2}}} \\
&\quad \times \|\underline{X}_{T-T^{l-1}}\|_F.
\end{aligned} \tag{24}$$

This proves Theorem 1.

The previous theorems relate to the changes in the unrecovered signal during the mixture process. In the following, we focus on examining the refinement process.

Then, we obtain the following theorem describing the relationship between the unrecovered signal after the mixture process and that after the refinement process in the
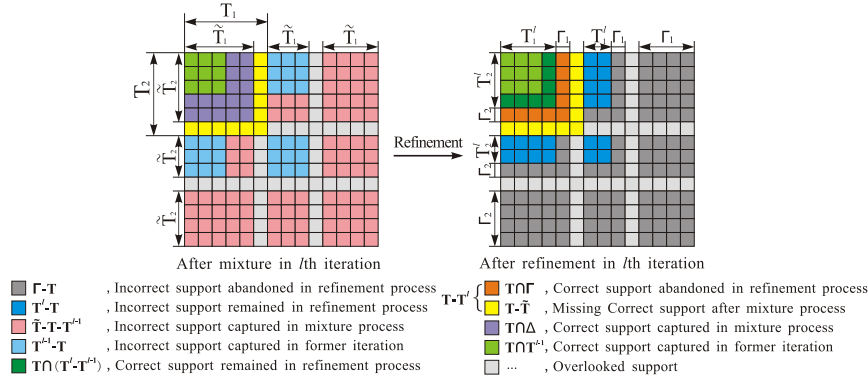
FIGURE 3. Graphical representation of notation in refinement process, with $d = 2$, $k_1 = k_2 = 6$.

l−th iteration. A graphical representation of the notation of the refinement process can be found in Fig. 3.

Theorem 2: Let $\underline{X}_{T-T^l}$ represent the unrecovered signal after the refinement process in the l−th iteration. The following inequality then holds:

$$\|\underline{X}_{T-T^l}\|_F \leq \Omega_2 \|\underline{X}_{T-\widetilde{T}}\|_F, \tag{25}$$

$$\Omega_2 = 1 + \frac{(2^d - 1)^{\frac{1}{2}}(d+1)\delta_{3k}}{(1 - \delta_{3k})^d}. \tag{26}$$

Proof: Initially, it is obvious that

$$\|\underline{X}_{T-T^l}\|_F \leq \|\underline{X}_{T \cap \Gamma}\|_F + \|\underline{X}_{T-\widetilde{T}}\|_F, \tag{27}$$

where Γ represents the global support abandoned during the refinement process. For an arbitrary dimension j, the abandoned support is denoted by Γ_j. Further detail is provided in Fig. 3. Hence, for the relationship between $\|\underline{X}_{T-T^l}\|_F$ and $\|\underline{X}_{T-\widetilde{T}}\|_F$, we are only required to study the relationship between $\|\underline{X}_{T \cap \Gamma}\|_F$ and $\|\underline{X}_{T-\widetilde{T}}\|_F$, which is given by the following two lemmas.

Lemma 4: By the definition of $\underline{E} = \underline{X}_{C,\widetilde{T}} - \underline{X}_{\widetilde{T}}$, we have

$$\|\underline{X}_{T \cap \Gamma}\|_F \leq (d+1)\|\underline{E}\|_F, \tag{28}$$

where $X_{C,\widetilde{T}}$ denotes the LS coefficient tensor with respect to $\underline{Y}$ and $\{\Phi_{\widetilde{T},j}\}$, and $j = 1, 2 \cdots d$.

Lemma 5: It holds that

$$\|\underline{E}\|_F \leq \frac{(2^d - 1)^{\frac{1}{2}}\delta_{3k}}{(1 - \delta_{2k})^d}\|\underline{X}_{T-\widetilde{T}}\|_F. \tag{29}$$

The proofs for Lemmas 4 and 5 are presented in the Appendix (parts V-D and V-E, respectively). By substituting (28) and (29) into (27), we have

$$\|\underline{X}_{T-T^l}\|_F \leq \left[1 + \frac{(2^d - 1)^{\frac{1}{2}}(d+1)\delta_{3k}}{(1 - \delta_{3k})^d}\right]\|\underline{X}_{T-\widetilde{T}}\|_F, \tag{30}$$

which completes the proof. □

Combining Theorems 1 and 2, we can obtain the relationship between the unrecovered signal after the $(l-1)$−iteration and that after the l−iteration.

Theorem 3: The following inequality is valid.

$$\|\underline{X}_{T-T^l}\|_F \leq \Omega\|\underline{X}_{T-T^{l-1}}\|_F = \Omega_1\Omega_2\|\underline{X}_{T-T^{l-1}}\|_F, \tag{31}$$

where Ω_1 and Ω_2 are given by (19) and (26), respectively.

In addition, a sufficient condition for exact reconstruction of a multi-dimensional block-sparse signal from finite iterations with the termination condition $\|\underline{Y}_{R^l}\|_F < \epsilon$ is $\Omega < 1$.

Remark: The condition $\Omega < 1$ implies that the unrecovered signal is always reduced after an iteration. Thus, after a finite number of iterations, there exists $\|\underline{X}_{T-T^l}\|_F \to 0$, demonstrating that all correct supports are determined. If and only if this situation occurs, we have $\|\underline{X}_{R^l}\|_F \to 0$ and $\|\underline{Y}_{R^l}\|_F \to 0$. In other words, once the termination condition $\|\underline{Y}_{R^l}\|_F < \epsilon$ is reached, the original signal is reconstructed successfully.
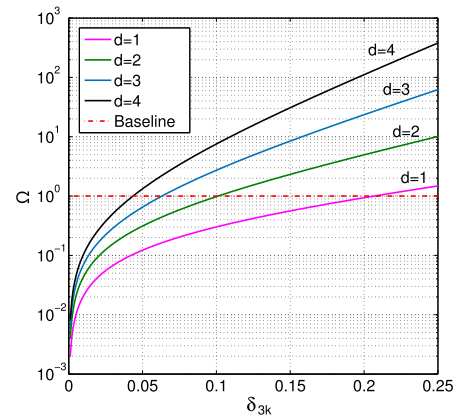

FIGURE 4. The scale factor Ω versus δ_{3k}.

The scale factor Ω is a function of d and δ_{3k}. A numerical analysis of Ω versus δ_{3k} with $d = 1 - 4$ is displayed in Fig. 4, where the area below the baseline $\Omega = 1$ corresponds to the sufficient condition for convergence. We can note that the critical point becomes smaller as the number of dimensions increases, indicating a stricter condition. The critical values of δ_{3k} for $\Omega < 1$ are summarized in Table 1. Below, we consider the sufficient condition with another termination condition, i.e., $\|\underline{Y}_{R^l}\|_F \geq \|\underline{Y}_{R^{l-1}}\|_F$.

TABLE 1. Critical values of δ_{3k} for $\Omega < 1$ and $\Omega_C < 1$.

d	1	2	3	4	$\cdots$
δ_{3k} for Ω	0.206	0.101	0.063	0.043	$\cdots$
δ_{3k} for Ω_C	0.163	0.081	0.051	0.036	$\cdots$

Theorem 4: A sufficient condition for exact reconstruction of a multi-dimensional block-sparse signal from finite iterations with the termination condition $\|\underline{Y}_{R^l}\|_F \geq \|\underline{Y}_{R^{l-1}}\|_F$ is

$$\Omega_C = \frac{\Omega_3 \Omega}{\Omega_4} < 1 \quad (\Omega_4 > 0), \tag{32}$$

where Ω is given by (31) and

$$\Omega_3 = (1 + \delta_{3k})^{\frac{d}{2}},$$
$$\Omega_4 = (1 - \delta_{3k})^{\frac{d}{2}} - \frac{\delta_{3k}(2^d - 1)^{\frac{1}{2}}(1 + \delta_{3k})^{\frac{d}{2}}}{(1 - \delta_{3k})^d}. \tag{33}$$

The proof is given in the Appendix (part V-F). The factor Ω_C is also a function of d and δ_{3k}, and its expansion may be complicated. One could visually comprehend this factor by observing Fig. 5, where the relationships between Ω_C and δ_{3k} with $d = 1 - 4$ are displayed. The critical values of δ_{3k} for $\Omega_C < 1$ are summarized in Table 1.
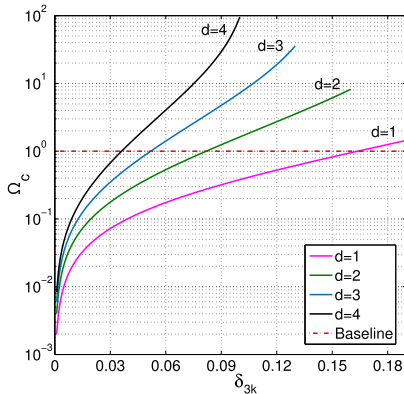
FIGURE 5. The scale factor Ω_C versus δ_{3k}.

Remark: When recovering a multi-dimensional block-sparse signal, $\Omega_C < 1$ means that there always exists $\|\underline{Y}_{R^l}\|_F < \|\underline{Y}_{R^{l-1}}\|_F$ after each iteration, unless $\|\underline{Y}_{R^{l-1}}\|_F = 0$. In other words, the termination condition is only reached if a signal is reconstructed successfully.

C. COMPLEXITY

The complexity of a greedy algorithm for CS reconstruction primarily arises from two aspects, the complexity during a single iteration and the required number of iterations. In this subsection, we present a detailed analysis of the computational complexity of the ArMOMP based on these two aspects.

First, we begin with the complexity during a single iteration. For convenience and without loss of generality, we set
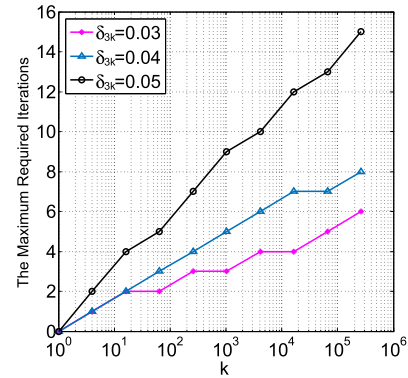
FIGURE 6. Maximum required iterations versus k for zero-one multi-dimensional block-sparse signals $d = 3$, $k_1 = k_2 = k_3 = k$.

$n_j = n$, $m_j = m$, and $k_j = k$; $n > m > k$ always holds. Obviously, by considering the ordinary condition for the major computational process only, the computational complexity for a single ArMOMP iteration is $O(dnm^d)$, which occurs at the computation for inner product tensors $\{\underline{P}_j^l\}$. In the following, we explore the required iterations for exact reconstruction of a multi-dimensional block-sparse signal using the ArMOMP algorithm.

