

# Block-Sparse Tensor Recovery

Liyang Lu<sup>1</sup>, Member, IEEE, Zhaocheng Wang<sup>2</sup>, Fellow, IEEE, Zhen Gao<sup>3</sup>, Member, IEEE, Sheng Chen<sup>4</sup>, Life Fellow, IEEE, and H. Vincent Poor<sup>5</sup>, Life Fellow, IEEE

**Abstract**—This work explores the fundamental problem of the recoverability of a sparse tensor being reconstructed from its compressed embodiment. We present a generalized model of block-sparse tensor recovery as a theoretical foundation, where concepts involving a holistic mutual incoherence property (MIP) of the measurement matrix set are defined. A representative algorithm based on the orthogonal matching pursuit (OMP) framework, called tensor generalized block OMP (T-GBOMP), is applied to the theoretical framework for analyzing both noiseless and noisy recovery conditions. Specifically, we present an exact recovery condition (ERC) and sufficient conditions for establishing it with consideration of different degrees of restriction. Reliable reconstruction conditions, in terms of the residual convergence, the estimated error and a signal-to-noise ratio bound, are established to reveal the computable theoretical interpretability based on the newly defined MIP. The flexibility of tensor recovery is highlighted, i.e., the reliable recovery can be guaranteed by optimizing the MIP of the measurement matrix set. Analytical comparisons demonstrate that the theoretical results developed are tighter and less restrictive than existing ones (if any). Further discussions provide tensor extensions for several classic greedy algorithms, indicating that the results derived are universal and applicable to all these tensorized variants.

**Index Terms**—Block sparsity, compressed sensing, tensor signal processing, mutual incoherence property, recovery condition.

Manuscript received 4 February 2024; revised 17 June 2024; accepted 15 August 2024. Date of publication 21 August 2024; date of current version 22 November 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62471275, in part by the Postdoctoral Fellowship Program of China Postdoctoral Science Foundation (CPSF) under Grant GZC20231374, in part by China Postdoctoral Science Foundation under Grant 2024M751680, and in part by Princeton Language and Intelligence. (Corresponding author: Zhaocheng Wang.)

Liyang Lu is with Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: lulyang@mail.tsinghua.edu.cn).

Zhaocheng Wang is with Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China, and also with Tsinghua Shenzhen International Graduate School, Shenzhen 518055, China (e-mail: zcwang@tsinghua.edu.cn).

Zhen Gao is with the Advanced Research Institute of Multidisciplinary Science (ARIMS), Beijing Institute of Technology, Beijing 100081, China (e-mail: gaozhen16@bit.edu.cn).

Sheng Chen is with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ Southampton, U.K. (e-mail: sqc@ecs.soton.ac.uk).

H. Vincent Poor is with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Communicated by M. Soltanolkotabi, Associate Editor for Signal Processing and Source Coding.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TIT.2024.3447050>.

Digital Object Identifier 10.1109/TIT.2024.3447050

## I. INTRODUCTION

RECOVERING an  $n$ -mode sparse tensor  $\mathcal{X} \in \mathbb{C}^{N_1 \times N_2 \times \dots \times N_n}$  from linear measurements [1], [2], [3]

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2 \times_3 \dots \times_n \mathbf{D}_n + \mathcal{N}, \quad (1)$$

occurs in many applications, such as channel estimation [4] and direction-of-arrival estimation [5], where  $\mathcal{Y} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_n}$  is the given measurement tensor,  $\mathbf{D}_t \in \mathbb{C}^{M_t \times N_t}$  ( $t \in \{1, 2, \dots, n\}$ ) are measurement matrices with  $M_t \ll N_t$ ,  $\mathcal{N}$  denotes the additive noise, and  $\times_i$  ( $i \in \{1, 2, \dots, n\}$ ) denotes the  $i$ th mode product of a tensor and a matrix. More specifically, the tensor product  $\mathcal{X} \times_i \mathbf{D}_i$  ( $i = 1, \dots, n$ ) is equivalent to first multiplying each slice of the  $i$ th mode of the tensor with the matrix  $\mathbf{D}_i$ , and then arranging the resulting matrices in sequence to form a new tensor. This model is called the generalized sparse recovery model since the conventional ones, e.g., single measurement vector (SMV) [6] and multiple measurement vector (MMV) [7] models, are one-mode tensor recovery models, i.e., the parameter  $n$  in (1) is equal to 1. As the exploitation of  $n$  modes provides instrumental correlation information from various domains and the confirmed potential to filter out noise inference, high-dimensional tensors exhibit more desirable performance than ones with low modes. Moreover, compared with the MMV problem, the intrinsic structure of the nonzero tensor supports is more flexible, since MMV assumes that the sparse patterns of each sparse signal are the same. These advantages, in terms of recoverability and flexibility, have sparked much interest in tensor recovery.

Representative tensor recovery algorithms, e.g., tensor orthogonal matching pursuit (T-OMP) [4], [8] and tensor block OMP (T-BOMP) [9], are promising due to their fast implementation through the greedy iterative mechanism. They gradually construct an estimated support of the sparse tensors by adding one index that is most strongly correlated with the residual into it per iteration, and then calculate the sparse approximation over the enlarged support. Different from T-OMP, T-BOMP considers the reconstruction of the sparse tensors that exhibit additional structure in the form of the nonzero atoms appearing in clusters, which are referred to as block-sparse [6], [10]. Making explicit use of block sparsity obtains better provable recovery characteristic than the schemes which treat the signals being randomly sparse [6]. In the applications, e.g., multi-band signal processing [11], [12], [13], the block sparsity occurs naturally.

A fundamental question in the analysis of tensor recovery is the characterization of its recoverability. The restricted isometry property (RIP) [14], [15], [16] is one of the main tools for measuring the recoverability of greedy algorithms in conventional SMV and MMV models. It indicates that if a measurement matrix satisfies the RIP with some suitable restricted isometry constants (RICs), then the sparse signals can be recovered reliably. However, calculating the RIC of a matrix is an NP-hard problem. By contrast, the mutual incoherence property (MIP) [6], [17], [18] is computable, and it provides a stronger condition than that of the RIP, i.e., meeting MIP implies that RIP holds but the converse is not true [19]. However, the existing MIP framework is not applicable to the tensor recovery scenario since it only represents the characteristic of a single measurement matrix and cannot materialize the cross coherence within a given matrix set.

Moreover, developing less restrictive theoretical guarantees is useful for practical decision-making, which is highly dependent on an enabling tensor model. The block-sparse tensor is regarded as a generalized case with further consideration of the internal structure characteristic [9], [20]. The authors of [20] consider the case in which one of the nonzero supports of a tensor consists of multiples indices, resulting in block sparsity, and the authors then reformulate the tensor recovery problem into the MMV model-based problem. The work [9] proposes to tensorize data in a block-sparse manner via utilizing its spatial distributions. In practical scenarios, the need for block-sparse tensor recovery arises in various applications, including spectrum sensing [11], [13], [21], channel estimation [4], [8], [22], [23], [24], [25], [26], [27] and sparse representation [28], [29], [30], [31], [32] across different modes. The signals of interest not only possess a multidimensional structure with each mode representing a specific physical attribute like time, frequency or space, leading to the tensor structure, but also exhibit block sparsity due to factors such as multiple-band spectrum utilization and energy spread effect [28], [33], [34]. By leveraging both the tensor structure and block sparsity, block-sparse tensor recovery holds promise in terms of enhancing recovery accuracy and simplifying implementation complexity. Despite these advancements, a comprehensive block-sparse tensor model elucidated through an associated MIP framework is yet to be formulated for computably feasible interpretability.

The focus of the present paper is on revealing theoretical recovery guarantees by exploiting the generalized tensor recovery model. The main contributions are summarized as follows.

- 1) We define concepts revealing the MIP of a given matrix set, as a foundation for the theoretical illumination of the tensor recovery problem. They are intended as indicators of whether the matrix set is sufficiently rich to allow for reliable reconstruction. The flexibility of tensor recovery is further highlighted, since these newly defined concepts can be optimized by adjusting the internal measurement matrices in the set. Analytical comparisons evince the newly defined concepts' stronger ability of tensor structured representation compared with conventional ones.

- 2) A generalized tensor recovery model with block sparsity is formulated. We model a new block structure inside the sparse tensor as shadow block sparsity. It quantifies the number of related blocks between each measurement matrix and the nonzero support blocks of the sparse tensor. Meanwhile, the cases in which the tensor has regular or irregular block structure, i.e., in which the block lengths in certain modes are the same or different, are considered. Correspondingly, we present analytical solutions for these distinct scenarios with the redefined MIP concepts.
- 3) As a generalized variant, a tensor generalized BOMP (T-GBOMP) stemming from the OMP framework that contains multiple tensor block selection mechanism, similar to those of compressive sampling matching pursuit (CoSaMP) [35] and multiple orthogonal least squares (MOLS) [36], is organized and summarized for the first time. The algorithm is exploited as an important tool to derive reliable reconstruction conditions of block-sparse tensor recovery. We further summarize and present several tensorized algorithm variants based on the existing greedy algorithms, and point out that the theoretical results derived are universal and applicable to all of these algorithm variants.
- 4) In-depth analyses are conducted based on the aforementioned preparations. The theoretical results contain two parts, i.e., the analyses on exact recovery and reliable reconstruction in the noiseless and noisy scenarios, respectively. Firstly, we present less restrictive exact recovery condition (ERC) for T-GBOMP (**Theorem 1** and *Remark 2*), and provide corresponding sufficient conditions, in terms of higher reconstructible sparsity level (**Theorem 2**, **Corollary 4**, *Remark 3*, *Remark 4*, and *Remark 5*). These results guarantee that T-GBOMP reconstructs the sparsest tensor from the measurements exactly if the sparsity level falls within an upper bound. Meanwhile, the exponential convergence of residual tensor has also been demonstrated (**Theorem 3**). We have presented the lower and upper bounds of shadow block sparsity in **Proposition 2**, along with specific tensor structures corresponding to these bounds, which facilitate a more intuitive analysis of the results presented in this study. By leveraging the concept of shadow block sparsity, we have obtained analytical results pertaining to crucial parameters, such as different coherences and tensor block length. These results showcase how a general tensor can align with the tensor structure that reaches the lower bound of shadow block sparsity (*Remark 6*). Secondly, noisy recovery conditions are derived based on MIP, which involves upper bounds on the error between the original tensor and the estimated one under various assumptions (**Theorem 4**, **Theorem 6** and **Theorem 8**), the lower bound of SNR required for reliable recovery (**Theorem 5**), and the norm of the residual tensor at specific iterations (**Theorem 7**). It is worth mentioning that the theoretical results derived are related to the number of tensor modes. Based on these findings and asymptotic analysis, it has been revealed that a larger

tensor mode allows for a lower bound on the required SNR for reliable recovery (*Remark 7*). Additionally, block orthogonality of measurement matrices can reduce the upper bound of reconstruction error, wherein a larger block length further increases the performance gap between block orthogonal and non block-orthogonal measurement matrices (*Remark 8*). The aforementioned results derived are tighter and less restrictive than the existing ones (if any and comparable), demonstrating the superiority of block-sparse tensor recovery over other methodologies.

## II. PRELIMINARIES

### A. Notation

We briefly summarize the notation used in this paper. Matrices are denoted by boldface uppercase letters, e.g.,  $\mathbf{D}$ , and tensors are denoted by calligraphic uppercase letters, e.g.,  $\mathcal{X}$ . Vectorization of a tensor  $\mathcal{X}$  is denoted by  $\text{vec}(\mathcal{X})$ . The element in the  $i$ -th row and  $j$ -th column of a matrix  $\mathbf{D}$  is denoted by  $\mathbf{D}_{(i,j)}$ , and  $\mathbf{D}_{(:,j)}$  denotes the  $j$ -th column of  $\mathbf{D}$ .  $\mathbf{D}_{\Theta}$  is a submatrix of  $\mathbf{D}$  that contains the columns indexed by the set  $\Theta$ .  $\mathbf{D} \setminus \mathbf{D}_{\Theta}$  is the residual matrix after all the columns indexed by  $\Theta$  are removed from  $\mathbf{D}$ .  $\mathbf{D}^T$  represents the transpose of  $\mathbf{D}$ , and  $\mathbf{W}^H$  denotes the conjugate transpose of  $\mathbf{W}$ . Given a block length, the  $j$ -th column-block submatrix of matrix  $\mathbf{D}$  is denoted as  $\mathbf{D}_{[j]}$ . The set  $\Theta_B$  consists of the block indices of the set  $\Theta$  based on a given block length.  $\langle \cdot, \cdot \rangle$  denotes the inner product.  $\inf\{f(x)\}$  denotes the infimum of the function  $f(x)$  with respect to the variable  $x$ .  $\mathbf{I}$  stands for the identity matrix. The operation  $\text{supp}(\mathcal{X})$  returns the index set of the nonzero support tensor blocks of  $\mathcal{X}$ . If  $\mathbf{D}_{\Theta}$  has full column rank,  $\mathbf{D}_{\Theta}^{\dagger} = (\mathbf{D}_{\Theta}^H \mathbf{D}_{\Theta})^{-1} \mathbf{D}_{\Theta}^H$  represents the pseudoinverse of  $\mathbf{D}_{\Theta}$ .  $\text{span}(\mathbf{D}_{\Theta})$  denotes the space spanned by the columns of  $\mathbf{D}_{\Theta}$ , and  $\mathbf{P}_{\mathbf{D}_{\Theta}} = \mathbf{D}_{\Theta} \mathbf{D}_{\Theta}^{\dagger}$  is the projection onto  $\text{span}(\mathbf{D}_{\Theta})$ , while  $\mathbf{P}_{\mathbf{D}_{\Theta}}^{\perp} = \mathbf{I} - \mathbf{P}_{\mathbf{D}_{\Theta}}$  is the projection onto the orthogonal complement of  $\text{span}(\mathbf{D}_{\Theta})$ . For a vector  $\mathbf{r}$ ,  $\text{diag}(\mathbf{r})$  denotes the diagonal matrix whose diagonal elements are the entries of  $\mathbf{r}$ . The all zero vector/matrix/tensor is uniformly denoted by  $\mathbf{0}$ .  $|\Theta|$  stands for the cardinality or the block cardinality that allows repetition of set  $\Theta$ , and  $|c|$  is the absolute value of scalar  $c$ . Give an  $n$ -mode tensor  $\mathcal{X}$ , denote its  $i$ th ( $i \in \{1, 2, \dots, n\}$ ) mode tensor index set as  $\Theta_i$  with  $|\Theta_i| = k$ . The corresponding tensor index set for  $\mathcal{X}$  can be represented as

$$\Theta = \{(\Theta_{1_1}, \Theta_{2_1}, \dots, \Theta_{n_1}), (\Theta_{1_2}, \Theta_{2_2}, \dots, \Theta_{n_2}), \dots, (\Theta_{1_k}, \Theta_{2_k}, \dots, \Theta_{n_k})\}$$

with  $|\Theta_i| = |\Theta| = k$  ( $i \in \{1, 2, \dots, n\}$ ), where  $\Theta_{i_j}$  denotes the  $j$ th element of the set  $\Theta_i$  ( $j \in \{1, 2, \dots, k\}$ ). The  $(i_1, \dots, i_n)$ th tensor block of  $\mathcal{X}$  with respect to  $\Theta$  is denoted by  $\mathcal{X}_{[\Theta_{i_1}, \dots, \Theta_{i_n}]} \in \mathbb{C}^{d_1 \times \dots \times d_n}$ . The Kronecker product is represented by  $\otimes$ . Based on a matrix set  $\Upsilon = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ) and the tensor index set  $\Theta$ , we denote the cascading matrix

$$\ddot{\mathbf{D}}_{\Theta} = \begin{bmatrix} \mathbf{D}_{n_{\Theta_{n[1]}}} \otimes \dots \otimes \mathbf{D}_{1_{\Theta_{1[1]}}}, \mathbf{D}_{n_{\Theta_{n[2]}}} \otimes \dots \otimes \mathbf{D}_{1_{\Theta_{1[2]}}}, \\ \dots, \mathbf{D}_{n_{\Theta_{n[|\Theta_n|]}}} \otimes \dots \otimes \mathbf{D}_{1_{\Theta_{1[|\Theta_1|]}}} \end{bmatrix},$$

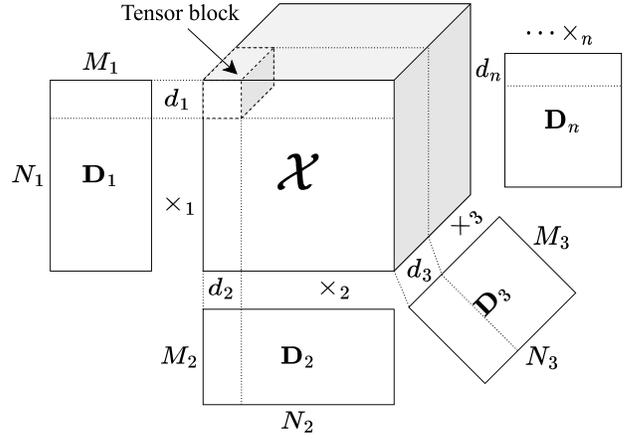


Fig. 1. Illustration of the block-sparse tensor recovery model.

where  $\mathbf{D}_{t_{\Theta_{t[j]}}}$ ,  $t \in \{1, \dots, n\}$  and  $j \in \{1, \dots, |\Theta_t|\}$ , is a column-block submatrix of  $\mathbf{D}_t$  indexed by  $\Theta_{t[j]}$ . The operation  $|\cdot|_u$  denotes the non-repeating cardinality of its argument.

### B. Block-Sparse Tensor Model and Useful Definitions

As illustrated in Fig. 1, the block-sparse tensor  $\mathcal{X} \in \mathbb{C}^{N_1 \times \dots \times N_n}$  is compressed by the  $n$  measurement matrices in the set  $\Upsilon = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times N_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ), where  $N_n = s_n d_n$ . If there exist  $k$  support tensor blocks with nonzero F-norms,  $\mathcal{X}$  is called  $k$  block-sparse tensor, and the corresponding indices constitute the set  $\Xi$  with  $|\Xi| = k$ . The measurement matrices  $\mathbf{D}_i$  for  $i \in \{1, \dots, n\}$  can be rewritten as a concatenation of the  $s_i$  column-block submatrices, i.e.,

$$\mathbf{D}_i = \begin{bmatrix} \underbrace{\mathbf{D}_{i(:,1)} \cdots \mathbf{D}_{i(:,d)}}_{\mathbf{D}_{i[1]}} \underbrace{\mathbf{D}_{i(:,d+1)} \cdots \mathbf{D}_{i(:,2d)}}_{\mathbf{D}_{i[2]}} \\ \dots \\ \underbrace{\mathbf{D}_{i(:,N-d+1)} \cdots \mathbf{D}_{i(:,N)}}_{\mathbf{D}_{i[s_i]}}, \end{bmatrix}$$

where  $\mathbf{D}_{i[j]} \in \mathbb{R}^{M_i \times d_i}$  is the  $j$ -th column-block submatrix of  $\mathbf{D}_i$ . Throughout the paper, the columns in the measurement matrix are normalized to have the unit  $\ell_2$ -norm.

Based on the model presented, we provide the MIP framework, which is an important tool to analyze the theoretical guarantees of compressed sensing (CS) approaches [19]. We first give the conventional MIP concept of a matrix  $\mathbf{D}$ , including the matrix coherence  $\underline{\mu}_{\mathbf{D}}$ , and the block-structure coherence  $\mu_{\mathbf{D}}$  and  $\nu_{\mathbf{D}}$ . Then, we propose the new concepts of mutual block coherence and mutual sub-coherence, which can be regarded as the generalized high-dimensional extensions of the conventional MIP.

*Definition 1 (Matrix coherence [17]):* The coherence of a matrix  $\mathbf{D}$ , which represents the similarity of its elements, is defined as

$$\underline{\mu}_{\mathbf{D}} = \max_{\forall i, j \neq i} |\langle \mathbf{D}_{(:,i)}, \mathbf{D}_{(:,j)} \rangle|.$$

*Definition 2 (Block-structure coherence [10]):* This block-structure coherence of a matrix  $\mathbf{D}$  consists of two individual

concepts, i.e., block coherence and sub-coherence. Let  $d$  be the block length. The block coherence of  $\mathbf{D}$  is defined as

$$\mu_{\mathbf{D}} = \max_{\forall i, j \neq i} \frac{\|\mathbf{M}_{[i,j]}\|_2}{d},$$

where  $\mathbf{M}_{[i,j]} = \mathbf{D}_{[i]}^H \mathbf{D}_{[j]}$ , and  $\mathbf{D}_{[i]}$  is the  $i$ -th column-block submatrix of  $\mathbf{D}$  with block length  $d$ . The sub-coherence of  $\mathbf{D}$  is defined as

$$\nu_{\mathbf{D}} = \max_{\forall l} \max_{\forall i, j \neq i} \left| \langle \mathbf{D}_{[l](:,i)}, \mathbf{D}_{[l](:,j)} \rangle \right|,$$

where  $\mathbf{D}_{[l](:,i)}$  is the  $i$ -th column of  $\mathbf{D}_{[l]}$ .

The block-structure coherence is a generalized extension of the matrix coherence, which is more capable of exploring the MIP-based performance guarantees offered by the block-structure characteristic of a matrix. Theoretical bounds based on the block-structure coherence are usually less restrictive than those based on conventional matrix coherence under the same assumptions [10]. When the block length  $d = 1$ , then block-structure coherence reduces to the matrix coherence, i.e.,  $\mu_{\mathbf{D}} = \underline{\mu}_{\mathbf{D}}$  and  $\nu_{\mathbf{D}} = 0$  [10]. For the tensor recovery case, there exists a matrix set, consisting of  $n$  measurement matrices. Neither the matrix coherence nor the block-structure coherence, which are both applicable only to single matrix, can accurately measure the cross coherence among the given multiple matrices. To represent the degree of coherence among multiple block-structured matrices, we propose new coherence concepts that provide the MIP among  $n$  measurement matrices as follows.

**Definition 3 (Mutual block coherence):** Given a matrix set  $\Upsilon = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ), the mutual block coherence of  $\Upsilon$  is defined by the following two formulations.

Firstly, considering that  $\forall (i_1, i_2, \dots, i_n), (j_1 \neq i_1, j_2 \neq i_2, \dots, j_n \neq i_n)$ , mutual block coherence  $\varpi_{\Upsilon}$  is given by

$$\varpi_{\Upsilon} = \max_{\forall (i_1, \dots, i_n), (j_1 \neq i_1, \dots, j_n \neq i_n)} \left( \frac{1}{\prod_{i=1}^n d_i} \left\| \left( \mathbf{D}_{n[i_n]} \otimes \dots \otimes \mathbf{D}_{1[i_1]} \right)^H \times \left( \mathbf{D}_{n[j_n]} \otimes \dots \otimes \mathbf{D}_{1[j_1]} \right) \right\|_2 \right)^{\frac{1}{n}}, \quad (2)$$

where  $d_i$  is the block length of the matrix  $\mathbf{D}_i$ .

Secondly, without loss of generality,  $\forall (i_1, i_2, \dots, i_n), (j_1 = i_1, j_2 = i_2, \dots, j_t = i_t, j_{t+1} \neq i_{t+1}, \dots, j_n \neq i_n)$  ( $t \geq 1$ ), mutual block coherence is defined as

$$\varpi_{\Upsilon} = \max_{\forall (i_1, \dots, i_n), (j_1 = i_1, \dots, j_t = i_t, j_{t+1} \neq i_{t+1}, \dots, j_n \neq i_n)} \left( \frac{1}{\prod_{i=1}^n d_i} \left\| \left( \mathbf{D}_{n[i_n]} \otimes \dots \otimes \mathbf{D}_{t+1[i_{t+1}]} \otimes \mathbf{D}_{t[i_t]} \otimes \dots \otimes \mathbf{D}_{1[i_1]} \right)^H \times \left( \mathbf{D}_{n[j_n]} \otimes \dots \otimes \mathbf{D}_{t+1[j_{t+1}]} \otimes \mathbf{D}_{t[i_t]} \otimes \dots \otimes \mathbf{D}_{1[i_1]} \right) \right\|_2 \right)^{\frac{1}{n}}. \quad (3)$$

The definition in (2) is more similar to the conventional sparse recovery case as defined in Definition 2, wherein the indices of the measurement matrix blocks used for coherence calculation differ. However, in high-dimensional tensor recovery scenarios, the measurement matrix labels corresponding to different tensor blocks may be partially the same, as long as these labels are not identical, which results in the definition in (3).

**Definition 4 (Mutual sub-coherence):** Given a matrix set  $\Upsilon = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ), the mutual sub-coherence of  $\Upsilon$  is given by the following two definitions.

Firstly, considering that  $\forall (i_1, i_2, \dots, i_n), (j_1 \neq i_1, j_2 \neq i_2, \dots, j_n \neq i_n)$ , mutual sub-coherence  $\tau_{\Upsilon}$  is given by

$$\tau_{\Upsilon} = \max_{\forall (i_1, \dots, i_n), (j_1 \neq i_1, \dots, j_n \neq i_n)} \left( \left\| \left( \mathbf{D}_{n(:,i_n)} \otimes \dots \otimes \mathbf{D}_{1(:,i_1)} \right)^H \times \left( \mathbf{D}_{n(:,j_n)} \otimes \dots \otimes \mathbf{D}_{1(:,j_1)} \right) \right\| \right)^{\frac{1}{n}}, \quad (4)$$

where  $d_i$  is the block length of the matrix  $\mathbf{D}_i$ , and  $\mathbf{D}_{l(:,i_l)}, \mathbf{D}_{l(:,j_l)} \in \mathbf{D}_{l[t_l]}$ ,  $t_l \in \{1, \dots, s_l\}$ ,  $l \in \{1, \dots, n\}$ .

Secondly, without loss of generality,  $\forall (i_1, i_2, \dots, i_n), (j_1 = i_1, j_2 = i_2, \dots, j_t = i_t, j_{t+1} \neq i_{t+1}, \dots, j_n \neq i_n)$  ( $t \geq 1$ ), mutual sub-coherence is defined as

$$\tau_{\Upsilon} = \max_{\forall (i_1, \dots, i_n), (j_1 = i_1, \dots, j_t = i_t, j_{t+1} \neq i_{t+1}, \dots, j_n \neq i_n)} \left( \left\| \left( \mathbf{D}_{n(:,i_n)} \otimes \dots \otimes \mathbf{D}_{t+1(:,i_{t+1})} \otimes \mathbf{D}_{t(:,i_t)} \otimes \dots \otimes \mathbf{D}_{1(:,i_1)} \right)^H \times \left( \mathbf{D}_{n(:,j_n)} \otimes \dots \otimes \mathbf{D}_{t+1(:,j_{t+1})} \otimes \mathbf{D}_{t(:,i_t)} \otimes \dots \otimes \mathbf{D}_{1(:,i_1)} \right) \right\| \right)^{\frac{1}{n-t}}. \quad (5)$$

More specially, the definition in (5) satisfies

$$\tau_{\Upsilon} = \max_{\forall (i_{t+1}, \dots, i_n), (j_{t+1} \neq i_{t+1}, \dots, j_n \neq i_n)} \left\| \left( \mathbf{D}_{n(:,i_n)}^H \mathbf{D}_{n(:,j_n)} \right) \otimes \dots \otimes \left( \mathbf{D}_{t(:,i_{t+1})}^H \mathbf{D}_{t(:,j_{t+1})} \right) \right\|_2^{\frac{1}{n-t}},$$

since the measurement matrices are normalized to have unit column norms.

Compared with the conventional and block-structure coherence, a significant advantage of the representativeness of the mutual coherence is that it can be optimized by adding a matrix with satisfactory coherence or replacing a matrix in the given set with a more desirable matrix. This indicates that the mutual coherence can provide the overall coherence variation within the matrix set. When  $n = 1$  and  $d \neq 1$ , the mutual block coherence and mutual sub-coherence degenerate into the block coherence and sub-coherence of Definition 2, respectively. On the other hand, if  $n = 1$  and  $d = 1$ , they naturally reduce to the matrix coherence of Definition 1. Moreover, for the

definition in (2), we have

$$\varpi_{\mathbf{Y}} = \max_{\forall (i_1, \dots, i_n), (j_1 \neq i_1, \dots, j_n \neq i_n)} \left( \frac{1}{\prod_{i=1}^n d_i} \left\| (\mathbf{D}_{n[i_n]}^H \mathbf{D}_{n[j_n]}) \otimes \dots \otimes (\mathbf{D}_{1[i_1]}^H \mathbf{D}_{1[j_1]}) \right\|_2 \right)^{\frac{1}{n}} \quad (6)$$

$$= \max_{\forall (i_1, \dots, i_n), (j_1 \neq i_1, \dots, j_n \neq i_n)} \left( \prod_{l=1}^n \frac{\|\mathbf{D}_{l[i_l]}^H \mathbf{D}_{l[j_l]}\|_2}{d_l} \right)^{\frac{1}{n}} \quad (7)$$

$$= \left( \prod_{l=1}^n \mu_{\mathbf{D}_l} \right)^{\frac{1}{n}} \quad (8)$$

$$\leq \left( \prod_{l=1}^n \underline{\mu}_{\mathbf{D}_l} \right)^{\frac{1}{n}} \quad (9)$$

$$\leq 1, \quad (10)$$

where (6) is from the definition of mutual block coherence, (7) is based on the property of the norm of Kronecker product, (8) is due to the definition of block coherence, (9) is from [10, Eq. (10)] which states that given a matrix  $\mathbf{D}$ ,  $\mu_{\mathbf{D}} \leq \underline{\mu}_{\mathbf{D}}$ , and (10) is because  $\underline{\mu}_{\mathbf{D}} \leq 1$ . If the matrices in the matrix set consist of orthonormal blocks, i.e.,  $\mathbf{D}_{t[j]}^H \mathbf{D}_{t[j]} = \mathbf{I}$  for  $t \in \{1, \dots, n\}$  and  $j \in \{1, \dots, s_t\}$ , we have

$$\varpi_{\mathbf{Y}} \leq \left( \prod_{l=1}^n \frac{1}{d_l} \right)^{\frac{1}{n}},$$

which is because given a matrix  $\mathbf{D}$  with orthonormal blocks and block length  $d$ ,  $\mu_{\mathbf{D}} \leq \frac{1}{d}$  [10]. As for the definition in (3), we have

$$\varpi_{\mathbf{Y}} = \max_{\forall (i_1, \dots, i_n), (j_1 = i_1, \dots, j_t = i_t, j_{t+1} \neq i_{t+1}, \dots, j_n \neq i_n)} \left( \prod_{l=1}^t \frac{\|\mathbf{D}_{l[i_l]}^H \mathbf{D}_{l[j_l]}\|_2}{d_l} \prod_{l=t+1}^n \frac{\|\mathbf{D}_{l[i_l]}^H \mathbf{D}_{l[j_l]}\|_2}{d_l} \right)^{\frac{1}{n}} \quad (11)$$

$$\leq \left( \prod_{l=1}^t \frac{1 + (d_l - 1)\nu_{\mathbf{D}_l}}{d_l} \prod_{l=t+1}^n \mu_{\mathbf{D}_l} \right)^{\frac{1}{n}} \quad (12)$$

where (11) is from the Geršgorin's disc theorem [37], and (12) is because  $\frac{1 + (d_l - 1)\nu_{\mathbf{D}_l}}{d_l} = \frac{1}{d_l}(1 - \nu_{\mathbf{D}_l}) + \nu_{\mathbf{D}_l} \leq 1$  and  $\mu_{\mathbf{D}_l} \leq 1$ .

Similarly, for the mutual sub-coherence defined in (4), we have

$$\tau_{\mathbf{Y}} = \max_{\forall (i_1, \dots, i_n), (j_1 \neq i_1, \dots, j_n \neq i_n)} \left( \left\| (\mathbf{D}_{n(\cdot, i_n)}^H \mathbf{D}_{n(\cdot, j_n)}) \otimes \dots \otimes (\mathbf{D}_{1(\cdot, i_1)}^H \mathbf{D}_{1(\cdot, j_1)}) \right\|_2 \right)^{\frac{1}{n}} \quad (13)$$

$$= \max_{\forall (i_1, \dots, i_n), (j_1 \neq i_1, \dots, j_n \neq i_n)} \left( \prod_{l=1}^n \left\| \mathbf{D}_{l(\cdot, i_l)}^H \mathbf{D}_{l(\cdot, j_l)} \right\|_2 \right)^{\frac{1}{n}} \quad (14)$$

$$= \left( \prod_{l=1}^n \nu_{\mathbf{D}_l} \right)^{\frac{1}{n}} \quad (15)$$

$$\leq \left( \prod_{l=1}^n \underline{\mu}_{\mathbf{D}_l} \right)^{\frac{1}{n}}, \quad (16)$$

where (16) is because given a matrix  $\mathbf{D}$ ,  $\nu_{\mathbf{D}} \leq \underline{\mu}_{\mathbf{D}}$ . Based on the similar derivations in (13)-(16), for the definition in (5), we have

$$\tau_{\mathbf{Y}} \leq \left( \prod_{l=t+1}^n \underline{\mu}_{\mathbf{D}_l} \right)^{\frac{1}{n-t}}.$$

*Remark 1:* The above definitions apply to the scenario where the block structure of the sparse tensor is regular, i.e., the block length of each block in each mode is the same. For the case in which the block lengths within the  $t$ th ( $t \in \{1, \dots, n\}$ ) mode are different, we present two useful solutions. The first one considers that irregular block lengths do not lead to a block structure, which results in a block length 1 for the modes where the block lengths are different. The second one considers this irregular block lengths case as a generalized block structure, and we provide extended MIP concepts serving for this generalized structural characteristic. Note that we consider the definitions in (2) and (4) as examples, and similar analyses can be easily extended to those of the definitions in (3) and (5).

We now present the first solution. The case in which the block lengths of the  $t$ th mode are different is provided as a concise example, which can be directly extended to the scenario where the block lengths of multiple modes are different. To this end, let the block length of the  $t$ th ( $t \in \{1, \dots, n\}$ ) mode be 1, i.e.,  $d_t = 1$ . From **Definitions 3** and **4**, we have

$$\varpi_{\mathbf{Y}} = \max_{(i_1, \dots, i_t, \dots, i_n), (j_1 \neq i_1, \dots, j_t \neq i_t, \dots, j_n \neq i_n)} \left( \frac{1}{\left( \prod_{l=1}^{t-1} d_l \right) d_t \left( \prod_{l=t+1}^n d_l \right)} \left\| (\mathbf{D}_{n[i_n]} \otimes \dots \otimes \mathbf{D}_{t[i_t]} \otimes \dots \otimes \mathbf{D}_{1[i_1]})^H \times (\mathbf{D}_{n[j_n]} \otimes \dots \otimes \mathbf{D}_{t[j_t]} \otimes \dots \otimes \mathbf{D}_{1[j_1]}) \right\|_2 \right)^{\frac{1}{n}}$$

$$= \max_{(i_1, \dots, i_n), (j_1 \neq i_1, \dots, j_n \neq i_n)} \left( \frac{1}{\left( \prod_{l=1}^{t-1} d_l \right) \left( \prod_{l=t+1}^n d_l \right)} \left\| (\mathbf{D}_{n[i_n]} \otimes \dots \otimes \mathbf{D}_{t(\cdot, i_t)} \otimes \dots \otimes \mathbf{D}_{1[i_1]})^H \times (\mathbf{D}_{n[j_n]} \otimes \dots \otimes \mathbf{D}_{t(\cdot, j_t)} \otimes \dots \otimes \mathbf{D}_{1[j_1]}) \right\|_2 \right)^{\frac{1}{n}},$$

$$\tau_{\mathbf{Y}} = \max_{(i_1, \dots, i_n), (j_1 \neq i_1, \dots, j_n \neq i_n)} \left( \left\| (\mathbf{D}_{n(\cdot, i_n)} \otimes \dots \otimes \mathbf{D}_{t(\cdot, i_t)} \otimes \dots \otimes \mathbf{D}_{1(\cdot, i_1)})^H \times (\mathbf{D}_{n(\cdot, j_n)} \otimes \dots \otimes \mathbf{D}_{t(\cdot, j_t)} \otimes \dots \otimes \mathbf{D}_{1(\cdot, j_1)}) \right\|_2 \right)^{\frac{1}{n}}.$$

Thus, similar to (8) and (15), we have

$$\begin{aligned}\varpi_{\mathbf{Y}} &= \underline{\mu}_{\mathbf{D}_t}^{\frac{1}{n}} \left( \prod_{l=1}^{t-1} \mu_{\mathbf{D}_l} \prod_{l=t+1}^n \mu_{\mathbf{D}_l} \right)^{\frac{1}{n}}, \\ \tau_{\mathbf{Y}} &= \underline{\nu}_{\mathbf{D}_t}^{\frac{1}{n}} \left( \prod_{l=1}^{t-1} \nu_{\mathbf{D}_l} \prod_{l=t+1}^n \nu_{\mathbf{D}_l} \right)^{\frac{1}{n}}.\end{aligned}$$

Under a special case where  $d_1 = \dots = d_n = 1$ , we define  $\tau_{\mathbf{Y}} = 0$ . In a second case, suppose that the block lengths of the  $t$ th ( $t \in \{1, \dots, n\}$ ) mode are different, and there are  $s_t$  tensor blocks in the  $t$ th mode with the block lengths being  $d_{t_i}$ ,  $i \in \{1, \dots, s_t\}$ . Then, we define

$$\begin{aligned}\varpi_{\mathbf{Y}} &= \max_{(i_1, \dots, i_n), (j_1 \neq i_1, \dots, j_n \neq i_n)} \\ &\left( \frac{1}{\prod_{t=1}^n \sqrt{d_{t_{i_n}} d_{t_{j_n}}}} \left\| (\mathbf{D}_{n_{[i_n]} \otimes \dots \otimes \mathbf{D}_{1_{[i_1]}})^H \right. \right. \\ &\quad \left. \left. \times (\mathbf{D}_{n_{[j_n]} \otimes \dots \otimes \mathbf{D}_{1_{[j_1]}}) \right\|_2 \right)^{\frac{1}{n}}, \\ \tau_{\mathbf{Y}} &= \max_{\forall (i_1, \dots, i_n), (j_1 \neq i_1, \dots, j_n \neq i_n)} \\ &\left( \left\| (\mathbf{D}_{n_{(\cdot, i_n)} \otimes \dots \otimes \mathbf{D}_{1_{(\cdot, i_1)}})^H \right. \right. \\ &\quad \left. \left. \times (\mathbf{D}_{n_{(\cdot, j_n)} \otimes \dots \otimes \mathbf{D}_{1_{(\cdot, j_1)}}) \right\|_2 \right)^{\frac{1}{n}}.\end{aligned}\quad (17)$$

It can be seen that  $\tau_{\mathbf{Y}}$  defined in (17) is the same as that in (4) with indices  $i_t$  and  $j_t$  ( $t \in \{1, \dots, n\}$ ) being within the same block of  $\mathbf{D}_t$ . It can be observed that the first solution does not require any additional input of structural information for reliable recovery analysis as the block length is regarded as 1, while the second solution relies on specific block length information of the irregular structure. For concise symbol representation, in the sequel, we consider the theoretical derivation of the regular block structure scenario, which can be directly extended to the irregular block length case by using the two solutions presented in this remark.