Lemma 6: When the sufficient condition for convergence is satisfied, after the initialization stage, it holds that

$$\|\underline{X}_{T-T^0}\|_F \leq \Omega_i \|\underline{X}\|_F, \tag{34}$$
$$\Omega_i = min(\Omega_0, 1), \tag{35}$$
$$\Omega_0 = \frac{d\left[(1+\delta_{3k})^{d+1} - [(1-\delta_{3k})^{\frac{d+1}{2}} - \delta_{3k}(1+\delta_{3k})^{\frac{d-1}{2}}]^2\right]^{\frac{1}{2}}}{(1+\delta_{3k})^{\frac{d+1}{2}}}. \tag{36}$$

The proof is provided in the Appendix (part V-G). Then, we define β as follows:

$$\beta = min_{i,j}\left[\|X_{(j)}(i_j, :)\|_F / \|\underline{X}\|_F\right],$$
$$X_{(j)} = unfold_j(\underline{X}), \tag{37}$$

with $j = 1, 2 \cdots, d$ and $i_j = 1, 2 \cdots n_j$. Note that $\beta\|\underline{X}\|_F$ represents the $F-$norm lower-bound of an arbitrary band of $\underline{X}$. Hence, once the unrecovered signal satisfies

$$\|\underline{X}_{T-T^l}\|_F \overset{T3,L6}{=} \Omega_i\Omega^l\|\underline{X}\|_F \leq \beta\|\underline{X}\|_F, \tag{38}$$

all correct supports are determined so that the signal is recovered successfully. Hence, we obtain the following theorem regarding the maximum required iterations.

Theorem 5: When the sufficient condition for convergence is satisfied, one requires l_{max} iterations at most to reconstruct a multi-dimensional block-sparse signal exactly, and

$$l_{max} = ceil\left[\frac{log(\beta/\Omega_i)}{log\Omega}\right]. \tag{39}$$

It is apparent that l_{max} is a function of δ_{3k} and β, while β depends on the original signal only. Obviously, β will be reduced as the number of entries increases; thus, more

iterations will be required for reconstruction. Here, we give an example in Fig. 6 to intuitively show the relationship between l_{max} and k using the zero-one signals. One can note that only a slight increase in the required iterations occurs for increased k, and high-performance sensing matrices, corresponding to smaller δ_{3k}, can reduce l_{max}.

D. NUMERICAL SIMULATIONS

We perform numerical simulations to compare the ArMOMP algorithm with the well-known NBOMP multiway greedy algorithm [35], and its extension, MaTOMP [37]. All three algorithms are used for TCS reconstruction by exploiting the Tucker structure of the measurement and sparsity operators, i.e., they work with the same model. The simulation setup is listed as follows.

- For simplicity and without loss of generality, we set the number of dimensions $d = 3$, with $n_j = n = 256$, $m_j = m = 128$, $k_1 = k$, $j \in \{1, 2, 3\}$; therefore, the total sampling ratio $r = (m/n)^3$. We are interested in the scenario $d = 3$, as many multi-dimensional applications of CS belong to this case, such as videos and hyperspectral images [29], [38].
- Given the parameters n and k, a (k, k, k)-block-sparse signal with Gaussian random supports is generated by following these steps:
  (1) First, a zero-tensor $\underline{X} = \underline{0} \in \mathbb{R}^{n \times n \times n}$ is initialized.
  (2) Second, three supports containing k indexes each are randomly selected for each dimension, denoted T_1, T_2, T_3.
  (3) Third, the entries of sub-tensor $\underline{X}(T_1, T_2, T_3)$ are set to 1 or a Gaussian-distributed number, referred to as a "zero-one signal" and "Gaussian random signal," respectively, in this paper.
- Three Gaussian random sensing matrices are generated with size $\Phi_j \in \mathbb{R}^{m \times n}, j \in \{1, 2, 3\}$.
- We regard the reconstruction as successful when $\|\hat{\underline{X}} - \underline{X}\|_F / \|\underline{X}\|_F < 0.001$. The simulation is repeated 100 times for each set of parameters and the average success ratios are calculated for comparison.
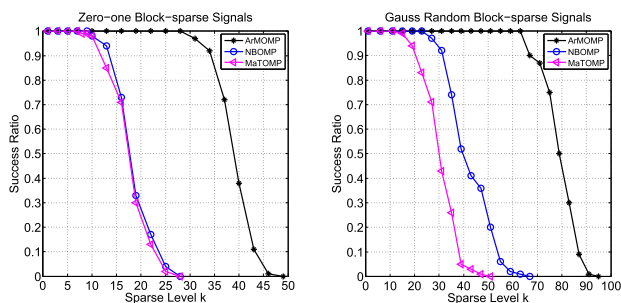

FIGURE 7. Comparison of reconstruction success ratios for zero-one and Gaussian random signals, and for ArMOMP, NBOMP, and MaTOMP algorithms.

The simulation results for the reconstruction accuracy are shown in Fig. 7. Regardless of the Gaussian random

or zero-one signals, the critical sparsity levels of the ArMOMP algorithm obviously exceed those of the NBOMP and MaTOMP algorithms. In the following, we discuss the improvement in reconstruction accuracy provided by the ArMOMP. Considering that a d-dimensional $(k, k, \cdots, k)$-block-sparse zero-one signal with regard to the $\{T_1, T_2, \cdots, T_d\}$ supports is reconstructed by NBOMP with a set of sensing matrices $\{\Phi_j\}$, the selected supports are correct in the first iteration only, if the following inequality holds:

$$\max_{\{\forall i_j \in T_j\}} |\underline{Y} \times_1 \phi_{i_1}^* \cdots \times_d \phi_{i_d}^*| > \max_{\{\exists i_j \notin T_j\}} |\underline{Y} \times_1 \phi_{i_1}^* \cdots \times_d \phi_{i_d}^*|, \tag{40}$$

where ϕ_{i_j} denotes the $i_j - th$ column of Φ_j and T_j represents the correct support of Φ_j. Without loss of generality, we assume that $\forall i_j, l_j \in T_j, i_j \neq l_j, \langle \phi_{i_j}, \phi_{l_j} \rangle = -\mu$. Then, we have

$$\max_{\{\forall i_j \in T_j\}} |\underline{Y} \times_1 \phi_{i_1}^* \cdots \times_d \phi_{i_d}^*| = |1 - (k-1)\mu|^d, \tag{41}$$

where μ denotes the coherence of the $\{\Phi_j\}$ sensing matrices. Hence, the left side of (40) tends to zero as k and d increase, demonstrating that inequality (40) is easy to break. In this scenario, the incorrect supports will be selected, yielding reconstruction failure. Further, the above discussion focuses on the first iteration, and the condition for correct support selection becomes harsher during subsequent iterations. For NBOMP and MaTOMP, incorrect choices are irreversible. On the other hand, the ArMOMP algorithm allows the existence of incorrect support choices, as they will certainly be removed via the refinement process if the sufficient condition for exact reconstruction is satisfied. Furthermore, even if this condition is not completely satisfied, the incorrect supports can also be refined with high probability as a result of the simulations. Hence, we believe that the sufficient condition for exact reconstruction can be further relaxed; this merits future study.

The enhancement in the reconstruction precision does not introduce a higher computational complexity. On the contrary, the ArMOMP algorithm yields improved reconstruction speed with the benefit that less computation is required. In fact, the computational complexities of ArMOMP during a single iteration are close to those of NBOMP and MaTOMP. However, the required numbers of iterations for NBOMP and MaTOMP are approximately linear with k, being significantly higher than those for ArMOMP, which is approximately linear with the logarithm of k. The iteration numbers and time consumption of ArMOMP rise considerably more slowly than those of NBOMP and MaTOMP as the signal size increases, as shown in Figs. 8 and 9. Further, it is interesting to note that the computational consumption for ArMOMP has a downward trend when k is sufficiently large. This is because the reconstruction success ratio begins to drop and, therefore, the termination condition is frequently reached ahead of time.

IV. ROBUSTNESS TO NOISE

In this section, we focus on the CS reconstruction using the ArMOMP algorithm under the noisy scenario. First, we
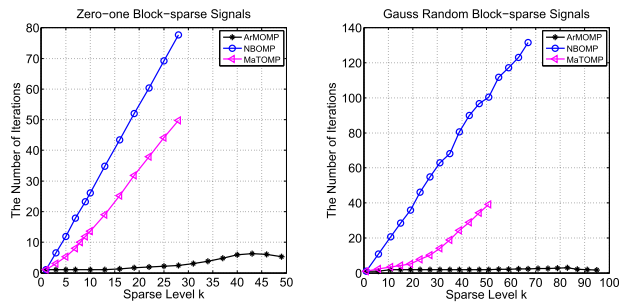
FIGURE 8. Comparison of required number of iterations for zero-one and Gaussian random signals, and for ArMOMP, NBOMP, and MaTOMP algorithms.
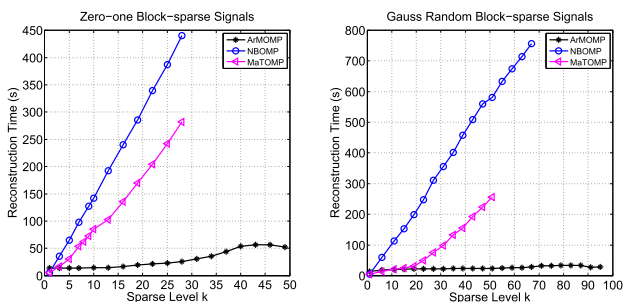

FIGURE 9. Comparison of reconstruction time for zero-one and Gaussian random signals, and for ArMOMP, NBOMP, and MaTOMP algorithms.

discuss several noisy scenarios and give a unified model for these cases. Then, we provide theorems related to the reconstruction precision and the corresponding numerical simulations.

A. NOISY MODEL
We primarily consider the following three scenarios, which include the most common scenarios in CS reconstruction.