Moreover, we denote the mutual block coherence and mutual sub-coherence as  $\varpi_{\mathbf{Y}, t}$  and  $\tau_{\mathbf{Y}, t}$ , respectively, when  $t$  indices ( $0 \leq t \leq n$ ) are the same in their definitions. Furthermore, when  $t = 0$ , the mutual block coherence and mutual sub-coherence can be simplified to be represented as  $\varpi_{\mathbf{Y}}$  and  $\tau_{\mathbf{Y}}$ , respectively. In the following, a new structural concept of the block-sparse tensor is presented, which can figuratively display the number of measurement matrix blocks that are related to the support tensor blocks in tensor recovery.

**Definition 5:** (Shadow block sparsity) As illustrated in Fig. 2, the  $i$ th ( $i = 1, 2, \dots, n$ ) mode shadow block sparsity  $k_i$  of a tensor  $\mathcal{X} \in \mathbb{C}^{N_1 \times \dots \times N_n}$  is defined as the number of nonzero blocks blocking the field of view in the  $i$ th mode direction.

Denoting the index set of nonzero tensor blocks in  $\mathcal{X}$  as  $\Xi = \{(\Xi_{1_1}, \Xi_{2_1}, \dots, \Xi_{n_1}), (\Xi_{1_2}, \Xi_{2_2}, \dots, \Xi_{n_2}), \dots, (\Xi_{1_k}, \Xi_{2_k}, \dots, \Xi_{n_k})\}$ , we have  $k_i = |\Xi_i|_u$  and  $k_i \in [1, k]$  ( $i \in [1, 2, \dots, n]$ ), since there exist duplicate indices in  $\Xi_i$ , where the elements in  $\Xi$  are  $n$ -dimensional. We further define a shadow extraction operation, which returns a neat subtensor consisting of the nonzero supports of the

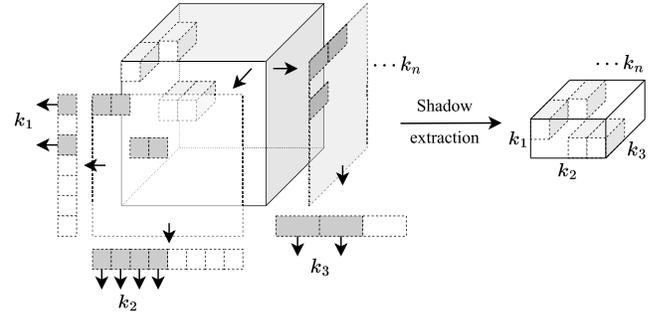


Fig. 2. Illustration of shadow block sparsity and shadow extraction.

original tensor but with sufficient gaps as shown in Fig. 2. In other words, the shadow extraction operation removes all of the zero slices of the tensor and returns the obtained subtensor, which is sized as  $k_1 \times k_2 \times \dots \times k_n$ . Mathematically, given an index set  $\Theta$ , the extraction operation on  $\mathcal{X}_{\Theta}$  returns the neat subtensor consisting of the tensor blocks indexed by  $\Theta$  and with sufficient gaps.

To aid the derivation of both exact and reliable recovery conditions in noiseless and noisy scenarios, we further present the definitions of block mixed norm, SNR and an extension of minimum-to-average ratio (MAR) [36] on block-sparse tensors as follows.

**Definition 6:** (Block mixed norm [10]) For a matrix  $\mathbf{D} \in \mathbb{C}^{M \times N}$  with  $M = md$  and  $N = nd$ , denote  $\mathbf{D}_{[i,j]}$  as the  $(i, j)$ th  $d \times d$  block of  $\mathbf{D}$ . The two components of the block mixed norm of  $\mathbf{D}$  are given by

$$\begin{aligned}\rho_r(\mathbf{D}) &= \max_i \sum_j \|\mathbf{D}_{[i,j]}\|_2, \\ \rho_c(\mathbf{D}) &= \max_j \sum_i \|\mathbf{D}_{[i,j]}\|_2.\end{aligned}$$

**Definition 7 (SNR):** The SNR is defined as the ratio of the signal component to the noise component, i.e.,

$$\text{SNR} = \frac{\|\mathcal{X} \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2 \times_3 \dots \times_n \mathbf{D}_n\|_F^2}{\|\mathcal{N}\|_F^2}.$$

**Definition 8 (MAR on block-sparse tensors):** Given a  $k$  block-sparse tensor  $\mathcal{X}$  whose nonzero support tensor blocks are indexed by the set  $\Xi$ , the MAR on  $\mathcal{X}$  is defined as

$$\text{MAR}_* = \frac{\sqrt{k} \min_{(i_1, \dots, i_n) \in \Xi} \|\mathcal{X}_{[i_1, \dots, i_n]}\|_F}{\|\mathcal{X}\|_F}.$$

### III. USEFUL LEMMAS AND BLOCK-SPARSE TENSOR RECOVERY ALGORITHMS

#### A. Some Useful Lemmas

Analyses related to the extreme eigenvalues of measurement matrices are useful, and they have been exploited in many works, such as [6] and [19]. Before presenting the theoretical results bounding the eigenvalues, a useful proposition is provided first for an intuitive expression of the eigenvalue bounds.

**Proposition 1:** Consider a matrix set  $\Upsilon = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ). Denote  $G_i = s_i - 1$ , and define the following generating function:

$$\begin{aligned} g(x) &= (x + G_1)(x + G_2) \cdots (x + G_n) \\ &= C_n x^n + C_{n-1} x^{n-1} + \cdots + C_0 x^0. \end{aligned}$$

For the block matrix  $\mathbf{T}^H \mathbf{T}$  whose blocks are  $(\mathbf{D}_{1[i_1]} \otimes \mathbf{D}_{2[i_2]} \otimes \cdots \otimes \mathbf{D}_{n[i_n]} \otimes \mathbf{D}_{n[i_{t+1}]} \otimes \cdots \otimes \mathbf{D}_{n[i_n]})^H (\mathbf{D}_{1[j_1]} \otimes \mathbf{D}_{2[j_2]} \otimes \cdots \otimes \mathbf{D}_{n[j_t]} \otimes \mathbf{D}_{n[j_{t+1}]} \otimes \cdots \otimes \mathbf{D}_{n[j_n]})$  sized as  $\prod_{i=1}^n d_i \times \prod_{i=1}^n d_i$ , where  $\mathbf{T} = \mathbf{D}_1 \otimes \mathbf{D}_2 \otimes \cdots \otimes \mathbf{D}_n$ , the number of blocks in the block matrix  $\mathbf{T}^H \mathbf{T}$  satisfying  $i_1 = j_1, i_2 = j_2, \dots, i_t = j_t$  is equal to  $C_t \prod_{l=1}^n s_l$  ( $1 \leq t \leq n$ ). Specifically, if there is no identical index, the corresponding number is equal to  $C_0 \prod_{l=1}^n s_l$ . If considering block matrices on a certain column or row, then the aforementioned numbers are equal to  $C_t$  and  $C_0$ , respectively.

Similarly, denote  $\underline{G}_i = d_i - 1$ , and define the following generating function:

$$\begin{aligned} \underline{g}(x) &= (x + \underline{G}_1)(x + \underline{G}_2) \cdots (x + \underline{G}_n) \\ &= \underline{C}_n x^n + \underline{C}_{n-1} x^{n-1} + \cdots + \underline{C}_0 x^0. \end{aligned}$$

For the matrix  $\underline{\mathbf{T}}^H \underline{\mathbf{T}} = (\mathbf{D}_{1[i_1]} \otimes \mathbf{D}_{2[i_2]} \otimes \cdots \otimes \mathbf{D}_{t[i_t]} \otimes \mathbf{D}_{t+1[i_{t+1}]} \otimes \cdots \otimes \mathbf{D}_{n[i_n]})^H (\mathbf{D}_{1[i_1]} \otimes \mathbf{D}_{2[i_2]} \otimes \cdots \otimes \mathbf{D}_{t[i_t]} \otimes \mathbf{D}_{t+1[i_{t+1}]} \otimes \cdots \otimes \mathbf{D}_{n[i_n]})$ , where  $\underline{\mathbf{T}} = \mathbf{D}_{1[i_1]} \otimes \mathbf{D}_{2[i_2]} \otimes \cdots \otimes \mathbf{D}_{t[i_t]} \otimes \mathbf{D}_{t+1[i_{t+1}]} \otimes \cdots \otimes \mathbf{D}_{n[i_n]}$ , whose elements are  $(\mathbf{D}_{1[i_1](:,j_1)} \otimes \mathbf{D}_{2[i_2](:,j_2)} \otimes \cdots \otimes \mathbf{D}_{t[i_t](:,j_t)} \otimes \mathbf{D}_{t+1[i_{t+1}](:,j_{t+1})} \otimes \cdots \otimes \mathbf{D}_{n[i_n](:,j_n)})^H (\mathbf{D}_{1[i_1](:,q_1)} \otimes \mathbf{D}_{2[i_2](:,q_2)} \otimes \cdots \otimes \mathbf{D}_{t[i_t](:,q_t)} \otimes \mathbf{D}_{t+1[i_{t+1}](:,q_{t+1})} \otimes \cdots \otimes \mathbf{D}_{n[i_n](:,q_n)})$ , the number of elements in  $\underline{\mathbf{T}}^H \underline{\mathbf{T}}$  satisfying  $j_1 = q_1, j_2 = q_2, \dots, j_t = q_t$  is equal to  $\underline{C}_t \prod_{l=1}^n s_l$  ( $1 \leq t \leq n$ ). Specifically, if there is no identical index, the corresponding number is equal to  $\underline{C}_0 \prod_{l=1}^n s_l$ . If considering elements on a certain column or row, then the aforementioned numbers are equal to  $\underline{C}_t$  and  $\underline{C}_0$ , respectively.

Note that  $\sum_{l=0}^n C_l = \prod_{l=1}^n s_l$  and  $\sum_{l=0}^n \underline{C}_l = \prod_{l=1}^n d_l$ . Meanwhile,  $C_t$  and  $\underline{C}_t$  can be formulated concretely as

$$\begin{aligned} C_t &= \sum_{1 \leq j_1 < j_2 < \cdots < j_{n-t} < \cdots \leq n} G_{j_1} G_{j_2} \cdots G_{j_{n-t}} \\ &= \sum_{1 \leq j_1 < j_2 < \cdots < j_{n-t} < \cdots \leq n} \prod_{i=1}^{n-t} G_{j_i}, \\ \underline{C}_t &= \sum_{1 \leq j_1 < j_2 < \cdots < j_{n-t} < \cdots \leq n} \underline{G}_{j_1} \underline{G}_{j_2} \cdots \underline{G}_{j_{n-t}} \\ &= \sum_{1 \leq j_1 < j_2 < \cdots < j_{n-t} < \cdots \leq n} \prod_{i=1}^{n-t} \underline{G}_{j_i}, \end{aligned}$$

respectively. For the sake of simplicity in notation, in the sequel, we denote  $C_t$  satisfying  $\sum_{l=0}^n C_l = \alpha$  and  $\underline{C}_t$  satisfying

$\sum_{l=0}^n \underline{C}_l = \beta$  as  $C_t^\alpha$  and  $\underline{C}_t^\beta$ , respectively. Based on **Proposition 1**, the following lemma reveals that the eigenvalues of the Kronecker product of the measurement matrices are bounded in terms of mutual block coherence and mutual sub-coherence.

**Lemma 1:** Consider a matrix set  $\Upsilon = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ). Let  $\mathbf{T} = \mathbf{D}_1 \otimes \cdots \otimes \mathbf{D}_n$ . Define  $\lambda_{\min}$  and  $\lambda_{\max}$  as the minimum and maximum eigenvalues of the matrix  $\mathbf{T}^H \mathbf{T}$ . If  $\underline{W}_{\Upsilon, \prod_{t=1}^n s_t} > 0$ , then the following inequality holds:

$$\underline{W}_{\Upsilon, \prod_{t=1}^n s_t} \leq \lambda_{\min} \leq \lambda_{\max} \leq \overline{W}_{\Upsilon, \prod_{t=1}^n s_t},$$

where

$$\begin{aligned} \underline{W}_{\Upsilon, \prod_{t=1}^n s_t} &= 1 - \sum_{t=0}^{n-1} \underline{C}_t \prod_{l=1}^n s_l \tau_{\Upsilon, t}^{n-t} - \sum_{t=0}^{n-1} C_t \prod_{l=1}^n s_l \varpi_{\Upsilon, t} \prod_{i=1}^n d_i, \\ \overline{W}_{\Upsilon, \prod_{t=1}^n s_t} &= 1 + \sum_{t=0}^{n-1} \underline{C}_t \prod_{l=1}^n s_l \tau_{\Upsilon, t}^{n-t} + \sum_{t=0}^{n-1} C_t \prod_{l=1}^n s_l \varpi_{\Upsilon, t} \prod_{i=1}^n d_i, \end{aligned}$$

in which  $\varpi_{\Upsilon, t}$  and  $\tau_{\Upsilon, t}$  are the mutual block coherence and mutual sub-coherence of the matrix set  $\Upsilon$ , respectively.

*Proof:* See Appendix A.  $\blacksquare$

It can be observed from **Lemma 1** that as the mutual block coherence  $\varpi_{\Upsilon, t}$  and the mutual sub-coherence  $\tau_{\Upsilon, t}$  decrease,  $\underline{W}_{\Upsilon, \prod_{t=1}^n s_t}$  and  $\overline{W}_{\Upsilon, \prod_{t=1}^n s_t}$  approach 1, indicating tighter eigenvalue bounds. Generally speaking, to recover the block-sparse tensor accurately, the degree of linear dependency among the entries of the corresponding measurement matrices needs to be small. This reveals that there is a significant amount of effective information in the measurement matrices. Assume that the measurement matrices in the set  $\Upsilon$  are block orthogonal, which results in the mutual sub-coherence  $\tau_{\Upsilon, t}$  being equal to 0, and we have

$$\begin{aligned} 1 - \sum_{t=0}^{n-1} C_t \prod_{l=1}^n s_l \varpi_{\Upsilon, t} \prod_{i=1}^n d_i &\leq \lambda_{\min} \\ &\leq \lambda_{\max} \leq 1 + \sum_{t=0}^{n-1} C_t \prod_{l=1}^n s_l \varpi_{\Upsilon, t} \prod_{i=1}^n d_i. \end{aligned} \quad (18)$$

The bounds in (18) becomes tighter than those in **Lemma 1**, which indicates that a block orthogonal structure can bring significant improvements in the eigenvalue boundary. Due to the orthogonality of the measurement matrix blocks, the atomic information contained within these blocks is relatively high, leading to a more reliable recovery performance.

**Lemma 1** is applicable to symmetric matrices, while, in the general case, the matrices to be analyzed are not symmetric matrices. Therefore, we present the following lemma in which an upper bound on the matrix spectral norm is provided, which is suitable for any given matrix.

**Lemma 2:** Given a matrix  $\mathbf{D} \in \mathbb{C}^{M \times N}$  with  $M = md$  and  $N = nd$ , we have

$$\|\mathbf{D}\|_2 \leq \sqrt{\rho_c(\mathbf{D}) \rho_r(\mathbf{D})} \quad (19)$$

$$\leq \max\{\rho_r(\mathbf{D}), \rho_c(\mathbf{D})\}, \quad (20)$$

where  $\mathbf{D}_{[i,j]}$  is the  $(i, j)$ th  $d \times d$  block of  $\mathbf{D}$ .

*Proof:* See Appendix B.  $\blacksquare$

It is interesting that (19) can be seen as an extension of the Geršgorin's disc theorem [37] for any block matrix and does not require the restriction that matrices need to be square. Moreover, the inequality (20) introduces a scaling, which is useful as it allows to bound the matrix norm based on just one

of the block mixed norms. In certain particular cases where the rows and columns of a matrix have different meanings, such as the rows of a measurement matrix represent coefficients from different candidate entries and the columns represent individual entries, the inequality (20) can provide intuitive bounds that reflect the characteristics of either the rows or the columns of the matrix.

Based on **Lemma 2**, for the case in which the matrix is square, the following corollary holds.

*Corollary 1:* Given a symmetric matrix  $\mathbf{D} \in \mathbb{C}^{M \times M}$  with  $M = md$ , we have

$$\|\mathbf{D}\|_2 \leq \min\{\rho_r(\mathbf{D}), \rho_c(\mathbf{D})\},$$

where  $\mathbf{D}_{[i,j]}$  is the  $(i, j)$ th  $d \times d$  block of  $\mathbf{D}$ .

The proof of Corollary 1 is direct since the spectral radius of a symmetric matrix is equal to its spectral norm, and the spectral radius of a square matrix is less than or equal to any other matrix norm. Now consider a symmetric matrix  $\mathbf{D}_*$ . Geršgorin's disc theorem on  $\mathbf{D}_*$  is given by

$$\|\mathbf{D}_*\|_2 \leq \min\{\|\mathbf{D}_*\|_\infty, \|\mathbf{D}_*\|_1\}. \quad (21)$$

Since

$$\begin{aligned} \|\mathbf{D}_*\|_\infty &= \max_l \sum_r |\mathbf{D}_{*(l,r)}| \\ &\geq \max_i \sum_j \|\mathbf{D}_{*[i,j]}\|_2 = \rho_r(\mathbf{D}_*), \\ \|\mathbf{D}_*\|_1 &= \max_r \sum_l |\mathbf{D}_{*(l,r)}| \\ &\geq \max_j \sum_i \|\mathbf{D}_{*[i,j]}\|_2 = \rho_c(\mathbf{D}_*), \end{aligned}$$

we have  $\min\{\rho_r(\mathbf{D}_*), \rho_c(\mathbf{D}_*)\} \leq \min\{\|\mathbf{D}_*\|_\infty, \|\mathbf{D}_*\|_1\}$ , which indicates that the bound in **Corollary 1** is tighter than (21).

Exploiting **Lemmas 1** and **2**, we present the following two corollaries. They are the building blocks for the feasible derivation of both exact and reliable recovery conditions in noiseless and noisy scenarios.

*Corollary 2:* Consider a matrix set  $\Upsilon = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ), and  $n$  pairs of column-block submatrices,  $\mathbf{D}_{1_{\Xi_1}}$  and  $\mathbf{D}_{1_{\Xi_1^*}}$  of  $\mathbf{D}_1, \dots, \mathbf{D}_{n_{\Xi_n}}$  and  $\mathbf{D}_{n_{\Xi_n^*}}$  of  $\mathbf{D}_n$  with  $\Xi_1 \cap \Xi_1^* = \emptyset, \dots, \Xi_n \cap \Xi_n^* = \emptyset$ ,  $|\Xi_1| = l_1, |\Xi_1^*| = l_1^*, \dots, |\Xi_n| = l_n$  and  $|\Xi_n^*| = l_n^*$ . Let  $\mathbf{T} = \mathbf{D}_{1_{\Xi_1}} \otimes \dots \otimes \mathbf{D}_{n_{\Xi_n}}$  and  $\mathbf{T}^* = \mathbf{D}_{1_{\Xi_1^*}} \otimes \dots \otimes \mathbf{D}_{n_{\Xi_n^*}}$ . Define  $\sigma_{\min}$  and  $\sigma_{\max}$  as the minimum and maximum singular values of  $\mathbf{T}^H \mathbf{T}^*$ . If  $\underline{W}_{\Upsilon, \prod_{t=1}^n l_t} > 0$  and  $\underline{W}_{\Upsilon, \prod_{t=1}^n l_t^*} > 0$ , then the following inequality holds:

$$\begin{aligned} \underline{W}_{\Upsilon, \prod_{t=1}^n l_t} \underline{W}_{\Upsilon, \prod_{t=1}^n l_t^*} &\leq \sigma_{\min} \\ &\leq \sigma_{\max} \leq \max \left\{ \varpi_{\Upsilon}^n \prod_{t=1}^n l_t d_t, \varpi_{\Upsilon}^n \prod_{t=1}^n l_t^* d_t \right\}, \end{aligned}$$

where  $\underline{W}_{\Upsilon, l_n}$  and  $\underline{W}_{\Upsilon, l_n^*}$  can be obtained by **Lemma 1**, and  $d_t$  ( $t \in \{1, 2, \dots, n\}$ ) is the block length of  $\mathbf{D}_t$ .

*Proof:* See Appendix C. ■

*Corollary 3:* Consider an  $n$ -mode  $k$  block-sparse tensor  $\mathcal{X} \in \mathbb{C}^{N_1 \times N_2 \times \dots \times N_n}$  with block length and block shadow

---

### Algorithm 1 T-GBOMP

---

**Input:**  $\mathbf{D}_1 \cdots \mathbf{D}_n$ ,  $\mathcal{Y}$ , block sparsity level  $k$ , block length  $d_1, d_2, \dots, d_n$ , residual tolerant  $\epsilon$  and selection parameter  $s \leq k$ .

**Output:**  $\hat{\mathcal{X}}, \hat{\Xi}$ .

- 1: **Initialization:**  $l = 0, \mathcal{R}^0 = \mathcal{Y}, \Xi^0 = \emptyset, \mathcal{X}^0 = \mathbf{0}$ .
  - 2: **while**  $l < k$  and  $\|\mathcal{R}^l\|_2 > \epsilon$  **do**
  - 3: Identify the support by  $\Theta^{l+1} = \arg \max_{\Theta: |\Theta|=s} \sum_{(i_1, i_2, \dots, i_n) \in \Theta} \|\mathcal{R}^l \times_1 \mathbf{D}_{1_{[i_1]}}^H \times_2 \cdots \times_n \mathbf{D}_{n_{[i_n]}}^H\|_F$ .
  - 4: Augment  $\Xi^{l+1} = \Xi^l \cup \Theta^{l+1}$ .
  - 5: Estimate  $\mathcal{X}^{l+1} = \arg \min_{\mathcal{X}: \text{supp}(\mathcal{X}) = \Xi^{l+1}} \|\text{vec}(\mathcal{Y}) - \sum_{(i_1, \dots, i_n) \in \Xi^{l+1}} (\mathbf{D}_{n_{[i_n]}} \otimes \cdots \otimes \mathbf{D}_{1_{[i_1]}}) \text{vec}(\mathcal{X}_{[i_1, \dots, i_n]})\|_F$ .
  - 6: Update  $\mathcal{R}^{l+1} = \mathcal{Y} - \mathcal{X}^{l+1} \times_1 \mathbf{D}_1 \times_2 \cdots \times_n \mathbf{D}_n$ .
  - 7: Update  $l = l + 1$ .
  - 8: **end while**
  - 9: **return**  $\hat{\Xi} = \arg \min_{\Theta: |\Theta|=k} \|\mathcal{X}^l - \mathcal{X}_{\Theta}^l\|_F$  and  $\hat{\mathcal{X}}$  with  $\hat{\mathcal{X}}_{\Omega \setminus \hat{\Xi}} = \mathbf{0}$ .
- 

sparsity of the  $n$ th mode equal to  $d_n$  and  $k_n$  respectively, and a measurement matrix set  $\Upsilon = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ). Then we have

$$\begin{aligned} \underline{W}_{\Upsilon, \prod_{t=1}^n k_t} \|\mathcal{X}\|_F &\leq \|\mathcal{X} \times_1 \mathbf{D}_1 \times_2 \cdots \times_n \mathbf{D}_n\|_F \\ &\leq \overline{W}_{\Upsilon, \prod_{t=1}^n k_t} \|\mathcal{X}\|_F, \\ \underline{W}_{\Upsilon, \prod_{t=1}^n k_t} \|\mathcal{X}\|_F &\leq \|\mathcal{X} \times_1 (\mathbf{D}_1^H \mathbf{D}_1) \times_2 (\mathbf{D}_2^H \mathbf{D}_2) \\ &\quad \times_3 \cdots \times_n (\mathbf{D}_n^H \mathbf{D}_n)\|_F \\ &\leq \overline{W}_{\Upsilon, \prod_{t=1}^n k_t} \|\mathcal{X}\|_F, \end{aligned} \quad (22)$$

where  $\underline{W}_{\Upsilon, k_n}$  and  $\overline{W}_{\Upsilon, k_n}$  can be obtained by **Lemma 1**.

For  $n$  pairs of column-block submatrices  $\mathbf{D}_{1_{\Xi_1}}$  and  $\mathbf{D}_{1_{\Xi_1^*}}$  of  $\mathbf{D}_1, \dots, \mathbf{D}_{n_{\Xi_n}}$  and  $\mathbf{D}_{n_{\Xi_n^*}}$  of  $\mathbf{D}_n$  with  $\Xi_1 \cap \Xi_1^* = \emptyset, \dots, \Xi_n \cap \Xi_n^* = \emptyset$ ,  $|\Xi_1| = l_1, |\Xi_1^*| = l_1^*, \dots, |\Xi_n| = l_n$  and  $|\Xi_n^*| = l_n^*$ , we have

$$\begin{aligned} \underline{W}_{\Upsilon, \prod_{t=1}^n l_t} \underline{W}_{\Upsilon, \prod_{t=1}^n l_t^*} \|\mathcal{X}\|_F &\leq \|\mathcal{X} \times_1 (\mathbf{D}_{1_{\Xi_1}}^H \mathbf{D}_{1_{\Xi_1^*}}) \times_2 \cdots \times_n (\mathbf{D}_{n_{\Xi_n}}^H \mathbf{D}_{n_{\Xi_n^*}})\|_F \\ &\leq \max \left\{ \varpi_{\Upsilon}^n \prod_{t=1}^n l_t d_t, \varpi_{\Upsilon}^n \prod_{t=1}^n l_t^* d_t \right\} \|\mathcal{X}\|_F. \end{aligned} \quad (23)$$

For any vector  $\mathbf{x} \in \mathbb{C}^{s_n \prod_{t=1}^n d_t}$ , we have

$$\underline{W}_{\Upsilon, s_n} \|(\mathbf{T}^\dagger)^H \mathbf{x}\|_F \leq \|\mathbf{x}\|_F \leq \overline{W}_{\Upsilon, s_n} \|(\mathbf{T}^\dagger)^H \mathbf{x}\|_F, \quad (24)$$

where  $\mathbf{T} = \left[ \mathbf{D}_{n_{\Xi_n[1]}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1[1]}} \mid \cdots \mid \mathbf{D}_{n_{\Xi_n[s_n]}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1[s_n]}} \right]$ .

*Proof:* See Appendix D. ■

### B. T-GBOMP and Relations With Other Existing Algorithms

As illustrated in Algorithm 1, the T-GBOMP algorithm chooses  $s$  tensor blocks that are most strongly correlated

with the residual tensor, and adds the selected indices to the support list per iteration. Then, it estimates the block-sparse tensor over the enlarged support matrix. As T-GBOMP incorporates multiple useful mechanisms from existing greedy algorithms, i.e., the multiple support selection [36], the structure calculation [6], the correlation discriminant [10], and the high-dimensional spatial mapping [4] mechanisms, T-GBOMP can be seen as a generalization of these earlier algorithms. Theoretical results derived for T-GBOMP can thus be applied/extended straightforwardly to the other algorithms. To lay the groundwork for this generalization, this section lists essential characteristics of several representative greedy algorithms, and provides the basic ideas for extending them. Note that OMP and OLS are the two classic frameworks for sparse recovery algorithms, and T-GBOMP is developed based on the OMP framework. We first analyze the relationship between T-GBOMP and other OMP-framework-based algorithms.

1) *OMP Framework-Based Algorithms*: Sparse recovery in vector form, i.e., the SMV model, is regarded as the 1-mode sparse tensor reconstruction. OMP is one of the most classical 1-mode recovery algorithms, which selects only one most correlated index per iteration [19]. Thus, T-GBOMP degenerates into OMP directly by setting the block lengths, the number of multiple supports selected and the number of modes as 1. The block version variant of OMP, i.e., BOMP [13], can be obtained by setting the number of multiple supports selected and the number of the modes in T-GBOMP to 1. CoSaMP is another representative algorithm that selects multiple entries per iteration, but only retains the most correlated nonzero atoms, the number of which is equal to the sparsity [35]. This mechanism is equivalent to that of T-GBOMP by setting the selection parameter as sparsity and the block length as 1. Furthermore, T-GBOMP reduces to the two high-dimensional tensor recovery algorithms, T-OMP [8] and T-BOMP [9], by setting the block length and the selection parameter to 1, and only setting the selection parameter to 1, respectively. Note that the sparse recovery based on the MMV model can also be regarded as the 1-mode tensor recovery problem. The mechanisms of other OMP-framework-based algorithms are similar and will not be elaborated here.

2) *OLS Framework-Based Algorithms*: The OLS and the OMP frameworks differ in the way of selecting the new support entry [18], [38]. More specifically, OLS chooses a column that minimizes the power of the new residual. It can be shown that compared with the OMP framework, the OLS framework exhibits better convergence behavior, at the expense of imposing higher computational complexity [36]. Mathematically, the main difference between the OMP-framework-based algorithm and the OLS-framework-based algorithm is that there exists a normalization factor in the selection mechanism of the OLS-framework-based algorithm [6]. For example, the following is the selection mechanism of OLS [6]:

$$i^{l+1} = \arg \max_{j \in \{1, \dots, N\} \setminus \Omega^l} \left\langle \frac{\mathbf{D}_{(:,j)}}{\|\mathbf{P}_{\Omega^l}^\perp \mathbf{D}_{(:,j)}\|_2}, \mathbf{r}^l \right\rangle, \quad (25)$$

where  $\|\mathbf{P}_{\Omega^l}^\perp \mathbf{D}_{(:,j)}\|_2$  is a normalization factor. As illustrated in existing works, e.g., [6] and [18], (25) can be lower and upper bounded, providing a means of applying the theoretical results

derived in the next section to OLS-framework-based algorithms. Similarly, popular OLS framework-based approaches, such as MOLS [36] and BOLS [6], have incorporated useful iterative mechanisms, resulting in desirable performance and complexity.

In a nutshell, OMP- and OLS-framework-based algorithms can be extended into the tensor case, and T-GBOMP acts as a generalized framework for these greedy algorithms. The analyses based on T-GBOMP in the sequel can thus serve as a theoretical basis for tensor greedy algorithms in general, and can be easily applied to the many existing algorithms by setting certain parameters to specific values as mentioned above.

#### IV. MAIN RESULTS

We now present our main results, namely, conditions that guarantee T-GBOMP can recover a block-sparse tensor from measurements. This section is divided into two parts. In the first part, an analysis under the noiseless case is conducted, including development of the ERC and reconstructible sparsity. In the second part, we present conditions for reliable recovery in the noisy scenario, including discussions of the difference between the original tensor and the estimated tensor, the residual convergence with respect to both the measurements and noise, and a lower bound of the SNR required for reliable recovery.

##### A. Noiseless Recovery Conditions

An exact recovery condition was first developed for OMP in the work [17], after which several ERCs applicable to various other algorithms, such as BOMP [10] and BOLS [6], were derived. These ERCs reveal that, if the measurement matrix satisfies certain conditions, then the corresponding algorithms will select all the correct supports during their sparsity iterations. The conditions apply only to 1-mode tensor recovery, and they are not suitable for the generalized  $n$ -mode tensor reconstruction scenarios, as discussed in Subsection III-B. Our first result is an ERC for an  $n$ -mode tensor recovery algorithm T-GBOMP to solve the exact sparse problem, which is a sufficient condition for T-GBOMP to identify the representation of the input tensor that exploits the least number of tensor blocks, i.e., the sparsest one.

*Theorem 1*: Let  $\mathcal{X}$  be a  $k$  block-sparse tensor, and for a given measurement matrix set  $\Upsilon = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ), let  $\mathcal{Y} = \mathcal{X} \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2 \times_3 \cdots \times_n \mathbf{D}_n$ . A sufficient condition for T-GBOMP to reconstruct  $\mathcal{X}$  is

$$\frac{Z_{\mathcal{R}^l}}{\sqrt{s}} \|\ddot{\mathbf{D}}_{\Xi}^\dagger \ddot{\mathbf{D}}_{\Psi}\|_2 < 1, \quad (26)$$

where

$$Z_{\mathcal{R}^l} = \frac{\|\ddot{\mathbf{D}}_{\Xi}^\dagger \text{vec}(\mathcal{R}^l)\|_F}{\|(\mathbf{D}_{n_{[i_n^*]}^\dagger}^\dagger \otimes \cdots \otimes \mathbf{D}_{1_{[i_1^*]}^\dagger}^\dagger) \text{vec}(\mathcal{R}^l)\|_F}, \quad (27)$$

$\Xi$  is the tensor index set corresponding to the nonzero supports of  $\mathcal{X}$ ,  $\Psi$  denotes the index set containing  $s$  largest elements in  $\{\|\mathcal{R}^l \times_1 \mathbf{D}_{1_{[i_1]}^\dagger}^\dagger \times_2 \cdots \times_n \mathbf{D}_{n_{[i_n]}^\dagger}^\dagger\|_F\}_{(i_1, i_2, \dots, i_n) \in \Omega \setminus \Xi}$ , and  $(i_1^*, \dots, i_n^*)$  is the tensor index that leads to the largest element in the set  $\{\|\mathcal{R}^l \times_1 \mathbf{D}_{1_{[i_1]}^\dagger}^\dagger \times_2 \cdots \times_n \mathbf{D}_{n_{[i_n]}^\dagger}^\dagger\|_F\}_{(i_1, i_2, \dots, i_n) \in \Xi}$ .

*Proof:* See Appendix E. ■

*Remark 2:* Note that  $Z_{\mathcal{R}^l}$  in (27) is bounded as  $1 \leq Z_{\mathcal{R}^l} \leq \sqrt{k}$ . The lower bound can be obtained by assuming that  $\mathbf{D}_{\Xi}^H$  is an all zero matrix except for the submatrix indexed by  $(i_1^*, \dots, i_n^*)$ , while the upper bound is derived by letting  $\mathbf{D}_{\Xi}^H$  be a cascade of  $k$   $\mathbf{D}_{n_{[i_1^*]}}^H \otimes \dots \otimes \mathbf{D}_{1_{[i_1^*]}}^H$  matrices. The ERC in (26) can thus be reformulated in terms of the least restricted variant and the most restricted variant, i.e.,

$$\frac{1}{\sqrt{s}} \|\mathbf{D}_{\Xi}^{\dagger} \mathbf{D}_{\Psi}\|_2 < 1, \quad (28)$$

$$\|\mathbf{D}_{\Xi}^{\dagger} \mathbf{D}_{\Psi}\|_2 < 1, \quad (29)$$

respectively. Compared with the existing results in [10], [17], and [6], (28) provides progress due to the existing of  $\frac{1}{\sqrt{s}}$ . For instance, the ERC for BOMP with a 1-mode tensor is [10]

$$\rho_c(\mathbf{D}_{\Xi}^{\dagger} \mathbf{D}_{\Omega \setminus \Xi}) < 1, \quad (30)$$

where  $\rho_c(\cdot)$  is the mixed matrix norm defined in **Definition 6**, and  $\mathbf{D} \in \mathbb{C}^{M \times N}$  is the 1-mode measurement matrix. It is worth noting that in most cases  $|\Omega \setminus \Xi| > |\Psi| = s$ . Hence  $\|\mathbf{D}_{\Xi}^{\dagger} \mathbf{D}_{\Psi}\|_2 \leq \rho_c(\mathbf{D}_{\Xi}^{\dagger} \mathbf{D}_{\Omega \setminus \Xi})$  under the reasonable assumption that the dimensions of measurement matrices are equivalent, i.e.,  $M = \prod_{t=1}^n M_t$  and  $N = \prod_{t=1}^n N_t$ . Since  $\|\mathbf{D}_{\Xi}^{\dagger} \mathbf{D}_{\Psi}\|_2$  in (28) is further multiplied by a parameter  $\frac{1}{\sqrt{s}} \leq 1$ , we obtain that the ERCs in (28) and (29) are less restrictive than the one in (30). Comparisons to other existing ERCs are similar. Thus, we can conclude that the T-GBOMP algorithm is more capable of achieving exact sparse recovery than its 1-mode counterparts.

Clearly (26) would not be very useful without the help of a method for checking if the ERC is satisfied. To this end, we present an intuitive result in terms of the sparsity bound required for the establishment of (26), named *reconstructible sparsity*, by exploiting the MIP. Note that in this subsection up to Theorem 3, we assume that  $\Xi \cup \Psi = \emptyset$  for intuitive results, which facilitates clear comparisons with existing results. The outcomes when  $\Xi \cup \Psi \neq \emptyset$  can be easily obtained based on the analyses conducted in this subsection.

*Theorem 2:* A sufficient condition for the establishment of (26) is

$$k < \frac{\sqrt{s} \left( 1 - \left( \prod_{t=1}^n d_t - 1 \right) \tau_{\mathbf{Y}}^n + \varpi_{\mathbf{Y}}^n \prod_{t=1}^n d_t \right)}{(Z_{\mathcal{R}^l} + \sqrt{s}) \varpi_{\mathbf{Y}}^n \prod_{t=1}^n d_t}. \quad (31)$$

*Proof:* See Appendix F. ■

It can be observed that **Theorem 2** provides an upper bound on the sparsity level, under which the T-GBOMP performs exact tensor recovery, i.e., the ERC in (26) holds. Based on *Remark 2*, we obtain the least restricted reconstructible sparsity by setting  $Z_{\mathcal{R}^l} = 1$ , as follows:

$$k < \frac{\sqrt{s} \left( 1 - \left( \prod_{t=1}^n d_t - 1 \right) \tau_{\mathbf{Y}}^n + \varpi_{\mathbf{Y}}^n \prod_{t=1}^n d_t \right)}{(1 + \sqrt{s}) \varpi_{\mathbf{Y}}^n \prod_{t=1}^n d_t}. \quad (32)$$

In the subsequent remark, we provide some useful comparisons between (32) and existing criteria.

*Remark 3:* Since the existing reconstructible sparsity levels are developed for 1-mode tensor recovery, we set  $n = 1$  in (32)

for fair comparison. Then, (32) is reformulated into

$$kd < \frac{\sqrt{s}d(1 - (d-1)\nu_{\mathbf{D}} + \mu_{\mathbf{D}}d)}{(1 + \sqrt{s})\mu_{\mathbf{D}}d}. \quad (33)$$

The classic reconstructible sparsity bounds in [10] and [17] are

$$K < \frac{1}{2} \left( \frac{1}{\mu_{\mathbf{D}}} + 1 \right), \quad (34)$$

$$kd < \frac{1}{2} \left( \frac{1}{\mu_{\mathbf{D}}} + d - \frac{d\nu_{\mathbf{D}}}{\mu_{\mathbf{D}}} + \frac{\nu_{\mathbf{D}}}{\mu_{\mathbf{D}}} \right), \quad (35)$$

respectively, where  $K = kd$  represents the total sparsity. Firstly, let  $d = 1$  in (33), and (33) thus becomes

$$K < \frac{\sqrt{s}}{1 + \sqrt{s}} \left( \frac{1}{\mu_{\mathbf{D}}} + 1 \right). \quad (36)$$

Since  $1 \leq s \leq k$ , we have  $\frac{\sqrt{s}}{1 + \sqrt{s}} \geq \frac{1}{2}$ . This indicates that when  $s = 1$ , (36) converges to (34), which is applicable for the conventional OMP, while in the case  $s > 1$ , the upper bound in (36) is better than that in (34), which reveals that T-GBOMP is capable of exactly recovering tensors with larger sparsity levels in the scenarios where  $Z_{\mathcal{R}^l} = 1$ . However, the condition  $Z_{\mathcal{R}^l} = 1$  is not achievable in practice, since the measurement matrix should be full rank for correct atom selection. Despite that the setting  $Z_{\mathcal{R}^l} = 1$  is impractical, the bound in (36) provides a non-negligible improvement over the existing results, unveiling possible potential performance gain on tensor recovery by T-GBOMP. Secondly, after some simplification operations, (33) becomes

$$kd < \frac{\sqrt{s}}{1 + \sqrt{s}} \left( \frac{1}{\mu_{\mathbf{D}}} + d - \frac{d\nu_{\mathbf{D}}}{\mu_{\mathbf{D}}} + \frac{\nu_{\mathbf{D}}}{\mu_{\mathbf{D}}} \right). \quad (37)$$

A similar conclusion can be made that the upper bound of reconstructible sparsity in (37) is equivalent to or better than that in (35).

Now, considering the most restricted case of **Theorem 2**, i.e.,  $Z_{\mathcal{R}^l} = \sqrt{k}$ , we provide the following corollary. Note that **Corollary 4** replaces the uncertainty factor  $Z_{\mathcal{R}^l}$  in (31) with a complex expression.

*Corollary 4:* If the following inequality

$$k < \left( \sqrt[3]{-\frac{Q}{2} + \sqrt{\Delta}} + \sqrt[3]{-\frac{Q}{2} - \sqrt{\Delta}} - \frac{\sqrt{s}}{3} \right)^2 \quad (38)$$

is satisfied, the ERC in (26) holds, where  $P = -\frac{s}{3}$ ,  $Q = \frac{27\delta + 2s\frac{3}{2}}{27}$ ,  $\Delta = \left(\frac{Q}{2}\right)^2 + \left(\frac{P}{3}\right)^3$ , and

$$\delta = -\frac{\sqrt{s} \left( 1 - \left( \prod_{t=1}^n d_t - 1 \right) \tau_{\mathbf{Y}}^n + \varpi_{\mathbf{Y}}^n \prod_{t=1}^n d_t \right)}{\varpi_{\mathbf{Y}}^n \prod_{t=1}^n d_t}. \quad (39)$$

*Proof:* See Appendix G. ■

*Remark 4:* We demonstrate that the most restricted reconstructible sparsity presented in **Corollary 4** can be superior to the existing results (34) and (35). For reliable recovery in CS, the matrix formed by the measurement matrix blocks corresponding to the nonzero indices of the sparse tensor must be of full rank, which has been stated in Remark 3. This indicates that  $Z_{\mathcal{R}^l}$  can be further bounded by  $Z_{\mathcal{R}^l} < \sqrt{k}$ . Since (38) that provides a complex reconstructible sparsity in closed-form

is derived from (31), we come to the inequality (31) for an intuitive explanation. Let  $Z_{\mathcal{R}^l} = \sqrt{k} - \theta$ , where  $\theta > 0$  is a scalar, and we have

$$k < \frac{\sqrt{s} \left( 1 - \left( \prod_{t=1}^n d_t - 1 \right) \tau_{\mathbf{Y}}^n + \varpi_{\mathbf{Y}}^n \prod_{t=1}^n d_t \right)}{(\sqrt{k} - \theta + \sqrt{s}) \varpi_{\mathbf{Y}}^n \prod_{t=1}^n d_t}, \quad (40)$$

which is the most restricted initial inequality for exporting the closed-form upper bound (38). By selecting a feasible  $s$  value as  $k$ , (40) changes into

$$k < \frac{\left( 1 - \left( \prod_{t=1}^n d_t - 1 \right) \tau_{\mathbf{Y}}^n + \varpi_{\mathbf{Y}}^n \prod_{t=1}^n d_t \right)}{\left( 2 - \frac{\theta}{\sqrt{k}} \right) \varpi_{\mathbf{Y}}^n \prod_{t=1}^n d_t}.$$

For the case of  $d = 1$  and  $n = 1$ , and for the case of  $d > 1$  and  $n = 1$ , (40) yields respectively

$$K < \frac{1}{2 - \frac{\theta}{\sqrt{k}}} \left( \frac{1}{\mu_{\mathbf{D}}} + 1 \right), \quad (41)$$

$$kd < \frac{1}{2 - \frac{\theta}{\sqrt{k}}} \left( \frac{1}{\mu_{\mathbf{D}}} + d - \frac{d\nu_{\mathbf{D}}}{\mu_{\mathbf{D}}} + \frac{\nu_{\mathbf{D}}}{\mu_{\mathbf{D}}} \right), \quad (42)$$

where  $K = kd = k \times 1$  implies the total sparsity in (41). Since  $\frac{1}{2 - \frac{\theta}{\sqrt{k}}} > \frac{1}{2}$ , the upper bounds in (41) and (42) provide an intuitive improvement compared to those in (34) and (35). This indicates that our derived reconstructible sparsity can be better than those in (34) and (35) even if the most restricted constraint is supposed in (40), which completes our demonstration.

*Remark 5:* By using the bound of (19) and based on the proof in Appendix F, we obtain the following conditions for reconstructible sparsities of the least and most restricted:

$$k < \left( -\frac{1}{2} + \sqrt{\frac{5}{4} - \frac{\left( \prod_{t=1}^n d_t - 1 \right) \tau_{\mathbf{Y}}^n - 1}{\varpi_{\mathbf{Y}}^n \prod_{t=1}^n d_t}} \right)^2,$$

$$k < \frac{\left( \frac{1 - \left( \prod_{t=1}^n d_t - 1 \right) \tau_{\mathbf{Y}}^n}{\varpi_{\mathbf{Y}}^n \prod_{t=1}^n d_t} \right)^2}{2 + \frac{2}{\varpi_{\mathbf{Y}}^n \prod_{t=1}^n d_t}} + 1 + \frac{2}{\varpi_{\mathbf{Y}}^n \prod_{t=1}^n d_t},$$

where  $\theta$  is similarly defined in Remark 4 which satisfies  $\theta > 0$ .

In the following theorem, we study the rate of convergence of T-GBOMP in the noiseless scenario, which provides useful insights, in particular, revealing that the residual power of T-GBOMP decreases exponentially as the number of iterations increases.

*Theorem 3:* Given a matrix set  $\mathbf{Y} = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ), if the MIP satisfies  $\underline{W}_{\mathbf{Y}, \prod_{t=1}^n k_n + s_k} > 0$ , then the residual of T-GBOMP satisfies

$$\|\mathcal{R}^{l+1}\|_F^2 \leq \left( 1 - \frac{s \underline{W}_{\mathbf{Y}, \prod_{t=1}^n k_n + s_l}}{k \overline{W}_{\mathbf{Y}, \prod_{t=1}^n k_n + s_l} \overline{W}_{\mathbf{Y}, s}} \right)^{l+1} \|\mathcal{Y}\|_F^2.$$

*Proof:* See Appendix H. ■

### B. Noisy Recovery Conditions

In practice, the tensor measurements are corrupted by noise. This subsection presents several important results guaranteeing the reliable recovery of the T-GBOMP algorithm with the help of the defined MIP concepts. As illustrated in Algorithm 1,

T-GBOMP stops the iteration procedure when  $l = k$  or  $\|\mathcal{R}^l\|_2 \leq \epsilon$ . The following theorem reveals an upper bound on the error between the original and the estimated tensors when T-GBOMP is terminated.

*Theorem 4:* Given a matrix set  $\mathbf{Y} = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ), if the iterative procedure of T-GBOMP satisfies  $\|\mathcal{R}^{l^*}\|_F \leq \epsilon$ ,  $\underline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t + k} > 0$  and  $\underline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t + s_{l^*}} > 0$  for  $l^* \leq k$ , then the following inequality with respect to  $\hat{\mathcal{X}}$  holds:

$$\|\mathcal{X} - \hat{\mathcal{X}}\|_F \leq \frac{2\overline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t + k}^{\frac{1}{2}} \epsilon}{\underline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t + k}^{\frac{1}{2}} \underline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t + s_{l^*}}^{\frac{1}{2}}} + \frac{2 \left( \overline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t + k}^{\frac{1}{2}} + \underline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t + s_{l^*}}^{\frac{1}{2}} \right) \|\mathcal{N}\|_F}{\underline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t + k}^{\frac{1}{2}} \underline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t + s_{l^*}}^{\frac{1}{2}}}, \quad (43)$$

where  $\overline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t + k}$ ,  $\underline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t + s_{l^*}}$  and  $\underline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t + k}$  can be obtained by Corollary 3.

*Proof:* See Appendix I. ■

Even if Theorem 4 is illustrated mainly based on the defined MIP framework, it seems to be complex. We now present intuitive explanations of Theorem 4 for adding the understanding of the theoretical reconstruction error. To begin with, we provide Proposition 2 that describes the lower and upper bounds of the multiplication of the shadow block sparsity.

*Proposition 2:* Given a  $k$  block-sparse tensor with  $n$  modes, its shadow block sparsity  $\{k_i\}$  ( $i \in \{1, 2, \dots, n\}$ ) satisfies

$$k \leq k_1 \times k_2 \times \dots \times k_n \leq k^n, \quad (44)$$

The proof of Proposition 2 is omitted, since it can be derived by considering the spatial geometric characteristic of the  $k$  nonzero tensor blocks. The lower bound in (44) can be achieved by assuming that the  $k$  nonzero tensor blocks are arranged regularly along one of the tensor modes, e.g., the  $t$ th ( $t \in \{1, 2, \dots, n\}$ ) tensor mode, resulting in  $k_t = k$ , while the remaining shadow block sparsity is equal to 1. On the other hand, if the  $k$  nonzero tensor blocks are arranged unobstructed throughout the entire tensor, the upper bound in (44) can be obtained with  $k_i = k$  ( $i \in \{1, 2, \dots, n\}$ ).

Based on Proposition 2 and the aforementioned discussions, the following remark presents results investigating some particular tensor structures for Theorem 4, which offer intuitive relationships between reconstruction error and MIP.

*Remark 6:* Considering that the tensor structure achieves the lower bound in Proposition 2 with  $d_1 = d_2 = \dots = d_n = d$ ,  $l^*$  satisfies  $\frac{k}{s} \leq l^* \leq k$  and  $\epsilon = \|\mathcal{N}\|_F$ , then (43) changes into

$$\|\mathcal{X} - \hat{\mathcal{X}}\|_F \leq \frac{\left( 4\overline{W}_{\mathbf{Y}, 2k(k+1)^{n-1}}^{\frac{1}{2}} + 2\underline{W}_{\mathbf{Y}, 2k(k+1)^{n-1}}^{\frac{1}{2}} \right) \|\mathcal{N}\|_F}{\underline{W}_{\mathbf{Y}, 2k(k+1)^{n-1}}^{\frac{1}{2}} \underline{W}_{\mathbf{Y}, (s+1)k(sk+1)^{n-1}}^{\frac{1}{2}}} \leq \frac{2\sqrt{10} \|\mathcal{N}\|_F}{1 - \sum_{t=0}^{n-1} \underline{C}_t^{d^n} \tau_{\mathbf{Y}, t}^{n-t} - \sum_{t=0}^{n-1} \underline{C}_t^{(s+1)k(sk+1)^{n-1}} \varpi_{\mathbf{Y}, t}^n d^n}, \quad (45)$$

where (45) is because  $\overline{W}_{\mathbf{r},x}$  and  $\underline{W}_{\mathbf{r},x}$  are monotonically increasing and decreasing functions of the variable  $x$ , respectively, and (46) follows from  $4\overline{W}_{\mathbf{r},2k(k+1)^{n-1}}^{\frac{1}{2}} + 2\underline{W}_{\mathbf{r},2k(k+1)^{n-1}}^{\frac{1}{2}} \leq 2\sqrt{10}$  and  $2k(k+1)^{n-1} \leq (s+1)k(sk+1)^{n-1}$ . Clearly, the upper bound of the reconstruction error  $\|\mathcal{X} - \hat{\mathcal{X}}\|_F$  decreases with the reductions in block sparsity, mutual block coherence and mutual sub-coherence, indicating an improved recovery performance. Compared with the existing results mainly based on the NP-hard RIP metric, e.g., [39] and [40], the derived bound in (46) provides a computable theoretical reconstruction error, making it more practically applicable.

Let us come to the tensor structure achieving the upper bound in **Proposition 2**, which seems to offer the most restricted results. Similar to the assumption and derivation of (46), the reconstruction error in **Theorem 4** can be further bounded by

$$\begin{aligned} & \|\mathcal{X} - \hat{\mathcal{X}}\|_F \\ & \leq \frac{\left(4\overline{W}_{\mathbf{r},2^nk^n}^{\frac{1}{2}} + 2\underline{W}_{\mathbf{r},2^nk^n}^{\frac{1}{2}}\right)\|\mathcal{N}\|_F}{\underline{W}_{\mathbf{r},2^nk^n}^{\frac{1}{2}}\underline{W}_{\mathbf{r},(s+1)^nk^n}^{\frac{1}{2}}} \\ & \leq \frac{2\sqrt{10}\|\mathcal{N}\|_F}{1 - \sum_{t=0}^{n-1} \underline{C}_t^{dn} \tau_{\mathbf{r},t}^{n-t} - \sum_{t=0}^{n-1} C_t^{(s+1)^nk^n} \varpi_{\mathbf{r},t}^n d^n}. \end{aligned}$$

Define the following two functions:

$$\begin{aligned} f(\varpi_{\mathbf{r}}, \tau_{\mathbf{r}}) &= \frac{2\sqrt{10}\|\mathcal{N}\|_F}{1 - \sum_{t=0}^{n-1} \underline{C}_t^{dn} \tau_{\mathbf{r},t}^{n-t} - \sum_{t=0}^{n-1} C_t^{(s+1)k(sk+1)^{n-1}} \varpi_{\mathbf{r},t}^n d^n}, \\ g(\varpi_{\mathbf{r}}, \tau_{\mathbf{r}}) &= \frac{2\sqrt{10}\|\mathcal{N}\|_F}{1 - \sum_{t=0}^{n-1} \underline{C}_t^{dn} \tau_{\mathbf{r},t}^{n-t} - \sum_{t=0}^{n-1} C_t^{(s+1)^nk^n} \varpi_{\mathbf{r},t}^n d^n}. \end{aligned}$$

Then, we obtain

$$\begin{aligned} & \frac{g(\varpi_{\mathbf{r}}, \tau_{\mathbf{r}})}{f(\varpi_{\mathbf{r}}, \tau_{\mathbf{r}})} \\ &= 1 + \frac{\sum_{t=0}^{n-1} (C_t^{(s+1)k(sk+1)^{n-1}} - C_t^{(s+1)^nk^n}) \varpi_{\mathbf{r},t}^n d^n}{1 - \sum_{t=0}^{n-1} \underline{C}_t^{dn} \tau_{\mathbf{r},t}^{n-t} - \sum_{t=0}^{n-1} C_t^{(s+1)^nk^n} \varpi_{\mathbf{r},t}^n d^n} \\ &= 1 + \Delta^*, \end{aligned} \quad (47)$$

where  $\Delta^* = \frac{\sum_{t=0}^{n-1} (C_t^{(s+1)k(sk+1)^{n-1}} - C_t^{(s+1)^nk^n}) \varpi_{\mathbf{r},t}^n d^n}{1 - \sum_{t=0}^{n-1} \underline{C}_t^{dn} \tau_{\mathbf{r},t}^{n-t} - \sum_{t=0}^{n-1} C_t^{(s+1)^nk^n} \varpi_{\mathbf{r},t}^n d^n}$ . For (47), we have

$$\begin{cases} \frac{\partial \Delta^*}{\partial \varpi_{\mathbf{r}}} \geq 0, & (48a) \\ \frac{\partial \Delta^*}{\partial \tau_{\mathbf{r}}} \geq 0, & (48b) \\ \frac{\partial \Delta^*}{\partial k} \leq 0. & (48c) \end{cases}$$

These results provide intuitive theoretical analyses on the question how can the general block-sparse tensor structure approaches the tensor structure that achieves the lower bound in (44), with respect to several key parameters in block-sparse

tensor recovery. Note that in order to achieve the approximation of the tensor structure mentioned above, the difference  $\Delta^*$  should be sufficiently small. Specially, the partial derivatives derived in (48a)-(48b) indicate that small coherence in terms of mutual block coherence and mutual sub-coherence can contribute to the desirable block-sparse tensor structure, i.e., achieving the lower bound in (44). The result (48c) indicates that when the block sparsity  $k$  is sufficiently small, the reconstruction error can be close to that of the satisfactory tensor structure mentioned above. Furthermore, when the total sparsity  $K = kd$  is fixed, we have  $\frac{\partial \Delta^*}{\partial d} \leq 0$ . This unveils an important conclusion related to the block structure characteristic of tensor recovery, i.e., the larger the block length, the closer the reconstruction error compared to that of the desirable tensor structure. It indicates that a stronger block structure can result in better recovery performance.

In the noisy scenario,  $\epsilon$  should be set appropriately to avoid late or early termination of the algorithm [36]. It is worth noting that early cessation leads to an insufficient recovery of the tensor, while stopping too late results in the tensor being disturbed by more noise, i.e., over-fitting. Note that in **Theorem 4**, we consider the case where  $\|\mathcal{R}^{l^*}\|_F \leq \epsilon$  for  $l^* \leq k$ . As given in Algorithm 1, T-GBOMP is forced to iterate for  $k$  iterations. However, the scenario where  $\|\mathcal{R}^{l^*}\|_F \leq \epsilon$  is not met when  $l^* \leq k$  naturally occurs since  $\epsilon$  is a preset parameter. The following theorem considers this scenario, and provides a recovery condition demonstrating that the algorithm can perform reliable recovery in at most  $k$  iterations.

**Theorem 5:** Given a matrix set  $\mathbf{r} = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ), suppose that  $\|\mathcal{R}^{l^*}\|_F \leq \epsilon$  is not met for  $l^* \leq k$ . If  $\underline{W}_{\mathbf{r},ks} > 0$ ,  $\underline{W}_{\mathbf{r},k^n} > 0$ , and the SNR satisfies

$$\begin{cases} \text{SNR} > \left( \sqrt{k} \overline{W}_{\mathbf{r},s}^{\frac{1}{2}} \overline{W}_{\mathbf{r},k}^{\frac{1}{2}} + \sqrt{ks} \overline{W}_{\mathbf{r},ks}^{\frac{1}{2}} \overline{W}_{\mathbf{r},k}^{\frac{1}{2}} \right)^2 \\ \quad \times \left( \left( \frac{\sqrt{s}}{\sqrt{k}} \underline{W}_{\mathbf{r},ks}^{\frac{1}{2}} \underline{W}_{\mathbf{r},k}^{\frac{1}{2}} - k \varpi_{\mathbf{r}} d_1 \right. \right. \\ \quad \left. \left. - \frac{(ks \varpi_{\mathbf{r}} d_1)^2}{\underline{W}_{\mathbf{r},ks}} \right) \text{MAR}_* \right)^{-2}, & n = 1, \quad (49a) \\ \text{SNR} > \left( \sqrt{k} \overline{W}_{\mathbf{r},s}^{\frac{1}{2}} \overline{W}_{\mathbf{r},\prod_{t=1}^n k_t}^{\frac{1}{2}} \right. \\ \quad \left. + \sqrt{ks} \overline{W}_{\mathbf{r},ks}^{\frac{1}{2}} \overline{W}_{\mathbf{r},\prod_{t=1}^n k_t}^{\frac{1}{2}} \right)^2 \\ \quad \times \left( \left( \frac{\sqrt{s}}{\sqrt{k}} \underline{W}_{\mathbf{r},ks}^{\frac{1}{2}} \underline{W}_{\mathbf{r},k^n}^{\frac{1}{2}} - k^n \varpi_{\mathbf{r},n-1} \prod_{t=1}^n d_t \right. \right. \\ \quad \left. \left. - \frac{k^{n+1} s \varpi_{\mathbf{r},n-1}^2 (\prod_{t=1}^n d_t)^2}{\underline{W}_{\mathbf{r},ks}} \right) \text{MAR}_* \right)^{-2}, & n > 1, \quad (49b) \end{cases}$$

then T-GBOMP selects all the support indices in at most  $k$  iterations.

*Proof:* See Appendix J.  $\blacksquare$

**Theorem 5** may not be intuitive in revealing the connection between various MIP metrics and the necessary SNR level for reliable recovery. To this end, the following remark is provided for a detailed descriptions.

*Remark 7:* Based on (20) in **Lemma 2**, (49a) and (49b) in **Theorem 5** can be integrated into the following inequality:

$$\begin{aligned} \sqrt{\text{SNR}} &> \left( \frac{1}{\sqrt{s}} \overline{W}_{\mathbf{Y},s}^{\frac{1}{2}} + \overline{W}_{\mathbf{Y},ks}^{\frac{1}{2}} \right) \sqrt{k} \overline{W}_{\mathbf{Y},\prod_{t=1}^n k_t}^{\frac{1}{2}} \\ &\times \left( \left( \frac{1}{\sqrt{k}} \overline{W}_{\mathbf{Y},ks}^{\frac{1}{2}} \overline{W}_{\mathbf{Y},k^n}^{\frac{1}{2}} - \frac{s \varpi_{\mathbf{Y},n-1}^{2n} k^{\frac{n}{2}+1} \left( \prod_{t=1}^n d_t \right)^2}{\overline{W}_{\mathbf{Y},ks}} \right. \right. \\ &\quad \left. \left. - \sqrt{k^n} \varpi_{\mathbf{Y},n-1}^n \prod_{t=1}^n d_t \right) \text{MAR}_* \right)^{-1}. \end{aligned} \quad (50)$$

Then, letting  $d_1 = d_2 = \dots = d_n = d$ , (50) changes into

$$\begin{aligned} &\sqrt{\text{SNR}} \\ &> \left( \frac{1}{\sqrt{s}} \sqrt{1 + \sum_{l=0}^{n-1} \underline{C}_l^{d^n} \tau_{\mathbf{Y},l}^{n-l} + \sum_{l=0}^{n-1} C_l^s \varpi_{\mathbf{Y},l}^n d^n} \right. \\ &\quad \left. + \sqrt{1 + \sum_{l=0}^{n-1} \underline{C}_l^{d^n} \tau_{\mathbf{Y},l}^{n-l} + \sum_{l=0}^{n-1} C_l^{ks} \varpi_{\mathbf{Y},l}^n d^n} \right) \\ &\times \sqrt{k} \sqrt{1 + \sum_{l=0}^{n-1} \underline{C}_l^{d^n} \tau_{\mathbf{Y},l}^{n-l} + \sum_{l=0}^{n-1} C_l^{\prod_{t=1}^n k_t} \varpi_{\mathbf{Y},l}^n d^n} \\ &\times \left( \left( \frac{1}{\sqrt{k}} \left( 1 - \sum_{l=0}^{n-1} \underline{C}_l^{d^n} \tau_{\mathbf{Y},l}^{n-l} - \sum_{l=0}^{n-1} C_l^{ks} \varpi_{\mathbf{Y},l}^n d^n \right)^{\frac{1}{2}} \right. \right. \\ &\quad \left. \times \left( 1 - \sum_{l=0}^{n-1} \underline{C}_l^{d^n} \tau_{\mathbf{Y},l}^{n-l} - \sum_{l=0}^{n-1} C_l^{k^n} \varpi_{\mathbf{Y},l}^n d^n \right)^{\frac{1}{2}} \right. \\ &\quad \left. - \frac{s \varpi_{\mathbf{Y},n-1}^{2n} k^{\frac{n}{2}+1} d^{2n}}{1 - \sum_{l=0}^{n-1} \underline{C}_l^{d^n} \tau_{\mathbf{Y},l}^{n-l} - \sum_{l=0}^{n-1} C_l^{ks} \varpi_{\mathbf{Y},l}^n d^n} \right. \\ &\quad \left. - k^{\frac{n}{2}} \varpi_{\mathbf{Y},n-1}^n d^n \right) \text{MAR}_* \Big)^{-1}. \end{aligned} \quad (51)$$

Since the prerequisites that  $\overline{W}_{\mathbf{Y},ks} > 0$  and  $\overline{W}_{\mathbf{Y},k^n} > 0$  are given in **Theorem 5**, we have

$$\sum_{l=0}^{n-1} \underline{C}_l^{d^n} \tau_{\mathbf{Y},l}^{n-l} + \sum_{l=0}^{n-1} C_l^{ks} \varpi_{\mathbf{Y},l}^n d^n < 1, \quad (52)$$

$$\sum_{l=0}^{n-1} \underline{C}_l^{d^n} \tau_{\mathbf{Y},l}^{n-l} + \sum_{l=0}^{n-1} C_l^{k^n} \varpi_{\mathbf{Y},l}^n d^n < 1. \quad (53)$$

Furthermore, (52) and (53) indicate that

$$\sum_{l=0}^{n-1} \underline{C}_l^{d^n} \tau_{\mathbf{Y},l}^{n-l} + \sum_{l=0}^{n-1} C_l^s \varpi_{\mathbf{Y},l}^n d^n < 1, \quad (54)$$

$$\sum_{l=0}^{n-1} \underline{C}_l^{d^n} \tau_{\mathbf{Y},l}^{n-l} + \sum_{l=0}^{n-1} C_l^{\prod_{t=1}^n k_t} \varpi_{\mathbf{Y},l}^n d^n < 1, \quad (55)$$

respectively. Thus, as for the right-hand side in (51), (52)-(55) lead to the following limits:

$$\lim_{n \rightarrow \infty} \sum_{l=0}^{n-1} \underline{C}_l^{d^n} \tau_{\mathbf{Y},l}^{n-l} + \sum_{l=0}^{n-1} C_l^{ks} \varpi_{\mathbf{Y},l}^n d^n = 0, \quad (56)$$

$$\lim_{n \rightarrow \infty} \sum_{l=0}^{n-1} \underline{C}_l^{d^n} \tau_{\mathbf{Y},l}^{n-l} + \sum_{l=0}^{n-1} C_l^{k^n} \varpi_{\mathbf{Y},l}^n d^n = 0, \quad (57)$$

$$\lim_{n \rightarrow \infty} \sum_{l=0}^{n-1} \underline{C}_l^{d^n} \tau_{\mathbf{Y},l}^{n-l} + \sum_{l=0}^{n-1} C_l^s \varpi_{\mathbf{Y},l}^n d^n = 0, \quad (58)$$

$$\lim_{n \rightarrow \infty} \sum_{l=0}^{n-1} \underline{C}_l^{d^n} \tau_{\mathbf{Y},l}^{n-l} + \sum_{l=0}^{n-1} C_l^{\prod_{t=1}^n k_t} \varpi_{\mathbf{Y},l}^n d^n = 0, \quad (59)$$

$$\lim_{n \rightarrow \infty} k^{\frac{n}{2}+1} \varpi_{\mathbf{Y},n-1}^{2n} d^{2n} = 0, \quad (60)$$

$$\lim_{n \rightarrow \infty} k^{\frac{n}{2}} \varpi_{\mathbf{Y},n-1}^n d^n = 0. \quad (61)$$

The condition of SNR presented in (51) can be reformulated intuitively based on the aforementioned limits, i.e.,

$$\lim_{n \rightarrow \infty} \sqrt{\text{SNR}} > \frac{\left( \frac{1}{\sqrt{s}} + 1 \right) k}{\text{MAR}_*}. \quad (62)$$

Denote the right-hand side of (62) as a function  $h(k, s, \text{MAR}_*)$ . The observation that  $\frac{\partial h(k, s, \text{MAR}_*)}{\partial s} < 0$  indicates that the larger the selection parameter  $s$ , the lower the SNR required for reliable recovery. This indicates that if it is possible to select more atoms in each iteration, then within a maximum of  $k$  iterations, the likelihood of the algorithm selecting all correct atoms increases, so as to achieve the desired performance with lower SNR. Meanwhile, the observation that  $\frac{\partial h(k, s, \text{MAR}_*)}{\partial k} > 0$  reveals that a larger block sparsity  $k$  requires a more stringent SNR for reliable recovery. This result is consistent with the upper bound of reconstruction error provided in **Theorem 4** which indicates that a lower reconstruction error may require a lower block sparsity level. Moreover, as  $\text{MAR}_*$  is defined in **Definition 8**, (62) can be adjusted into

$$\lim_{n \rightarrow \infty} \sqrt{\text{SNR}} > \frac{\left( \frac{1}{\sqrt{s}} + 1 \right) \sqrt{k} \|\mathcal{X}\|_F}{\min_{(i_1, \dots, i_n) \in \Xi} \|\mathcal{X}_{[i_1, \dots, i_n]}\|_F}.$$

It can be seen that  $k$  and  $s$  are on the same scale that contributes to  $h(k, s, \text{MAR}_*)$ , implying that an increase in the selection parameter  $s$  can effectively counteract the performance loss resulting from the increased block sparsity. Since  $\frac{\partial h(k, s, \text{MAR}_*)}{\partial \text{MAR}_*} < 0$ , a lower SNR bound is acquired as the support tensor blocks become more powerful. The results related to the limits in (56)-(61) indicate a monotonic decreasing property of the bound in the right-hand side of (51) with respect to the number of tensor modes  $n$ . Therefore, this analysis unveils that a high-dimensional tensor recovery can provide strong robustness.

The following theorem is an extension of **Theorem 4**, which considers the scenario where the SNR satisfies the conditions in **Theorem 5**, and  $\|\mathcal{R}^{l^*}\|_F \leq \epsilon$  is not met for  $l^* \leq k$ .

*Theorem 6:* Given a matrix set  $\mathbf{Y} = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ), suppose that  $\|\mathcal{R}^{l^*}\|_F \leq \epsilon$  is not met for  $l^* \leq k$ . If  $\overline{W}_{\mathbf{Y},ks} > 0$ ,  $\overline{W}_{\mathbf{Y},\prod_{t=1}^n k_t+k} > 0$ , and the SNR

satisfies (49a) and (49b), then the output  $\hat{\mathcal{X}}$  satisfies

$$\begin{cases} \|\hat{\mathcal{X}} - \mathcal{X}\|_F \leq \frac{\|\mathcal{N}\|_F}{W_{\mathbf{Y},k}^{\frac{1}{2}}}, s = 1, \\ \|\hat{\mathcal{X}} - \mathcal{X}\|_F \leq \left(1 + \frac{\bar{W}_{\mathbf{Y},\prod_{t=1}^n k_t+k}}{W_{\mathbf{Y},k,s}^{\frac{1}{2}}}\right) \\ \times \frac{2\|\mathcal{N}\|_F}{W_{\mathbf{Y},\prod_{t=1}^n k_t+k}^{\frac{1}{2}}}, s > 1. \end{cases} \quad (63a)$$

*Proof:* See Appendix K. ■

We now present an insight to the block orthogonal tensor recovery model, where the measurement matrix satisfying block orthogonal has been proved to offer more assured theoretical guarantees.

*Remark 8:* Consider that the measurement matrices satisfy block orthogonality among the measurement matrix set  $\mathbf{Y}$ , i.e.,  $\nu_{\mathbf{D}_t} = 0$  ( $t \in \{1, 2, \dots, n\}$ ), and thus  $\tau_{\mathbf{Y},l} = 0$  ( $l \in \{0, 1, \dots, n\}$ ). Let  $s = 1$ . Then, the condition in **Theorem 6** becomes

$$\|\hat{\mathcal{X}} - \mathcal{X}\|_F \leq \frac{\|\mathcal{N}\|_F}{\sqrt{1 - \sum_{l=0}^{n-1} C_l^k \varpi_{\mathbf{Y},l}^n \prod_{t=1}^n d_t}}. \quad (64)$$

By defining the right-hand sides of (63a) and (64) as functions  $p(\varpi_{\mathbf{Y}}, \tau_{\mathbf{Y}})$  and  $q(\varpi_{\mathbf{Y}}, \tau_{\mathbf{Y}})$ , respectively, we have

$$\begin{aligned} & \frac{q(\varpi_{\mathbf{Y}}, \tau_{\mathbf{Y}})}{p(\varpi_{\mathbf{Y}}, \tau_{\mathbf{Y}})} \\ &= \left( \frac{1 - \sum_{l=0}^{n-1} C_l^k \varpi_{\mathbf{Y},l}^n \prod_{t=1}^n d_t}{1 - \sum_{l=0}^{n-1} C_l^{\prod_{t=1}^n d_t} \tau_{\mathbf{Y},l}^{n-l} - \sum_{l=0}^{n-1} C_l^k \varpi_{\mathbf{Y},l}^n \prod_{t=1}^n d_t} \right)^{\frac{1}{2}} \\ &= \left( 1 + \frac{\sum_{l=0}^{n-1} C_l^{\prod_{t=1}^n d_t} \tau_{\mathbf{Y},l}^{n-l}}{1 - \sum_{l=0}^{n-1} C_l^{\prod_{t=1}^n d_t} \tau_{\mathbf{Y},l}^{n-l} - \sum_{l=0}^{n-1} C_l^k \varpi_{\mathbf{Y},l}^n \prod_{t=1}^n d_t} \right)^{\frac{1}{2}} \\ &= (1 + \underline{\Delta})^{\frac{1}{2}}, \end{aligned}$$

where  $\underline{\Delta} = \frac{\sum_{l=0}^{n-1} C_l^{\prod_{t=1}^n d_t} \tau_{\mathbf{Y},l}^{n-l}}{1 - \sum_{l=0}^{n-1} C_l^{\prod_{t=1}^n d_t} \tau_{\mathbf{Y},l}^{n-l} - \sum_{l=0}^{n-1} C_l^k \varpi_{\mathbf{Y},l}^n \prod_{t=1}^n d_t}$ .