1) APPROXIMATE MULTI-DIMENSIONAL BLOCK-SPARSE SIGNAL
It is known that some multi-dimensional signals such as videos and hyperspectral images are not naturally multi-dimensional block-sparse. For CS reconstruction, these signals must be represented as approximate multi-dimensional block-sparse signals by sparsifying bases [30], [35]. Let $\underline{X_A} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_d}$ be an approximate signal that can be split into two parts as follows

$$\underline{X}(i_1, i_2 \cdots i_d) = \begin{cases} \underline{X_A}(i_1, i_2 \cdots i_d), & i_j \in T_j, \\ 0, & otherwise, \end{cases} \quad (42)$$

$$\underline{Z}(i_1, i_2 \cdots i_d) = \begin{cases} 0, & i_j \in T_j, \\ \underline{X_A}(i_1, i_2 \cdots i_d), & otherwise, \end{cases} \quad (43)$$

where $\underline{X} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_d}$ is a $(k_1, k_2, \cdots, k_d)$-block-sparse tensor with supports $\{T_1, T_2, \cdots, T_d\}$ and $\underline{Z} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_d}$ denotes the approximate component. Therefore, one can

express the compressive sampling model as

$$\underline{Y} = \underline{X_A} \times_1 \mathbf{\Phi}_1 \cdots \times_d \mathbf{\Phi}_d = (\underline{X} + \underline{Z}) \times_1 \mathbf{\Phi}_1 \cdots \times_d \mathbf{\Phi}_d,$$
$$= \underline{X} \times_1 \mathbf{\Phi}_1 \cdots \times_d \mathbf{\Phi}_d + \underline{N} = \underline{Y_0} + \underline{N}. \quad (44)$$

The tensor $\underline{N}$ can be regarded as the noise introduced by the sampling process in terms of $\underline{Z}$ and the $\{\mathbf{\Phi}_j\}$ sensing matrices. The magnitude of $\underline{N}$ is upper-bounded by $\|\underline{N}\|_F \leq \sigma_1 \sigma_2 \cdots \sigma_d \|\underline{Z}\|_F$, where σ_j denotes the largest singular value of $\mathbf{\Phi}_j$. Hence, the magnitude of the resulting noise depends on the sparsity level and the sensing matrix performance. Considering that $\|\underline{Z}\|_F$ is generally several orders of magnitude lower than $\|\underline{X}\|_F$, $\|\underline{N}\|_F$ is always considerably smaller than $\|\underline{Y_0}\|_F$.

2) NOISED SIGNAL
Let $\underline{X} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_d}$ be a $(k_1, k_2, \cdots, k_d)$-block-sparse signal with supports $\{T_1, T_2, \cdots, T_d\}$ and $\underline{X_N} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_d}$ be the additive noise tensor. The sampling model is expressed as

$$\underline{Y} = (\underline{X} + \underline{X_N}) \times_1 \mathbf{\Phi}_1 \cdots \times_d \mathbf{\Phi}_d,$$
$$= \underline{X} \times_1 \mathbf{\Phi}_1 \cdots \times_d \mathbf{\Phi}_d + \underline{N} = \underline{Y_0} + \underline{N}. \quad (45)$$

3) NOISED MEASUREMENT
This scenario considers the noise from a compressive sampling process. This noise is primarily generated by the sampling mechanisms and signal transmission systems. The model can be expressed as

$$\underline{Y} = \underline{X} \times_1 \mathbf{\Phi}_1 \times_2 \mathbf{\Phi}_2 \cdots \times_d \mathbf{\Phi}_d + \underline{N} = \underline{Y_0} + \underline{N}. \quad (46)$$

The equations (44), (45), and (46) imply that all three scenarios mentioned above can be unified in (46). For convenience and unity, in the following, we present theories based on (46) only, where $\underline{N}$ denotes the noise introduced by all three scenarios.

B. THEORIES FOR NOISY SCENARIO
Theorem 6: In the noisy scenario, the following inequality holds

$$\|\underline{X_{T-T^l}}\|_F \leq \Omega \|\underline{X_{T-T^{l-1}}}\|_F + \Omega_N \|\underline{N}\|_F,$$
$$= \Omega \|\underline{X_{T-T^{l-1}}}\|_F + (\Omega_2 \Omega_{N1} + \Omega_{N2}) \|\underline{N}\|_F, \quad (47)$$
$$\Omega_{N1} = \frac{2d(1+\delta_{3k})^{\frac{1}{2}}}{(1-\delta_{3k})^{\frac{d+1}{2}}}, \quad \Omega_{N2} = \frac{(d+1)(1+\delta_{3k})^{\frac{d}{2}}}{(1-\delta_{3k})^d}, \quad (48)$$

where Ω_2 and Ω are given by Theorems 2 and 3, respectively. This theorem implies the upper bound of the unrecovered signal with regard to the noise after the l−th iteration. The proof is provided in the Appendix (part V-H). Quite evidently, the reconstruction error caused by the noise cannot be eliminated within a finite number of iterations. To evaluate the reconstruction precision, we employ the following theorem.

Theorem 7: When the termination condition $\|\underline{Y}_{R^l}\|_F \geq \|\underline{Y}_{R^{l-1}}\|_F$ is reached, there exists

$$\|\underline{X} - \hat{\underline{X}}\|_F \leq \left[\frac{\Omega_5(\Omega_3\Omega_N + 2)}{\Omega_4 - \Omega_3\Omega} + \Omega_{N3} \right] \|\underline{N}\|_F, \quad (49)$$

when $\Omega_4 - \Omega_3\Omega > 0$. The notation $\hat{\underline{X}}$ represents the final reconstruction, Ω_3, Ω_4, Ω_N, and Ω are given in (33), (47), and (31), respectively, and

$$\Omega_5 = 1 + \frac{\delta_{3k}(2^d - 1)^{\frac{1}{2}}}{(1 - \delta_{3k})^d}, \quad \Omega_{N3} = \frac{(1 + \delta_{3k})^{\frac{d}{2}}}{(1 - \delta_{3k})^d}. \quad (50)$$

The proof is given in the Appendix (part V-I).

Remark: The relationships for the theoretical upper-bound factor versus δ_{3k} are given for different d in Fig. 10. As the noise levels are generally several orders of magnitude lower than the original signal, we can always obtain reconstructions with high accuracy.

FIGURE 11. Simulations for approximate multi-dimensional block-sparse signal, relative reconstruction error versus sampling ratio.

FIGURE 10. Theoretical upper-bound factor versus δ_{3k} for reconstruction in noisy scenario.

FIGURE 12. Simulations for noised signal, relative reconstruction error versus relative signal error.

C. NUMERICAL SIMULATIONS FOR NOISY SCENARIOS

Besides the theoretical analysis, we also simulate reconstructions for noised scenarios. For convenience, we continue to set $d = 3$, $n_1 = n_2 = n_3 = n = 256$, $m_1 = m_2 = m_3 = m$ and $k_1 = k_2 = k_3 = k$; therefore, the total sampling ratio $r = m^3/n^3$. Gaussian random sensing matrices are used for all simulations. The Gaussian noise is added to the measurements obtained from zero-one multi-dimensional block-sparse signals.

1) APPROXIMATE MULTI-DIMENSIONAL BLOCK-SPARSE SIGNAL

The approximate multi-dimensional block-sparse signal is also referred to as the multi-dimensional power-law decaying signal, defined as

$$\underline{X}_A(i_1, i_2 \cdots i_d) = ci_m^{-p},$$
$$i_m = max(i_j), \quad j = 1, 2 \cdots d, \quad (51)$$

where the constants $c > 0$ and $p > 1$. One can understand this signal as a tensor for which most of the energy is concentrated
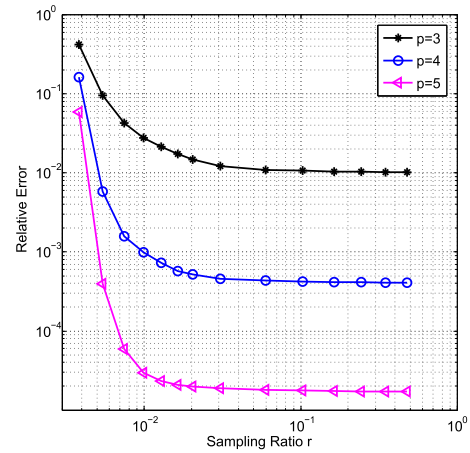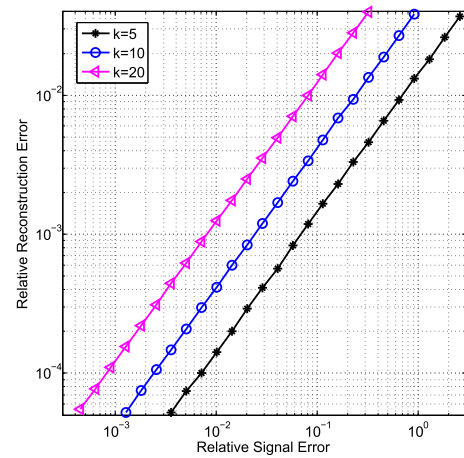
on a sub-tensor (block) while the entries decay away from the block. A larger p can help to concentrate the energy and improve the sparsity level. We select the tensor-based power-law decaying signal for the simulations because this is closest to reality. Most multi-dimensional signals can be represented in the form of a power-law decaying based on sparsifying bases. The lengths of the supports for each dimension are set to $k = 20$, meaning that a $20 \times 20 \times 20$-sized block is recovered as the reconstruction, while the other entries are regarded as noise. For each set of parameters p and m, we repeat the simulations 100 times and calculate the average relative reconstruction error $\|\hat{\underline{X}} - \underline{X}_A\|_F / \|\underline{X}_A\|_F$. Note that the constant c represents a scaling factor, which has no effect on the results. Hence, we set $c = 1$ for simplicity. The results are displayed in Fig. 11. We can see that the reconstruction error reduces rapidly and becomes stable at a very low sampling ratio (approximately $r = 0.03$). In addition, the stabilized reconstruction error is primarily determined by the sparsity level.
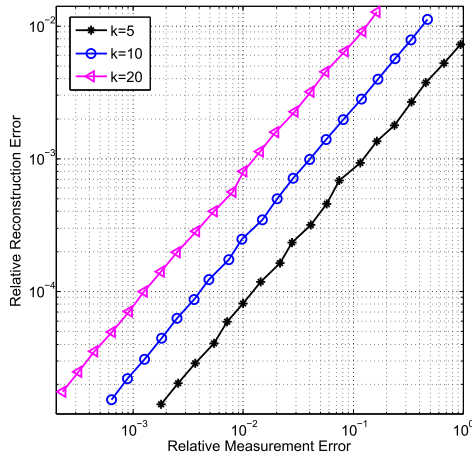
FIGURE 13. Simulations for noised measurements, relative reconstruction error versus relative signal error.

2) NOISED SIGNAL

We randomly generate zero-one signals and then add Gaussian noise at different levels. The parameter m is set to 128. The simulation is repeated 100 times and the average values of the relative reconstruction error $\|\hat{\underline{X}} - \underline{X}\|_F / \|\underline{X}\|_F$ and relative signal error $\|\underline{X}_N - \underline{X}\|_F / \|\underline{X}\|_F$ are taken into consideration. The results are plotted in Fig. 12, where we can see that the robustness of the ArMOMP algorithm is satisfactory.

3) NOISED MEASUREMENTS

In the next set of simulations, the Gaussian noise is added to the measurements obtained from zero-one multi-dimensional block-sparse signals. We set $m = 128$ and also repeat each simulation 100 times to obtain $\|\hat{\underline{X}} - \underline{X}\|_F / \|\underline{X}\|_F$ and the relative measurement error $\|\underline{N}\|_F / \|\underline{Y}\|_F$. The results are presented in Fig. 13. As mentioned above, the reconstruction error is considerably lower than its theoretical upper bound.

V. CONCLUSION

In this paper, we have proposed a novel multiway greedy algorithm named ArMOMP for TCS reconstruction. We have determined the sufficient conditions for convergence and the maximum required number of iterations for exact reconstruction in the absence of noise. We have also provided the upper-bound for the F-norm of reconstruction error under the noise condition. Simulation results demonstrate that the ArMOMP algorithm has advantages over existing multiway greedy algorithms as regards reconstruction speed and precision. We believe that the proposed algorithm can be used not only for TCS, but also in other fields that require computation of Tucker-model-based sparse representation like hyperspectral image processing and tensor-based dictionary learning.

APPENDIX
A. PROOF OF LEMMA 1

Here, we prove Inequality (10) only, because (11) and (12) can be easily obtained by employing this proof and the consequences of the RIP. Because of the definition of block

sparsity, the equation $\|\underline{X} \times_1 \mathbf{\Phi}_1 \cdots \times_d \mathbf{\Phi}_d\|_F = \|\underline{X_A} \times_1 \mathbf{\Phi}_{A,1} \cdots \times_d \mathbf{\Phi}_{A,d}\|_F$ obviously holds, as $\underline{X}(i_1, i_2 \cdots, i_d) = 0 \ \forall \underline{X}(i_1, i_2 \cdots, i_d) \notin \underline{X_A}$. Without losing generality, we decompose the TTM operation for the 1D case first

$$\|\underline{X_A} \times_1 \mathbf{\Phi}_{A,1}\|_F^2$$
$$= \sum_{i_2=1}^{k_2} \cdots \sum_{i_d=1}^{k_d} \|\mathbf{\Phi}_{A,1}(\underline{X_A}(:, i_2, \cdots, i_d))\|_F^2,$$
$$\leq \sum_{i_2=1}^{k_2} \cdots \sum_{i_d=1}^{k_d} (1 + \delta_{k_1}) \|\underline{X_A}(:, i_2, \cdots, i_d)\|_F^2,$$
$$= (1 + \delta_{k_1}) \|\underline{X_A}\|_F^2. \tag{52}$$

Using the same strategy, we have

$$\|\underline{X_A} \times_1 \mathbf{\Phi}_{A,1} \times_2 \mathbf{\Phi}_{A,2}\|_F^2 \leq (1 + \delta_{k_1})(1 + \delta_{k_2}) \|\underline{X_A}\|_F^2. \tag{53}$$

Therefore,

$$\|\underline{X_A} \times_1 \mathbf{\Phi}_{A,1} \cdots \times_d \mathbf{\Phi}_{A,d}\|_F \leq \prod_{j=1}^{d} (1 + \delta_{k_j})^{\frac{1}{2}} \|\underline{X_A}\|_F. \tag{54}$$

Similarly, we can obtain the lower bound

$$\prod_{j=1}^{d} (1 - \delta_{k_j})^{\frac{1}{2}} \|\underline{X_A}\|_F \leq \|\underline{X_A} \times_1 \mathbf{\Phi}_{A,1} \cdots \times_d \mathbf{\Phi}_{A,d}\|_F, \tag{55}$$

which completes the proof.

B. PROOF OF LEMMA 2

This lemma claims the relationship between $\|X_{R^{l-1}}\|_F$ and $\|X_{T-T^{l-1}}\|_F$. According to the composition of $\|\overline{X_{R^{l-1}}}\|_F$, it is obviously

$$\|X_{R^{l-1}}\|_F \leq \|X_P\|_F + \|X_{T-T^{l-1}}\|_F. \tag{56}$$

Therefore, we need to derive the relationship between $\|X_P\|_F$ and $\|X_{T-T^{l-1}}\|_F$. Note that $\underline{Y_{T-T^{l-1}}}$ is generated from $X_{T-T^{l-1}}$, which is irregularly shaped. Thus, we split this term as follows.

$$\underline{Y_{T-T^{l-1}}} = \sum_{\xi \subseteq \{1,2 \cdots d\}} \|X_{T-T^{l-1}}^{\xi} \cdots \times_j \mathbf{\Phi}_{T-T^{l-1},j} \cdots$$
$$\times_h \mathbf{\Phi}_{T \cap T^{l-1},h} \cdots\|_F, \tag{57}$$

where $X_{T-T^{l-1}}^{\xi}$ represents a sub-tensor having $j-$dimension supported by $T_j - T_j^{l-1}$, while $h-$dimension is supported by $T_h \cap T_h^{l-1}, \forall j \in \xi, h \notin \xi$. We provide a graphical representation in Fig. 14, taking $d = 2$ as an example. Substituting (57) into (23), we obtain

$$\|\underline{X_P}\|_F \leq \sum_{\xi \subseteq \{1,2 \cdots d\}} \|X_{T-T^{l-1}}^{\xi} \cdots \times_j \mathbf{\Phi}_{T^{l-1},j}^{\dagger} \mathbf{\Phi}_{T-T^{l-1},j} \cdots$$
$$\times_h \mathbf{\Phi}_{T^{l-1},h}^{\dagger} \mathbf{\Phi}_{T \cap T^{l-1},h} \cdots\|_F, \tag{58}$$

where $j \in \xi$, and $h \notin \xi$. Then, we expand $X_{T-T^{l-1}}^{\xi}$ to $\tilde{X}_{T-T^{l-1}}^{\xi}$, which is also represented in Fig. 14. Compared with $X_{T-T^{l-1}}^{\xi}$,

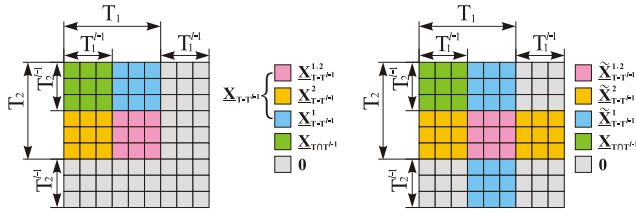**FIGURE 14.** Example of splitting and expanding of unrecovered signal $\underline{X}_{T-T^{l-1}}$, with $d = 2$, $k_1 = k_2 = 6$.

FIGURE 15. Graphical representation of notations used in proof of Lemma 3, for $d = 2$, $k_1 = k_2 = 6$.

the supports of the $\xi-$dimension remain the same, while the others are expanded to T_h^{l-1}, $h \notin \xi$. Considering the fact that the extensions contain zero-entries only, and $\mathbf{\Phi}_{T^{l-1},j}^{\dagger} = (\mathbf{\Phi}_{T^{l-1},j}^{*}\mathbf{\Phi}_{T^{l-1},j})^{-1}\mathbf{\Phi}_{T^{l-1},j}^{*}$, we rewrite Inequality (58) as

$$\|\underline{X}_P\|_F \leq \sum_{\xi\subseteq\{1,2\cdots d\}} \|\tilde{\underline{X}}_{T-T^{l-1}}^{\xi} \cdots \times_j (\mathbf{\Phi}_{T^{l-1},j}^{*}\mathbf{\Phi}_{T^{l-1},j})^{-1}$$
$$\times \mathbf{\Phi}_{T^{l-1},j}^{*}\mathbf{\Phi}_{T-T^{l-1},j}\cdots \times_h \mathbf{I}_h\cdots\|_F, \quad (59)$$

where $\mathbf{I}_h$ denotes the unit matrix, $j \in \xi$, and $h \notin \xi$. Then, using Lemma 1, and the fact $\|\underline{X}_{T-T^{l-1}}^{\xi}\|_F = \|\tilde{\underline{X}}_{T-T^{l-1}}^{\xi}\|_F$, we have

$$\|\underline{X}_P\|_F \leq \frac{\delta_{2k}}{(1-\delta_k)^d} \mathbf{\Sigma}_{\xi\subseteq\{1,2\cdots d\}} \|\underline{X}_{T-T^{l-1}}^{\xi}\|_F,$$
$$\overset{@}{\leq} \frac{\delta_{2k}}{(1-\delta_k)^d} (C_d^1 + C_d^2 \cdots + C_d^d)^{\frac{1}{2}}$$
$$\times (\mathbf{\Sigma}_{\xi\subseteq\{1,2\cdots d\}} \|\underline{X}_{T-T^{l-1}}^{\xi}\|_F^2)^{\frac{1}{2}},$$
$$= \frac{\delta_{2k}(2^d-1)^{\frac{1}{2}}}{(1-\delta_k)^d} \|\underline{X}_{T-T^{l-1}}\|_F, \quad (60)$$

where (*a*) follows from the average inequality and C_d^j denotes a composite number. By substituting (60) into (56), we can obtain

$$\|\underline{X}_{R^{l-1}}\|_F \leq \left[1 + \frac{\delta_{2k}(2^d-1)^{\frac{1}{2}}}{(1-\delta_k)^d}\right]\|\underline{X}_{T-T^{l-1}}\|_F, \quad (61)$$

which completes the proof.