In general, as the block lengths increase, the dimensions of the measurement matrix blocks are increased correspondingly, leading to the increases of  $\varpi_{\mathbf{Y},l}^n \prod_{t=1}^n d_t$  and  $\tau_{\mathbf{Y},l}^{n-l}$ . Then, we obtain that the  $\underline{\Delta}$  is increased with the increase of block lengths. It reveals that the reconstruction error difference between the block orthogonal tensor recovery and non-orthogonal tensor recovery is further exacerbated as the block structure property becomes stronger. This necessitates the design of measurement matrix that obeys block orthogonality for better recovery performance guarantees, which can be achieved through methodologies like Schmidt orthogonalization [41].

**Theorems 4 to 6** present reliable recovery conditions of T-GBOMP by considering the level of residual power within  $k$  iterations as a prerequisite. Now we naturally turn to the study of bounding the residual power. To begin with, we first provide the following lemma that characterizes the relationship between the residuals in different numbers of iterations. It

is worth mentioning that the formulation of this lemma is provided in Appendix M for the sake of conciseness.

*Lemma 3:* Given a matrix set  $\mathbf{Y} = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ), if  $W_{\mathbf{Y},|\Lambda^l \cup \Xi^{c+\Delta c-1}|} > 0$ ,  $c \geq l$  and  $\Delta c > 0$  for any integer  $c$ , and  $\varrho \in \{1, \dots, \max\{0, \lceil \log \frac{e|\Lambda^l|}{s} \rceil\} + 1\}$ , the residual  $\mathcal{R}^{c+\Delta c}$  in the  $(c + \Delta c)$  iteration satisfies

$$\begin{aligned} & \|\mathcal{R}^{c+\Delta c}\|_F^2 \\ & - \|\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_n^l}} \otimes \dots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_1^l}} \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^l}) + \text{vec}(\mathcal{N})\|_F^2 \\ & \leq G_{\mathbf{Y},\varrho,c,\Delta c} \left( \|\text{vec}(\mathcal{R}^c)\|_F^2 \right. \\ & \quad \left. - \|\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_n^l}} \otimes \dots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_1^l}} \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^l}) + \text{vec}(\mathcal{N})\|_F^2 \right), \end{aligned}$$

where

$$G_{\mathbf{Y},\varrho,c,\Delta c} = \exp \left( - \frac{\Delta c W_{\mathbf{Y},|\Lambda^l \cup \Xi^{c+\Delta c-1}|}}{\left\lceil \frac{|\Lambda^l|}{s} \right\rceil \bar{W}_{\mathbf{Y},s}} \right).$$

*Proof:* See Appendix L. ■

In the following remark, we present the results related to the Frobenius norm of residual tensors in different iterations, which is regarded as the true residual error.

*Remark 9:* For the  $l$ th ( $0 < l < k$ ) iteration, denote the tensor index sets selected in the  $(l+1)$ th iterations as  $\Xi^{l+1}$ . Following the proof in Appendix H, we have

$$\begin{aligned} \|\mathcal{R}^l - \mathcal{R}^{l+1}\|_F^2 &= \|\mathbf{P}_{\mathbf{D}_{\Xi^{l+1}}} \text{vec}(\mathcal{R}^l)\|_F^2 \leq \|\text{vec}(\mathcal{R}^l)\|_F^2 \\ &\leq \left( 1 - \frac{s W_{\mathbf{Y},\prod_{t=1}^n k_n+s(l-1)}}{k \bar{W}_{\mathbf{Y},\prod_{t=1}^n k_n+s(l-1)} \bar{W}_{\mathbf{Y},s}} \right)^l \|\mathcal{Y}\|_F^2. \end{aligned}$$

For a more intuitive representation, let  $d_1 = d_2 = \dots = d_n = d$ , and we have

$$\begin{aligned} \|\mathcal{R}^l - \mathcal{R}^{l+1}\|_F^2 &\leq \left( 1 - \frac{s(1 - \sum_{l=0}^{n-1} C_l^{\prod_{t=1}^n d_t} \tau_{\mathbf{Y},l}^{n-l} - \sum_{l=0}^{n-1} C_l^{\prod_{t=1}^n k_n+s(l-1)} \varpi_{\mathbf{Y},l}^n d^n)}{k(1 + \sum_{l=0}^{n-1} C_l^{\prod_{t=1}^n d_t} \tau_{\mathbf{Y},l}^{n-l} + \sum_{l=0}^{n-1} C_l^{\prod_{t=1}^n k_n+s(l-1)} \varpi_{\mathbf{Y},l}^n d^n)} \right)^l \\ &\quad \times \frac{1}{1 + \sum_{l=0}^{n-1} C_l^{\prod_{t=1}^n d_t} \tau_{\mathbf{Y},l}^{n-l} + \sum_{l=0}^{n-1} C_l^s \varpi_{\mathbf{Y},l}^n d^n} \|\mathcal{Y}\|_F^2. \end{aligned}$$

It can be observed that as the number of iteration  $l$  increases, the upper bound of the residual tensor error gradually approaches 0, which indicates that the significant tensor blocks are being subtracted from the measurement tensor. It can be roughly assumed that all the correct tensor blocks have been selected if  $\|\mathcal{R}^l - \mathcal{R}^{l+1}\|_F^2$  tends very close to 0. Similar to the analysis in **Remark 7**, we obtain

$$\lim_{n \rightarrow \infty} \|\mathcal{R}^l - \mathcal{R}^{l+1}\|_F^2 \leq \left( 1 - \frac{s}{k} \right)^l \|\mathcal{Y}\|_F^2.$$

This result indicates that as  $s$  increases, the significant tensor blocks are more likely to be selected correctly, resulting in a faster convergence speed that equalizes  $\mathcal{R}^l$  and  $\mathcal{R}^{l+1}$ . A special case arises when  $s = k$ , wherein we observe  $\lim_{n \rightarrow \infty} \|\mathcal{R}^l - \mathcal{R}^{l+1}\|_F^2 = 0$ , which indicates that all the

correct tensor blocks are selected during the first iteration. Moreover, it is worth mentioning that the convergence speed increases with the increase of the tensor mode  $n$  when the other parameters are fixed. This also unveils that the derived bound of the residual error becomes smaller as the tensor mode  $n$  increases. That is, the high-dimensional tensor recovery provides tight theoretical guarantees.

Based on **Lemma 3**, we develop the following theorem, which indicates that after a specified number of iterations related to the number of remaining supports to be selected, the residual is upper bounded by a quantity with respect to the MIP and noise power.

**Theorem 7:** Given a matrix set  $\mathbf{Y} = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ), let  $\Lambda^l$  be the set consisting of the remaining support block tensors after  $l$  iterations with  $|\Lambda^l| = \theta$ , and  $\alpha$  be an integer such that  $\alpha = \inf \left\{ x \in \left\{ 1, \dots, \max \left\{ 0, \left\lceil \log \frac{e|\Lambda^l|}{s} \right\rceil \right\} + 1 \right\}, \frac{\exp(x-2)s(e-1)}{\exp(x-1)-1+(e-1)x} \right\}$ . If the MIP of  $\mathbf{Y}$  satisfies  $\underline{W}_{\mathbf{Y},k} > 0$ ,  $\underline{W}_{\mathbf{Y},sl+s+\theta} > 0$  and  $\underline{W}_{\mathbf{Y},sl+s+\theta} > 0$ , then we have

$$\|\mathcal{R}^{l+|\Lambda^l}\|_F \leq \xi \|\mathcal{N}\|_F,$$

where

$$\begin{aligned} \xi &= \left( \omega(1+\gamma)\overline{W}_{\mathbf{Y},\theta^n} + \frac{W_{\mathbf{Y},sl+s+\theta}}{\overline{W}_{\mathbf{Y},s}}(1+\gamma)\overline{W}_{\mathbf{Y},\theta^n}\eta^{-1} \right)^{\frac{1}{2}} \\ &\quad \times \left( 1 + \left( \frac{1}{1-\beta}(1+\gamma^{-1}) \right)^{\frac{1}{2}} \right) \\ &\quad \times \underline{W}_{\mathbf{Y},k}^{-\frac{1}{2}} \left( 1 - \left( \frac{\omega(1+\gamma)\overline{W}_{\mathbf{Y},\theta^n} + \frac{W_{\mathbf{Y},sl+s+\theta}}{\overline{W}_{\mathbf{Y},s}}(1+\gamma)\overline{W}_{\mathbf{Y},\theta^n}\eta^{-1}}{\overline{W}_{\mathbf{Y},k}} \right)^{\frac{1}{2}} \right)^{-1} \\ &\quad + \frac{(1+\gamma^{-1})^{\frac{1}{2}}}{(1-\beta)^{\frac{1}{2}}}, \\ \omega &= \max \left\{ \frac{1}{\eta(1-\eta\beta)}, 1 - \frac{W_{\mathbf{Y},sl+s+\theta}}{\overline{W}_{\mathbf{Y},s}} \right\}, \\ \beta &= \exp \left( - \frac{\alpha W_{\mathbf{Y},sl+s+\theta}}{\overline{W}_{\mathbf{Y},s}} \right), \end{aligned}$$

with  $\eta\beta < 1$  and  $\gamma > 0$ .

*Proof:* See Appendix M.  $\blacksquare$

It can be observed that if T-GBOMP has already operated  $l$  iterations, then the algorithm iterates at most  $|\Lambda^l|$  additional iterations to guarantee that the residual falls below  $\xi \|\mathcal{N}\|_F$  with the MIP satisfying the conditions in **Theorem 7**. Note that when  $l = 0$ , we have  $|\Lambda^0| = k$ . Based on these discussions, we present the following corollary.

**Corollary 5:** Given a matrix set  $\mathbf{Y} = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ), let  $\alpha$  be an integer such that  $\alpha = \inf \left\{ x \in \left\{ 1, \dots, \max \left\{ 0, \left\lceil \log \frac{ek}{s} \right\rceil \right\} + 1 \right\}, \frac{\exp(x-2)s(e-1)}{\exp(x-1)-1+(e-1)x} \right\}$ . If the MIP of  $\mathbf{Y}$  satisfies  $\underline{W}_{\mathbf{Y},sk+k} > 0$ , then we have

$$\|\mathcal{R}^k\|_F \leq \xi^* \|\mathcal{N}\|_F,$$

where

$$\xi^* = \left( \frac{1}{\eta(1-\frac{1}{e})}(1+\gamma)\overline{W}_{\mathbf{Y},k^n} \right)^{\frac{1}{2}}$$

$$\begin{aligned} &+ \frac{W_{\mathbf{Y},s+k}}{\overline{W}_{\mathbf{Y},s}}(1+\gamma)\overline{W}_{\mathbf{Y},k^n}\eta^{-1} \right)^{\frac{1}{2}} \\ &\times \left( 1 + \left( \frac{1}{1-\beta}(1+\gamma^{-1}) \right)^{\frac{1}{2}} \right) \\ &\times \underline{W}_{\mathbf{Y},k}^{-\frac{1}{2}} \left( 1 - \left( \frac{\frac{1}{\eta(1-\frac{1}{e})}(1+\gamma)\overline{W}_{\mathbf{Y},k^n} + \frac{W_{\mathbf{Y},s+k}}{\overline{W}_{\mathbf{Y},s}}(1+\gamma)\overline{W}_{\mathbf{Y},k^n}\eta^{-1}}{\overline{W}_{\mathbf{Y},k}} \right)^{\frac{1}{2}} \right)^{-1} \\ &\quad + \frac{(1+\gamma^{-1})^{\frac{1}{2}}}{(1-\beta)^{\frac{1}{2}}}, \end{aligned} \quad (65)$$

$$\eta = \frac{1}{e} \exp \left( \frac{\alpha W_{\mathbf{Y},sk+k}}{\overline{W}_{\mathbf{Y},s}} \right), \quad \beta = \exp \left( - \frac{\alpha W_{\mathbf{Y},sk+k}}{\overline{W}_{\mathbf{Y},s}} \right),$$

with  $\eta\beta < 1$  and  $\gamma > 0$ .

Note that in **Corollary 5**,  $\eta\beta = \frac{1}{e} \leq 1$ . Thus,

$$\begin{aligned} \frac{1}{\eta(1-\eta\beta)} &= \frac{1}{\frac{1}{e} \exp \left( \frac{\alpha W_{\mathbf{Y},sk+k}}{\overline{W}_{\mathbf{Y},s}} \right) (1-\frac{1}{e})} \\ &\geq 1 - \frac{W_{\mathbf{Y},sl+s+\theta}}{\overline{W}_{\mathbf{Y},s}}, \end{aligned}$$

and we have  $\omega = \frac{1}{\eta(1-\frac{1}{e})}$ .

**Remark 10:** The existing result given in [42] for generalized OMP indicates that when the RIP is constrained to a derived degree, the number of iterations should be  $\max \{ K, \lceil \frac{8K}{s} \rceil \}$  for the residual to be less than a product of the noise power and a constant related to the RIP indicator, where  $K = kd$  is the total sparsity. However, since the calculation of the RIP is NP-hard, the result of [42] is not verifiable. By contrast, the MIP is computable, and our result in **Corollary 5** reveals that when the number of iterations is equal to  $k$ , the residual can be smaller than  $\xi^* \|\mathcal{N}\|_F$  with  $\underline{W}_{\mathbf{Y},sk+k} > 0$ . Intuitively, due to the consideration of the block structure, we have  $k < K$  for  $d > 1$ . The T-GBOMP exhibits significant superiority in terms of lower iterative complexity over that of the generalized OMP. In particular, setting  $s = 1$  and  $d = 1$ , we obtain that the number of iterations for the generalized OMP derived in [42] is  $8K$  and the parallel result in **Corollary 5** is  $K$ . We present two explanations for this phenomenon. The first is that the MIP condition  $\underline{W}_{\mathbf{Y},sk+k} > 0$  in **Corollary 5** is much more restrictive than the RIP constraint in [42], and hence the number of iterations in **Corollary 5** is much smaller than that in [42]. Since computing RIP is NP-hard, this viewpoint cannot be verified. The second is that the result given in **Corollary 5** is indeed better than the one given in [42], which provides a lower number of iterations for reliable recovery, resulting in less consumption of computing resources. In any case, since MIP is computationally friendly, **Corollary 5** is useful in presenting a verifiable prerequisite for determining whether the signal components in the residual are fully extracted.

Similar to **Theorem 6**, the following theorem demonstrates that the estimated error is upper bounded by a more relaxed quantity with respect to MIP and noise power under a less restrictive MIP condition.

**Theorem 8:** Given a matrix set  $\Upsilon = \{\mathbf{D}_i \in \mathbb{C}^{M_i \times s_i d_i}, 1 \leq i \leq n\}$  ( $n \geq 1$ ), let  $\alpha$  be an integer such that  $\alpha = \inf \left\{ x \in \left\{ 1, \dots, \max \left\{ 0, \left\lceil \log \frac{ek}{s} \right\rceil \right\} + 1 \right\}, \frac{\exp(x-2)s(e-1)}{\exp(x-1)-1+(e-1)x} \right\}$ . If the MIP of  $\Upsilon$  satisfies that  $\underline{W}_{\Upsilon, s_{k+k}} > 0$ , then we have

$$\begin{aligned} & \|\hat{\mathcal{X}} - \mathcal{X}\|_F \\ & \leq \frac{2 \left( \underline{W}_{\Upsilon, \prod_{t=1}^n k_t+k_s} (\xi^* + 1) \overline{W}_{\Upsilon, \prod_{t=1}^n k_t+k_s} + 1 \right)}{\underline{W}_{\Upsilon, \prod_{t=1}^n k_t+k_s}} \|\mathcal{N}\|_F, \end{aligned} \quad (66)$$

where  $\xi^*$  is given in (65).

*Proof:* See Appendix N.  $\blacksquare$

**Remark 11:** From **Theorems 6** and **8**, when  $\|\mathcal{N}\|_F = 0$ , we have  $\hat{\mathcal{X}} = \mathcal{X}$ , indicating that the block-sparse tensor can be recovered exactly in the noiseless scenario. The variables on the right-hand sides of equations (63a), (63b) and (66) are the computable MIP, which not only presents strong theoretical interpretability for the feasibility of the block-sparse tensor recovery with low error reconstruction, but also provides a basis for making decisions regarding reliable performance guarantees before operation of the algorithm.

## V. SIMULATION RESULTS

We empirically investigate the recovery performance of T-GBOMP in reconstructing block-sparse tensors under both noiseless and noisy conditions. It is important to note that we not only compare the recovery performance using various metrics but also demonstrate the resilience of different algorithms to suboptimal coherence of measurement matrices. We adopt similar testing metrics as those used in [43] and [36]. In particular, the exact recovery ratio (ERR) of block-sparse tensor reconstruction is examined as a function of block sparsity and the number of measurements in the noise-free scenario. In the noisy scenario, false alarm/mis-detection ratio and the normalized mean square error (NMSE) are employed to assess the recovery performance of various approaches. To explore the robustness against unsatisfactory MIP, an effortless method is to use the hybrid dictionary as presented in [38] and [44], where the conventional matrix coherence is close to 1. This measurement matrix substantially diminishes the performance of the compared algorithms, with the extent of performance degradation varying across different algorithms, thereby enabling a more comprehensive insight into the merits and demerits of each algorithm. In a similar fashion, we also produce measurement matrices with conventional matrix coherence nearly 1, but by generating small-dimensional Gaussian matrices. The number of tensor modes is fixed to 3. The measurement matrices satisfy  $\prod_{i=1}^3 M_i = 144$  or  $\prod_{i=1}^3 M_i = 216$ , and  $\prod_{i=1}^3 N_i = 512$  or  $\prod_{i=1}^3 N_i = 1024$  with each element drawn independently and identically from a Gaussian distribution. When  $\prod_{i=1}^3 M_i = 216$  and  $M_1 = M_2 = M_3$ , then we have  $M_1 = M_2 = M_3 = 6$ , which are small-dimensional measurement matrices whose conventional matrix coherence is about 0.75. Meanwhile, measurement matrices with low matrix coherence are also generated for performance comparison, and they have

a coherence value of approximately 0.23. Moreover, these matrices are created as block orthogonal matrices with  $\tau_{\Upsilon} = 0$  and are normalized to have unit column norm. For each realization of a block-sparse tensor, its block support is chosen uniformly at random, and nonzero elements are 1) drawn independently from a standard Gaussian distribution, or 2) drawn independently from the set  $\{\pm 1\}$ . These two types of tensors are referred to as the block-sparse Gaussian tensor and the block-sparse 2-ary pulse amplitude modulation (2-PAM) tensor, respectively [36]. In our experiments, we perform 1,000 independent trials for each point of the approaches, and consider the following recovery algorithms: 1) BOMP [10]; 2) BOLS [6]; 3) T-OMP [8]; 4) T-GOMP ( $s=2$ ); 5) T-GOMP ( $s=3$ ); 6) T-BOMP; 7) T-GBOMP ( $s=2$ ); and 8) T-GBOMP ( $s=3$ ).

### A. The Noiseless Scenario

In the noiseless case, the ERR is used as a metric to evaluate the performance of various algorithms, which is defined as [43]

$$\text{ERR} = \frac{\text{number of exact reconstructions}}{\text{number of total trials}}.$$

In Fig. 3, we depict the ERR as a function of the block sparsity  $k$ . The block lengths are fixed at  $d_1 = 2$ ,  $d_2 = 2$  and  $d_3 = 1$ , resulting in a total sparsity of  $K = d_1 d_2 d_3 k = 4k$ . The conventional matrix coherence of the measurement matrices is around 0.75, which is unsatisfactory for recovery and can significantly degrade the performance of the algorithms. The mutual block coherence, as defined in **Definition 3**, is about 0.6. It is evident that due to the unsatisfactory MIP of the measurement matrices, none of the compared approaches can consistently achieve exact recovery. However, it can be seen that T-GBOMP ( $s=2$ ) and T-GBOMP ( $s=3$ ) provide better assurance of exact recovery, outperforming other algorithms when the block sparsity is less than or equal to 2, wherein  $k=1$  is the only possible point for the algorithms to achieve exact recovery, i.e.,  $\text{ERR} = 1$ . As can be seen from Fig. 3(a), T-GBOMP ( $s=2$ ) and T-GBOMP ( $s=3$ ) perform worse than BOLS in recovering the block-sparse Gaussian tensors with  $\prod_{i=1}^3 M_i = 144$  and  $\prod_{i=1}^3 N_i = 512$  when a block sparsity is greater than 3, which indicates that the OLS-framework may be highly competitive in providing reliable recovery in the scenarios with unsatisfactory MIP and high block sparsity. For 2-PAM tensors, Figs. 3(b) and 3(d) show that T-BOMP, T-GBOMP ( $s=2$ ) and T-GBOMP ( $s=3$ ) exhibit significant performance improvements over the other algorithms. This suggests that tensor structure information is beneficial for reliable recovery, as T-BOMP outperforms BOMP and BOLS. Moreover, exploiting block sparsity yields more assured results, since T-GBOMP ( $s=2$ ) and T-GBOMP ( $s=3$ ) outperform T-GOMP ( $s=2$ ) and T-GOMP ( $s=3$ ). As T-GBOMP performs better than T-BOMP, and T-BOMP performs better than BOMP, it becomes evident that the larger the tensor mode, the better the recovery performance of the algorithms. This aligns with the statement in Section IV, indicating that a larger tensor mode can offer more reliable recovery performance. Furthermore, in general, T-GBOMP ( $s=3$ ) outperforms T-GBOMP ( $s=2$ ), indicating that a

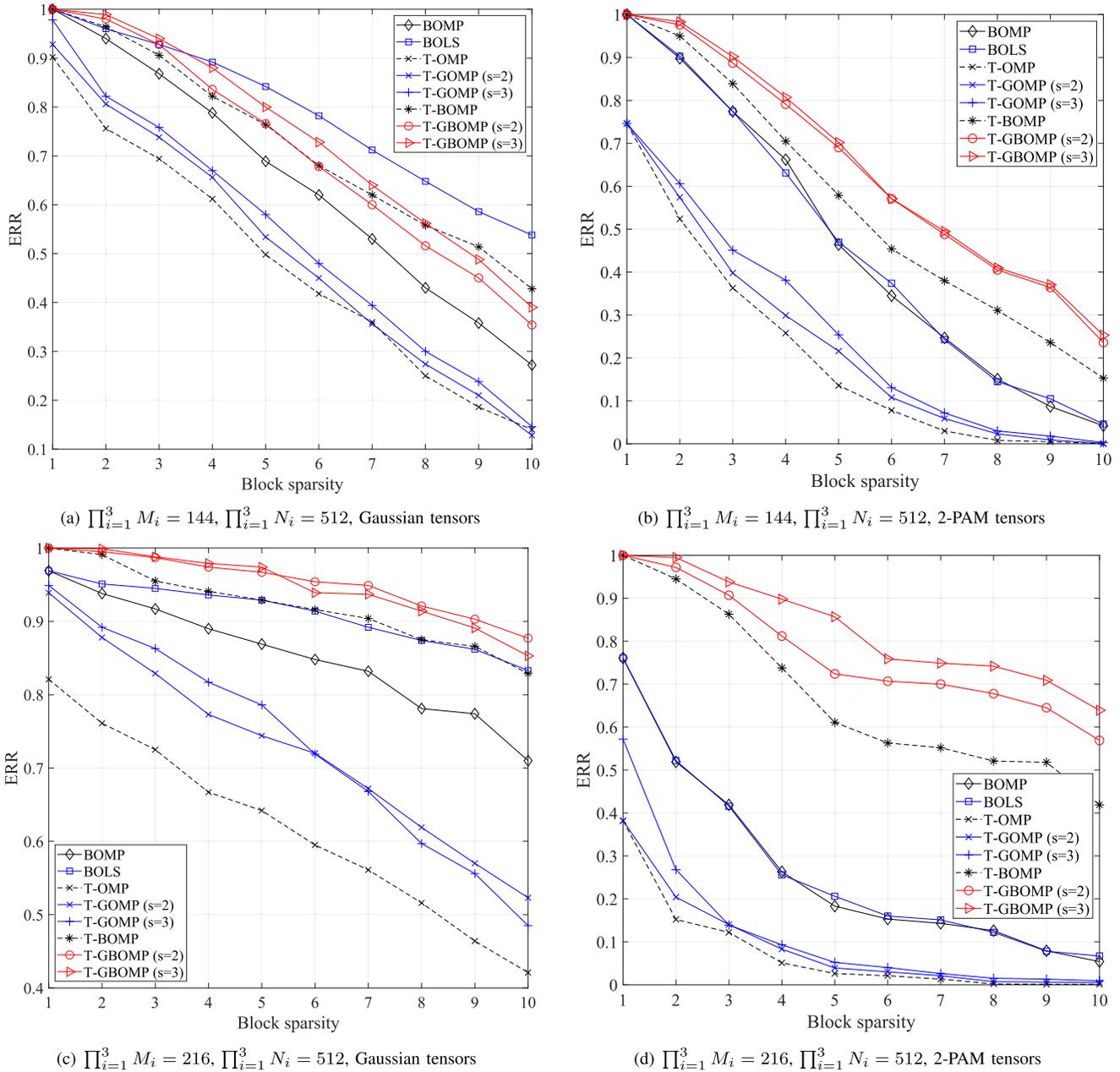


Fig. 3. ERR for recovering block-sparse tensors as a function of the block sparsity.

larger selection parameter can indeed enhance the accuracy of recovery by increasing the likelihood of selecting all the correct atoms. We now verify the analytic results related to reconstructible sparsity derived in Subsection IV-A. As illustrated in Remark 5, based on the parameter settings in Fig. 3, the most restricted reconstructible sparsity can be calculated by  $\left(-\frac{1}{2} + \sqrt{\frac{5}{4} - \frac{-1}{0.6^3 \times 4}}\right)^2 \approx 1$ , which is consistent with the results given in Fig. 3. Observe that T-GBOMP ( $s = 2$ ) and T-GBOMP ( $s = 3$ ) can achieve 100% ERR when the block sparsity is less than or equal to 1.

In Fig. 4, where measurement matrices are generated to have low matrix coherence, the ERR is plotted as a function of the number of measurements. It can be observed that T-GBOMP ( $s = 2$ ) and T-GBOMP ( $s = 3$ ) outperform the other compared methods, with T-GBOMP ( $s = 3$ ) achieving the best ERR

when the number of measurements exceeds 144. When the total sparsity is fixed, i.e.,  $K = k \prod_{i=1}^3 d_i = 16$ , the recovery performance of T-GBOMP ( $s = 2$ ) and T-GBOMP ( $s = 3$ ) are improved as the block length increases. This indicates that a stronger block-structure characteristic leads to a more assured recoverability of the algorithms. It is worth noting that the block-structure characteristic is crucial for the reliability of recovery. For instance, in Fig. 4(a), for T-GBOMP ( $s = 3$ ) to attain the 100% ERR, the number of measurements is larger than 288. In contrast, as can be seen from Fig. 4(b), T-GBOMP ( $s = 3$ ) can provide 100% ERR performance when the number of measurements exceeds 216. Furthermore, since the total sparsity remains constant, the corresponding ERRs of T-GOMP ( $s = 2$ ) and T-GOMP ( $s = 3$ ) are similar given a same number of measurements, as can be seen from

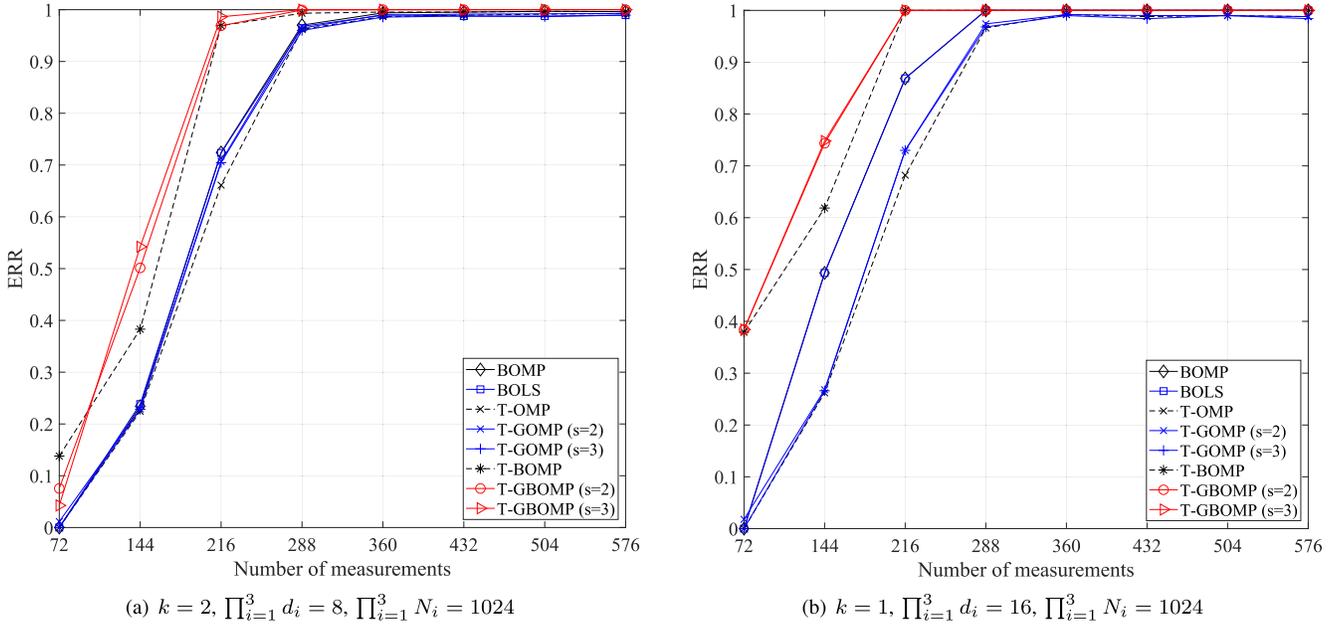


Fig. 4. ERR for recovering block-sparse 2-PAM tensors as a function of the number of measurements.

Figs. 4(a) and 4(b). This finding aligns with the expectation that these algorithms do not leverage the block structure of the tensors being recovered, and therefore are not significantly influenced by changes in the block structure given a fixed total sparsity.

### B. The Noisy Scenario

The false alarm/miss-detection and NMSE are used as metrics to evaluate recovery performance of different approaches. The false alarm ratio is expressed as

$$\text{False alarm ratio} = \frac{|\hat{\Xi} \setminus \Xi|}{k \prod_{i=1}^3 d_i},$$

which is equivalent to the miss-detection ratio since  $|\hat{\Xi}| = |\Xi| = k \prod_{i=1}^3 d_i$ , where  $\hat{\Xi}$  is the estimated set consisting of the indices corresponding to nonzero tensor elements, and  $\Xi$  denotes the correct index set. The NMSE is expressed as

$$\text{NMSE} = \frac{\|\hat{\mathcal{X}} - \mathcal{X}\|_F^2}{\|\mathcal{X}\|_F^2},$$

where  $\hat{\mathcal{X}}$  represents the estimated block-sparse tensor, and  $\mathcal{X}$  denotes the true block-sparse tensor.

As illustrated in Fig. 5, T-GBOMP ( $s = 3$ ) achieves the best false alarm/miss-detection ratio among all the compared algorithms. This indicates that a larger selection parameter  $s$  can provide more assured performance in identifying the correct tensor supports. When the compression ratio becomes larger but still remains below  $\frac{1}{2}$  as depicted in Fig. 5(b), T-GBOMP ( $s = 2$ ) and T-GBOMP ( $s = 3$ ) achieve exceptionally good detection performance, surpassing the other algorithms, with the false alarm/miss-detection ratios of T-GBOMP ( $s = 2$ ) and T-GBOMP ( $s = 3$ ) reaching 0 across all compared SNR values. In general, algorithms that exploit block sparsity outperform those that do not utilize block

sparsity, indicating that the block-structure characteristic is indeed a beneficial factor for reliable recovery. Furthermore, it can be observed that the methods using block-sparse tensor structure, i.e., T-BOMP, T-GBOMP ( $s = 2$ ) and T-GBOMP ( $s = 3$ ), provide better false alarm/miss-detection ratios than BOMP and BOLS, which only utilize block sparsity without employing tensor structure. This is due to the vectorization in conventional non-tensor algorithms which causes the loss of certain high-dimensional feature information in sparse recovery.

As can be observed from Fig. 6, where measurement matrices are generated with low matrix coherence, T-GBOMP ( $s = 3$ ) generally achieves the lowest NMSE among all the compared methods. The results of Fig. 6 show that as the block sparsity increases, the performance of T-BOMP, T-GBOMP ( $s = 2$ ) and T-GBOMP ( $s = 3$ ) deteriorate, but they still outperform the other approaches significantly. We can now discuss the theoretical results derived in Subsection IV-B. Firstly, regarding the lower bound of SNR required for reliable recovery, we consider the intuitive result presented in Remark 7, taking into account the parameter settings of T-GBOMP ( $s = 3$ ) in Fig. 6(a). Since block-sparse 2-PAM tensors are generated in the simulation of Fig. 6(a), the MAR is equal to 1 as defined in Definition 8. The lower bound of SNR required for reliable recovery in Remark 7 is thus equal to  $10 \log_{10} \left( 2 \left( \frac{1}{\sqrt{3}} + 1 \right) \right)^2 \approx 10$  dB. This holds true, as the NMSE is relatively low, and remains under  $10^{-2}$  when the SNR exceeds 10 dB in Fig. 6(a). It is worth mentioning that the result given in Remark 7 is less restrictive than those in Theorem 5 due to the asymptotic analysis, which indicates that the correctness of the result in Remark 7 also demonstrates the truth in Theorem 5. Secondly, regarding the reconstruction error which can be reflected by the NMSE, we consider the results presented in Theorem 6. Similarly, we assume a less restricted scenario for calculating the reconstruction error

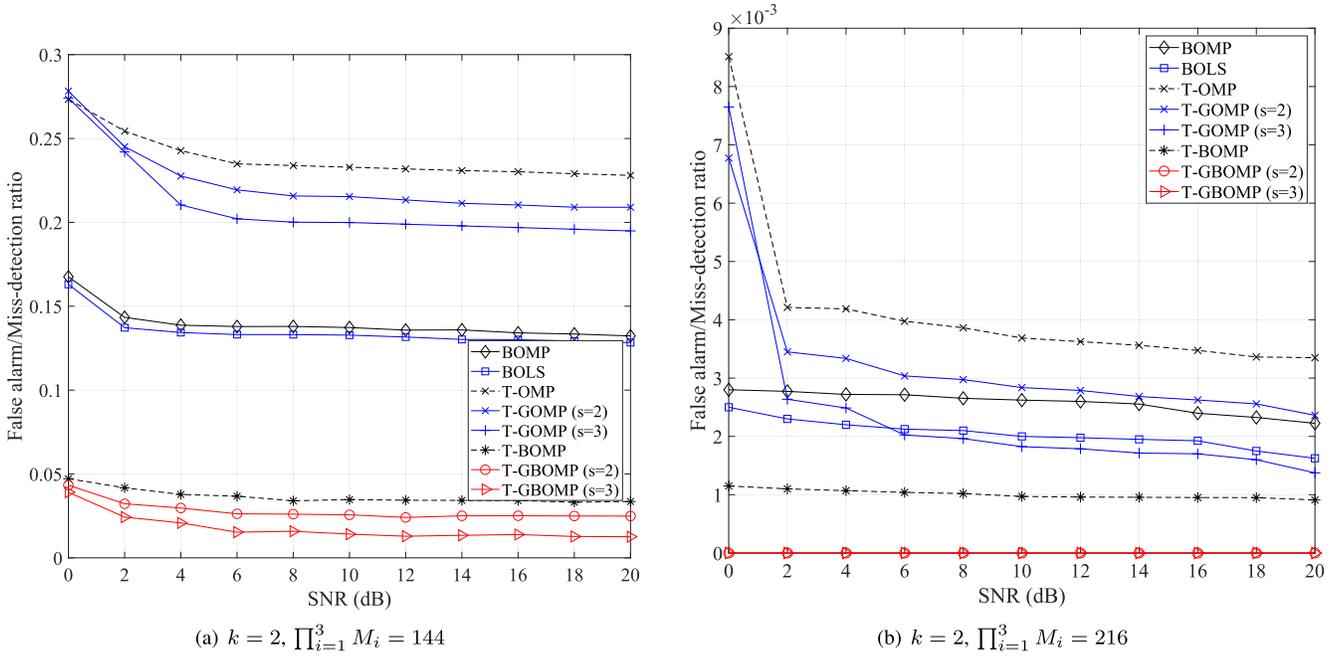


Fig. 5. False alarm/miss-detection ratio for recovering block-sparse 2-PAM tensors as a function of SNR.