C. PROOF OF LEMMA 3

According to the algorithm flow, we have

$$\|\underline{Y}_{R^{l-1}} \times_j \mathbf{\Phi}_{\Delta,j}^{*}\|_F \geq \|\underline{Y}_{R^{l-1}} \times_j \mathbf{\Phi}_{T-T^{l-1},j}^{*}\|_F. \quad (62)$$

Without loss of generality, we begin with $j = 1$, and remove the same parts from both sides of (62), so that

$$\|\underline{Y}_{R^{l-1}} \times_1 \mathbf{\Phi}_{\Delta-T,1}^{*}\|_F \geq \|\underline{Y}_{R^{l-1}} \times_1 \mathbf{\Phi}_{T-\tilde{T},1}^{*}\|_F. \quad (63)$$

Now, we extend $\underline{X}_{R^{l-1}}$ to obtain a complete and regular tensor $\underline{X}_{R^{l-1}}^{ext}$, which is presented in Fig. 15. Note that the extension contains zero-entries only; thus,

$$\|\underline{Y}_{R^{l-1}} \times_1 \mathbf{\Phi}_{\Delta-T,1}^{*}\|_F = \|\underline{X}_{R^{l-1}}^{ext} \times_1 \mathbf{\Phi}_{\Delta-T,1}^{*}\mathbf{\Phi}_{T\cup T^{l-1},1}$$
$$\times_2 \mathbf{\Phi}_{T\cup T^{l-1},2}\cdots\times_d \mathbf{\Phi}_{T\cup T^{l-1},d}\|_F. \quad (64)$$

Then, we enlarge the left side of (63) by employing Lemma 1, such that

$$\|\underline{Y}_{R^{l-1}} \times_1 \mathbf{\Phi}_{\Delta-T,1}\|_F \overset{L1}{\leq} \delta_{3k_1}\left[(1+\delta_{2k_2})\cdots(1+\delta_{2k_d})\right]^{\frac{1}{2}}$$
$$\times \|\underline{X}_{R^{l-1}}\|_F. \quad (65)$$

On the other hand, a split of the right side of (63) is given by

$$\|\underline{Y}_{R^{l-1}} \times_1 \mathbf{\Phi}_{T-\tilde{T},1}^{*}\|_F$$
$$= \|\underline{X}_{R^{l-1},T-\tilde{T}}^{1} \times_1 \mathbf{\Phi}_{T-\tilde{T},1}^{*}\mathbf{\Phi}_{T-\tilde{T},1}$$
$$\times_2 \mathbf{\Phi}_{T,2}\cdots\times_d \mathbf{\Phi}_{T,d} + \underline{X}_{R^{l-1},(T\cup T^{l-1})-(T-\tilde{T})}^{1}$$
$$\times_1 \mathbf{\Phi}_{T-\tilde{T},1}^{*}\mathbf{\Phi}_{(T\cup T^{l-1})-(T-\tilde{T}),1}$$
$$\times_2 \mathbf{\Phi}_{T\cup T^{l-1},2}\cdots\times_d \mathbf{\Phi}_{T\cup T^{l-1},d}\|_F, \quad (66)$$

where $\underline{X}_{R^{l-1},T-\tilde{T}}^{1}$ represents a sub-tensor, the $1-$dimension of which is supported by $T_1 - \tilde{T}_1$, while the other dimensions are supported by $T_h, h \neq 1$. $\underline{X}_{R^{l-1},(T\cup T^{l-1})-(T-\tilde{T})}^{1}$ is obtained by extending $\underline{X}_{R^{l-1}}$ without $\underline{X}_{R^{l-1},T-\tilde{T}}^{1}$ to a complete and regular sub-tensor. Likewise, the extension is composed of zero-entries only. Then, the lower-bound of (66) can be expressed as

$$\|\underline{Y}_{R^{l-1}} \times_1 \mathbf{\Phi}_{T-\tilde{T},1}^{*}\|_F$$
$$\geq \|\underline{X}_{R^{l-1},T-\tilde{T}}^{1} \times_1 \mathbf{\Phi}_{T-\tilde{T},1}^{*}\mathbf{\Phi}_{T-\tilde{T},1}$$
$$\times_2 \mathbf{\Phi}_{T,2}\cdots\times_d \mathbf{\Phi}_{T,d}\|_F - \|\underline{X}_{R^{l-1},(T\cup T^{l-1})-(T-\tilde{T})}^{1}$$
$$\times_1 \mathbf{\Phi}_{T-\tilde{T},1}^{*}\mathbf{\Phi}_{(T\cup T^{l-1})-(T-\tilde{T}),1}$$
$$\times_2 \mathbf{\Phi}_{T\cup T^{l-1},2}\cdots\times_d \mathbf{\Phi}_{T\cup T^{l-1},d}\|_F. \quad (67)$$

By employing Lemma 1 and $\|\underline{X}_{R^{l-1},(T\cup T^{l-1})-(T-\tilde{T})}^{1}\|_F \leq \|\underline{X}_{R^{l-1}}^{1}\|_F$, we can further simplify (67) to

$$\|\underline{Y}_{R^{l-1}} \times_1 \mathbf{\Phi}_{T-\tilde{T},1}^{*}\|_F \geq (1-\delta_{k_1})\left[(1-\delta_{k_2})\cdots(1-\delta_{k_d})\right]^{\frac{1}{2}}$$
$$\|\underline{X}_{R^{l-1},T-\tilde{T}}^{1}\|_F - \delta_{2k_1}\left[(1+\delta_{2k_2})\cdots(1+\delta_{2k_d})\right]^{\frac{1}{2}} \|\underline{X}_{R^{l-1}}\|_F. \quad (68)$$

Combining (63), (65), (68), and considering that the relationship is valid for each $j \in \{1, 2, \cdots, d\}$, we obtain

$$\|\underline{X}_{R^{l-1},T-\tilde{T}}^{j}\|_F \leq \frac{2\delta_{3k}(1+\delta_{2k})^{\frac{d-1}{2}}}{(1-\delta_k)^{\frac{d+1}{2}}} \|\underline{X}_{R^{l-1}}\|_F. \quad (69)$$
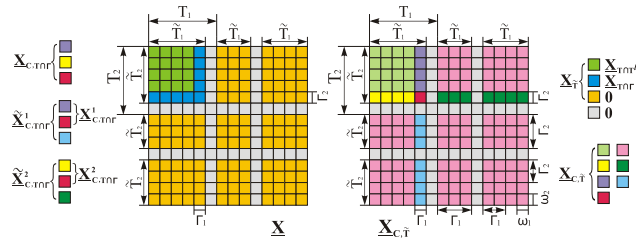
FIGURE 16. Graphical representation of notations used in proof of Lemma 4, with $d = 2$, $k_1 = k_2 = 6$.

FIGURE 17. Graphical representation of notations used in proof of Lemma 5, with $d = 2$, $k_1 = k_2 = 6$.

Now, we return to $\underline{X}_{T-\widetilde{T}}$ with reference to Figs. 1 and 15. We have

$$
\begin{aligned}
\|\underline{X}_{T-\widetilde{T}}\|_F &\leq \sum_{j=1}^{d} \|\underline{X}^j_{R^{l-1}, T-\widetilde{T}}\|_F \\
&\leq \frac{2d\delta_{3k}(1+\delta_{2k})^{\frac{d-1}{2}}}{(1-\delta_k)^{\frac{d+1}{2}}} \|\underline{X}_{R^{l-1}}\|_F,
\end{aligned}
\tag{70}
$$

which completes the proof.

D. PROOF OF LEMMA 4

From the definition of $\underline{E}$, one can directly obtain

$$
\|\underline{X}_{T\cap\Gamma}\|_F \leq \|\underline{X}_{C,T\cap\Gamma}\|_F + \|\underline{E}\|_F,
\tag{71}
$$

by which the problem is translated into proving $\|\underline{X}_{C,T\cap\Gamma}\|_F \leq d\|\underline{E}\|_F$. For an arbitrary dimension j, it holds that

$$
\|\underline{X}^j_{C,T\cap\Gamma}\|_F \leq \|\underline{\widetilde{X}}^j_{C,T\cap\Gamma}\|_F,
\tag{72}
$$

where $\underline{X}^j_{C,T\cap\Gamma}$ denotes a sub-tensor of $\underline{X}_{C,\widetilde{T}}$, the j-dimension of which is supported by Γ_j while the others are supported by $T_j \cap \widetilde{T}_j$. Another sub-tensor $\underline{\widetilde{X}}_{C,T\cap\Gamma}$ can be regarded as the extension of $\underline{X}^j_{C,\widetilde{T}}$, supported by Γ in the $j-$dimension and $\widetilde{T}$ in other dimensions. We also provide a graphical illustration of this notation in Fig. 16. Obviously, there always exists a set of supports $\{\omega_j\}$ for each dimension that satisfies

$$
\|\underline{\widetilde{X}}^j_{C,T\cap\Gamma}\|_F \leq \|\underline{\widetilde{X}}^j_{C,\omega}\|_F,
\tag{73}
$$

with $\omega_j \subseteq \widetilde{T}_j$, $\omega_j \cap T_j = \emptyset$, and $|\omega_j| = |T_j \cap \Gamma_j|$. Similarly, the notation $\underline{\widetilde{X}}^j_{C,\omega}$ represents a sub-tensor of $\underline{X}_{C,\widetilde{T}}$, the $j-$dimension of which is supported by ω_j, while the other dimensions are supported by $\widetilde{T}_h$, $h \neq j$. For example, $\underline{\widetilde{X}}^1_{C,\omega} = \underline{\widetilde{X}}_{C,\widetilde{T}}(\omega_1, :, \cdots, :)$. Therefore, we have

$$
\begin{aligned}
\|\underline{\widetilde{X}}^j_{C,\omega}\|_F &= \|\underline{X}^j_\omega + \underline{E}^j_\omega\|_F \leq \|\underline{X}^j_\omega\|_F + \|\underline{E}^j_\omega\|_F, \\
&\overset{(a)}{=} 0 + \|\underline{E}^j_\omega\|_F \leq \|\underline{E}\|_F,
\end{aligned}
\tag{74}
$$

where (a) holds because $\omega_j \cap T_j = \emptyset$, and $\underline{X}^j_\omega$ and $\underline{E}^j_\omega$ share the same support with $\underline{\widetilde{X}}^j_{C,\omega}$. Then, $\underline{X}_{C,T\cap\Gamma}$ is split with reference

to Fig. 16, and we obtain

$$
\begin{aligned}
\|\underline{X}_{C,T\cap\Gamma}\|_F & \\
&\leq \sum_{j=1}^{d} \|\underline{X}^j_{C,T\cap\Gamma}\|_F \overset{(72)}{\leq} \sum_{j=1}^{d} \|\underline{\widetilde{X}}^j_{C,T\cap\Gamma}\|_F \\
&\overset{(73)}{\leq} \sum_{j=1}^{d} \|\underline{\widetilde{X}}^j_{C,\omega}\|_F \overset{(74)}{\leq} \sum_{j=1}^{d} \|\underline{E}\|_F = d\|\underline{E}\|_F.
\end{aligned}
\tag{75}
$$

By substituting (75) into (71), we find

$$
\|\underline{X}_{T\cap\Gamma}\|_F \leq (d+1)\|\underline{E}\|_F,
\tag{76}
$$

which completes this proof.

E. PROOF OF LEMMA 5

We begin this proof with the definition of $\underline{X}_{C,\widetilde{T}}$:

$$
\underline{X}_{C,\widetilde{T}} = \underline{X}_T \times_1 \Phi^\dagger_{\widetilde{T},1} \Phi_{T,1} \times_2 \Phi^\dagger_{\widetilde{T},2} \Phi_{T,2} \cdots \times_d \Phi^\dagger_{\widetilde{T},d} \Phi_{T,d}.
\tag{77}
$$

By splitting $\underline{X}_T$, we obtain

$$
\begin{aligned}
\underline{X}_{C,\widetilde{T}} &= \underline{X}_{T\cap\widetilde{T}} \times_1 \Phi^\dagger_{\widetilde{T},1} \Phi_{T\cap\widetilde{T},1} \cdots \times_d \Phi^\dagger_{\widetilde{T},d} \Phi_{T\cap\widetilde{T},d} \\
&+ \sum_{\xi\subseteq\{1,2\cdots d\}} \underline{X}^\xi_{T\cap\widetilde{T}} \cdots \times_j \Phi^\dagger_{\widetilde{T},j} \Phi_{T-\widetilde{T},j} \cdots \\
&\times_h \Phi^\dagger_{\widetilde{T},h} \Phi_{T\cap\widetilde{T},h} \cdots,
\end{aligned}
\tag{78}
$$

where $\underline{X}^\xi_{T\cap\widetilde{T}}$ denotes a sub-tensor of $\underline{X}$, the $\xi-$dimensions of which are supported by $T_j-\widetilde{T}_j$ while the other dimensions are supported by $T_h \cap \widetilde{T}_h$, $\forall j \in \xi$, $h \notin \xi$. Then, we extend $\underline{X}^\xi_{T\cap\widetilde{T}}$ to $\underline{X}^\xi_{\widetilde{T}}$. The supports for the $\xi-$dimensions remain $T_j - \widetilde{T}_j$, while those for the other dimensions are expanded to $\widetilde{T}_h$, $j \in \xi$, $h \notin \xi$. It should be noted that the expansion only contains zero-entries. It holds that

$$
\underline{X}_{C,\widetilde{T}} = \underline{X}_{\widetilde{T}} + \sum_{\xi\subseteq\{1,2\cdots d\}} \underline{X}^\xi_{\widetilde{T}} \cdots \times_j \Phi^\dagger_{\widetilde{T},j} \Phi_{T-\widetilde{T},j} \cdots \times_h \underline{I}_h \cdots,
\tag{79}
$$

where $\underline{I}_h$ denotes the unit matrix, $j \in \xi$, and $h \notin \xi$. Next, we substitute (79) into the definition $\underline{E} = \underline{X}_{C,\widetilde{T}} - \underline{X}_{\widetilde{T}}$

to obtain

$$\|\underline{E}\|_F = \| \sum_{\xi \subseteq \{1,2\cdots d\}} X_{\widetilde{T}}^{\xi} \cdots \times_j \Phi_{\widetilde{T},j}^{\dagger} \Phi_{T-\widetilde{T},j} \cdots \times_h I_h \cdots \|_F. \quad (80)$$

Then, by employing Lemma 1, we obtain

$$\|\underline{E}\|_F \leq \frac{\delta_{3k}}{(1-\delta_{2k})^d} \sum_{\xi \subseteq \{1,2\cdots d\}} \|X_{\widetilde{T}}^{\xi}\|_F,$$

$$= \frac{\delta_{3k}}{(1-\delta_{2k})^d} \sum_{\xi \subseteq \{1,2\cdots d\}} \|X_{T \cap \widetilde{T}}^{\xi}\|_F$$

$$\overset{(a)}{\leq} \frac{(2^d-1)^{\frac{1}{2}}\delta_{3k}}{(1-\delta_{2k})^d} \|X_{T-\widetilde{T}}\|_F, \quad (81)$$

where (a) follows from the average inequality, which is similar to (60). Thus, we have completed the proof of Lemma 5.

F. PROOF OF THEOREM 4
By reference to (22), one can write Y_{R^l} as

$$\|Y_R^l\|_F = \|res(Y_{T-T^l}, \{\Phi_{T^l,j}\})\|_F \leq \|Y_{T-T^l}\|_F, \quad (82)$$

where Y_{T-T^l} denotes a component of $\underline{Y}$ generated by X_{T-T^l}. The graphical illustration is similar to that presented in Fig. 14. Now, we define a new tensor $X_{T-T^l}^{ext} \in \mathbb{R}^{k_1 \times k_2 \cdots \times k_d}$ as

$$X_{T-T^l}^{ext}(i_1, i_2 \cdots i_d) = \begin{cases} 0, & i_j \in T_j \cap T_j^l, \\ X_T(i_1, i_2 \cdots i_d), & otherwise. \end{cases} \quad (83)$$

It is clear that $\|X_{T-T^l}^{ext}\|_F = \|X_{T-T^l}\|_F$. Thus, we have

$$\|Y_R^l\|_F$$
$$\leq \|Y_{T-T^l}\|_F$$
$$= \|X_{T-T^l}^{ext} \times_1 \Phi_{T,1} \times_2 \Phi_{T,2} \cdots \times_d \Phi_{T,d}\|_F,$$
$$\overset{L1}{\leq} (1+\delta_{3k})^{\frac{d}{2}} \|X_{T-T^l}\|_F \overset{(31)}{\leq} (1+\delta_{3k})^{\frac{d}{2}} \Omega \|X_{T-T^{l-1}}\|_F,$$
$$= \Omega_3 \Omega \|X_{T-T^{l-1}}\|_F. \quad (84)$$

On the other hand, the lower bound for $\|Y_R^{l-1}\|_F$ is expressed as

$$\|Y_R^{l-1}\|_F$$
$$\overset{(a)}{\geq} \|X_{T-T^{l-1}}^{ext} \times_1 \Phi_{T,1} \times_2 \Phi_{T,2} \cdots \times_d \Phi_{T,d}\|_F$$
$$\quad - \|X_{P,T^{l-1}} \times_1 \Phi_{T^{l-1},1} \times_2 \Phi_{T^{l-1},2} \cdots \times_d \Phi_{T^{l-1},d}\|_F,$$
$$\overset{(60),L1}{\geq} \left[(1-\delta_{3k})^{\frac{d}{2}} - \frac{\delta_{3k}(2^d-1)^{\frac{1}{2}}(1+\delta_{3k})^{\frac{d}{2}}}{(1-\delta_{3k})^d} \right] \|X_{T-T^{l-1}}\|_F,$$
$$= \Omega_4 \|X_{T-T^{l-1}}\|_F \quad (85)$$

where (a) holds based on the split of Y_R^{l-1}, shown in Fig. 15. The definition of $X_{T-T^{l-1}}^{ext}$ is identical to $X_{T-T^l}^{ext}$ and $\|X_{T-T^{l-1}}^{ext}\|_F = \|X_{T-T^{l-1}}\|_F$. Combining (84) and (85), we obtain

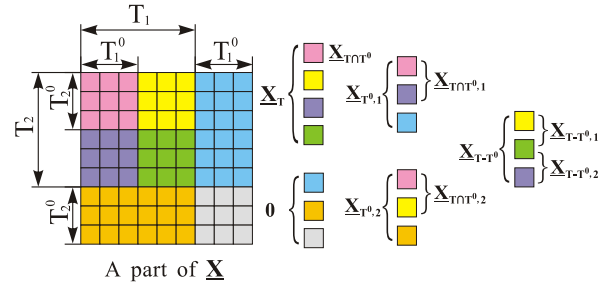$$\|Y_R^l\|_F \leq \frac{\Omega_3 \Omega}{\Omega_4} \|Y_R^{l-1}\|_F = \Omega_C \|Y_R^{l-1}\|_F. \quad (86)$$

FIGURE 18. Graphical representation of notations used in proof of Lemma 6, with $d = 2$, $k_1 = k_2 = 6$.

Obviously, there always exists $\|Y_R^l\|_F < \|Y_R^{l-1}\|_F$ when $\Omega_C < 1$. Hence, this proof is complete.

G. PROOF OF LEMMA 6
Obviously, it always holds that

$$\|X_{T-T^0}\|_F \leq \|X_T\|_F = \|\underline{X}\|_F. \quad (87)$$

Then, we derive another upper-bound for $\|X_{T-T^0}\|_F$. Without loss of generality, we start the proof considering the 1−dimension in the initialization stage:

$$\|\underline{Y} \times_1 \Phi_{T^0,1}^*\|_F \geq \|\underline{Y} \times_1 \Phi_{T,1}^*\|_F \overset{L1}{\geq} (1-\delta_{3k})^{\frac{d+1}{2}} \|\underline{X}\|_F. \quad (88)$$

On the other hand, an upper bound can be obtained:

$$\|\underline{Y} \times_1 \Phi_{T^0,1}^*\|_F$$
$$\leq \|X_{T-T^0,1} \times_1 \Phi_{T^0,1}^* \Phi_{T-T^0,1} \cdots \times_d \Phi_{T,d}\|_F$$
$$\quad + \|X_{T \cap T^0,1} \times_1 \Phi_{T^0,1}^* \Phi_{T \cap T^0,1} \cdots \times_d \Phi_{T,d}\|_F,$$
$$\overset{(a)}{=} \|X_{T-T^0,1} \times_1 \Phi_{T^0,1}^* \Phi_{T-T^0,1} \cdots \times_d \Phi_{T,d}\|_F$$
$$\quad + \|X_{T^0,1} \times_1 \Phi_{T^0,1}^* \Phi_{T^0,1} \cdots \times_d \Phi_{T,d}\|_F, \quad (89)$$

where the notation is graphically represented in Fig. 18. Note that (a) holds as the extension from $X_{T \cap T^0,1}$ to $X_{T^0,1}$ is composed of zero-entries only.