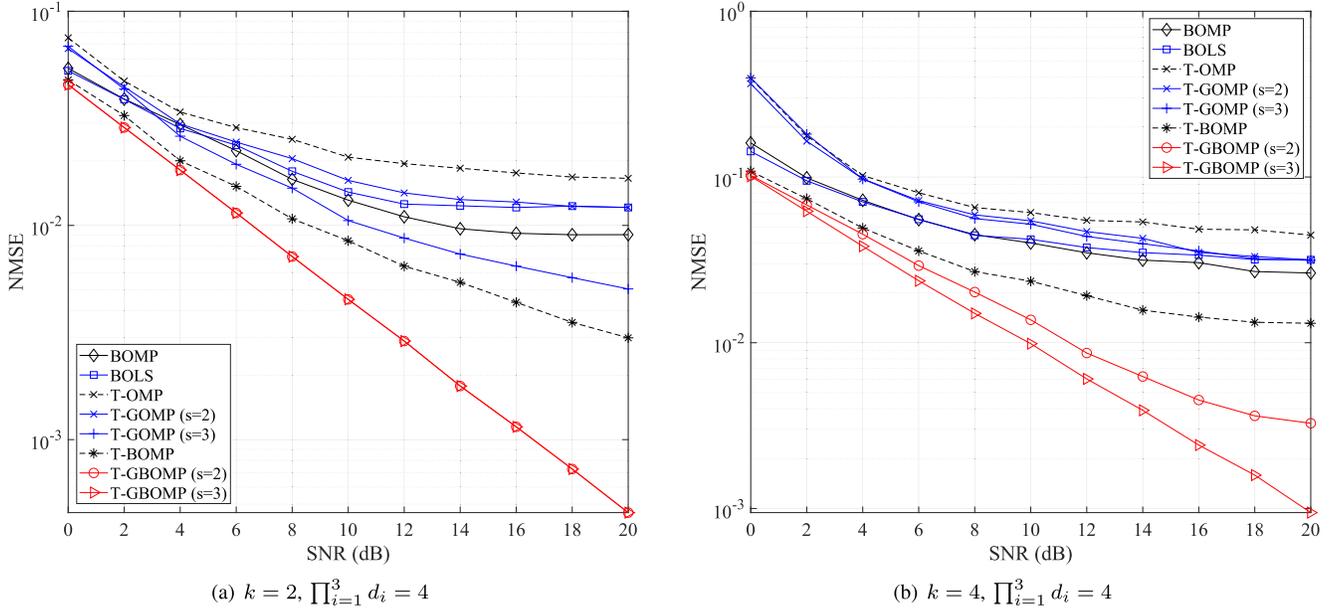


Fig. 6. NMSE for recovering block-sparse 2-PAM tensors as a function of SNR.

bound, where each element in the measurement matrix of Fig. 6(a) follows an independent and identically distributed Gaussian distribution with zero mean and variance  $\frac{1}{\prod_{i=1}^3 M_i}$ . In Fig. 6(b), when  $\text{SNR} = 10 \text{ dB}$  and  $K = k \prod_{i=1}^3 d_i = 8$ , we have  $\mathbb{E}(\|\mathcal{N}\|_F) = \sqrt{K} \times 10^{-\frac{10}{10}} = \frac{2\sqrt{5}}{5}$ . Therefore, according to **Theorem 6**, we have  $\text{NMSE} = \frac{\|\mathcal{X} - \hat{\mathcal{X}}\|_F^2}{\|\mathcal{X}\|_F^2} \leq \frac{(4\|\mathcal{N}\|_F)^2}{K^2} = \frac{16(\frac{2\sqrt{5}}{5})^2}{64} = \frac{1}{5}$ . This result is consistent with the findings in Fig. 6(a), indicating that the other theoretical results derived regarding the reconstruction error are also valid. Note that the analytical results presented are more pessimistic than the numerical results, since our analyses

based on the MIP framework essentially consider worst-case scenarios.

## VI. CONCLUSION

In this paper, we have defined new MIP concepts measuring the uncertainty relations in a matrix set, formulated block-sparse tensor recovery modeling, and elaborated T-GBOMP as a generalized algorithmic basis for theoretical derivation. Then, we have provided in-depth theoretical analyses for T-GBOMP, in terms of exact recovery conditions in the noiseless case and reliable reconstruction conditions in noisy scenarios. Specifically, we have presented an exact recovery

condition and closed-form sufficient conditions that ensure T-GBOMP can recover the block-sparse tensor exactly if the block sparsity is smaller than a derived upper bound. We have further derived error bounds and the residual convergence level of T-GBOMP after a specified number of iterations, in the presence of noise. Moreover, the SNR required for reliable recovery with respect to the newly defined version of MIP has also been developed. We have further highlighted that the T-GBOMP includes the existing greedy algorithms as special cases and it can be extended to other tensor variants, indicating that the theoretical results developed in this paper are quite broad.

## APPENDIX

### A. Proof of Lemma 1

*Proof:* Use elementary column transformation to form  $\mathbf{T} \in \mathbb{C}^{\prod_{i=1}^n M_t \times s_t d_t}$  as a concatenation of  $\prod_{t=1}^n s_t$  column-block submatrices, of sizes  $\prod_{t=1}^n M_t \times \prod_{t=1}^n d_t$ , where the  $c$ th ( $c \in \{1, \dots, \prod_{t=1}^n s_t\}$ ) block of the transformed matrix is  $\mathbf{D}_{1[i_1]} \otimes \dots \otimes \mathbf{D}_{n[i_n]}$ , and  $i_1 \in \{1, \dots, s_1\}, \dots, i_n \in \{1, \dots, s_n\}$ . We denote this new transformed matrix as  $\mathbf{Q}$ , and the eigenvalues of  $\mathbf{Q}^H \mathbf{Q}$  are equal to those of  $\mathbf{T}^H \mathbf{T}$ . To lower bound  $\lambda_{\min}$ , we prove that the matrix  $\mathbf{Q}^H \mathbf{Q} - \lambda \mathbf{I}$  is nonsingular under the condition  $\lambda < 1 - (d-1)\nu_D - (s-1)d\mu_D$  when  $(d-1)\nu_D + (s-1)d\mu_D < 1$ . This is equivalent to proving that for any nonzero vector  $\mathbf{f} = [\mathbf{f}_{[1]}^T, \mathbf{f}_{[2]}^T, \dots, \mathbf{f}_{[s]}^T]^T \in \mathbb{C}^{\prod_{t=1}^n s_t \prod_{t=1}^n d_t}$ ,  $(\mathbf{Q}^H \mathbf{Q} - \lambda \mathbf{I})\mathbf{f} \neq \mathbf{0}$ . Without loss of generality, we assume  $\|\mathbf{f}_{[1]}\|_2 \geq \|\mathbf{f}_{[2]}\|_2 \geq \dots \geq \|\mathbf{f}_{[s]}\|_2$ . The  $\ell_2$ -norm of the first block of  $(\mathbf{Q}^H \mathbf{Q} - \lambda \mathbf{I})\mathbf{f}$  satisfies

$$\begin{aligned} & \left\| [(\mathbf{Q}^H \mathbf{Q} - \lambda \mathbf{I})\mathbf{f}]_{[1]} \right\|_2 \\ &= \left\| (\mathbf{Q}_{[1]}^H \mathbf{Q}_{[1]} - \lambda \mathbf{I})\mathbf{f}_{[1]} + \sum_{i=2}^{\prod_{t=1}^n s_t} \mathbf{Q}_{[1]}^H \mathbf{Q}_{[i]} \mathbf{f}_{[i]} \right\|_2 \\ &\geq \left\| \mathbf{Q}_{[1]}^H \mathbf{Q}_{[1]} \mathbf{f}_{[1]} \right\|_2 - \lambda \|\mathbf{f}_{[1]}\|_2 - \left\| \sum_{i=2}^{\prod_{t=1}^n s_t} \mathbf{Q}_{[1]}^H \mathbf{Q}_{[i]} \mathbf{f}_{[i]} \right\|_2 \\ &\geq \left\| \mathbf{Q}_{[1]}^H \mathbf{Q}_{[1]} \mathbf{f}_{[1]} \right\|_2 - \lambda \|\mathbf{f}_{[1]}\|_2 \\ &\quad - \sum_{t=0}^{n-1} C_t \varpi_{\mathbf{r},t}^n \prod_{i=1}^n d_i \max_{l \in \{C_t\}} \|\mathbf{f}_{[l]}\|_2, \end{aligned} \quad (67)$$

where the set  $\{C_t\}$  consists of the indices corresponding to  $C_t$  as defined in **Proposition 1**. Denote  $\mathbf{Q}_{[1]}^H \mathbf{Q}_{[1]} = \mathbf{I} + \mathbf{A}$ , where  $\forall i, \mathbf{A}_{(i,i)} = 0$ . Based on Geršgorin's disc theorem [37]

$$\|\mathbf{A}\|_2 \leq \sum_{t=0}^{n-1} C_t \tau_{\mathbf{r},t}^{n-t}.$$

Thus,

$$\left\| \mathbf{Q}_{[1]}^H \mathbf{Q}_{[1]} \right\|_2 \geq \|\mathbf{I}\|_2 - \|\mathbf{A}\|_2 \geq 1 - \sum_{t=0}^{n-1} C_t \tau_{\mathbf{r},t}^{n-t}. \quad (68)$$

Combining (67) and (68) yields

$$\left\| [(\mathbf{Q}^H \mathbf{Q} - \lambda \mathbf{I})\mathbf{f}]_{[1]} \right\|_2$$

$$\begin{aligned} &\geq \left( 1 - \sum_{t=0}^{n-1} C_t \tau_{\mathbf{r},t}^{n-t} - \lambda \right) \|\mathbf{f}_{[1]}\|_2 \\ &\quad - \sum_{t=0}^{n-1} C_t \varpi_{\mathbf{r},t}^n \prod_{i=1}^n d_i \max_{l \in \{C_t\}} \|\mathbf{f}_{[l]}\|_2 \\ &> \sum_{t=0}^{n-1} C_t \varpi_{\mathbf{r},t}^n \prod_{i=1}^n d_i \|\mathbf{f}_{[1]}\|_2 \\ &\quad - \sum_{t=0}^{n-1} C_t \varpi_{\mathbf{r},t}^n \prod_{i=1}^n d_i \max_{l \in \{C_t\}} \|\mathbf{f}_{[l]}\|_2 \\ &= \sum_{t=0}^{n-1} C_t \varpi_{\mathbf{r},t}^n \prod_{i=1}^n d_i \left( \|\mathbf{f}_{[1]}\|_2 - \max_{l \in \{C_t\}} \|\mathbf{f}_{[l]}\|_2 \right) \geq 0. \end{aligned}$$

This reveals that  $(\mathbf{Q}^H \mathbf{Q} - \lambda \mathbf{I})\mathbf{f} \neq \mathbf{0}$  and  $\lambda_{\min} \geq 1 - \sum_{t=0}^{n-1} C_t \tau_{\mathbf{r},t}^{n-t} - \sum_{t=0}^{n-1} C_t \varpi_{\mathbf{r},t}^n \prod_{i=1}^n d_i$  holds.

The inequality  $\lambda_{\max} \leq 1 + \sum_{t=0}^{n-1} C_t \tau_{\mathbf{r},t}^{n-t} + \sum_{t=0}^{n-1} C_t \varpi_{\mathbf{r},t}^n \prod_{i=1}^n d_i$  can be proved similarly. ■

### B. Proof of Lemma 2

*Proof:* Based on matrix norm property, we have

$$\begin{aligned} \|\mathbf{D}\|_2 &= \sqrt{\|\mathbf{D}^H \mathbf{D}\|_2} = \sqrt{\rho(\mathbf{D}^H \mathbf{D})} \\ &\leq \sqrt{\min\{\rho_r(\mathbf{D}^H \mathbf{D}), \rho_c(\mathbf{D}^H \mathbf{D})\}} \end{aligned} \quad (69)$$

$$\leq \sqrt{\min\{\rho_r(\mathbf{D}^H) \rho_r(\mathbf{D}), \rho_c(\mathbf{D}^H) \rho_c(\mathbf{D})\}} \quad (70)$$

$$\begin{aligned} &= \sqrt{\min\{\rho_c(\mathbf{D}) \rho_r(\mathbf{D}), \rho_r(\mathbf{D}) \rho_c(\mathbf{D})\}} \\ &= \sqrt{\rho_c(\mathbf{D}) \rho_r(\mathbf{D})}, \end{aligned} \quad (71)$$

where  $\rho(\cdot)$  denotes the spectral radius on its objective, (69) is because  $\rho_r(\mathbf{D})$  and  $\rho_c(\mathbf{D})$  are matrix norms as discussed in [10], and the spectral radius of a square matrix is less than or equal to any other matrix norm, (70) follows from the sub-multiplicativity of matrix norm, and (71) is because  $\rho_r(\mathbf{D}^H) = \rho_c(\mathbf{D})$  and  $\rho_c(\mathbf{D}^H) = \rho_r(\mathbf{D})$  [10]. ■

### C. Proof of Corollary 2

*Proof:* Observe that

$$\left\| \mathbf{T}^H \mathbf{T}_* \right\|_2 \leq \min\{\rho_r(\mathbf{T}^H \mathbf{T}_*), \rho_c(\mathbf{T}^H \mathbf{T}_*)\} \quad (72)$$

$$\leq \min\left\{ \varpi_{\mathbf{r}}^n \prod_{t=1}^n l_t d_t, \varpi_{\mathbf{r}}^n \prod_{t=1}^n l_t^* d_t \right\}, \quad (73)$$

where (72) is from **Lemma 2**, and (73) is from the definition of mutual block coherence. On the other hand,

$$\left\| \mathbf{T}^H \mathbf{T}_* \right\|_2$$

$$\geq \sigma_{\min}(\mathbf{T}) \sigma_{\min}(\mathbf{T}_*)$$

$$\geq \left( 1 - \left( \prod_{t=1}^n d_t - 1 \right) \tau_{\mathbf{r}}^n - \left( \prod_{t=1}^n l_t - 1 \right) \varpi_{\mathbf{r}}^n \prod_{t=1}^n d_t \right)^{\frac{1}{2}}$$

$$\times \left( 1 - \left( \prod_{t=1}^n d_t - 1 \right) \tau_{\mathbf{r}}^n - \left( \prod_{t=1}^n l_t^* - 1 \right) \varpi_{\mathbf{r}}^n \prod_{t=1}^n d_t \right)^{\frac{1}{2}}$$

$$= \underline{W}_{\mathbf{r}, \prod_{t=1}^n l_t}^{\frac{1}{2}} \underline{W}_{\mathbf{r}, \prod_{t=1}^n l_t}^{\frac{1}{2}}, \quad (74)$$

where (74) is derived based on **Lemma 2**, and  $\sigma_{\min}(\cdot)$  represents the minimum singular value of its target. ■

#### D. Proof of Corollary 3

*Proof:* Observe that

$$\begin{aligned} & \|\mathcal{X} \times_1 \mathbf{D}_1 \times_2 \cdots \times_n \mathbf{D}_n\|_F \\ &= \|\mathcal{X}_{\Xi} \times_1 \bar{\mathbf{D}}_1 \times_2 \cdots \times_n \bar{\mathbf{D}}_n\|_F \\ &= \|(\bar{\mathbf{D}}_n \otimes \bar{\mathbf{D}}_{n-1} \otimes \cdots \otimes \bar{\mathbf{D}}_1) \text{vec}(\mathcal{X}_{\Xi})\|_F \\ &\leq \|\bar{\mathbf{D}}_n \otimes \bar{\mathbf{D}}_{n-1} \otimes \cdots \otimes \bar{\mathbf{D}}_1\|_2 \|\text{vec}(\mathcal{X}_{\Xi})\|_F \\ &\leq \bar{W}_{\mathbf{r}, \prod_{t=1}^n k_t}^{\frac{1}{2}} \|\mathcal{X}\|_F, \end{aligned} \quad (75)$$

where  $\bar{\mathbf{D}}_i$  ( $i \in [1, 2, \dots, n]$ ) are the column-block submatrices of  $\mathbf{D}_i$  whose column indices correspond to the shadow block sparsity of  $\mathcal{X}$ ,  $\text{vec}(\mathcal{X}) \in \mathbb{C}^{k_1 d_1 k_2 d_2 \cdots k_n d_n}$  is the vectorized  $\mathcal{X}_{\Xi}$ , and (75) is based on **Lemma 1** and due to the fact that  $\|\text{vec}(\mathcal{X}_{\Xi})\|_F = \|\mathcal{X}\|_F$ . On the other hand,

$$\begin{aligned} & \|\mathcal{X} \times_1 \mathbf{D}_1 \times_2 \cdots \times_n \mathbf{D}_n\|_F \\ &\geq \sigma_{\min}(\bar{\mathbf{D}}_n \otimes \bar{\mathbf{D}}_{n-1} \otimes \cdots \otimes \bar{\mathbf{D}}_1) \|\text{vec}(\mathcal{X}_{\Xi})\|_F \\ &\geq \underline{W}_{\mathbf{r}, \prod_{t=1}^n k_t}^{\frac{1}{2}} \|\mathcal{X}\|_F, \end{aligned} \quad (76)$$

where  $\sigma_{\min}(\cdot)$  denotes the smallest singular value of its argument, and (76) is based on **Lemma 1**.

The proofs of (22) and (23) are similar, and the proof of (24) is essentially an extension of those of the [35, Proposition 3.1] and [36, Lemma 6]. ■

#### E. Proof of Theorem 1

*Proof:* Note that

$$\begin{aligned} & \|\mathcal{R}^l \times_1 \mathbf{D}_{1[i_1^*]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n^*]}^H\|_F \\ &= \max_{(i_1, i_2, \dots, i_n) \in \Xi} \|\mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H\|_F. \end{aligned} \quad (77)$$

Letting  $q_s$  be the  $s$ th largest element in  $\Psi$ , we have

$$q_s \leq \frac{1}{\sqrt{s}} \sqrt{\sum_{(i_1, \dots, i_n) \in \Psi} \|\mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H\|_F^2}. \quad (78)$$

If the bound in (78) is less than (77), i.e.,

$$\frac{\sqrt{\sum_{(i_1, \dots, i_n) \in \Psi} \|\mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H\|_F^2}}{\sqrt{s} \|\mathcal{R}^l \times_1 \mathbf{D}_{1[i_1^*]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n^*]}^H\|_F} < 1, \quad (79)$$

then in the next iteration at least  $c$  correct indices are selected. Since  $\bar{\mathbf{D}}_{\Xi} \bar{\mathbf{D}}_{\Xi}^{\dagger}$  is the orthogonal projector onto  $\text{span}(\bar{\mathbf{D}}_{\Xi})$  [6], we have  $\bar{\mathbf{D}}_{\Xi} \bar{\mathbf{D}}_{\Xi}^{\dagger} \text{vec}(\mathcal{R}^l) = \text{vec}(\mathcal{R}^l)$ . Thus, the left-hand side of (79) becomes

$$\frac{\sqrt{\sum_{(i_1, \dots, i_n) \in \Psi} \|(\mathbf{D}_{n[i_n]}^H \otimes \cdots \otimes \mathbf{D}_{1[i_1]}^H) \bar{\mathbf{D}}_{\Xi} \bar{\mathbf{D}}_{\Xi}^{\dagger} \text{vec}(\mathcal{R}^l)\|_F^2}}{\sqrt{s} \|(\mathbf{D}_{n[i_n^*]}^H \otimes \cdots \otimes \mathbf{D}_{1[i_1^*]}^H) \text{vec}(\mathcal{R}^l)\|_F}$$

$$= \frac{\sqrt{\sum_{(i_1, \dots, i_n) \in \Psi} \|(\mathbf{D}_{n[i_n]}^H \otimes \cdots \otimes \mathbf{D}_{1[i_1]}^H) (\bar{\mathbf{D}}_{\Xi}^{\dagger})^H \bar{\mathbf{D}}_{\Xi}^H \text{vec}(\mathcal{R}^l)\|_F^2}}{\sqrt{s} \|(\mathbf{D}_{n[i_n^*]}^H \otimes \cdots \otimes \mathbf{D}_{1[i_1^*]}^H) \text{vec}(\mathcal{R}^l)\|_F} \quad (80)$$

$$\leq \frac{1}{\sqrt{s}} \|\bar{\mathbf{D}}_{\Psi}^H (\bar{\mathbf{D}}_{\Xi}^{\dagger})^H\|_2 \frac{\|\bar{\mathbf{D}}_{\Xi}^H \text{vec}(\mathcal{R}^l)\|_F}{\|(\mathbf{D}_{n[i_n^*]}^H \otimes \cdots \otimes \mathbf{D}_{1[i_1^*]}^H) \text{vec}(\mathcal{R}^l)\|_F} \quad (81)$$

$$= \frac{Z_{\mathcal{R}^l}}{\sqrt{s}} \|\bar{\mathbf{D}}_{\Xi}^{\dagger} \bar{\mathbf{D}}_{\Psi}\|_2, \quad (82)$$

where (80) is because  $\bar{\mathbf{D}}_{\Xi} \bar{\mathbf{D}}_{\Xi}^{\dagger}$  is Hermitian matrix, (81) is from the sub-multiplicativity of matrix norm, and (82) is due to the definition of the function  $Z_{\mathcal{R}^l}$ . Then, following the similar proofs in [6], [10], we can prove that T-GBOMP selects a new tensor block participating in the unique solution of  $\mathcal{Y} = \mathcal{X} \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2 \times_3 \cdots \times_n \mathbf{D}_n$ . ■

#### F. Proof of Theorem 2

*Proof:* Observe that

$$\begin{aligned} & \frac{Z_{\mathcal{R}^l}}{\sqrt{s}} \|\bar{\mathbf{D}}_{\Xi}^{\dagger} \bar{\mathbf{D}}_{\Psi}\|_2 \\ &= \frac{Z_{\mathcal{R}^l}}{\sqrt{s}} \|(\bar{\mathbf{D}}_{\Xi}^H \bar{\mathbf{D}}_{\Xi})^{-1} \bar{\mathbf{D}}_{\Xi}^H \bar{\mathbf{D}}_{\Psi}\|_2 \\ &\leq \frac{Z_{\mathcal{R}^l}}{\sqrt{s}} \|(\bar{\mathbf{D}}_{\Xi}^H \bar{\mathbf{D}}_{\Xi})^{-1}\|_2 \|\bar{\mathbf{D}}_{\Xi}^H \bar{\mathbf{D}}_{\Psi}\|_2 \end{aligned} \quad (83)$$

$$\leq \frac{Z_{\mathcal{R}^l}}{\sqrt{s}} \|(\bar{\mathbf{D}}_{\Xi}^H \bar{\mathbf{D}}_{\Xi})^{-1}\|_2 \times \max \left\{ \varpi_{\mathbf{r}}^n k \prod_{t=1}^n d_t, \varpi_{\mathbf{r}}^n s \prod_{t=1}^n d_t \right\} \quad (84)$$

$$\leq \frac{Z_{\mathcal{R}^l}}{\sqrt{s}} \|(\bar{\mathbf{D}}_{\Xi}^H \bar{\mathbf{D}}_{\Xi})^{-1}\|_2 \varpi_{\mathbf{r}}^n k \prod_{t=1}^n d_t, \quad (85)$$

where (83) is from the sub-multiplicativity, (84) is based on **Corollary 3**, and (85) is because  $s \leq k$ .

It remains to derive an upper bound on  $\|(\bar{\mathbf{D}}_{\Xi}^H \bar{\mathbf{D}}_{\Xi})^{-1}\|_2$ . Let  $\bar{\mathbf{D}}_{\Xi}^H \bar{\mathbf{D}}_{\Xi} = \mathbf{I} + \mathbf{A}$ , where  $\mathbf{A} \in \mathbb{C}^{k \prod_{t=1}^n d_t \times k \prod_{t=1}^n d_t}$  is a symmetric matrix with  $\mathbf{A}_{(i,i)} = 0, \forall i$ . Thus, based on **Corollary 1**, we have

$$\|\mathbf{A}\|_2 \leq \left( \prod_{t=1}^n d_t - 1 \right) \tau_{\mathbf{r}}^n + (k-1) \varpi_{\mathbf{r}}^n \prod_{t=1}^n d_t. \quad (86)$$

The assumption now implies that  $\left( \prod_{t=1}^n d_t - 1 \right) \tau_{\mathbf{r}}^n + (k-1) \varpi_{\mathbf{r}}^n \prod_{t=1}^n d_t < 1$ . Based on [37, Corollary 5.6.16], we obtain

$$\|(\bar{\mathbf{D}}_{\Xi}^H \bar{\mathbf{D}}_{\Xi})^{-1}\|_2 = \|(\mathbf{I} + \mathbf{A})^{-1}\|_2 = \left\| \sum_{t=0}^{\infty} (-\mathbf{A})^t \right\|_2 \quad (87)$$

$$\leq \sum_{t=0}^{\infty} \|\mathbf{A}\|_2^t \quad (88)$$

$$= \frac{1}{1 - \|\mathbf{A}\|_2} \quad (89)$$

$$\leq \frac{1}{1 - \left( \prod_{t=1}^n d_t - 1 \right) \tau_{\mathbf{r}}^n - (k-1) \varpi_{\mathbf{r}}^n \prod_{t=1}^n d_t}, \quad (90)$$

where (88) is from the triangle inequality, (89) is the sum of proportional sequence, and (90) follows by using (86).

Combining (90) and (85) yields

$$\frac{Z_{\mathcal{R}^l}}{\sqrt{s}} \|\ddot{\mathbf{D}}_{\Xi}^{\dagger} \ddot{\mathbf{D}}_{\Psi}\|_2 \leq \frac{Z_{\mathcal{R}^l}}{\sqrt{s}} \frac{\varpi_{\mathbf{r}}^n k \prod_{t=1}^n d_t}{1 - \left( \prod_{t=1}^n d_t - 1 \right) \tau_{\mathbf{r}}^n - (k-1) \varpi_{\mathbf{r}}^n \prod_{t=1}^n d_t} \quad (91)$$

$$< 1. \quad (92)$$

Then, (31) follows from (91) and (92), and this completes the proof.  $\blacksquare$

### G. Proof of Corollary 4

*Proof:* Letting  $Z_{\mathcal{R}^l} = \sqrt{k}$ , we obtain the following cubic inequality with respect to  $k$ :

$$\left(k^{\frac{1}{2}}\right)^3 + \sqrt{s} \left(k^{\frac{1}{2}}\right)^2 + \delta < 0, \quad (93)$$

where  $\delta$  is given by (39) in Corollary 4. Setting  $k^{\frac{1}{2}} = x - \frac{\sqrt{s}}{3}$ , (93) changes into

$$x^3 + Px + Q < 0. \quad (94)$$

The real-domain solution of (94) based on the Cardano formula is

$$x < \sqrt[3]{-\frac{Q}{2} + \sqrt{\Delta}} + \sqrt[3]{-\frac{Q}{2} - \sqrt{\Delta}}. \quad (95)$$

(38) is the consequence of combining (95) and  $k^{\frac{1}{2}} = x - \frac{\sqrt{s}}{3}$ .  $\blacksquare$

### H. Proof of Theorem 3

*Proof:* Note that

$$\|\mathcal{R}^l - \mathcal{R}^{l+1}\|_F = \left\| \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}^{\perp} \text{vec}(\mathcal{Y}) - \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^{l+1}}}^{\perp} \text{vec}(\mathcal{Y}) \right\|_F \quad (96)$$

$$= \left\| (\mathbf{I} - \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}} \text{vec}(\mathcal{Y})) - (\mathbf{I} - \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^{l+1}}}) \text{vec}(\mathcal{Y}) \right\|_F \quad (97)$$

$$= \left\| \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^{l+1}}} \text{vec}(\mathcal{Y}) - \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^{l+1}}} \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}} \text{vec}(\mathcal{Y}) \right\|_F \quad (98)$$

$$= \left\| \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^{l+1}}} (\mathbf{I} - \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}) \text{vec}(\mathcal{Y}) \right\|_F = \left\| \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^{l+1}}} \text{vec}(\mathcal{R}^l) \right\|_F \quad (99)$$

$$\geq \left\| \mathbf{P}_{\ddot{\mathbf{D}}_{\Theta^{l+1}}} \text{vec}(\mathcal{R}^l) \right\|_F \quad (100)$$

$$= \left\| (\mathbf{P}_{\ddot{\mathbf{D}}_{\Theta^{l+1}}})^H \text{vec}(\mathcal{R}^l) \right\|_F \quad (101)$$

$$= \left\| (\ddot{\mathbf{D}}_{\Theta^{l+1}} \ddot{\mathbf{D}}_{\Theta^{l+1}}^{\dagger})^H \text{vec}(\mathcal{R}^l) \right\|_F$$

$$= \left\| (\ddot{\mathbf{D}}_{\Theta^{l+1}}^{\dagger})^H \ddot{\mathbf{D}}_{\Theta^{l+1}}^H \text{vec}(\mathcal{R}^l) \right\|_F$$

$$\geq \frac{\left\| \ddot{\mathbf{D}}_{\Theta^{l+1}}^H \text{vec}(\mathcal{R}^l) \right\|_F}{\overline{W}_{\mathbf{r},s}^{\frac{1}{2}}}, \quad (101)$$

where  $\Theta^{l+1}$  is the selected index set in the  $(l+1)$ th iteration as given in Algorithm 1, (96) and (97) are due to the definitions of the residual matrix and orthogonal projection matrix, respectively, (98) and (99) are because  $\text{span}(\ddot{\mathbf{D}}_{\Xi^l}) \subseteq \text{span}(\ddot{\mathbf{D}}_{\Xi^{l+1}})$

and  $\text{span}(\ddot{\mathbf{D}}_{\Theta^{l+1}}) \subseteq \text{span}(\ddot{\mathbf{D}}_{\Xi^{l+1}})$ , respectively, (100) is from the property of projection matrix, and (101) is based on Corollary 3.

It remains to develop the lower bound on  $\left\| \ddot{\mathbf{D}}_{\Theta^{l+1}}^H \text{vec}(\mathcal{R}) \right\|_F$ . Let  $\Theta^* = \arg \max_{\Theta: |\Theta|=s} \sum_{(i_1, i_2, \dots, i_n) \in \Theta} \left\| \mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \right\|_F^2$ . Then, we have

$$\begin{aligned} & \left\| \ddot{\mathbf{D}}_{\Theta^{l+1}}^H \text{vec}(\mathcal{R}) \right\|_F^2 \\ &= \sum_{(i_1, i_2, \dots, i_n) \in \Theta^{l+1}} \left\| \mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \right\|_F^2 \\ &= \max_{\Theta: |\Theta|=s} \sum_{(i_1, i_2, \dots, i_n) \in \Theta} \left\| \mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \right\|_F^2 \\ &\geq \sum_{(i_1, i_2, \dots, i_n) \in \Theta^*} \left\| \mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \right\|_F^2 \\ &= \max_{\Theta: |\Theta|=s} \sum_{(i_1, i_2, \dots, i_n) \in \Theta} \left\| \mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \right\|_F^2. \end{aligned} \quad (102)$$

Combining (101) and (102), we have

$$\begin{aligned} & \left\| \mathcal{R}^l \right\|_F^2 - \left\| \mathcal{R}^{l+1} \right\|_F^2 \\ &= \left\| \mathcal{R}^l - \mathcal{R}^{l+1} \right\|_F^2 \\ &\geq \frac{\max_{\Theta: |\Theta|=s} \sum_{(i_1, i_2, \dots, i_n) \in \Theta} \left\| \mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \right\|_F^2}{\overline{W}_{\mathbf{r},s}}. \end{aligned} \quad (103)$$

Moreover, since  $s \leq k$ ,

$$\begin{aligned} & \max_{\Theta: |\Theta|=s} \sum_{(i_1, i_2, \dots, i_n) \in \Theta} \left\| \mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \right\|_F^2 \\ &\geq \frac{s}{k} \sum_{(i_1, \dots, i_n) \in \Xi} \left\| \mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \right\|_F^2 \\ &\geq \frac{s}{k} \sum_{(i_1, \dots, i_n) \in \Xi \cup \Xi^l} \left\| \mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \right\|_F^2 \end{aligned} \quad (104)$$

$$= \frac{s}{k} \left\| \ddot{\mathbf{D}}_{\Xi \cup \Xi^l}^H \left( \text{vec}(\mathcal{Y}) - (\mathbf{D}_{n_{\Xi^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi^l}}) \text{vec}(\mathcal{X}_{\Xi^l}) \right) \right\|_F^2 \quad (105)$$

$$= \frac{s}{k} \left\| \ddot{\mathbf{D}}_{\Xi \cup \Xi^l}^H \left( (\mathbf{D}_{n_{\Xi}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi}}) \text{vec}(\mathcal{X}) - (\mathbf{D}_{n_{\Xi^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi^l}}) \text{vec}(\mathcal{X}_{\Xi^l}) \right) \right\|_F^2$$

$$= \frac{s}{k} \left\| \ddot{\mathbf{D}}_{\Xi \cup \Xi^l}^H \ddot{\mathbf{D}}_{\Xi \cup \Xi^l} \left( \text{vec}(\mathcal{X}_{\Xi}) - \text{vec}(\mathcal{X}_{\Xi^l}) \right) \right\|_F^2 \quad (106)$$

$$\geq \frac{s}{k} \overline{W}_{\mathbf{r}, \prod_{t=1}^n k_n + s} \left\| \text{vec}(\mathcal{X}_{\Xi}) - \text{vec}(\mathcal{X}_{\Xi^l}) \right\|_F^2 \quad (107)$$

$$\geq \frac{s \overline{W}_{\mathbf{r}, \prod_{t=1}^n k_n + s}}{k \overline{W}_{\mathbf{r}, \prod_{t=1}^n k_n + s}^{\frac{1}{2}}} \left\| \ddot{\mathbf{D}}_{\Xi \cup \Xi^l} \left( \text{vec}(\mathcal{X}_{\Xi}) - \text{vec}(\mathcal{X}_{\Xi^l}) \right) \right\|_F^2 \quad (108)$$

$$= \frac{s \overline{W}_{\mathbf{r}, \prod_{t=1}^n k_n + s}}{k \overline{W}_{\mathbf{r}, \prod_{t=1}^n k_n + s}^{\frac{1}{2}}} \left\| \mathcal{R}^l \right\|_F^2, \quad (109)$$

where (104) is because  $\sum_{(i_1, \dots, i_n) \in \Xi^l} \|\mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \dots \times_n \mathbf{D}_{n[i_n]}^H\|_F^2 = 0$ , (105) is due to (1), (106) is because  $(\mathbf{D}_{n_{\Xi_n}} \otimes \dots \otimes \mathbf{D}_{1_{\Xi_1}}) \text{vec}(\mathcal{X}_{\Xi}) = (\mathbf{D}_{n_{\Xi_n \cup \Xi_1^l}} \otimes \dots \otimes \mathbf{D}_{1_{\Xi_1 \cup \Xi_1^l}}) \text{vec}(\mathcal{X}_{\Xi \cup \Xi^l})$ , (107) and (108) are from **Corollary 3**. Combining (103) and (109) yields

$$\|\mathcal{R}^l\|_F^2 - \|\mathcal{R}^{l+1}\|_F^2 \geq \frac{sW_{\mathbf{r}, \prod_{t=1}^n k_n + sl}}{k\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_n + sl} \overline{W}_{\mathbf{r}, s}} \|\mathcal{R}^l\|_F^2.$$

Thus, we have

$$\begin{aligned} \|\mathcal{R}^{l+1}\|_F^2 &\leq \left(1 - \frac{sW_{\mathbf{r}, \prod_{t=1}^n k_n + sl}}{k\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_n + sl} \overline{W}_{\mathbf{r}, s}}\right) \|\mathcal{R}^l\|_F^2 \\ &\leq \prod_{j=0}^l \left(1 - \frac{sW_{\mathbf{r}, \prod_{t=1}^n k_n + sj}}{k\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_n + sj} \overline{W}_{\mathbf{r}, s}}\right) \|\mathcal{R}^0\|_F^2 \\ &\leq \left(1 - \frac{sW_{\mathbf{r}, \prod_{t=1}^n k_n + sl}}{k\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_n + sl} \overline{W}_{\mathbf{r}, s}}\right)^{l+1} \|\mathcal{Y}\|_F^2. \end{aligned}$$

### I. Proof of Theorem 4

*Proof:* Denoting  $\mathcal{Q}^{l*}$  as the best  $k$  block-sparse tensor approximation of  $\mathcal{X}^{l*}$ , and  $k_n$  as the shadow block sparsity in the  $n$ th mode, we have

$$\begin{aligned} &\|\mathcal{Q}^{l*} - \mathcal{X}\|_F \\ &= \|\mathcal{Q}^{l*} - \mathcal{X}^{l*} + \mathcal{X}^{l*} - \mathcal{X}\|_F \\ &\leq \|\mathcal{Q}^{l*} - \mathcal{X}^{l*}\|_F + \|\mathcal{X}^{l*} - \mathcal{X}\|_F \end{aligned} \quad (110)$$

$$\leq 2\|\mathcal{X}^{l*} - \mathcal{X}\|_F \quad (111)$$

$$\leq \frac{2\|(\mathcal{X}^{l*} - \mathcal{X}) \times_1 \mathbf{D}_1 \times_2 \dots \times_n \mathbf{D}_n\|_F}{\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + sl^*}} \quad (112)$$

$$\leq \frac{2(\|\mathcal{R}^{l*}\|_F + \|\mathcal{N}\|_F)}{\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + sl^*}} \quad (113)$$

$$\leq \frac{2(\epsilon + \|\mathcal{N}\|_F)}{\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + sl^*}}, \quad (114)$$

where (110) is due to the triangle inequality, (111) is because  $\mathcal{Q}^{l*}$  is the best  $k$  block-sparse tensor approximation to  $\mathcal{X}^{l*}$ , (112) is from **Corollary 3** and the fact that each shadow block sparsity of the tensor  $\mathcal{X}^{l*} - \mathcal{X}$  can increase by up to  $sl^*$ , and (113) is because  $\mathcal{R}^{l*} = \mathcal{Y} - \mathcal{X}^{l*} \times_1 \mathbf{D}_1 \times_2 \dots \times_n \mathbf{D}_n = (\mathcal{X}^{l*} - \mathcal{X}) \times_1 \mathbf{D}_1 \times_2 \dots \times_n \mathbf{D}_n + \mathcal{N}$ .

Further,

$$\begin{aligned} &\|\mathcal{Q}^{l*} - \mathcal{X}\|_F \\ &\geq \frac{\|(\mathcal{Q}^{l*} - \mathcal{X}) \times_1 \mathbf{D}_1 \times_2 \dots \times_n \mathbf{D}_n\|_F}{\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k}} \end{aligned} \quad (115)$$

$$\begin{aligned} &= \frac{\|\mathcal{Q}^{l*} \times_1 \mathbf{D}_1 \times_2 \dots \times_n \mathbf{D}_n - \mathcal{Y} + \mathcal{N}\|_F}{\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k}} \\ &\geq \frac{\|\mathcal{Q}^{l*} \times_1 \mathbf{D}_1 \times_2 \dots \times_n \mathbf{D}_n - \mathcal{Y}\|_F - \|\mathcal{N}\|_F}{\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k}} \end{aligned} \quad (116)$$

$$\geq \frac{\|\hat{\mathcal{X}} \times_1 \mathbf{D}_1 \times_2 \dots \times_n \mathbf{D}_n - \mathcal{Y}\|_F - \|\mathcal{N}\|_F}{\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k}} \quad (117)$$

$$= \frac{\|(\hat{\mathcal{X}} - \mathcal{X}) \times_1 \mathbf{D}_1 \times_2 \dots \times_n \mathbf{D}_n - \mathcal{N}\|_F - \|\mathcal{N}\|_F}{\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k}}$$

$$\geq \frac{\|(\hat{\mathcal{X}} - \mathcal{X}) \times_1 \mathbf{D}_1 \times_2 \dots \times_n \mathbf{D}_n\|_F - 2\|\mathcal{N}\|_F}{\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k}} \quad (118)$$

$$\geq \frac{W_{\mathbf{r}, \prod_{t=1}^n k_t + k} \|(\hat{\mathcal{X}} - \mathcal{X})\|_F - 2\|\mathcal{N}\|_F}{\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k}}, \quad (119)$$

where (115) and (119) are derived by **Corollary 3** and due to the fact that the  $n$ th mode shadow block sparsity of  $\mathcal{Q}^{l*} - \mathcal{X}$  is up to  $k_n + k$ , (116) and (118) are based on the triangle inequality, and (117) is because  $\mathcal{Q}^{l*}$  is supported on  $\Xi^{l*}$  and  $\hat{\mathcal{X}} = \arg \min_{\mathcal{X}: \text{supp}(\mathcal{X}) = \Xi^{l*}} \|\text{vec}(\mathcal{Y}) - \sum_{(i_1, i_2, \dots, i_n) \in \Xi^{l*}} (\mathbf{D}_{n[i_n]} \otimes \dots \otimes \mathbf{D}_{1[i_1]}) \text{vec}(\mathcal{X})_{[i_1, \dots, i_n]}\|_F$ . Finally, the proof is completed by combining (114) and (119). ■

### J. Proof of Theorem 5

*Proof:* The proof is based on induction of selecting at least one correct tensor block for each iteration. Note that this proof technique is similar in spirit to those in [36] and [43].

a) First, we prove that T-GBOMP succeeds at choosing the correct index in the first iteration. According to Algorithm 1, T-GBOMP chooses the index set

$$\begin{aligned} &\Xi^1 \\ &= \arg \max_{\Theta: |\Theta|=s} \sum_{(i_1, \dots, i_n) \in \Theta} \|\mathcal{R}^0 \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \dots \times_n \mathbf{D}_{n[i_n]}^H\|_F \end{aligned} \quad (120)$$

$$= \arg \max_{\Theta: |\Theta|=s} \sum_{(i_1, \dots, i_n) \in \Theta} \|\mathcal{Y} \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \dots \times_n \mathbf{D}_{n[i_n]}^H\|_F, \quad (121)$$

where

$$\begin{aligned} \Theta = \{ &(\Theta_{1_1}, \Theta_{2_1}, \dots, \Theta_{n_1}), (\Theta_{1_2}, \Theta_{2_2}, \dots, \Theta_{n_2}), \\ &\dots, (\Theta_{1_k}, \Theta_{2_k}, \dots, \Theta_{n_k}) \}. \end{aligned}$$

Since  $s \leq k$ , we have

$$\begin{aligned} &\sum_{(i_1, \dots, i_n) \in \Xi^1} \|\mathcal{Y} \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \dots \times_n \mathbf{D}_{n[i_n]}^H\|_F \\ &\geq \sqrt{\frac{s}{k}} \sum_{(i_1, \dots, i_n) \in \Xi} \|\mathcal{Y} \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \dots \times_n \mathbf{D}_{n[i_n]}^H\|_F \end{aligned} \quad (120)$$

$$\begin{aligned} &= \sqrt{\frac{s}{k}} \sum_{(i_1, \dots, i_n) \in \Xi} \|(\mathcal{X} \times_1 \mathbf{D}_1 \times_2 \dots \times_n \mathbf{D}_n + \mathcal{N}) \\ &\quad \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \dots \times_n \mathbf{D}_{n[i_n]}^H\|_F \end{aligned} \quad (121)$$

$$\begin{aligned} &\geq \sqrt{\frac{s}{k}} \sum_{(i_1, \dots, i_n) \in \Xi} \left( \|(\mathcal{X} \times_1 \mathbf{D}_1 \times_2 \dots \times_n \mathbf{D}_n) \right. \\ &\quad \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \dots \times_n \mathbf{D}_{n[i_n]}^H\|_F \\ &\quad \left. - \|\mathcal{N} \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \dots \times_n \mathbf{D}_{n[i_n]}^H\|_F \right) \end{aligned} \quad (122)$$

$$\begin{aligned}
&= \sqrt{\frac{s}{k}} \sum_{(i_1, \dots, i_n) \in \Xi} \left( \|\mathcal{X}_{\Xi} \times_1 (\mathbf{D}_{1[i_1]}^H \mathbf{D}_{1\Xi_1}) \right. \\
&\quad \times_2 \cdots \times_n (\mathbf{D}_{n[i_n]}^H \mathbf{D}_{n\Xi_n}) \|_F \\
&\quad \left. - \|\mathcal{N} \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \|_F \right) \\
&\geq \sqrt{sk} \left( \overline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t}^{\frac{1}{2}} \overline{W}_{\mathbf{Y}, 1}^{\frac{1}{2}} \|\mathcal{X}\|_F - \overline{W}_{\mathbf{Y}, 1}^{\frac{1}{2}} \|\mathcal{N}\|_F \right), \quad (123)
\end{aligned}$$

where  $\mathbf{D}_{1\Xi_1} \in \mathbb{C}^{M_1 \times sd}$ ,  $\mathbf{D}_{2\Xi_2} \in \mathbb{C}^{M_2 \times sd}$ ,  $\dots$ ,  $\mathbf{D}_{n\Xi_n} \in \mathbb{C}^{M_n \times sd}$  indexed by  $\Xi_1, \Xi_2, \dots, \Xi_n$  are the column-block submatrices of  $\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_n$ , respectively, (120) is due to the fact that the average of the correlation between  $s$  selected blocks and the residual in the first iteration is the largest and should be no less than that of any other subset with cardinality being no less than  $s$ , (121) is from (1), (122) follows from the triangle inequality, and (123) is derived based on **Corollary 3**.

Consider the case in which no correct index is selected at the first iteration, i.e., there exists at least one tensor index set satisfying that  $\Xi_i \cap \Xi_i^1 = \emptyset$  ( $i \in \{1, 2, \dots, n\}$ ); we have

$$\begin{aligned}
&\sum_{(i_1, \dots, i_n) \in \Xi^1} \|\mathcal{Y} \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \|_F \\
&= \sum_{(i_1, \dots, i_n) \in \Xi^1} \left\| (\mathcal{X}_{\Xi} \times_1 \mathbf{D}_{1\Xi_1} \times_2 \cdots \times_n \mathbf{D}_{n\Xi_n} + \mathcal{N}) \right. \\
&\quad \left. \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \|_F \right. \\
&\leq \sum_{(i_1, \dots, i_n) \in \Xi^1} \left( \|\mathcal{X}_{\Xi} \times_1 \mathbf{D}_{1\Xi_1} \times_2 \cdots \times_n \mathbf{D}_{n\Xi_n} \right. \\
&\quad \left. \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \|_F \right. \\
&\quad \left. + \|\mathcal{N} \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \|_F \right) \\
&= \sum_{(i_1, \dots, i_n) \in \Xi^1} \left( \|\mathcal{X}_{\Xi} \times_1 (\mathbf{D}_{1[i_1]}^H \mathbf{D}_{1\Xi_1}) \right. \\
&\quad \times_2 \cdots \times_n (\mathbf{D}_{n[i_n]}^H \mathbf{D}_{n\Xi_n}) \|_F \\
&\quad \left. + \|\mathcal{N} \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \|_F \right) \\
&\leq s \left( \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n k_t d_t \right) \|\mathcal{X}\|_F + s \overline{W}_{\mathbf{Y}, 1}^{\frac{1}{2}} \|\mathcal{N}\|_F, \quad (124)
\end{aligned}$$

where  $s$  is the number of selected indices during the first iteration of T-GBOMP,  $\sum_{t=0}^{n-1} \mathcal{C}_t = \prod_{i=1}^n d_i - 1$ ,  $\overline{W}_{\mathbf{Y}, 1}$  can be obtained by **Corollary 3**, and (124) is derived from the scenario where only one tensor index set is selected incorrectly, which could present the strictest upper bound in (124).

Then, if

$$\begin{aligned}
&\sqrt{sk} \left( \overline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t}^{\frac{1}{2}} \overline{W}_{\mathbf{Y}, 1}^{\frac{1}{2}} \|\mathcal{X}\|_F - \overline{W}_{\mathbf{Y}, 1}^{\frac{1}{2}} \|\mathcal{N}\|_F \right) \\
&> s \left( \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n k_t d_t \right) \|\mathcal{X}\|_F + s \overline{W}_{\mathbf{Y}, 1}^{\frac{1}{2}} \|\mathcal{N}\|_F, \quad (125)
\end{aligned}$$

(123) is contradicted.

Reorganizing (125), we have

$$\frac{\|\mathcal{X}\|_F}{\|\mathcal{N}\|_F} > \frac{(s + \sqrt{sk}) \overline{W}_{\mathbf{Y}, 1}^{\frac{1}{2}}}{\sqrt{sk} \overline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t}^{\frac{1}{2}} \overline{W}_{\mathbf{Y}, 1}^{\frac{1}{2}} - s \left( \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n k_t d_t \right)}. \quad (126)$$

On the other hand,

$$\frac{\|\mathcal{X}\|_F}{\|\mathcal{N}\|_F} \geq \frac{\|\mathcal{X} \times_1 \mathbf{D}_1 \times_2 \cdots \times_n \mathbf{D}_n\|_F}{\overline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t}^{\frac{1}{2}} \|\mathcal{N}\|_F} \quad (127)$$

$$= \frac{\sqrt{\text{SNR}}}{\overline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t}^{\frac{1}{2}}}, \quad (128)$$

where (127) is from **Corollary 3** and (128) is due to the definition of SNR.

Therefore, combining (126) and (128), we obtain that if

$$\sqrt{\text{SNR}} > \frac{(\frac{\sqrt{k}}{\sqrt{s}} + 1) \overline{W}_{\mathbf{Y}, 1}^{\frac{1}{2}} \overline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t}^{\frac{1}{2}}}{\frac{\sqrt{k}}{\sqrt{s}} \overline{W}_{\mathbf{Y}, \prod_{t=1}^n k_t}^{\frac{1}{2}} \overline{W}_{\mathbf{Y}, 1}^{\frac{1}{2}} - \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n k_t d_t}}, \quad (129)$$

then at least one correct tensor block is selected at T-GBOMP's first iteration.

b) Second, supposing that T-GBOMP selects at least one correct tensor block at each of the previous  $l$  ( $1 \leq l < k$ ) iterations, we prove that the algorithm selects at least one correct tensor block at the  $(l+1)$ -th iteration. Denote  $p_1$  as the largest element in the set  $\{ \|\mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \|_F \}_{(i_1, \dots, i_n) \in \Xi \setminus \Xi^l}$ . The following inequality holds:

$$\begin{aligned}
p_1 &\geq \frac{1}{\sqrt{|\Xi \setminus \Xi^l|}} \left( \sum_{(i_1, i_2, \dots, i_n) \in \Xi \setminus \Xi^l} \|\mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \right. \\
&\quad \left. \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H \|_F^2 \right)^{\frac{1}{2}} \quad (130)
\end{aligned}$$

$$\begin{aligned}
&\geq \frac{1}{\sqrt{k-c}} \left( \|\ddot{\mathbf{D}}_{\Xi \setminus \Xi^l} \mathbf{P}_{\mathbf{D}_{\Xi^l}}^{\frac{1}{2}} \text{vec}(\mathcal{Y}) \|_F \right) \\
&= \frac{1}{\sqrt{k-c}} \left( \|\ddot{\mathbf{D}}_{\Xi \setminus \Xi^l} \mathbf{P}_{\mathbf{D}_{\Xi^l}}^{\frac{1}{2}} (\mathbf{D}_{n_{\Xi_n \setminus \Xi_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1 \setminus \Xi_1^l}}) \right. \\
&\quad \left. \times \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^l}) + \ddot{\mathbf{D}}_{\Xi \setminus \Xi^l} \mathbf{P}_{\mathbf{D}_{\Xi^l}}^{\frac{1}{2}} \text{vec}(\mathcal{N}) \|_F \right) \quad (131)
\end{aligned}$$

$$\begin{aligned}
&\geq \frac{1}{\sqrt{k-c}} \left( \|\ddot{\mathbf{D}}_{\Xi \setminus \Xi^l} \mathbf{P}_{\mathbf{D}_{\Xi^l}}^{\frac{1}{2}} (\mathbf{D}_{n_{\Xi_n \setminus \Xi_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1 \setminus \Xi_1^l}}) \right. \\
&\quad \left. \times \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^l}) \|_F \right. \\
&\quad \left. - \|\ddot{\mathbf{D}}_{\Xi \setminus \Xi^l} \mathbf{P}_{\mathbf{D}_{\Xi^l}}^{\frac{1}{2}} \text{vec}(\mathcal{N}) \|_F \right) \quad (132)
\end{aligned}$$

$$\begin{aligned}
&\geq \frac{1}{\sqrt{k-c}} \left( \sigma_{\min} \left( \ddot{\mathbf{D}}_{\Xi \setminus \Xi^l} (\mathbf{D}_{n_{\Xi_n \setminus \Xi_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1 \setminus \Xi_1^l}}) \right) \right. \\
&\quad \left. \times \|\mathcal{X}_{\Xi \setminus \Xi^l}\|_F - \sigma_{\max} \ddot{\mathbf{D}}_{\Xi \setminus \Xi^l} \|\mathcal{N}\|_F \right) \quad (133)
\end{aligned}$$

$$\begin{aligned}
&\geq \frac{1}{\sqrt{k-c}} \left( \overline{W}_{\mathbf{Y}, ls}^{\frac{1}{2}} \overline{W}_{\mathbf{Y}, (k-c)^n}^{\frac{1}{2}} \|\mathcal{X}_{\Xi \setminus \Xi^l}\|_F \right. \\
&\quad \left. - \overline{W}_{\mathbf{Y}, ls}^{\frac{1}{2}} \|\mathcal{N}\|_F \right), \quad (134)
\end{aligned}$$

where  $c = |\Xi \cap \Xi^l| \geq l$ ,  $\overline{W}_{\mathbf{Y}, k_n}$  and  $\overline{W}_{\mathbf{Y}, k_n}$  can be obtained by **Corollary 3**, (130) follows from the Cauchy-Schwarz inequality, (131) is from (1), (132) is based on the triangle

inequality, (133) is based on [19, Lemma 5], and (134) follows from **Corollary 3**.

Denoting  $\Psi$  as the index set corresponding to the  $s$  largest elements in  $\{\|\mathcal{R}^l \times_1 \mathbf{D}_{1_{[i_1]}^H} \times_2 \cdots \times_n \mathbf{D}_{n_{[i_n]}^H}\|_F\}_{(i_1, \dots, i_n) \in \Omega \setminus (\Xi \cup \Xi^l)}$  and letting  $q_s$  be the  $s$ th largest element in  $\Psi$ , we have

$$\begin{aligned}
 q_s &\leq \frac{1}{\sqrt{s}} \sqrt{\sum_{(i_1, \dots, i_n) \in \Psi} \|\mathcal{R}^l \times_1 \mathbf{D}_{1_{[i_1]}^H} \times_2 \cdots \times_n \mathbf{D}_{n_{[i_n]}^H}\|_F^2} \\
 &= \frac{1}{\sqrt{s}} \|\ddot{\mathbf{D}}_{\Psi} \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}^{\perp} \text{vec}(\mathcal{Y})\|_F \\
 &= \frac{1}{\sqrt{s}} \|\ddot{\mathbf{D}}_{\Psi} \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}^{\perp} (\mathbf{D}_n \otimes \cdots \otimes \mathbf{D}_1) \text{vec}(\mathcal{X}) \\
 &\quad + \ddot{\mathbf{D}}_{\Psi} \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}^{\perp} \text{vec}(\mathcal{N})\|_F \\
 &\leq \frac{1}{\sqrt{s}} \left( \|\ddot{\mathbf{D}}_{\Psi} \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}^{\perp} (\mathbf{D}_{n_{\Xi_n \setminus \Xi_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1 \setminus \Xi_1^l}}) \right. \\
 &\quad \left. \times \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^l})\|_F + \|\ddot{\mathbf{D}}_{\Psi} \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}^{\perp} \text{vec}(\mathcal{N})\|_F \right) \\
 &\leq \frac{1}{\sqrt{s}} \left( \|\ddot{\mathbf{D}}_{\Psi} \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}^{\perp} (\mathbf{D}_{n_{\Xi_n \setminus \Xi_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1 \setminus \Xi_1^l}}) \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^l})\|_F \right. \\
 &\quad + \|\ddot{\mathbf{D}}_{\Psi} (\mathbf{D}_{n_{\Xi_n \setminus \Xi_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1 \setminus \Xi_1^l}}) \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^l})\|_F \\
 &\quad \left. + \|\ddot{\mathbf{D}}_{\Psi} \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}^{\perp} \text{vec}(\mathcal{N})\|_F \right), \quad (135)
 \end{aligned}$$

where (135) is because  $\mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}^{\perp} = \mathbf{I} - \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}$ . In the following, we derive upper bounds of the three terms in (135), i.e.,  $\|\ddot{\mathbf{D}}_{\Psi} \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}^{\perp} (\mathbf{D}_{n_{\Xi_n \setminus \Xi_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1 \setminus \Xi_1^l}}) \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^l})\|_F$ ,  $\|\ddot{\mathbf{D}}_{\Psi} (\mathbf{D}_{n_{\Xi_n \setminus \Xi_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1 \setminus \Xi_1^l}}) \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^l})\|_F$  and  $\|\ddot{\mathbf{D}}_{\Psi} \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}^{\perp} \text{vec}(\mathcal{N})\|_F$ .

For the first term, note that

$$\begin{aligned}
 &\|\ddot{\mathbf{D}}_{\Psi} \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}^{\perp} (\mathbf{D}_{n_{\Xi_n \setminus \Xi_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1 \setminus \Xi_1^l}}) \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^l})\|_F \\
 &= \|\ddot{\mathbf{D}}_{\Psi} \ddot{\mathbf{D}}_{\Xi^l} (\ddot{\mathbf{D}}_{\Xi^l}^H \ddot{\mathbf{D}}_{\Xi^l})^{-1} \ddot{\mathbf{D}}_{\Xi^l}^H \\
 &\quad \times (\mathbf{D}_{n_{\Xi_n \setminus \Xi_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1 \setminus \Xi_1^l}}) \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^l})\|_F \\
 &\leq \|\ddot{\mathbf{D}}_{\Psi} \ddot{\mathbf{D}}_{\Xi^l}\|_2 \|(\ddot{\mathbf{D}}_{\Xi^l}^H \ddot{\mathbf{D}}_{\Xi^l})^{-1}\|_2 \\
 &\quad \times \|\ddot{\mathbf{D}}_{\Xi^l}^H (\mathbf{D}_{n_{\Xi_n \setminus \Xi_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1 \setminus \Xi_1^l}})\|_2 \|\mathcal{X}_{\Xi \setminus \Xi^l}\|_F, \quad (136)
 \end{aligned}$$

where (136) is from the sub-multiplicativity. Based on **Corollary 3**, we have

$$\begin{aligned}
 \|\ddot{\mathbf{D}}_{\Psi} \ddot{\mathbf{D}}_{\Xi^l}\|_2 &\leq \max \left\{ s \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t, l s \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t \right\} \\
 &= l s \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t, \quad (137)
 \end{aligned}$$

$$\begin{aligned}
 &\|\ddot{\mathbf{D}}_{\Xi^l}^H (\mathbf{D}_{n_{\Xi_n \setminus \Xi_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1 \setminus \Xi_1^l}})\|_2 \\
 &\leq \max \left\{ l s \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t, (k-c)^n \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t \right\}. \quad (138)
 \end{aligned}$$

It remains to derive an upper bound on  $\|(\ddot{\mathbf{D}}_{\Xi^l}^H \ddot{\mathbf{D}}_{\Xi^l})^{-1}\|_2$ . Denote  $\ddot{\mathbf{D}}_{\Xi^l}^H \ddot{\mathbf{D}}_{\Xi^l}$  as  $\mathbf{I} + \mathbf{A}$  since the diagonal elements of  $\ddot{\mathbf{D}}_{\Xi^l}^H \ddot{\mathbf{D}}_{\Xi^l}$  are all 1, where  $\mathbf{A}$  is a matrix of size  $l s \prod_{t=1}^n d_t \times l s \prod_{t=1}^n d_t$  with blocks of size  $\prod_{t=1}^n d_t \times \prod_{t=1}^n d_t$ . Then, by using **Corollary 1**, we obtain

$$\|\mathbf{A}\|_2 \leq \sum_{t=0}^{n-1} \underline{C}_t \tau_{\mathbf{Y}, t}^{n-t} + \sum_{t=0}^{n-1} C_t \varpi_{\mathbf{Y}, t}^n \prod_{i=1}^n d_i,$$

where  $\sum_{t=0}^{n-1} \underline{C}_t = \prod_{t=1}^n d_t - 1$ ,  $\sum_{t=0}^{n-1} C_t = l s - 1$ , and  $\underline{C}_t$  and  $C_t$  can be obtained by **Proposition 1**.

Suppose that  $\sum_{t=0}^{n-1} \underline{C}_t \tau_{\mathbf{Y}, t}^{n-t} + \sum_{t=0}^{n-1} C_t \varpi_{\mathbf{Y}, t}^n \prod_{i=1}^n d_i < 1$ . Similar to (87) to (90), we have

$$\begin{aligned}
 &\|(\ddot{\mathbf{D}}_{\Xi^l}^H \ddot{\mathbf{D}}_{\Xi^l})^{-1}\|_2 \\
 &\leq \frac{1}{1 - \sum_{t=0}^{n-1} \underline{C}_t \tau_{\mathbf{Y}, t}^{n-t} - \sum_{t=0}^{n-1} C_t \varpi_{\mathbf{Y}, t}^n \prod_{i=1}^n d_i}. \quad (139)
 \end{aligned}$$

Combining (136), (137), (138) and (139) leads to

$$\begin{aligned}
 &\|\ddot{\mathbf{D}}_{\Psi} \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}^{\perp} (\mathbf{D}_{n_{\Xi_n \setminus \Xi_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1 \setminus \Xi_1^l}}) \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^l})\|_F \\
 &\leq (l s \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t) \\
 &\quad \times \max \left\{ l s \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t, (k-c)^n \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t \right\} \\
 &\quad \times \frac{1}{1 - \sum_{t=0}^{n-1} \underline{C}_t \tau_{\mathbf{Y}, t}^{n-t} - \sum_{t=0}^{n-1} C_t \varpi_{\mathbf{Y}, t}^n \prod_{i=1}^n d_i} \|\mathcal{X}_{\Xi \setminus \Xi^l}\|_F \\
 &= (l s \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t) \\
 &\quad \times \max \left\{ l s \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t, (k-c)^n \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t \right\} \\
 &\quad \times \frac{1}{\underline{W}_{\mathbf{Y}, l s}} \|\mathcal{X}_{\Xi \setminus \Xi^l}\|_F. \quad (140)
 \end{aligned}$$

For the second term, we have

$$\begin{aligned}
 &\|\ddot{\mathbf{D}}_{\Psi} (\mathbf{D}_{n_{\Xi_n \setminus \Xi_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Xi_1 \setminus \Xi_1^l}}) \text{vec}(\overline{\mathcal{X}})\|_F \\
 &\leq \max \left\{ s \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t, (k-c)^n \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t \right\} \\
 &\quad \times \|\mathcal{X}_{\Xi \setminus \Xi^l}\|_F, \quad (141)
 \end{aligned}$$

where (141) follows from **Corollary 3**.

For the third term, we have

$$\begin{aligned}
 \|\ddot{\mathbf{D}}_{\Psi} \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}^{\perp} \text{vec}(\mathcal{N})\|_F &\leq \|\ddot{\mathbf{D}}_{\Psi}\|_2 \|\mathcal{N}\|_F \\
 &\leq \overline{W}_{\mathbf{Y}, s}^{\frac{1}{2}} \|\mathcal{N}\|_F. \quad (142)
 \end{aligned}$$

Combining (135), (140), (141) and (142), we have

$$\begin{aligned}
 q_s &\leq \frac{1}{\sqrt{s}} \left( (l s \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t) \right. \\
 &\quad \times \max \left\{ l s \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t, (k-c)^n \varpi_{\mathbf{Y}, n-1}^n \prod_{t=1}^n d_t \right\} \\
 &\quad \left. + \|\ddot{\mathbf{D}}_{\Psi} \mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^l}}^{\perp} \text{vec}(\mathcal{N})\|_F \right)
 \end{aligned}$$

$$\begin{aligned} & \times \frac{1}{W_{\mathbf{r},ls}} \|\mathcal{X}_{\Xi \setminus \Xi^l}\|_F \\ & + \max \left\{ s \varpi_{\mathbf{r},n-1}^n \prod_{t=1}^n d_t, (k-c)^n \varpi_{\mathbf{r},n-1}^n \prod_{t=1}^n d_t \right\} \\ & \times \left( \|\mathcal{X}_{\Xi \setminus \Xi^l}\|_F + \overline{W}_{\mathbf{r},s}^{\frac{1}{2}} \|\mathcal{N}\|_F \right). \end{aligned} \quad (143)$$

Note that if  $p_1 > q_s$ ,  $p_1$  belongs to the set of  $s$  largest elements among all the elements in  $\{\|\mathcal{R}^l \times_1 \mathbf{D}_{1[i_1]}^H \times_2 \cdots \times_n \mathbf{D}_{n[i_n]}^H\|_F\}_{(i_1, \dots, i_n) \in \Omega \setminus \Xi^l}$ , then at least one correct index is selected in the  $(l+1)$ th iteration. Exploiting (134) and (143),  $p_1 > q_s$  holds true if the following inequality is satisfied:

$$\begin{aligned} & \frac{1}{\sqrt{k-c}} \left( W_{\mathbf{r},ls}^{\frac{1}{2}} W_{\mathbf{r},(k-c)^n}^{\frac{1}{2}} \|\mathcal{X}_{\Xi \setminus \Xi^l}\|_F - \overline{W}_{\mathbf{r},ls}^{\frac{1}{2}} \|\mathcal{N}\|_F \right) \\ & > \frac{1}{\sqrt{s}} \left( l s \varpi_{\mathbf{r},n-1}^n \prod_{t=1}^n d_t \right) \\ & \times \max \left\{ l s \varpi_{\mathbf{r},n-1}^n \prod_{t=1}^n d_t, (k-c)^n \varpi_{\mathbf{r},n-1}^n \prod_{t=1}^n d_t \right\} \\ & \times \frac{1}{W_{\mathbf{r},ls}} \|\mathcal{X}_{\Xi \setminus \Xi^l}\|_F \\ & + \max \left\{ s \varpi_{\mathbf{r},n-1}^n \prod_{t=1}^n d_t, (k-c)^n \varpi_{\mathbf{r},n-1}^n \prod_{t=1}^n d_t \right\} \\ & \times \left( \|\mathcal{X}_{\Xi \setminus \Xi^l}\|_F + \overline{W}_{\mathbf{r},s}^{\frac{1}{2}} \|\mathcal{N}\|_F \right). \end{aligned} \quad (144)$$

Note that

$$\begin{aligned} \frac{\|\overline{\mathcal{X}}\|_F}{\|\mathcal{N}\|_F} & \geq \frac{\sqrt{|\Xi \setminus \Xi^l|} \min_{(i_1, \dots, i_n) \in \Xi} \|\mathcal{X}_{[i_1, \dots, i_n]}\|_F}{\|\mathcal{N}\|_F} \\ & = \frac{\text{MAR}_* \sqrt{k-c} \|\mathcal{X}\|_F}{\sqrt{k} \|\mathcal{N}\|_F} \end{aligned} \quad (145)$$

$$\geq \frac{\text{MAR}_* \sqrt{k-c} \|\mathcal{X} \times_1 \mathbf{D}_1 \times_2 \cdots \times_n \mathbf{D}_n\|_F}{\sqrt{k} \overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t} \|\mathcal{N}\|_F} \quad (146)$$

$$= \frac{\text{MAR}_* \sqrt{(k-c) \text{SNR}}}{\sqrt{k} \overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t}}, \quad (147)$$

where (145) is from the definition of  $\text{MAR}_*$ , (146) is based on **Corollary 3**, and (147) follows from the definition of SNR. Combining (144) and (147), we conclude that if

$$\begin{aligned} & \sqrt{\text{SNR}} \\ & > \left( \frac{1}{\sqrt{s}} \overline{W}_{\mathbf{r},s}^{\frac{1}{2}} + \frac{1}{\sqrt{k-c}} \overline{W}_{\mathbf{r},ls}^{\frac{1}{2}} \right) \sqrt{k} \overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t}^{\frac{1}{2}} \\ & \times \left( \left( \frac{1}{\sqrt{k-c}} W_{\mathbf{r},ls}^{\frac{1}{2}} W_{\mathbf{r},(k-c)^n}^{\frac{1}{2}} \right. \right. \\ & \left. \left. - \frac{1}{\sqrt{s}} \left( l s \varpi_{\mathbf{r},n-1}^n \prod_{t=1}^n d_t \right) \right) \right) \end{aligned}$$

$$\begin{aligned} & \times \max \left\{ l s \varpi_{\mathbf{r},n-1}^n \prod_{t=1}^n d_t, (k-c)^n \varpi_{\mathbf{r},n-1}^n \prod_{t=1}^n d_t \right\} \frac{1}{W_{\mathbf{r},ls}} \\ & - \frac{1}{\sqrt{s}} \max \left\{ s \varpi_{\mathbf{r},n-1}^n \prod_{t=1}^n d_t, (k-c)^n \varpi_{\mathbf{r},n-1}^n \prod_{t=1}^n d_t \right\} \\ & \times \text{MAR}_* \sqrt{k-c} \Big)^{-1} \end{aligned} \quad (148)$$

holds, the T-GBOMP chooses at least one correct index in the  $l+1$ th iteration.

c) So far, we have proved that under the condition (129) T-GBOMP succeeds in the first iteration, and under the condition (148) T-GBOMP succeeds in the general iteration. We now present a condition under which T-GBOMP succeeds in every iteration. Due to the fact that  $1 \leq k-c < k$ ,  $1 \leq l < k$  and  $1 \leq s \leq k$ , (148) is transformed into

$$\begin{cases} \sqrt{\text{SNR}} > \left( \frac{\sqrt{k}}{\sqrt{s}} \overline{W}_{\mathbf{r},s}^{\frac{1}{2}} \overline{W}_{\mathbf{r},k}^{\frac{1}{2}} + \sqrt{k} \overline{W}_{\mathbf{r},ks}^{\frac{1}{2}} \overline{W}_{\mathbf{r},k}^{\frac{1}{2}} \right) \\ \times \left( \frac{1}{\sqrt{s}} \left( \frac{\sqrt{s}}{\sqrt{k}} W_{\mathbf{r},ks}^{\frac{1}{2}} W_{\mathbf{r},k}^{\frac{1}{2}} - k \varpi_{\mathbf{r}} d_1 \right. \right. \\ \left. \left. - \frac{(k s \varpi_{\mathbf{r}} d_1)^2}{W_{\mathbf{r},ks}} \right) \text{MAR}_* \right)^{-1}, n=1, \end{cases} \quad (149a)$$

$$\begin{cases} \sqrt{\text{SNR}} > \left( \frac{\sqrt{k}}{\sqrt{s}} \overline{W}_{\mathbf{r},s}^{\frac{1}{2}} \overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t}^{\frac{1}{2}} \right. \\ \left. + \sqrt{k} \overline{W}_{\mathbf{r},ks}^{\frac{1}{2}} \overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t}^{\frac{1}{2}} \right) \\ \times \left( \frac{1}{\sqrt{s}} \left( \frac{\sqrt{s}}{\sqrt{k}} W_{\mathbf{r},ks}^{\frac{1}{2}} W_{\mathbf{r},k}^{\frac{1}{2}} - k^n \varpi_{\mathbf{r},n-1}^n \prod_{t=1}^n d_t \right. \right. \\ \left. \left. - \frac{k^{n+1} s \varpi_{\mathbf{r},n-1}^n \left( \prod_{t=1}^n d_t \right)^2}{W_{\mathbf{r},ks}} \right) \text{MAR}_* \right)^{-1}, n > 1. \end{cases} \quad (149b)$$

Since  $\frac{\sqrt{k}}{\sqrt{s}} \geq \frac{\sqrt{s}}{\sqrt{k}}$ ,  $\frac{1}{\sqrt{s}} \leq 1$  and  $\text{MAR}_* \leq 1$ , it can be obtained that (149a) and (149b) are more restrictive than the bound in (129). Therefore, (149a) and (149b) guarantee that T-GBOMP chooses all correct indices in  $k$  iterations. Simplifying (149a) and (149b) leads to (49a) and (49b). ■

### K. Proof of Theorem 6

*Proof:* Denote  $l^*$  as the number of iterations of T-GBOMP, where  $l^* \leq k$ . In the case that  $s=1$ ,  $l^*=k$ . Based on **Theorem 5**, we have  $\hat{\Xi} = \Xi^k = \Xi$  and  $\hat{\mathcal{X}} = \mathcal{X}^k$ , since  $\hat{\mathcal{X}} = \arg \min_{\mathcal{X}: \text{supp}(\mathcal{X}) = \Xi} \|\text{vec}(\mathcal{Y}) - \sum_{(i_1, i_2, \dots, i_n) \in \Xi} (\mathbf{D}_{n[i_n]} \otimes \cdots \otimes \mathbf{D}_{1[i_1]}) \text{vec}(\mathcal{X})_{[i_x]}\|_F$  and  $\hat{\mathcal{X}}_{\Omega \setminus \Xi} = \mathbf{0}$ . Therefore,

$$\begin{aligned} \|\hat{\mathcal{X}} - \mathcal{X}\|_F & = \|\ddot{\mathbf{D}}_{\Xi}^{\dagger} \text{vec}(\mathcal{Y}) - \text{vec}(\mathcal{X}_{\Xi})\|_F \\ & = \|\ddot{\mathbf{D}}_{\Xi}^{\dagger} (\text{vec}(\mathcal{Y}) - \ddot{\mathbf{D}}_{\Xi} \text{vec}(\mathcal{X}_{\Xi}))\|_F \end{aligned} \quad (150)$$

$$= \|\ddot{\mathbf{D}}_{\Xi}^{\dagger} \text{vec}(\mathcal{N})\|_F \quad (151)$$

$$\leq \frac{\|\ddot{\mathbf{D}}_{\Xi} \ddot{\mathbf{D}}_{\Xi}^{\dagger} \text{vec}(\mathcal{N})\|_F}{W_{\mathbf{r},k}^{\frac{1}{2}}} \quad (152)$$

$$= \frac{\|\mathbf{P}_{\ddot{\mathbf{D}}_{\Xi}} \text{vec}(\mathcal{N})\|_F}{W_{\mathbf{r},k}^{\frac{1}{2}}} \quad (153)$$

$$\leq \frac{\|\mathcal{N}\|_F}{W_{\mathbf{r},k}^{\frac{1}{2}}}, \quad (154)$$

where (150) is because  $\ddot{\mathbf{D}}_{\Xi} \ddot{\mathbf{D}}_{\Xi}^{\dagger} = \mathbf{I}$ , (151) follows from (1), (152) is due to **Corollary 3**, (153) is derived by  $\mathbf{P}_{\ddot{\mathbf{D}}_{\Xi}} = \ddot{\mathbf{D}}_{\Xi} \ddot{\mathbf{D}}_{\Xi}^{\dagger}$ , and (154) is because  $\|\mathbf{P}_{\ddot{\mathbf{D}}_{\Xi}}\|_2 \leq 1$ .