By employing Lemma 1 for (89), we have

$$\|\underline{Y} \times_1 \Phi_{T^0,1}^*\|_F$$
$$\leq \delta_{3k}(1+\delta_{3k})^{\frac{d-1}{2}} \|\underline{X}\|_F + (1+\delta_{3k})^{\frac{d+1}{2}} \|X_{T \cap T^0,1}\|_F. \quad (90)$$

Combining (88) and (90), we obtain

$$(1-\delta_{3k})^{\frac{d+1}{2}} \|\underline{X}\|_F$$
$$\leq \delta_{3k}(1+\delta_{3k})^{\frac{d-1}{2}} \|\underline{X}\|_F + (1+\delta_{3k})^{\frac{d+1}{2}} \|X_{T \cap T^0,1}\|_F. \quad (91)$$

Elementary calculations show that $(1-\delta_{3k})^{\frac{d+1}{2}} > \delta_{3k}(1+\delta_{3k})^{\frac{d-1}{2}}$ when the sufficient condition for convergence is satisfied; therefore,

$$\|X_{T \cap T^0,1}\|_F \geq \frac{(1-\delta_{3k})^{\frac{d+1}{2}} - \delta_{3k}(1+\delta_{3k})^{\frac{d-1}{2}}}{(1+\delta_{3k})^{\frac{d+1}{2}}} \|\underline{X}\|_F. \quad (92)$$

Note that the relationship holds for each dimension j, so that

$$\|\underline{X}_{T-T^0,j}\|_F$$
$$= (\|\underline{X}\|_F^2 - \|\underline{X}_{T \cap T^0,j}\|_F^2)^{\frac{1}{2}}$$
$$\leq \frac{\left[(1+\delta_{3k})^{d+1} - [(1-\delta_{3k})^{\frac{d+1}{2}} - \delta_{3k}(1+\delta_{3k})^{\frac{d-v1}{2}}]^2\right]^{\frac{1}{2}}}{(1+\delta_{3k})^{\frac{d+1}{2}}} \|\underline{X}\|_F. \tag{93}$$

With reference to Fig. 18, we have

$$\|\underline{X}_{T-T^0}\|_F$$
$$\leq \sum_{j=1}^{d} \|\underline{X}_{T-T^0,j}\|_F$$
$$\leq \frac{d\left[(1+\delta_{3k})^{d+1} - [(1-\delta_{3k})^{\frac{d+1}{2}} - \delta_{3k}(1+\delta_{3k})^{\frac{d-1}{2}}]^2\right]^{\frac{1}{2}}}{(1+\delta_{3k})^{\frac{d+1}{2}}} \|\underline{X}\|_F,$$
$$= \Omega_0 \|\underline{X}\|_F. \tag{94}$$

By combining (87) and (94), we finally complete this proof.

H. PROOF OF THEOREM 6
First, we derive the relationship between $\|\underline{X}_{T-\tilde{T}}\|_F$ and $\|\underline{X}_{T-T^{l-1}}\|_F$. During the mixture process, $\forall j \in \{1, 2 \cdots d\}$, there exists

$$\|\underline{Y}_{R^{l-1}} \times_j \Phi^*_{\Delta-T,j}\|_F \geq \|\underline{Y}_{R^{l-1}} \times_j \Phi^*_{T-\tilde{T},j}\|_F. \tag{95}$$

As $\underline{Y}_{R^{l-1}}$ can be expressed as

$$\underline{Y}_{R^{l-1}} = res(\underline{Y}_0, \{\Phi_{T^{l-1},j}\}) + res(\underline{N}, \{\Phi_{T^{l-1},j}\}),$$
$$= \underline{Y}_{0R^{l-1}} + \underline{N}_{R^{l-1}}, \tag{96}$$

the left side of (95) can be enlarged to

$$\|\underline{Y}_{R^{l-1}} \times_j \Phi^*_{\Delta-T,j}\|_F$$
$$\leq \|\underline{Y}_{0R^{l-1}} \times_j \Phi^*_{\Delta-T,j}\|_F + \|\underline{N}_{R^{l-1}} \times_j \Phi^*_{\Delta-T,j}\|_F,$$
$$\overset{(a)}{\leq} \|\underline{Y}_{0R^{l-1}} \times_j \Phi^*_{\Delta-T,j}\|_F + (1+\delta_{k_j})^{\frac{1}{2}} \|\underline{N}_{R^{l-1}}\|_F,$$
$$\overset{(65)}{\leq} \delta_{3k}(1+\delta_{3k})^{\frac{d-1}{2}} \|\underline{X}_{R^{l-1}}\|_F + (1+\delta_{3k})^{\frac{1}{2}} \|\underline{N}\|_F, \tag{97}$$

where $\underline{N}_{R^{l-1}}$ denotes the residual of $\underline{N}$, (a) holds because of Lemma 1 and the fact that $\forall \alpha \in \mathbb{R}^{m_j}$, $\|\Phi^*_{\Delta-T,j}\alpha\|_2 \leq \|\Phi^*_{\Delta-T,j}\|_2\|\alpha\|_2 = \|\Phi_{\Delta-T,j}\|_2\|\alpha\|_2 = (1+\delta_{k_j})^{\frac{1}{2}}\|\alpha\|_2$. The concepts behind the notation used here can be found in Fig. 15. On the other hand, the lower bound for the right side of (95) is given by

$$\|\underline{Y}_{R^{l-1}} \times_j \Phi^*_{T-\tilde{T},j}\|_F$$
$$\geq \|\underline{Y}_{0R^{l-1}} \times_j \Phi^*_{T-\tilde{T},j}\|_F - \|\underline{N}_{R^{l-1}} \times_j \Phi^*_{T-\tilde{T},j}\|_F,$$
$$\overset{(68)}{\geq} (1-\delta_{3k})^{\frac{d+1}{2}} \|\underline{X}^j_{R^{l-1},T-\tilde{T}}\|_F - \delta_{3k}(1+\delta_{3k})^{\frac{d-1}{2}} \|\underline{X}_{R^{l-1}}\|_F$$
$$- (1+\delta_{3k})^{\frac{1}{2}} \|\underline{N}\|_F. \tag{98}$$

By combining (95), (97) and (98), we have

$$\|\underline{X}_{T-\tilde{T}}\|_F$$
$$\leq \sum_{j=1}^{d} \|\underline{X}^j_{R^{l-1},T-\tilde{T}}\|_F,$$
$$\leq \frac{2d\delta_{3k}(1+\delta_{2k})^{\frac{d-1}{2}}}{(1-\delta_k)^{\frac{d+1}{2}}} \|\underline{X}_{R^{l-1}}\|_F + \frac{2d(1+\delta_{3k})^{\frac{1}{2}}}{(1-\delta_{3k})^{\frac{d+1}{2}}} \|\underline{N}\|_F,$$
$$\overset{L2,T1}{\leq} \Omega_1 \|\underline{X}_{T-T^{l-1}}\|_F + \frac{2d(1+\delta_{3k})^{\frac{1}{2}}}{(1-\delta_{3k})^{\frac{d+1}{2}}} \|\underline{N}\|_F,$$
$$= \Omega_1 \|\underline{X}_{T-T^{l-1}}\|_F + \Omega_{N1} \|\underline{N}\|_F. \tag{99}$$

Next, we derive the relationship between $\|\underline{X}_{T-\tilde{T}}\|_F$ and $\|\underline{X}_{T-T^l}\|_F$. Obviously, it holds that

$$\|\underline{X}_{T-T^l}\|_F \leq \|\underline{X}_{T \cap \Gamma^l}\|_F + \|\underline{X}_{T-\tilde{T}}\|_F,$$
$$\overset{T5}{\leq} (d+1)\|\underline{E}\|_F + \|\underline{X}_{T-\tilde{T}}\|_F, \tag{100}$$

where the notation is graphically represented in Figs. 16 and 17. From the definition of $\underline{E}$, we have

$$\|\underline{E}\|_F = \|\underline{X}_{C,\tilde{T}} - \underline{X}_{\tilde{T}}\|_F$$
$$= \|\underline{Y} \times_1 \Phi^\dagger_{\tilde{T},1} \cdots \times_d \Phi^\dagger_{\tilde{T},d} - \underline{X}_{\tilde{T}}\|_F,$$
$$\leq \|\underline{Y}_0 \times_1 \Phi^\dagger_{\tilde{T},1} \cdots \times_d \Phi^\dagger_{\tilde{T},d} - \underline{X}_{\tilde{T}}\|_F$$
$$+ \|\underline{N} \times_1 \Phi^\circ_{\tilde{T},1} \cdots \times_d \Phi^\dagger_{\tilde{T},d}\|_F. \tag{101}$$

Note that, here, $\underline{Y}_0$ is equivalent to the $\underline{Y}$ in Eq. (77), and $\Phi^\dagger_{\tilde{T},j} = (\Phi^*_{\tilde{T},j} \Phi_{\tilde{T},j})^{-1} \Phi^*_{\tilde{T},j}$. Hence, we can rewrite (101) as

$$\|\underline{E}\|_F \overset{L5,L1}{\leq} \frac{(2^d-1)^{\frac{1}{2}}\delta_{3k}}{(1-\delta_{3k})^d} \|\underline{X}_{T-\tilde{T}}\|_F + \frac{(1+\delta_{3k})^{\frac{d}{2}}}{(1-\delta_{3k})^d} \|\underline{N}\|_F. \tag{102}$$

Combining (102), (100), and (99), we obtain

$$\|\underline{X}_{T-T^l}\|_F \leq \Omega_2\Omega_1 \|\underline{X}_{T-T^{l-1}}\|_F$$
$$+ (\Omega_2\Omega_{N1} + \Omega_{N2})\|\underline{N}\|_F,$$
$$\overset{T7}{=} \Omega \|\underline{X}_{T-T^{l-1}}\|_F + \Omega_N \|\underline{N}\|_F, \tag{103}$$

which completes this proof.