In the case that  $s > 1$ , we have  $\Xi \subseteq \Xi^k$ . Letting  $\mathcal{Q}^k$  be the best  $k$  block-sparse approximation of  $\mathcal{X}^k$ , we obtain

$$\begin{aligned} & \|\mathcal{Q}^k - \mathcal{X}\|_F \\ &= \|\mathcal{Q}^k - \mathcal{X}^k + \mathcal{X}^k - \mathcal{X}\|_F \\ &\leq \|\mathcal{Q}^k - \mathcal{X}^k\|_F + \|\mathcal{X}^k - \mathcal{X}\|_F \end{aligned} \quad (155)$$

$$\leq 2\|\mathcal{X}^k - \mathcal{X}\|_F \quad (156)$$

$$= 2\|\ddot{\mathbf{D}}_{\Xi^k}^{\dagger} \text{vec}(\mathcal{Y}) - \text{vec}(\mathcal{X}_{\Xi^k})\|_F \quad (157)$$

$$= 2\|\ddot{\mathbf{D}}_{\Xi^k}^{\dagger} \text{vec}(\mathcal{X}_{\Xi} \times_1 \mathbf{D}_{1_{\Xi_1}} \times_2 \cdots \times_n \mathbf{D}_{n_{\Xi_n}} + \mathcal{N}) - \text{vec}(\mathcal{X}_{\Xi^k})\|_F \quad (158)$$

$$= 2\|\ddot{\mathbf{D}}_{\Xi^k}^{\dagger} (\text{vec}(\mathcal{X}_{\Xi^k} \times_1 \mathbf{D}_{1_{\Xi_1^k}} \times_2 \cdots \times_n \mathbf{D}_{n_{\Xi_n^k}} + \mathcal{N}) - \mathbf{T}_k \text{vec}(\mathcal{X}_{\Xi^k}))\|_F$$

$$= 2\|\ddot{\mathbf{D}}_{\Xi^k}^{\dagger} \text{vec}(\mathcal{N})\|_F \leq \frac{2\|\ddot{\mathbf{D}}_{\Xi^k} \ddot{\mathbf{D}}_{\Xi^k}^{\dagger} \text{vec}(\mathcal{N})\|_F}{W_{\mathbf{r},sk}^{\frac{1}{2}}} \quad (159)$$

$$= \frac{2\|\mathbf{P}_{\ddot{\mathbf{D}}_{\Xi^k}} \text{vec}(\mathcal{N})\|_F}{W_{\mathbf{r},sk}^{\frac{1}{2}}} \leq \frac{2\|\mathcal{N}\|_F}{W_{\mathbf{r},sk}^{\frac{1}{2}}}, \quad (160)$$

where (155) is from the triangle inequality, (156) is because  $\mathcal{Q}^k$  is the best  $k$  block-sparse approximation to  $\mathcal{X}^k$ , (157) is because the nonzero tensor blocks of  $\mathcal{X}^k$  are estimated by  $\ddot{\mathbf{D}}_{\Xi^k}^{\dagger} \text{vec}(\mathcal{Y})$  and  $\Xi \subseteq \Xi^k$ , (158) is based on (1), and (159) is derived by **Corollary 3**.

From Appendix I, we know that

$$\|\mathcal{Q}^k - \mathcal{X}\|_F \geq \frac{W_{\mathbf{r}, \prod_{t=1}^n k_t + k}^{\frac{1}{2}} \|\hat{\mathcal{X}} - \mathcal{X}\|_F - 2\|\mathcal{N}\|_F}{W_{\mathbf{r}, \prod_{t=1}^n k_t + k}^{\frac{1}{2}}}. \quad (161)$$

The proof is completed by combining (160) and (161).  $\blacksquare$

### L. Proof of Lemma 3

*Proof:* From (103), for any integer  $c$  with  $c \geq l$ , we have

$$\begin{aligned} & \|\mathcal{R}^c\|_F^2 - \|\mathcal{R}^{c+1}\|_F^2 \\ &\geq \frac{\sum_{(i_1, i_2, \dots, i_n) \in \Theta^{c+1}} \|\mathcal{R}^c \times_1 \mathbf{D}_{1_{[i_1]}^H} \times_2 \cdots \times_n \mathbf{D}_{n_{[i_n]}^H}\|_F^2}{\overline{W}_{\mathbf{r},s}}. \end{aligned} \quad (162)$$

Denote  $\mathcal{Z} \in \mathbb{C}^{N_1 \times \dots \times N_n}$  as a tensor with  $\text{vec}(\mathcal{Z}_{\Xi \cap \Xi^l \cup \Lambda_{\theta}^l}) = \text{vec}(\mathcal{X}_{\Xi \cap \Xi^l \cup \Lambda_{\theta}^l})$  and

$\text{vec}(\mathcal{Z}_{\Omega \setminus (\Xi \cap \Xi^l \cup \Lambda_{\theta}^l)}) = \mathbf{0}$ . From [42, C. 12], we have

$$\begin{aligned} & \langle (\mathbf{D}_n^H \otimes \cdots \otimes \mathbf{D}_1^H) \text{vec}(\mathcal{R}^c), \text{vec}(\mathcal{Z}) \rangle \\ &\leq \left( \left[ \frac{|\Lambda_{\theta}^l|}{s} \right] \right)^{\frac{1}{2}} \|\ddot{\mathbf{D}}_{\Theta^{c+1}}^H \text{vec}(\mathcal{R}^c)\|_F \|\text{vec}(\mathcal{Z}_{\Omega \setminus \Xi^c})\|_F. \end{aligned} \quad (163)$$

On the other hand,

$$\begin{aligned} & \langle (\mathbf{D}_n^H \otimes \cdots \otimes \mathbf{D}_1^H) \text{vec}(\mathcal{R}^c), \text{vec}(\mathcal{Z}) \rangle \\ &\geq \|(\mathbf{D}_n \otimes \cdots \otimes \mathbf{D}_1) (\text{vec}(\mathcal{Z}) - \text{vec}(\hat{\mathcal{X}}^c))\|_F \\ &\quad \times \left( \|\text{vec}(\mathcal{R}^c)\|_F^2 - \|(\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_e}^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_e}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_{\theta}^l}) + \text{vec}(\mathcal{N})\|_F^2 \right)^{\frac{1}{2}} \end{aligned} \quad (164)$$

$$\begin{aligned} & \geq W_{\mathbf{r}, |\Lambda_{\theta}^l \cup \Xi^c|}^{\frac{1}{2}} \|\text{vec}(\mathcal{Z}) - \text{vec}(\hat{\mathcal{X}}^c)\|_F \\ &\quad \times \left( \|\text{vec}(\mathcal{R}^c)\|_F^2 - \|(\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_e}^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_e}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_{\theta}^l}) + \text{vec}(\mathcal{N})\|_F^2 \right)^{\frac{1}{2}} \end{aligned} \quad (165)$$

$$\begin{aligned} & \geq W_{\mathbf{r}, |\Lambda_{\theta}^l \cup \Xi^c|}^{\frac{1}{2}} \|\text{vec}(\mathcal{Z}_{\Omega \setminus \Xi^c} - \hat{\mathcal{X}}_{\Omega \setminus \Xi^c}^c)\|_F \\ &\quad \times \left( \|\text{vec}(\mathcal{R}^c)\|_F^2 - \|(\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_e}^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_e}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_{\theta}^l}) + \text{vec}(\mathcal{N})\|_F^2 \right)^{\frac{1}{2}} \\ & \geq W_{\mathbf{r}, |\Lambda_{\theta}^l \cup \Xi^c|}^{\frac{1}{2}} \|\text{vec}(\mathcal{Z}_{\Omega \setminus \Xi^c})\|_F \\ &\quad \times \left( \|\text{vec}(\mathcal{R}^c)\|_F^2 - \|(\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_e}^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_e}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_{\theta}^l}) + \text{vec}(\mathcal{N})\|_F^2 \right)^{\frac{1}{2}}, \end{aligned} \quad (166)$$

where (164) is based on [42, C. 13], (165) is from **Corollary 3**, and (166) is because  $\text{vec}(\hat{\mathcal{X}}_{\Omega \setminus \Xi^c}^c) = \mathbf{0}$ . Combining (163) and (166) yields

$$\begin{aligned} & \|\ddot{\mathbf{D}}_{\Theta^{c+1}}^H \text{vec}(\mathcal{R}^c)\|_F \\ &\geq \frac{W_{\mathbf{r}, |\Lambda_{\theta}^l \cup \Xi^c|}^{\frac{1}{2}}}{\left( \left[ \frac{|\Lambda_{\theta}^l|}{s} \right] \right)^{\frac{1}{2}}} \left( \|\text{vec}(\mathcal{R}^c)\|_F^2 - \|(\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_e}^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_e}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_{\theta}^l}) + \text{vec}(\mathcal{N})\|_F^2 \right)^{\frac{1}{2}}. \end{aligned} \quad (167)$$

Based on (162) and (167) we have

$$\begin{aligned} & \|\mathcal{R}^c\|_F^2 - \|\mathcal{R}^{c+1}\|_F^2 \\ &\geq \frac{W_{\mathbf{r}, |\Lambda_{\theta}^l \cup \Xi^c|}}{\left[ \frac{|\Lambda_{\theta}^l|}{s} \right] \overline{W}_{\mathbf{r},s}} \left( \|\text{vec}(\mathcal{R}^c)\|_F^2 - \|(\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_e}^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_e}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_{\theta}^l}) + \text{vec}(\mathcal{N})\|_F^2 \right) \end{aligned}$$

$$+ \text{vec}(\mathcal{N})\|_F^2 \Big).$$

Then, for  $c^* \in \{c, \dots, c + \Delta c - 1\}$ , we have

$$\begin{aligned} & \|\mathcal{R}^{c^*}\|_F^2 - \|\mathcal{R}^{c^*+1}\|_F^2 \\ & \geq \frac{W_{\mathbf{r}, |\Lambda_\rho^l \cup \Xi^{c^*}|}}{\left\lceil \frac{|\Lambda_\rho^l|}{s} \right\rceil \overline{W}_{\mathbf{r}, s}} \left( \left\| \text{vec}(\mathcal{R}^{c^*}) \right\|_F^2 \right. \\ & \quad \left. - \left\| (\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_\rho}^l}} \otimes \dots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_\rho}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\rho^l}) \right. \right. \\ & \quad \left. \left. + \text{vec}(\mathcal{N}) \right\|_F^2 \right) \end{aligned} \quad (169)$$

$$\begin{aligned} & \geq \left( 1 - \exp \left( - \frac{W_{\mathbf{r}, |\Lambda_\rho^l \cup \Xi^{c^*}|}}{\left\lceil \frac{|\Lambda_\rho^l|}{s} \right\rceil \overline{W}_{\mathbf{r}, s}} \right) \right) \left( \left\| \text{vec}(\mathcal{R}^{c^*}) \right\|_F^2 \right. \\ & \quad \left. - \left\| (\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_\rho}^l}} \otimes \dots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_\rho}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\rho^l}) \right. \right. \\ & \quad \left. \left. + \text{vec}(\mathcal{N}) \right\|_F^2 \right) \\ & \geq \left( 1 - \exp \left( - \frac{W_{\mathbf{r}, |\Lambda_\rho^l \cup \Xi^{c+\Delta c-1}|}}{\left\lceil \frac{|\Lambda_\rho^l|}{s} \right\rceil \overline{W}_{\mathbf{r}, s}} \right) \right) \left( \left\| \text{vec}(\mathcal{R}^{c^*}) \right\|_F^2 \right. \\ & \quad \left. - \left\| (\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_\rho}^l}} \otimes \dots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_\rho}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\rho^l}) \right. \right. \\ & \quad \left. \left. + \text{vec}(\mathcal{N}) \right\|_F^2 \right), \end{aligned} \quad (170)$$

where (169) is from (168), and (170) is due to the monotonic decreasing property of  $\overline{W}_{\mathbf{r}, t}$  with respect to  $t$ . Subtract both sides of (170) by  $\|\text{vec}(\mathcal{R}^{c^*})\|_F^2 - \left\| (\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_\rho}^l}} \otimes \dots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_\rho}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\rho^l}) + \text{vec}(\mathcal{N}) \right\|_F^2$ , we obtain

$$\begin{aligned} & \|\mathcal{R}^{c^*+1}\|_F^2 \\ & - \left\| (\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_\rho}^l}} \otimes \dots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_\rho}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\rho^l}) + \text{vec}(\mathcal{N}) \right\|_F^2 \\ & \leq \exp \left( - \frac{W_{\mathbf{r}, |\Lambda_\rho^l \cup \Xi^{c+\Delta c-1}|}}{\left\lceil \frac{|\Lambda_\rho^l|}{s} \right\rceil \overline{W}_{\mathbf{r}, s}} \right) \\ & \quad \times \left( \left\| \text{vec}(\mathcal{R}^{c^*}) \right\|_F^2 \right. \\ & \quad \left. - \left\| (\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_\rho}^l}} \otimes \dots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_\rho}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\rho^l}) \right. \right. \\ & \quad \left. \left. + \text{vec}(\mathcal{N}) \right\|_F^2 \right). \end{aligned} \quad (171)$$

Substituting  $c^* = c, \dots, c + \Delta c - 1$  into (171), we have

$$\begin{aligned} & \|\mathcal{R}^{c+1}\|_F^2 \\ & - \left\| (\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_\rho}^l}} \otimes \dots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_\rho}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\rho^l}) + \text{vec}(\mathcal{N}) \right\|_F^2 \\ & \leq \exp \left( - \frac{W_{\mathbf{r}, |\Lambda_\rho^l \cup \Xi^{c+\Delta c-1}|}}{\left\lceil \frac{|\Lambda_\rho^l|}{s} \right\rceil \overline{W}_{\mathbf{r}, s}} \right) \\ & \quad \times \left( \left\| \text{vec}(\mathcal{R}^c) \right\|_F^2 \right. \\ & \quad \left. - \left\| (\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_\rho}^l}} \otimes \dots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_\rho}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\rho^l}) \right. \right. \\ & \quad \left. \left. + \text{vec}(\mathcal{N}) \right\|_F^2 \right), \end{aligned} \quad (172)$$

$$\|\mathcal{R}^{c+2}\|_F^2$$

$$\begin{aligned} & - \left\| (\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_\rho}^l}} \otimes \dots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_\rho}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\rho^l}) + \text{vec}(\mathcal{N}) \right\|_F^2 \\ & \leq \exp \left( - \frac{W_{\mathbf{r}, |\Lambda_\rho^l \cup \Xi^{c+\Delta c-1}|}}{\left\lceil \frac{|\Lambda_\rho^l|}{s} \right\rceil \overline{W}_{\mathbf{r}, s}} \right) \\ & \quad \times \left( \left\| \text{vec}(\mathcal{R}^{c+1}) \right\|_F^2 \right. \\ & \quad \left. - \left\| (\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_\rho}^l}} \otimes \dots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_\rho}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\rho^l}) \right. \right. \\ & \quad \left. \left. + \text{vec}(\mathcal{N}) \right\|_F^2 \right), \end{aligned} \quad (173)$$

$$\begin{aligned} & \vdots \\ & \|\mathcal{R}^{c+\Delta c}\|_F^2 \\ & - \left\| (\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_\rho}^l}} \otimes \dots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_\rho}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\rho^l}) + \text{vec}(\mathcal{N}) \right\|_F^2 \\ & \leq \exp \left( - \frac{W_{\mathbf{r}, |\Lambda_\rho^l \cup \Xi^{c+\Delta c-1}|}}{\left\lceil \frac{|\Lambda_\rho^l|}{s} \right\rceil \overline{W}_{\mathbf{r}, s}} \right) \\ & \quad \times \left( \left\| \text{vec}(\mathcal{R}^{c+\Delta c-1}) \right\|_F^2 \right. \\ & \quad \left. - \left\| (\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_\rho}^l}} \otimes \dots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_\rho}^l}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\rho^l}) \right. \right. \\ & \quad \left. \left. + \text{vec}(\mathcal{N}) \right\|_F^2 \right). \end{aligned} \quad (174)$$

The desired result is obtained by combining the results of (172) to (174). ■

#### M. Proof of Theorem 7

*Proof:* We first present some definitions used in the proof.  $\Lambda^l = \Xi \setminus \Xi^l$  is the index set of the remaining support tensor blocks after  $l$  iterations. Without loss of generality, assume that  $\Lambda^l = \{1, 2, \dots, |\Lambda^l|\}$  corresponding to the nonzero blocks of  $\text{vec}(\mathcal{X})$ , and  $\text{vec}(\mathcal{X})$  is arranged in descending order of their Frobenius norms, i.e.,  $\|\text{vec}(\mathcal{X})_{[1]}\|_F \geq \|\text{vec}(\mathcal{X})_{[2]}\|_F \geq \dots \geq \|\text{vec}(\mathcal{X})_{[|\Lambda^l|]}\|_F$ . The subset  $\Lambda_\rho^l$  of  $\Lambda^l$  is given by

$$\Lambda_\rho^l = \begin{cases} \emptyset, & \rho = 0, \\ \{1, \dots, \lceil \exp(\rho - 1) \rceil s\}, & \rho = 1, \dots, \max \left\{ 0, \left\lceil \log \frac{e|\Lambda^l|}{s} \right\rceil \right\}, \end{cases} \quad (175a)$$

$$\Lambda_\rho^l = \begin{cases} \emptyset, & \rho = 0, \\ \{1, \dots, \lceil \exp(\rho - 1) \rceil s\}, & \rho = 1, \dots, \max \left\{ 0, \left\lceil \log \frac{e|\Lambda^l|}{s} \right\rceil \right\}, \end{cases} \quad (175b)$$

$$\Lambda_\rho^l = \begin{cases} \emptyset, & \rho = 0, \\ \{1, \dots, \lceil \exp(\rho - 1) \rceil s\}, & \rho = 1, \dots, \max \left\{ 0, \left\lceil \log \frac{e|\Lambda^l|}{s} \right\rceil \right\}, \\ \Lambda^l, & \rho = \max \left\{ 0, \left\lceil \log \frac{e|\Lambda^l|}{s} \right\rceil \right\} + 1. \end{cases} \quad (175c)$$

For a given set  $\Lambda^l$  and a constant  $\eta \geq 2$ , let  $E \in \{1, 2, \dots, \max \left\{ 0, \left\lceil \log \frac{e|\Lambda^l|}{s} \right\rceil \right\} + 1\}$  be a positive integer that

$$\left\| \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_0^l}) \right\|_F^2 < \eta \left\| \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_1^l}) \right\|_F^2, \quad (176a)$$

$$\left\| \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_1^l}) \right\|_F^2 < \eta \left\| \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_2^l}) \right\|_F^2, \quad (176b)$$

$$\left\| \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_{E-1}^l}) \right\|_F^2 \geq \eta \left\| \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_E^l}) \right\|_F^2, \quad (176c)$$

It is worth noting that  $E$  always exists according to [42]. Moreover, based on (176a)-(176c), there exist two conclusions for  $E \geq 2$  [42]:

$$\left\| \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\rho^l}) \right\|_F^2 \leq \eta^{E-1-\rho} \left\| \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_{E-1}^l}) \right\|_F^2, \quad \rho = 0, 1, \dots, E, \quad (177)$$

$$|\Lambda^l| > \left(\frac{2\eta - 1}{2\eta - 2}\right) 2^{E-2} s, \quad (178)$$

where (178) can be proved by considering the block structure in  $\text{vec}(\mathcal{X})$  and using the similar proof in [42, Appendix B]. Based on these preparations, we now turn to our proof of **Theorem 7**, which is derived based on the mathematical induction.

a) Consider the case in which  $|\Lambda^l| = 0$ . This indicates that all the support tensor blocks are selected such that

$$\|\mathcal{R}^l\|_F = \|\mathcal{Y} - \hat{\mathcal{X}}^l \times_1 \mathbf{D}_1 \times_2 \cdots \times_n \mathbf{D}_n\|_F \quad (179)$$

$$= \min_{\text{supp}(\mathcal{X}) = \Xi^l} \left\| \text{vec}(\mathcal{Y}) - \sum_{(i_1, \dots, i_n) \in \Xi^l} (\mathbf{D}_{n_{[i_n]}} \otimes \cdots \otimes \mathbf{D}_{1_{[i_1]}}) \times \text{vec}(\mathcal{X}_{[i_1, \dots, i_n]}) \right\|_F \quad (180)$$

$$\leq \|\mathcal{Y} - \mathcal{X} \times_1 \mathbf{D}_1 \times_2 \cdots \times_n \mathbf{D}_n\|_F \quad (181)$$

$$= \|\mathcal{N}\|_F \quad (182)$$

$$\leq \xi \|\mathcal{N}\|_F,$$

where (179) is due to the definition of the residual, (180) is from Algorithm 1, (181) is because  $|\Lambda^l| = 0$ , and (182) is based on (1).

b) Assume that the argument holds up to an integer  $\theta - 1$  ( $\theta \geq 1$ ), i.e., whenever  $|\Lambda^l| \leq \theta - 1$ , the algorithm requires at most  $l_E + \max\{|\Lambda^l| - \lceil \exp(E - 2) \rceil s\}$  additional iterations to choose all the support tensor blocks.

c) Under this inductive assumption, we prove that if  $|\Lambda| = \theta$ , the algorithm also needs at most  $l_E + \max\{|\Lambda^l| - \lceil \exp(E - 2) \rceil s\}$  additional iterations to choose the remaining  $\theta$  support tensor blocks in  $\Xi$ .

Consider the case in which  $E \geq 2$ . Observe that

$$\begin{aligned} \|\mathcal{R}^{l_E}\|_F &= \|\mathbf{P}_{\mathbf{D}_{\Xi^{l_E}}}^\perp (\ddot{\mathbf{D}}_{\Xi \setminus \Xi^{l_E}} \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^{l_E}}) + \text{vec}(\mathcal{N}))\|_F \\ &\geq \|\mathbf{P}_{\mathbf{D}_{\Xi^{l_E}}}^\perp \ddot{\mathbf{D}}_{\Xi \setminus \Xi^{l_E}} \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^{l_E}})\|_F \\ &\quad - \|\mathbf{P}_{\mathbf{D}_{\Xi^{l_E}}}^\perp \text{vec}(\mathcal{N})\|_F \end{aligned} \quad (183)$$

$$\geq \sigma_{\min}(\mathbf{P}_{\mathbf{D}_{\Xi^{l_E}}}^\perp \ddot{\mathbf{D}}_{\Xi \setminus \Xi^{l_E}}) \|\text{vec}(\mathcal{X}_{\Xi \setminus \Xi^{l_E}})\|_F - \|\text{vec}(\mathcal{N})\|_F \quad (184)$$

$$\geq \sigma_{\min}(\ddot{\mathbf{D}}_{\Xi}) \|\text{vec}(\mathcal{X}_{\Xi \setminus \Xi^{l_E}})\|_F - \|\text{vec}(\mathcal{N})\|_F \quad (185)$$

$$\geq \underline{W}_{\mathbf{r}, k}^{\frac{1}{2}} \|\text{vec}(\mathcal{X}_{\Lambda^{l_E}})\|_F - \|\text{vec}(\mathcal{N})\|_F, \quad (186)$$

where (183) is based on the triangle inequality, (184) is because  $\|\mathbf{P}_{\mathbf{D}_{\Xi^{l_E}}}^\perp \text{vec}(\mathcal{N})\|_F \leq \|\text{vec}(\mathcal{N})\|_F$ , (185) is from [19, Lemma 5], and (186) is based on **Corollary 3**.

Let  $l_0 = l$  and  $l_i = l + \alpha \sum_{\varrho=1}^i \left\lceil \frac{|\Lambda_\varrho^l|}{s} \right\rceil$  for each of  $i \in \{1, \dots, E\}$ . Based on **Lemma 3** with  $\varrho = i$ ,  $c = l_{i-1}$  and  $\Delta c = l_i - l_{i-1}$ , we obtain

$$\begin{aligned} \|\mathcal{R}^{l_i}\|_F^2 &- \left\| (\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_n^i}} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_1^i}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^i}) + \text{vec}(\mathcal{N}) \right\|_F^2 \\ &\leq G_{\mathbf{r}, i, l_{i-1}, l_i - l_{i-1}} \left( \|\text{vec}(\mathcal{R}^{l_{i-1}})\|_F^2 \right. \end{aligned}$$

$$\begin{aligned} &- \left\| (\mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_n^i}} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_1^i}}) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^i}) \right. \\ &\quad \left. + \text{vec}(\mathcal{N}) \right\|_F^2 \Big), \end{aligned} \quad (187)$$

for each of  $i \in \{1, \dots, E\}$ . Note that

$$\begin{aligned} G_{\mathbf{r}, i, l_{i-1}, l_i - l_{i-1}} &= \exp\left(-\frac{(l_i - l_{i-1}) \underline{W}_{\mathbf{r}, |\Lambda_\varrho^l \cup \Xi^{l_i-1}|}}{\left\lceil \frac{|\Lambda_\varrho^l|}{s} \right\rceil \overline{W}_{\mathbf{r}, s}}\right) \\ &= \exp\left(-\frac{\left\lceil \frac{\alpha |\Lambda_\varrho^l|}{s} \right\rceil \underline{W}_{\mathbf{r}, |\Lambda_\varrho^l \cup \Xi^{l_i-1}|}}{\left\lceil \frac{|\Lambda_\varrho^l|}{s} \right\rceil \overline{W}_{\mathbf{r}, s}}\right) \\ &\leq \exp\left(-\frac{\alpha \underline{W}_{\mathbf{r}, |\Lambda_\varrho^l \cup \Xi^{l_i-1}|}}{\overline{W}_{\mathbf{r}, s}}\right), \end{aligned}$$

for each of  $i \in \{1, 2, \dots, E\}$ . Since

$$|\Lambda_\varrho^l \cup \Xi^{l_i-1}| \leq |\Lambda^l \cup \Xi^{l_i-1}| \quad (188)$$

$$\leq |\Xi \cup \Xi^{l_i-1}| \quad (189)$$

$$\leq |\Xi \cup \Xi^{l_E}|, \quad (190)$$

where (188) and (189) are from the definitions of  $\Lambda_\varrho^l$  and  $\Lambda^l$ , respectively, and (190) is because  $l_i - 1 \leq l_E$ . Thus, we have

$$G_{\mathbf{r}, i, l_{i-1}, l_i - l_{i-1}} \leq \exp\left(-\frac{\alpha \underline{W}_{\mathbf{r}, |\Xi \cup \Xi^{l_E}|}}{\overline{W}_{\mathbf{r}, s}}\right). \quad (191)$$

From (175a) to (175c), we know that  $|\Lambda_\varrho^l| \leq \lceil \exp(\varrho - 1) \rceil s$  for  $\varrho = 1, 2, \dots, E$ . Hence we obtain

$$l_E = l + \alpha \sum_{\varrho=1}^E \left\lceil \frac{|\Lambda_\varrho^l|}{s} \right\rceil \quad (192)$$

$$\leq l + \alpha \sum_{\varrho=1}^E \lceil \exp(\varrho - 1) \rceil \quad (193)$$

$$\leq l + \alpha \sum_{\varrho=1}^E (\exp(\varrho - 1) + 1)$$

$$= l + \alpha \left( \frac{\exp(E) - 1}{e - 1} + E \right) \leq l + \lceil \exp(E - 2) \rceil s \quad (194)$$

$$\leq l + |\Lambda^l| \quad (195)$$

$$\leq l + \theta, \quad (196)$$

where (192) follows from the definition of  $l_E$ , (193) is from (175a) to (175c), (194) is because  $\alpha = \inf \left\{ x \in \left\{ 1, \dots, \max \left\{ 0, \left\lceil \log \frac{e|\Lambda^l|}{s} \right\rceil \right\} + 1 \right\}, \frac{\exp(x-2)s(e-1)}{\exp(x-1)-1+(e-1)x} \right\}$ , and (195) is from (175a) to (175c). Combining (191) and (195), we have

$$\begin{aligned} G_{\mathbf{r}, i, l_{i-1}, l_i - l_{i-1}} &\leq \exp\left(-\frac{\alpha \underline{W}_{\mathbf{r}, |\Xi \cup \Xi^{l_i+\theta}|}}{\overline{W}_{\mathbf{r}, s}}\right) \\ &\leq \exp\left(-\frac{\alpha \underline{W}_{\mathbf{r}, sl + s\theta + \theta}}{\overline{W}_{\mathbf{r}, s}}\right), \end{aligned}$$

since  $|\Xi \cup \Xi^{l_i+\theta}| = |\Xi^{l_i+\theta}| + |\Lambda^{l_i+\theta}| \leq s(l + \theta) + |\Lambda^l| \leq sl + s\theta + \theta$ .

Letting  $\beta = \exp\left(-\frac{\alpha W_{\mathbf{r},sl+s\theta+\theta}}{W_{\mathbf{r},s}}\right)$ , the following inequality holds:

$$\begin{aligned} & \|\mathcal{R}^{l_E}\|_F^2 \\ & \leq \beta^E \|\text{vec}(\mathcal{R}^{l_E})\|_F^2 \\ & \quad + (1-\beta) \sum_{\varrho=1}^E \beta^{E-\varrho} \left\| \left( \mathbf{D}_{n_{\Lambda_n^l} \setminus \Lambda_n^l} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l} \setminus \Lambda_1^l} \right) \right. \\ & \quad \left. \times \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^l} + \text{vec}(\mathcal{N})) \right\|_F^2 \quad (197) \end{aligned}$$

$$\begin{aligned} & \leq \beta^E \left\| \left( \mathbf{D}_{n_{\Lambda_n^l} \setminus \Lambda_n^l} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l} \setminus \Lambda_1^l} \right) \text{vec}(\mathcal{X}_{\Lambda^l}) + \text{vec}(\mathcal{N}) \right\|_F^2 \\ & \quad + (1-\beta) \sum_{\varrho=1}^E \beta^{E-\varrho} \left\| \left( \mathbf{D}_{n_{\Lambda_n^l} \setminus \Lambda_n^l} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l} \setminus \Lambda_1^l} \right) \right. \\ & \quad \left. \times \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^l} + \text{vec}(\mathcal{N})) \right\|_F^2 \quad (198) \end{aligned}$$

$$\begin{aligned} & \leq \beta^E \left( (1+\gamma) \left\| \left( \mathbf{D}_{n_{\Lambda_n^l} \setminus \Lambda_n^l} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l} \setminus \Lambda_1^l} \right) \text{vec}(\mathcal{X}_{\Lambda^l}) \right\|_F^2 \right. \\ & \quad \left. + (1+\gamma^{-1}) \|\text{vec}(\mathcal{N})\|_F^2 \right) + (1-\beta) \sum_{\varrho=1}^E \beta^{E-\varrho} \\ & \quad \times \left( (1+\gamma) \left\| \left( \mathbf{D}_{n_{\Lambda_n^l} \setminus \Lambda_n^l} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l} \setminus \Lambda_1^l} \right) \right. \right. \\ & \quad \left. \left. \times \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^l} + \text{vec}(\mathcal{N})) \right\|_F + (1+\gamma^{-1}) \beta^{E-\varrho} \|\text{vec}(\mathcal{N})\|_F^2 \right) \quad (199) \end{aligned}$$

$$\begin{aligned} & \leq \beta^E \left( (1+\gamma) \overline{W}_{\mathbf{r},|\Lambda^l|^n} \|\text{vec}(\mathcal{X}_{\Lambda^l})\|_F^2 \right. \\ & \quad \left. + (1+\gamma^{-1}) \|\text{vec}(\mathcal{N})\|_F^2 \right) \\ & \quad + (1-\beta) \sum_{\varrho=1}^E \beta^{E-\varrho} \left( (1+\gamma) \overline{W}_{\mathbf{r},|\Lambda^l|^n} \|\text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^l} + \text{vec}(\mathcal{N}))\|_F \right. \\ & \quad \left. + (1+\gamma^{-1}) \beta^{E-\varrho} \|\text{vec}(\mathcal{N})\|_F^2 \right) \quad (200) \end{aligned}$$

$$\begin{aligned} & = \left( \beta^E (1+\gamma) \overline{W}_{\mathbf{r},|\Lambda^l|^n} \|\text{vec}(\mathcal{X}_{\Lambda^l})\|_F^2 \right. \\ & \quad \left. + (1-\beta) \sum_{\varrho=1}^E \beta^{E-\varrho} (1+\gamma) \overline{W}_{\mathbf{r},|\Lambda^l|^n} \|\text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^l} + \text{vec}(\mathcal{N}))\|_F \right) \\ & \quad + \left( \beta^E (1+\gamma^{-1}) \right. \\ & \quad \left. + (1-\beta) \sum_{\varrho=1}^E \beta^{E-\varrho} (1+\gamma^{-1}) \right) \|\text{vec}(\mathcal{N})\|_F^2 \quad (201) \end{aligned}$$

$$\begin{aligned} & \leq \left( \eta^{E-1} \beta^E + (1-\beta) \sum_{\varrho=1}^E \eta^{E-1-\varrho} \beta^{E-\varrho} \right) \\ & \quad \times (1+\gamma) \overline{W}_{\mathbf{r},|\Lambda^l|^n} \|\text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^l} + \text{vec}(\mathcal{N}))\|_F \\ & \quad + \left( \beta^E + (1-\beta) \sum_{\varrho=1}^E \beta^{E-\varrho} \right) (1+\gamma^{-1}) \|\text{vec}(\mathcal{N})\|_F^2 \quad (202) \end{aligned}$$

$$< \left( (\eta\beta)^E + \sum_{\varrho=0}^{E-1} (\eta\beta)^\varrho \right)$$

$$\begin{aligned} & \times \eta^{-1} (1+\gamma) \overline{W}_{\mathbf{r},|\Lambda^l|^n} \|\text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^l} + \text{vec}(\mathcal{N}))\|_F \\ & \quad + \left( \beta^E + \sum_{\varrho=0}^{E-1} \beta^\varrho \right) (1+\gamma^{-1}) \|\text{vec}(\mathcal{N})\|_F^2 \quad (203) \end{aligned}$$

$$\begin{aligned} & < \left( \sum_{\varrho=0}^{\infty} (\eta\beta)^\varrho \right) \eta^{-1} (1+\gamma) \overline{W}_{\mathbf{r},|\Lambda^l|^n} \|\text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^l} + \text{vec}(\mathcal{N}))\|_F \\ & \quad + \left( \sum_{\varrho=0}^{\infty} \beta^\varrho \right) (1+\gamma^{-1}) \|\text{vec}(\mathcal{N})\|_F^2 \quad (204) \\ & = \frac{1}{\eta(1-\eta\beta)} (1+\gamma) \overline{W}_{\mathbf{r},\theta^n} \|\text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^l} + \text{vec}(\mathcal{N}))\|_F^2 \\ & \quad + \frac{1}{1-\beta} (1+\gamma^{-1}) \|\text{vec}(\mathcal{N})\|_F^2. \quad (205) \end{aligned}$$

Here (197) is derived by first substituting  $\beta$  into (187) to obtain

$$\begin{aligned} & \|\mathcal{R}^{l_E}\|_F^2 \leq \beta \|\text{vec}(\mathcal{R}^{l_E})\|_F^2 \\ & \quad + (1-\beta) \left\| \left( \mathbf{D}_{n_{\Lambda_n^l} \setminus \Lambda_n^l} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l} \setminus \Lambda_1^l} \right) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^l} + \text{vec}(\mathcal{N})) \right\|_F^2, \quad (206) \end{aligned}$$

for each  $i \in \{1, \dots, E\}$ , and (197) is obtained after some additional calculations on (206), while (198) is from [42, Proposition 1], (199) is due to the fact that for  $\gamma > 0$ ,  $\|\mathbf{x} + \mathbf{y}\|_F^2 \leq (1+\gamma)\|\mathbf{x}\|_F^2 + (1+\gamma^{-1})\|\mathbf{y}\|_F^2$  and  $\mathbf{x}, \mathbf{y}$  are vectors, (200) is based on Corollary 3, (201) is from direct simplification, (202) is because (177), (203) is from  $\beta < 1$ , (204) is similar to the derivation of [42, (49)], and (205) holds because  $\eta\beta < 1$ .

Combining (186) and (205) yields

$$\begin{aligned} & \|\text{vec}(\mathcal{X}_{\Lambda^l E})\|_F \\ & \leq \left( \frac{(1+\gamma) \overline{W}_{\mathbf{r},\theta^n}}{\eta(1-\eta\beta) \overline{W}_{\mathbf{r},k}} \right)^{\frac{1}{2}} \|\text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^l} + \text{vec}(\mathcal{N}))\|_F \\ & \quad + \frac{1}{\overline{W}_{\mathbf{r},k}^{\frac{1}{2}}} \left( 1 + \left( \frac{1}{1-\beta} (1+\gamma^{-1}) \right)^{\frac{1}{2}} \right) \|\text{vec}(\mathcal{N})\|_F. \end{aligned}$$

By choosing an appropriate  $\gamma$ , we have  $\left( \frac{(1+\gamma) \overline{W}_{\mathbf{r},\theta^n}}{\eta(1-\eta\beta) \overline{W}_{\mathbf{r},k}} \right)^{\frac{1}{2}} < 1$ .

Then, consider the following two cases. First, if

$$\begin{aligned} & \frac{1}{\overline{W}_{\mathbf{r},k}^{\frac{1}{2}}} \left( 1 + \left( \frac{1}{1-\beta} (1+\gamma^{-1}) \right)^{\frac{1}{2}} \right) \|\text{vec}(\mathcal{N})\|_F \\ & < \left( 1 - \left( \frac{(1+\gamma) \overline{W}_{\mathbf{r},\theta^n}}{\eta(1-\eta\beta) \overline{W}_{\mathbf{r},k}} \right)^{\frac{1}{2}} \right) \|\text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda^l} + \text{vec}(\mathcal{N}))\|_F, \end{aligned}$$

then we directly have

$$\|\mathcal{X}_{\Lambda^l E}\|_F^2 \leq \|\mathcal{X}_{\Lambda^l \setminus \Lambda^l} + \text{vec}(\mathcal{N})\|_F^2. \quad (207)$$

Recall that  $\Lambda^l E = \Xi \setminus \Xi^{l_E}$  and  $l_E = l + \alpha \sum_{\varrho=1}^E \lceil \frac{\Lambda^l E}{s} \rceil$ . Then, (207) indicates that  $|\Lambda^l E| \leq |\Lambda^l \setminus \Lambda^l| = \theta - \lceil \exp(E-2) \rceil s$ , due to the similar analysis in [43, Sec. V-B]. Thus, there remain at most  $\theta - \lceil \exp(E-2) \rceil s$  support tensor blocks after the  $l_E$ th iteration. Then, based on the induction hypothesis, we have

$$\|\mathcal{R}^{l_E + \theta - \lceil \exp(E-2) \rceil s}\|_F \leq \xi \|\mathcal{N}\|_F. \quad (208)$$

Note that

$$\begin{aligned}
 & l_E + \theta - \lceil \exp(E - 2) \rceil s \\
 &= l + \alpha \sum_{\varrho=1}^E \left\lceil \frac{|\Lambda_\varrho^l|}{s} \right\rceil + \theta - \lceil \exp(E - 2) \rceil s \\
 &\leq l + \alpha \sum_{\varrho=1}^E \lceil \exp(\varrho - 1) \rceil + \theta - \lceil \exp(E - 2) \rceil s \\
 &\leq l + \alpha \left( \frac{1 - \exp(E - 1)}{1 - e} + E \right) + \theta - \lceil \exp(E - 2) \rceil s \\
 &\leq l + \theta, \tag{209}
 \end{aligned}$$

where (209) follows from that  $\alpha = \inf \left\{ x \in \left\{ 1, \dots, \max \left\{ 0, \left\lceil \log \frac{e|\Lambda^l|}{s} \right\rceil \right\} + 1 \right\}, \frac{\exp(x-2)s(e-1)}{\exp(x-1)-1+(e-1)x} \right\}$ , and hence  $\alpha \left( \frac{1 - \exp(E-1)}{1-e} + E \right) - \lceil \exp(E - 2) \rceil s \leq 0$ . Thus, we have

$$\|\mathcal{R}^{l+\theta}\|_F \leq \|\mathcal{R}^{l_E + \theta - \lceil \exp(E-2) \rceil s}\|_F \tag{210}$$

$$\leq \xi \|\mathcal{N}\|_F. \tag{211}$$

where (210) is because the residual is non-increasing, and (211) is from (208).

Second, if

$$\begin{aligned}
 & \frac{1}{W_{\mathbf{r},k}^{\frac{1}{2}}} \left( 1 + \left( \frac{1}{1-\beta} (1 + \gamma^{-1}) \right)^{\frac{1}{2}} \right) \|\text{vec}(\mathcal{N})\|_F \\
 & \geq \left( 1 - \left( \frac{(1+\gamma)\overline{W}_{\mathbf{r},\theta^n}}{\eta(1-\eta\beta)\overline{W}_{\mathbf{r},k}} \right)^{\frac{1}{2}} \right) \|\text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_{E-1}^l})\|_F, \tag{212}
 \end{aligned}$$

then we have

$$\|\mathcal{R}^{l+\theta}\|_F \leq \|\mathcal{R}^{l_E}\|_F \tag{213}$$

$$\begin{aligned}
 & \leq \left( \frac{1}{\eta(1-\eta\beta)} (1 + \gamma) \overline{W}_{\mathbf{r},\theta^n} \right)^{\frac{1}{2}} \|\text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_{E-1}^l})\|_F \\
 & \quad + \left( \frac{1}{1-\beta} (1 + \gamma^{-1}) \right)^{\frac{1}{2}} \|\text{vec}(\mathcal{N})\|_F \tag{214}
 \end{aligned}$$

$$\begin{aligned}
 & \leq \left( \frac{\left( \frac{1}{\eta(1-\eta\beta)} (1 + \gamma) \overline{W}_{\mathbf{r},\theta^n} \right)^{\frac{1}{2}} \left( 1 + \left( \frac{1}{1-\beta} (1 + \gamma^{-1}) \right)^{\frac{1}{2}} \right)}{W_{\mathbf{r},k}^{\frac{1}{2}} \left( 1 - \left( \frac{(1+\gamma)\overline{W}_{\mathbf{r},\theta^n}}{\eta(1-\eta\beta)\overline{W}_{\mathbf{r},k}} \right)^{\frac{1}{2}} \right)} \right. \\
 & \quad \left. + \left( \frac{1}{1-\beta} (1 + \gamma^{-1}) \right)^{\frac{1}{2}} \right) \|\text{vec}(\mathcal{N})\|_F \tag{215}
 \end{aligned}$$

$$\leq \xi \|\mathcal{N}\|_F, \tag{216}$$

where (213) is from (196), (214) is due to (205), and (215) is because of (212).

Now consider the case in which  $E = 1$ . Similar to (186), we have

$$\begin{aligned}
 & \|\mathcal{R}^{l+1}\|_F \\
 &= \left\| \mathbf{P}_{\mathbf{D}_{\Xi^{l+1}}}^{\perp} \left( \ddot{\mathbf{D}}_{\Xi \setminus \Xi^{l+1}} \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^{l+1}}) + \text{vec}(\mathcal{N}) \right) \right\|_F \\
 &\geq \left\| \mathbf{P}_{\mathbf{D}_{\Xi^{l+1}}}^{\perp} \ddot{\mathbf{D}}_{\Xi \setminus \Xi^{l+1}} \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^{l+1}}) \right\|_F - \left\| \mathbf{P}_{\mathbf{D}_{\Xi^{l+1}}}^{\perp} \text{vec}(\mathcal{N}) \right\|_F
 \end{aligned}$$

$$\begin{aligned}
 & \geq \sigma_{\min} \left( \mathbf{P}_{\mathbf{D}_{\Xi^{l+1}}}^{\perp} \ddot{\mathbf{D}}_{\Xi \setminus \Xi^{l+1}} \right) \left\| \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^{l+1}}) \right\|_F - \|\text{vec}(\mathcal{N})\|_F \\
 & \geq \sigma_{\min}(\ddot{\mathbf{D}}_{\Xi}) \left\| \text{vec}(\mathcal{X}_{\Xi \setminus \Xi^{l+1}}) \right\|_F - \|\text{vec}(\mathcal{N})\|_F \\
 & \geq W_{\mathbf{r},k}^{\frac{1}{2}} \left\| \text{vec}(\mathcal{X}_{\Lambda^{l+1}}) \right\|_F - \|\text{vec}(\mathcal{N})\|_F. \tag{217}
 \end{aligned}$$

Then, from (168), we obtain

$$\begin{aligned}
 & \|\mathcal{R}^l\|_F^2 - \|\mathcal{R}^{l+1}\|_F^2 \\
 & \geq \frac{W_{\mathbf{r},|\Lambda_\varrho^l \cup \Xi^l|}}{\left\lceil \frac{|\Lambda_\varrho^l|}{s} \right\rceil \overline{W}_{\mathbf{r},s}} \left( \|\text{vec}(\mathcal{R}^l)\|_F^2 \right. \\
 & \quad \left. - \left\| \left( \mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_e}^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_e}^l}} \right) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\varrho^l}) \right. \right. \\
 & \quad \left. \left. + \text{vec}(\mathcal{N}) \right\|_F^2 \right) \\
 & = \frac{W_{\mathbf{r},|\Lambda_\varrho^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r},s}} \left( \|\text{vec}(\mathcal{R}^l)\|_F^2 \right. \\
 & \quad \left. - \left\| \left( \mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_e}^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_e}^l}} \right) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\varrho^l}) \right. \right. \\
 & \quad \left. \left. + \text{vec}(\mathcal{N}) \right\|_F^2 \right), \tag{218}
 \end{aligned}$$

where (218) is because  $\left\lceil \frac{|\Lambda_\varrho^l|}{s} \right\rceil \leq 1$ . Thus, for the same  $\gamma$  as given in (199), we have

$$\begin{aligned}
 & \|\mathcal{R}^{l+1}\|_F^2 \\
 & \leq \left( 1 - \frac{W_{\mathbf{r},|\Lambda_\varrho^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r},s}} \right) \|\mathcal{R}^l\|_F^2 \\
 & \quad + \frac{W_{\mathbf{r},|\Lambda_\varrho^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r},s}} \left\| \left( \mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_e}^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_e}^l}} \right) \right. \\
 & \quad \left. \times \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\varrho^l}) + \text{vec}(\mathcal{N}) \right\|_F^2 \tag{219}
 \end{aligned}$$

$$\begin{aligned}
 & \leq \left( 1 - \frac{W_{\mathbf{r},|\Lambda_\varrho^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r},s}} \right) \left\| \left( \mathbf{D}_{n_{\Lambda_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l}} \right) \right. \\
 & \quad \left. \times \text{vec}(\mathcal{X}_{\Lambda^l}) + \text{vec}(\mathcal{N}) \right\|_F^2 \\
 & \quad + \frac{W_{\mathbf{r},|\Lambda_\varrho^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r},s}} \left\| \left( \mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_1}^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_1}^l}} \right) \right. \\
 & \quad \left. \times \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\varrho^l}) + \text{vec}(\mathcal{N}) \right\|_F^2 \tag{220}
 \end{aligned}$$

$$\begin{aligned}
 & \leq \left( 1 - \frac{W_{\mathbf{r},|\Lambda_\varrho^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r},s}} \right) \\
 & \quad \times \left( (1 + \gamma) \left\| \left( \mathbf{D}_{n_{\Lambda_n^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l}} \right) \text{vec}(\mathcal{X}_{\Lambda^l}) \right\|_F^2 \right. \\
 & \quad \left. + (1 + \gamma^{-1}) \|\text{vec}(\mathcal{N})\|_F^2 \right) \\
 & \quad + \frac{W_{\mathbf{r},|\Lambda_\varrho^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r},s}} \\
 & \quad \times \left( (1 + \gamma) \left\| \left( \mathbf{D}_{n_{\Lambda_n^l \setminus \Lambda_{n_e}^l}} \otimes \cdots \otimes \mathbf{D}_{1_{\Lambda_1^l \setminus \Lambda_{1_1}^l}} \right) \text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_\varrho^l}) \right\|_F^2 \right. \\
 & \quad \left. + (1 + \gamma^{-1}) \|\text{vec}(\mathcal{N})\|_F^2 \right) \tag{221}
 \end{aligned}$$

$$\begin{aligned}
 & \leq \left( 1 - \frac{W_{\mathbf{r},|\Lambda_\varrho^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r},s}} \right) \times \left( (1 + \gamma) \overline{W}_{\mathbf{r},\theta^n} \|\text{vec}(\mathcal{X}_{\Lambda^l})\|_F^2 \right. \\
 & \quad \left. + (1 + \gamma^{-1}) \|\text{vec}(\mathcal{N})\|_F^2 \right)
 \end{aligned}$$

$$+ \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} \left( (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} \|\text{vec}(\mathcal{X}_{\Lambda^l \setminus \Lambda_1^l})\|_F^2 + (1 + \gamma^{-1}) \|\text{vec}(\mathcal{N})\|_F^2 \right) \quad (222)$$

$$\leq \left( 1 - \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} \right) \left( (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} \|\text{vec}(\mathcal{X}_{\Lambda^l})\|_F^2 + (1 + \gamma^{-1}) \|\text{vec}(\mathcal{N})\|_F^2 \right) + \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} \left( (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} \eta^{-1} \|\text{vec}(\mathcal{X}_{\Lambda^l})\|_F^2 + (1 + \gamma^{-1}) \|\text{vec}(\mathcal{N})\|_F^2 \right) \quad (223)$$

$$= \left( \left( 1 - \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} \right) (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} + \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} \eta^{-1} \right) \|\text{vec}(\mathcal{X}_{\Lambda^l})\|_F^2 + (1 + \gamma^{-1}) \|\text{vec}(\mathcal{N})\|_F^2, \quad (224)$$

where (219) is from direct simplification, (220) is derived based on [42, Proposition 1], (221) is similar to the proof of (199), (222) is from **Corollary 3**, and (223) is due to (176c).

Combining (217) and (224) yields

$$\begin{aligned} & \|\text{vec}(\mathcal{X}_{\Lambda^{l+1}})\|_F \\ & \leq \left( \left( 1 - \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} \right) (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} + \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} \right)^{\frac{1}{2}} \frac{1}{\underline{W}_{\mathbf{r}, k}^{\frac{1}{2}}} \|\text{vec}(\mathcal{X}_{\Lambda^l})\|_F \\ & + \frac{1 + (1 + \gamma^{-1})^{\frac{1}{2}}}{\underline{W}_{\mathbf{r}, k}^{\frac{1}{2}}} \|\text{vec}(\mathcal{N})\|_F. \end{aligned} \quad (225)$$

For an appropriate  $\gamma$ , we have

$$\begin{aligned} & \left( \left( 1 - \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} \right) (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} + \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} \right)^{\frac{1}{2}} \frac{1}{\underline{W}_{\mathbf{r}, k}^{\frac{1}{2}}} < 1. \end{aligned}$$

Then, consider the following two cases. On one hand, if

$$\begin{aligned} & \frac{1 + (1 + \gamma^{-1})^{\frac{1}{2}}}{\underline{W}_{\mathbf{r}, k}^{\frac{1}{2}}} \|\text{vec}(\mathcal{N})\|_F \\ & < \left( 1 - \left( \left( 1 - \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} \right) (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} + \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} \right)^{\frac{1}{2}} \frac{1}{\underline{W}_{\mathbf{r}, k}^{\frac{1}{2}}} \right) \|\text{vec}(\mathcal{X}_{\Lambda^l})\|_F \end{aligned}$$

holds, then (225) reveals that

$$\|\mathcal{X}_{\Lambda^{l+1}}\|_F^2 \leq \|\mathcal{X}_{\Lambda^l}\|_F^2.$$

This indicates that  $|\Lambda^{l+1}| < \theta$ . Similarly, there remain at most  $\theta - 1$  support tensor blocks after the  $(l + 1)$ th iteration. Since

$$l + 1 + \theta - 1 = l + \theta,$$

we have

$$\|\mathcal{R}^{l+1+\theta-1}\|_F = \|\mathcal{R}^{l+\theta}\|_F \leq \xi \|\mathcal{N}\|_F.$$

On the other hand, if

$$\begin{aligned} & \frac{1 + (1 + \gamma^{-1})^{\frac{1}{2}}}{\underline{W}_{\mathbf{r}, k}^{\frac{1}{2}}} \|\text{vec}(\mathcal{N})\|_F \\ & \geq \left( 1 - \left( \left( 1 - \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} \right) (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} + \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} \eta^{-1} \right)^{\frac{1}{2}} \frac{1}{\underline{W}_{\mathbf{r}, k}^{\frac{1}{2}}} \right) \|\text{vec}(\mathcal{X}_{\Lambda^l})\|_F \end{aligned} \quad (226)$$

holds, then we have

$$\begin{aligned} & \|\mathcal{R}^{l+\theta}\|_F \leq \|\mathcal{R}^{l+1}\|_F \quad (227) \\ & \leq \left( \left( 1 - \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} \right) (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} + \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} \eta^{-1} \right)^{\frac{1}{2}} \|\text{vec}(\mathcal{X}_{\Lambda^l})\|_F \\ & + (1 + \gamma^{-1})^{\frac{1}{2}} \|\text{vec}(\mathcal{N})\|_F \quad (228) \\ & \leq \left( \left( \left( 1 - \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} \right) (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} + \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} \eta^{-1} \right)^{\frac{1}{2}} (1 + (1 + \gamma^{-1})^{\frac{1}{2}}) \times \underline{W}_{\mathbf{r}, k}^{-\frac{1}{2}} \left( 1 - \left( \left( 1 - \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} \right) (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} + \frac{W_{\mathbf{r}, |\Lambda_1^l \cup \Xi^l|}}{\overline{W}_{\mathbf{r}, s}} (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} \eta^{-1} \right)^{\frac{1}{2}} \frac{1}{\underline{W}_{\mathbf{r}, k}} \right)^{-1} + (1 + \gamma^{-1})^{\frac{1}{2}} \right) \|\text{vec}(\mathcal{N})\|_F \quad (229) \\ & \leq \left( \left( \left( 1 - \frac{W_{\mathbf{r}, sl+s+\theta}}{\overline{W}_{\mathbf{r}, s}} \right) (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} + \frac{W_{\mathbf{r}, sl+s+\theta}}{\overline{W}_{\mathbf{r}, s}} (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} \eta^{-1} \right)^{\frac{1}{2}} (1 + (1 + \gamma^{-1})^{\frac{1}{2}}) \times \underline{W}_{\mathbf{r}, k}^{-\frac{1}{2}} \left( 1 - \left( \left( 1 - \frac{W_{\mathbf{r}, sl+s+\theta}}{\overline{W}_{\mathbf{r}, s}} \right) (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} + \frac{W_{\mathbf{r}, sl+s+\theta}}{\overline{W}_{\mathbf{r}, s}} (1 + \gamma) \overline{W}_{\mathbf{r}, \theta^n} \eta^{-1} \right)^{\frac{1}{2}} \frac{1}{\underline{W}_{\mathbf{r}, k}} \right)^{-1} + (1 + \gamma^{-1})^{\frac{1}{2}} \right) \|\text{vec}(\mathcal{N})\|_F \quad (230) \\ & \leq \xi \|\mathcal{N}\|_F, \end{aligned}$$

where (227) is because the residual is non-increasing, (228) is based on (224), (229) is derived by (226), and (230) is because  $|\Lambda_1^l \cup \Xi^l| \leq |\Xi \cup \Xi^{l+1}| = |\Xi^{l+1}| + |\Lambda^{l+1}|$ . ■

## N. Proof of Theorem 8

*Proof:* Observe that

$$\begin{aligned} & \|\mathcal{R}^k\|_F \\ &= \|\mathcal{Y} - \hat{\mathcal{X}}^k \times_1 \mathbf{D}_1 \times_2 \cdots \times_n \mathbf{D}_n\|_F \end{aligned} \quad (231)$$

$$\begin{aligned} &= \|\mathcal{X} \times_1 \mathbf{D}_1 \times_2 \cdots \times_n \mathbf{D}_n + \mathcal{N} \\ &\quad - \hat{\mathcal{X}}^k \times_1 \mathbf{D}_1 \times_2 \cdots \times_n \mathbf{D}_n\|_F \end{aligned} \quad (232)$$

$$\begin{aligned} &= \|(\mathcal{X} - \hat{\mathcal{X}}^k) \times_1 \mathbf{D}_1 \times_2 \cdots \times_n \mathbf{D}_n + \mathcal{N}\|_F \\ &\geq \|(\mathcal{X} - \hat{\mathcal{X}}^k) \times_1 \mathbf{D}_1 \times_2 \cdots \times_n \mathbf{D}_n\|_F - \|\mathcal{N}\|_F \end{aligned} \quad (233)$$

$$\geq \underline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k_s}^{-\frac{1}{2}} \|(\mathcal{X} - \hat{\mathcal{X}}^k)\|_F - \|\mathcal{N}\|_F, \quad (234)$$

where (231) is from the definition of residual, (232) is based on (1), (233) is derived by the triangle inequality, and (234) is from Corollary 3.

Thus, based on (234) and Corollary 5, we have

$$\begin{aligned} \|\mathcal{X} - \hat{\mathcal{X}}^k\|_F &\leq \underline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k_s}^{-\frac{1}{2}} (\|\mathcal{R}^k\|_F + \|\mathcal{N}\|_F) \\ &\leq \underline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k_s}^{-\frac{1}{2}} (\xi^* + 1) \|\mathcal{N}\|_F. \end{aligned} \quad (235)$$

On the other hand, similar to the proof in Appendix I, using  $\hat{\mathcal{Q}}^k$  as the best  $k$  block-sparse tensor approximation of  $\hat{\mathcal{X}}^k$ , we have

$$\|\hat{\mathcal{Q}}^k - \mathcal{X}\|_F = \|\hat{\mathcal{Q}}^k - \hat{\mathcal{X}}^k + \hat{\mathcal{X}}^k - \mathcal{X}\|_F \quad (236)$$

$$\leq \|\hat{\mathcal{Q}}^k - \hat{\mathcal{X}}^k\|_F + \|\hat{\mathcal{X}}^k - \mathcal{X}\|_F \quad (237)$$

$$\leq 2\|\hat{\mathcal{X}}^k - \mathcal{X}\|_F$$

$$\leq 2\underline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k_s}^{-\frac{1}{2}} (\xi^* + 1) \|\mathcal{N}\|_F, \quad (238)$$

where (236) is because  $\hat{\mathcal{Q}}^k$  is the best approximation, (237) is from the triangle inequality, and (238) is based on (235).

From (119), we have

$$\|\hat{\mathcal{Q}}^k - \mathcal{X}\|_F \geq \frac{\underline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k_s}^{-\frac{1}{2}} \|(\hat{\mathcal{X}}^k - \mathcal{X})\|_F - 2\|\mathcal{N}\|_F}{\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k_s}^{-\frac{1}{2}}}. \quad (239)$$

Then, using (235) and (239), we have

$$\begin{aligned} & \frac{\underline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k_s}^{-\frac{1}{2}} \|(\hat{\mathcal{X}}^k - \mathcal{X})\|_F - 2\|\mathcal{N}\|_F}{\overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k_s}^{-\frac{1}{2}}} \\ & \leq 2\underline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k_s}^{-\frac{1}{2}} (\xi^* + 1) \|\mathcal{N}\|_F, \end{aligned}$$

i.e.,

$$\begin{aligned} & \|\hat{\mathcal{X}}^k - \mathcal{X}\|_F \\ & \leq \frac{2\left(\underline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k_s}^{-\frac{1}{2}} (\xi^* + 1) \overline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k_s}^{-\frac{1}{2}} + 1\right)}{\underline{W}_{\mathbf{r}, \prod_{t=1}^n k_t + k_s}^{-\frac{1}{2}}} \|\mathcal{N}\|_F. \end{aligned}$$

## REFERENCES

- [1] P. M. Hoang, H. D. Tuan, T. T. Son, H. V. Poor, and L. Hanzo, "Learning unbalanced and sparse low-order tensors," *IEEE Trans. Signal Process.*, vol. 70, pp. 5624–5638, 2022.
- [2] C. Cai, G. Li, H. V. Poor, and Y. Chen, "Nonconvex low-rank tensor completion from noisy data," *Oper. Res.*, vol. 70, no. 2, pp. 1219–1237, Mar. 2022.
- [3] C. Cai, H. V. Poor, and Y. Chen, "Uncertainty quantification for nonconvex tensor completion: Confidence intervals, heteroscedasticity and optimality," *IEEE Trans. Inf. Theory*, vol. 69, no. 1, pp. 407–452, Jan. 2023.
- [4] C.-H. Lo and P.-Y. Tsai, "Tensor-based orthogonal matching pursuit with phase rotation for channel estimation in hybrid beamforming MIMO-OFDM systems," in *Proc. ICASSP*, May 2022, pp. 101–105.
- [5] H. Zheng, C. Zhou, Z. Shi, Y. Gu, and Y. D. Zhang, "Coarray tensor direction-of-arrival estimation," *IEEE Trans. Signal Process.*, vol. 71, pp. 1128–1142, 2023.
- [6] L. Lu, W. Xu, Y. Wang, and Z. Tian, "Recovery conditions of sparse signals using orthogonal least squares-type algorithms," *IEEE Trans. Signal Process.*, vol. 70, pp. 4727–4741, 2022.
- [7] H. Zhang, M. Slawski, and P. Li, "The benefits of diversity: Permutation recovery in unlabeled sensing from multiple measurement vectors," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, pp. 2509–2529, Apr. 2022.
- [8] D. C. Araújo, A. L. F. de Almeida, J. P. C. L. D. Costa, and R. T. de Sousa, "Tensor-based channel estimation for massive MIMO-OFDM systems," *IEEE Access*, vol. 7, pp. 42133–42147, 2019.
- [9] S. Zubair and W. Wang, "Signal classification based on block-sparse tensor representation," in *Proc. ICDCSP*, Hong Kong, Aug. 2014, pp. 361–365.
- [10] Y. C. Eldar, P. Kuppinger, and H. Boleskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.
- [11] L. Lu, W. Xu, Y. Wang, and Z. Tian, "Supervised dictionary learning for block threshold feature in compressive spectrum sensing," *IEEE Trans. Cognit. Commun. Netw.*, vol. 8, no. 4, pp. 1632–1646, Dec. 2022.
- [12] M. Azghani, A. Abtahi, and F. Marvasti, "Simultaneous block iterative method with adaptive thresholding for cooperative spectrum sensing," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 5598–5605, Jun. 2019.
- [13] L. Lu, W. Xu, Y. Wang, and Z. Tian, "Compressive spectrum sensing using sampling-controlled block orthogonal matching pursuit," *IEEE Trans. Commun.*, vol. 71, no. 2, pp. 1096–1111, Feb. 2023.
- [14] J. Wen, J. Wang, and Q. Zhang, "Nearly optimal bounds for orthogonal least squares," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5347–5356, Oct. 2017.
- [15] J. Kim, J. Wang, and B. Shim, "Optimal restricted isometry condition of normalized sampling matrices for exact sparse recovery with orthogonal least squares," *IEEE Trans. Signal Process.*, vol. 69, pp. 1521–1536, 2021.
- [16] C. Herzet, C. Soussen, J. Idier, and R. Gribonval, "Exact recovery conditions for sparse representations with partial support information," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7509–7524, Nov. 2013.
- [17] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [18] C. Herzet, A. Drémeau, and C. Soussen, "Relaxed recovery conditions for OMP/OLS by exploiting both coherence and decay," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 459–470, Jan. 2016.
- [19] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.
- [20] R. Boyer and M. Haardt, "Noisy compressive sampling based on block-sparse tensors: Performance limits and beamforming techniques," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6075–6088, Dec. 2016.
- [21] F. Li and X. Zhao, "Block-structured compressed spectrum sensing with Gaussian mixture noise distribution," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1183–1186, Aug. 2019.
- [22] A. Pizzo, T. L. Marzetta, and L. Sanguinetti, "Spatially-stationary model for holographic MIMO small-scale fading," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 1964–1979, Sep. 2020.
- [23] L. Wei et al., "Multi-user holographic MIMO surfaces: Channel modeling and spectral efficiency analysis," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 1112–1124, Aug. 2022.
- [24] A. Pizzo, L. Sanguinetti, and T. L. Marzetta, "Fourier plane-wave series expansion for holographic MIMO communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 6890–6905, Sep. 2022.

- [25] Y. Guo, Y. Chen, and Y. Wang, "Channel estimation for holographic MIMO: Wavenumber-domain sparsity inspired approaches," 2024, *arXiv:2405.05641*.
- [26] L. Lu, K. Ma, and Z. Wang, "Block-dominant compressed sensing for near-field communications: Fundamentals, solutions and future directions," 2024, *arXiv:2403.12369*.
- [27] H. Wang, J. Fang, H. Duan, and H. Li, "Near/far-field channel estimation for terahertz systems with ELAAs: A block-sparse-aware approach," 2024, *arXiv:2404.05544*.
- [28] C. F. Caiafa and A. Cichocki, "Block sparse representations of tensors using Kronecker bases," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 2709–2712.
- [29] N. Li, N. Pfeifer, and C. Liu, "Tensor-based sparse representation classification for urban airborne LiDAR points," *Remote Sens.*, vol. 9, no. 12, p. 1216, Nov. 2017.
- [30] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3336–3351, Aug. 2014.
- [31] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Dictionary optimization for block-sparse representations," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2386–2395, May 2012.
- [32] Y. Wang, Y. Y. Tang, L. Li, and X. Zheng, "Block sparse representation for pattern classification: Theory, extensions and applications," *Pattern Recognit.*, vol. 88, pp. 198–209, Apr. 2019.
- [33] Z. Wang et al., "A tutorial on extremely large-scale MIMO for 6G: Fundamentals, signal processing, and applications," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 3, pp. 1560–1605, 3rd Quart., 2024.
- [34] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [35] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.
- [36] J. Wang and P. Li, "Recovery of sparse signals using multiple orthogonal least squares," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2049–2062, Apr. 2017.
- [37] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY, USA: Cambridge Univ. Press, 1985.
- [38] C. Soussen, R. Gribonval, J. Idier, and C. Herzet, "Joint  $k$ -step analysis of orthogonal matching pursuit and orthogonal least squares," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3158–3174, May 2013.
- [39] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, Nov. 2009.
- [40] Y. Du et al., "Block-sparsity-based multiuser detection for uplink grant-free NOMA," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 7894–7909, Dec. 2018.
- [41] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 2, no. 2, pp. 302–309, Mar. 1991.
- [42] J. Wang, S. Kwon, P. Li, and B. Shim, "Recovery of sparse signals via generalized orthogonal matching pursuit: A new analysis," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 1076–1089, Feb. 2016.
- [43] J. Kim, J. Wang, L. T. Nguyen, and B. Shim, "Joint sparse recovery using signal space matching pursuit," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 5072–5096, Aug. 2020.
- [44] L. Lu, W. Xu, Y. Wang, and Z. Tian, "Blind orthogonal least squares based compressive spectrum sensing," *IEEE Trans. Veh. Technol.*, vol. 72, no. 12, pp. 16772–16776, Dec. 2023.

**Liyang Lu** (Member, IEEE) received the B.S. degree from the School of Information Engineering, Beijing University of Posts and Telecommunications (BUPT), China, in 2017, and the Ph.D. degree from the School of Artificial Intelligence, BUPT, in 2022. He is currently a Post-Doctoral Fellow with Tsinghua University. His research interests include compressed sensing, cognitive radios, and MIMO communications.

**Zhaocheng Wang** (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees from Tsinghua University in 1991, 1993, and 1996, respectively. From 1996 to 1997, he was a Post-Doctoral Fellow with Nanyang Technological University, Singapore. From 1997 to 1999, he was a Research Engineer/a Senior Engineer with OKI Techno Centre (Singapore) Pte., Ltd., Singapore. From 1999 to 2009, he was a Senior Engineer/a Principal Engineer with Sony Deutschland GmbH, Germany. Since 2009, he has been a Professor with the Department of Electronic Engineering, Tsinghua University, where he is currently the Director of the Broadband Communication Key Laboratory, Beijing National Research Center for Information Science and Technology (BNRist). He has authored or co-authored two books, which have been selected by the IEEE Press Series on Digital and Mobile Communication (Wiley IEEE Press). He has also authored/co-authored more than 200 peer-reviewed journal articles. He holds 60 U.S./EU granted patents (23 of them as the first inventor). His research interests include wireless communications, millimeter wave communications, and optical wireless communications. He is a fellow of the Institution of Engineering and Technology. He was a recipient of the ICC2013 Best Paper Award, the OECC2015 Best Student Paper Award, the 2016 IEEE Scott Helt Memorial Award, the 2016 IET Premium Award, the 2016 National Award for Science and Technology Progress (First Prize), the ICC2017 Best Paper Award, the 2018 IEEE ComSoc AsiaPacific Outstanding Paper Award, and the 2020 IEEE ComSoc Leonard G. Abraham Prize.

**Zhen Gao** (Member, IEEE) received the B.S. degree in information engineering from Beijing Institute of Technology, Beijing, China, in 2011, and the Ph.D. degree in communication and signal processing from Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, China, in 2016. He is currently an Associate Professor with Beijing Institute of Technology. His current research interests include wireless communications, with a focus on multi-carrier modulations, multiple antenna systems, and sparse signal processing. He was a recipient of the 2016 IEEE Broadcast Technology Society Scott Helt Memorial Award (Best Paper), the Exemplary Reviewer Award of IEEE COMMUNICATIONS LETTERS in 2016, the IET Electronics Letters Premium Award (Best Paper) in 2016, and the Young Elite Scientists Sponsorship Program from China Association for Science and Technology from 2018 to 2021.

**Sheng Chen** (Life Fellow, IEEE) received the B.Eng. degree from East China Petroleum Institute, Dongying, China, in January 1982, the Ph.D. degree from the City, University of London, London, in September 1986, both in control engineering, and the Doctor of Sciences (D.Sc.) degree from the University of Southampton, Southampton, U.K., in August 2005. From 1986 to 1999, he held research and academic appointments at The University of Sheffield, The University of Edinburgh, and the University of Portsmouth; U.K. Since 1999, he has been with the School of Electronics and Computer Science, University of Southampton, where he holds the post of a Professor of intelligent systems and signal processing. He has published over 700 research articles. He has more than 20 000 Web of Science citations with H-index 61 and more than 39 000 Google Scholar citations with H-index 83. His research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural network and machine learning, and evolutionary computation methods and optimization. He is a fellow of the Royal Academy of Engineering, U.K.; Asia-Pacific Artificial Intelligence Association; and IET. He is one of the original ISI highly cited researchers in engineering (March 2004).

**H. Vincent Poor** (Life Fellow, IEEE) received the Ph.D. degree in EECS from Princeton University in 1977. From 1977 to 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990, he has been on the faculty at Princeton University, where he is currently the Michael Henry Strater University Professor. From 2006 to 2016, he served as the Dean of Princeton's School of Engineering and Applied Science, and he held visiting appointments at several other universities, including most recently at the University of California, Berkeley, and the University of Cambridge. His research interests are in the areas of information theory, machine learning and network science, and their applications in wireless networks, energy systems, and related fields. Among his publications in these areas is the book *Machine Learning and Wireless Communications* (Cambridge University Press, 2022). He is a member of the National Academy of Engineering and the National Academy of Sciences and a foreign member of the Royal Society and other national and international academies. He received the IEEE Alexander Graham Bell Medal in 2017.