I. PROOF OF THEOREM 7
One can give an upper bound for $\|\underline{Y}_{R^l}\|_F$:

$$\|\underline{Y}_{R^l}\|_F \leq \|res(\underline{Y}_0, \{\Phi_{T^l,j}\})\|_F + \|res(\underline{N}, \{\Phi_{T^l,j}\})\|_F. \tag{104}$$

Note that, here, $\underline{Y}_0$ is equivalent to the $\underline{Y}$ in Eq. (82). Thus, we have

$$\|\underline{Y}_{R^l}\|_F \overset{(84)}{\leq} (1+\delta_{3k})^{\frac{d}{2}} \|\underline{X}_{T-T^l}\|_F + \|\underline{N}\|_F,$$
$$= \Omega_3 \|\underline{X}_{T-T^l}\|_F + \|\underline{N}\|_F. \tag{105}$$

The lower bound for $\|\underline{Y}_{R^{l-1}}\|_F$ can be given as

$$
\begin{aligned}
\|\underline{Y}_{R^{l-1}}\|_F &\geq \|res(\underline{Y}_0, \{\boldsymbol{\Phi}_{T^{l-1},j}\})\|_F \\
&\quad - \|res(\underline{N}, \{\boldsymbol{\Phi}_{T^{l-1},j}\})\|_F, \\
&\stackrel{(85)}{\geq} \Omega_4 \|\underline{X}_{T-T^{l-1}}\|_F - \|\underline{N}\|_F. \quad (106)
\end{aligned}
$$

It is clear that, once the termination condition is reached, there must exist:

$$
\Omega_4 \|\underline{X}_{T-T^{l-1}}\|_F - \|\underline{N}\|_F \leq \Omega_3 \|\underline{X}_{T-T^l}\|_F + \|\underline{N}\|_F. \quad (107)
$$

We substitute (103) into (107) and note that support T^{l-1} is selected as the final support for the reconstruction $\hat{T}$. We obtain

$$
\|\underline{X}_{T-\hat{T}}\|_F \leq \frac{\Omega_3\Omega_N + 2}{\Omega_4 - \Omega_3\Omega} \|\underline{N}\|_F, \quad (108)
$$

when $\Omega_4 - \Omega_3\Omega > 0$. Thus, we have claimed an upper bound for $\|\underline{X}_{T-\hat{T}}\|_F$ with respect to the noise level $\|\underline{N}\|_F$. In the following, we derive the relationship between the reconstruction error $\|\underline{X} - \underline{\hat{X}}\|_F$ and $\|\underline{X}_{T-\hat{T}}\|_F$. With reference to the strategy used in the proof of Lemma 2,

$$
\begin{aligned}
&\|\underline{X} - \underline{\hat{X}}\|_F \\
&\leq \|\underline{X}_{T-\hat{T}}\|_F + \|\underline{X}_{\hat{T}} - \underline{X}_{C,\hat{T}}\|_F, \\
&\leq \|\underline{X}_{T-\hat{T}}\|_F + \|\underline{X}_{\hat{T}} - \underline{Y}_0 \times_1 \boldsymbol{\Phi}^{\dagger}_{\hat{T},1} \times_2 \boldsymbol{\Phi}^{\dagger}_{\hat{T},2} \cdots \times_d \boldsymbol{\Phi}^{\dagger}_{\hat{T},d}\|_F \\
&\quad + \|\underline{N} \times_1 \boldsymbol{\Phi}^{\dagger}_{\hat{T},1} \times_2 \boldsymbol{\Phi}^{\dagger}_{\hat{T},2} \cdots \times_d \boldsymbol{\Phi}^{\dagger}_{\hat{T},d}\|_F, \\
&\stackrel{(60),L1}{\leq} \left[1 + \frac{\delta_{3k}(2^d - 1)^{\frac{1}{2}}}{(1-\delta_{3k})^d}\right] \|\underline{X}_{T-T^{l-1}}\|_F + \frac{(1+\delta_{3k})^{\frac{d}{2}}}{(1-\delta_{3k})^d} \|\underline{N}\|_F, \\
&= \Omega_5 \|\underline{X}_{T-T^{l-1}}\|_F + \Omega_{N3} \|\underline{N}\|_F. \quad (109)
\end{aligned}
$$

Substituting (108) into (109), we have

$$
\|\underline{X} - \underline{\hat{X}}\|_F \leq \left[\frac{\Omega_5(\Omega_3\Omega_N + 2)}{\Omega_4 - \Omega_3\Omega} + \Omega_{N3}\right] \|\underline{N}\|_F, \quad (110)
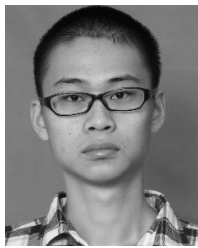$$

when $\Omega_4 - \Omega_3\Omega > 0$. Hence, this proof is completed.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] M. Rani, S. Dhok, and R. Deshmukh, "A systematic review of compressive sensing: Concepts, implementations and applications," *IEEE Access*, vol. 6, pp. 4875–4894, 2018.

[3] Q. Shen, W. Liu, W. Cui, and S. Wu, "Underdetermined DOA estimation under the compressive sensing framework: A review," *IEEE Access*, vol. 4, pp. 8865–8878, 2016.

[4] D. Craven, B. McGinley, L. Kilmartin, M. Glavin, and E. Jones, "Compressed sensing for bioelectric signals: A review," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 2, pp. 529–540, Mar. 2015.

[5] Y. Zhang, L. Y. Zhang, J. Zhou, L. Liu, F. Chen, and X. He, "A review of compressive sensing in information security field," *IEEE Access*, vol. 4, pp. 2507–2519, 2016.

[6] A. Massa, P. Rocca, and G. Oliveri, "Compressive sensing in electromagnetics—A review," *IEEE Antennas Propag. Mag.*, vol. 57, no. 1, pp. 224–238, Feb. 2015.

[7] E. J. Candès, "Compressive sampling," in *Proc. Int. Congr. Math.*, Madrid, Spain, vol. 3, 2006, pp. 1433–1452.

[8] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[9] 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.

[10] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[11] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.

[12] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.

[13] D. Needell and J. A. Trop, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comp. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.

[14] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Found. Comput. Math.*, vol. 9, no. 3, pp. 317–334, Jun. 2009.

[15] X. Wang, G. Zhang, F. Wen, D. Ben, and W. Liu, "Angle estimation for bistatic MIMO radar with unknown mutual coupling based on three-way compressive sensing," *J. Syst. Eng. Electron.*, vol. 28, no. 2, pp. 257–266, May 2017.

[16] M. Takahiro *et al.*, "Multi-dimensional wireless tomography using tensor-based compressed sensing," *Wireless Pers. Commun.*, vol. 96, no. 3, pp. 3361–3384, Oct. 2017.

[17] Y. Wang *et al.*, "Compressive sensing of hyperspectral images via joint tensor Tucker decomposition and weighted total variation regularization," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2457–2461, Dec. 2017.

[18] W. Qiu, J. Zhou, H. Zhao, and Q. Fu, "Three-dimensional sparse turntable microwave imaging based on compressive sensing," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 826–830, Apr. 2015.

[19] Y. Li, W. Dai, J. Zou, H. Xiong, and Y. F. Zheng, "Structured sparse representation with union of data-driven linear and multilinear subspaces model for compressive video sampling," *IEEE Trans. Signal Process.*, vol. 65, no. 19, pp. 5062–5077, Oct. 2017.

[20] G. Chen, G. Li, and J. Zhang, "Tensor compressed video sensing reconstruction by combination of fractional-order total variation and sparsifying transform," *Signal Process. Image*, vol. 55, pp. 146–156, Jul. 2017.

[21] O. Michailovich, Y. Rathi, and S. Dolui, "Spatially regularized compressed sensing for high angular resolution diffusion imaging," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1100–1115, May 2011.

[22] X. Fan, Q. Lian, and B. Shi, "Compressed sensing MRI with phase noise disturbance based on adaptive tight frame and total variation," *IEEE Access*, vol. 5, pp. 19311–19321, 2017.

[23] A. Cichocki *et al.*, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.

[24] S. Friedland, Q. Li, and D. Schonfeld, "Compressive sensing of sparse tensors," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4438–4447, Oct. 2014.

[25] R. Boyer and M. Haardt, "Noisy compressive sampling based on block-sparse tensors: Performance limits and beamforming technique," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6075–6088, Dec. 2016.

[26] W. Cao *et al.*, "Total variation regularized tensor RPCA for background subtraction from compressive measurements," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4075–4090, Sep. 2016.

[27] M. F. Duarte *et al.*, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.

[28] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 47, no. 10, pp. B44–B51, Apr. 2008.

[29] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 105–115, Jan. 2014.

[30] M. F. Duarte and R. G. Baraniuk, "Kronecker compressive sensing," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 494–504, Feb. 2012.

[31] Y. August, C. Bachman, Y. Rivenson, and A. Stern, "Compressive hyperspectral imaging by random separable projections in both the patial and the spectral domains," *Appl. Opt.*, vol. 52, no. 10, pp. D46–D54, Apr. 2013.

[32] X. Ding, W. Chen, and I. J. Wassell, "Joint sensing matrix and sparsifying dictionary optimization for tensor compressive sensing," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3632–3646, Jul. 2017.

[33] Y. Guo, X. Song, N. Li, and D. Fang, "An efficient missing data prediction method based on Kronecker compressive sensing in multivariable time series," *IEEE Access*, vol. 6, pp. 57239–57248, 2018.

[34] D. Jiang, L. Nie, Z. Lv, and H. Song, "Spatio-temporal Kronecker compressive sensing for traffic matrix recovery," *IEEE Access*, vol. 4, pp. 3046–3053, 2016.

[35] C. F. Caiafa and A. Cichocki, "Computing sparse representations of multidimensional signals using Kronecker bases," *Neural Comput.*, vol. 25, no. 1, pp. 186–220, 2013.

[36] N. D. Sidiropoulos and A. Kyrillidis, "Multi-way compressed sensing for sparse low-rank tensors," *IEEE Signal Process. Lett.*, vol. 19, no. 11, pp. 757–760, Nov. 2012.

[37] R. Zhao, Q. Wang, Y. Shen, and J. Li, "Multiatom tensor orthogonal matching pursuit algorithm for compressive-sensing-based hyperspectral image reconstruction," *J. Appl. Remote Sens*, vol. 10, p. 045002, Oct. 2016.

[38] R. G. Baraniuk, T. Goldstein, and A. C. Sankaranarayanan, "Compressive video sensing: Algorithms, architectures, and applications," *IEEE Signal Process. Mag.*, vol. 34, no. 1, pp. 52–66, Jan. 2017.

JUN FU received the Ph.D. degree in agricultural engineering from Jilin University, Changchun, China, in 2016, where he is currently an Associate Professor with the Key Laboratory of Bionic Engineering. His current research interests include intelligent agriculture and remote sensing.

LUQUAN REN received the M.S. degree from the Jilin University of Technology, Changchun, China, in 1981. He is currently an Academician of the Chinese Academy of Sciences and a Professor with the Key Laboratory of Bionic Engineering, Jilin University. His current research interests include bionic theory, intelligent agriculture, and remote sensing.

RONGQIANG ZHAO received the B.S. degree in agricultural engineering from Jilin University, Changchun, China, in 2013, and the Ph.D. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2019. He is with the Key Laboratory of Bionic Engineering, Ministry of Education, and the College of Biological and Agricultural Engineering, Jilin University. His current research interests include compressive sensing, sparse representation, and image processing.

QIANG WANG received the Ph.D. degree in control science and engineering from the Harbin Institute of Technology, China, in 2004, where he is currently a Professor with the Department of Control Science and Engineering. His current research interests include data fusion, wireless sensor networks, signal and image processing, and compressive sensing.

• • •