

Randomized Algorithms for Computation of Tucker Decomposition and Higher Order SVD (HOSVD)

SALMAN AHMADI-ASL¹, **STANISLAV ABUKHOVICH¹**,
MAAME G. ASANTE-MENASH¹, (Student Member, IEEE),
ANDRZEJ CICHOCKI¹, (Fellow, IEEE), **ANH-HUY PHAN¹**, (Member, IEEE),
TOSHIHISA TANAKA², (Senior Member, IEEE), AND **IVAN OSELEDETS¹**

¹CDISE, Skolkovo Institute of Science and Technology (SKOLTECH), 121205 Moscow, Russia

²Institute of Global Innovation and Research, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan

Corresponding author: Salman Ahmadi-Asl (s.asl@skoltech.ru)

This work was supported in part by the Ministry of Education and Science of the Russian Federation through the Mega Grant Project 14.756.31.0001.

ABSTRACT Big data analysis has become a crucial part of new emerging technologies such as the internet of things, cyber-physical analysis, deep learning, anomaly detection, etc. Among many other techniques, dimensionality reduction plays a key role in such analyses and facilitates feature selection and feature extraction. Randomized algorithms are efficient tools for handling big data tensors. They accelerate decomposing large-scale data tensors by reducing the computational complexity of deterministic algorithms and the communication among different levels of memory hierarchy, which is the main bottleneck in modern computing environments and architectures. In this article, we review recent advances in randomization for computation of Tucker decomposition and Higher Order SVD (HOSVD). We discuss random projection and sampling approaches, single-pass and multi-pass randomized algorithms and how to utilize them in the computation of the Tucker decomposition and the HOSVD. Simulations on synthetic and real datasets are provided to compare the performance of some of best and most promising algorithms.

INDEX TERMS Randomized algorithm, tensor decomposition, random projection, sampling, unfolding, Tucker decomposition, HOSVD.

I. INTRODUCTION

Real world data often are multidimensional and naturally are represented as tensors or multidimensional (multiway) arrays. It is crucial to preserve the multidimensional structure of the data tensors in order to extract meaningful latent variables and reveal the hidden structures of the data tensors. Tucker decomposition [1]–[3] is a natural generalization of the SVD to higher-order data tensors and has found various applications such as reducing the number of parameters in deep neural networks [4], [5], handwritten digit classification [6], computer vision [7], recommender systems [8]–[11], signal processing [12]–[14], etc. Deterministic algorithms for decomposing large-scale data tensors into the Tucker

format are prohibitive and require high memory and computational complexity or only applicable for structured data tensors [15], [16]. It has been proved that randomized algorithms can tackle this difficulty by exploiting only a part of the data tensors with applications in tensor regression [17], tensor completion [18], [19] and deep learning [20]. Because of this property, they scale quite well to the tensor dimensions and orders. Two main features of the randomized algorithms which make them suitable approaches for handling large-scale data tensors are: 1) they are parallelizable and 2) they reduce the communication among different levels of memories, which is the main bottleneck in modern computing environments and architectures. In particular, the second property is important for the data tensors stored out-of-cores, where, the communication cost may exceed our main computations. Due to the two mentioned benefits, in recent years, there is

The associate editor coordinating the review of this manuscript and approving it for publication was Zhipeng Cai¹.

a growing interest in developing randomized algorithms for the Tucker decomposition and Higher Order SVD¹ (HOSVD) [21]. Randomized algorithms for the Tucker decomposition are mainly categorized into four groups as follows (see Figure 1 and Table 1)

- Random projection
- Sampling
- Count-sketch
- Random least-squares

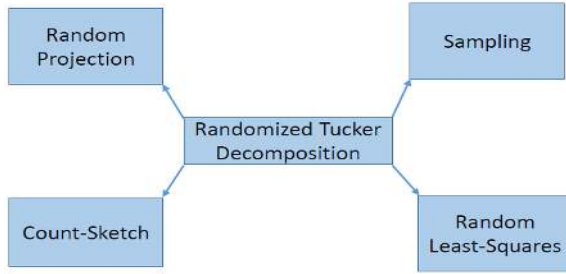


FIGURE 1. Taxonomy of randomized algorithms for low-rank Tucker approximation.

TABLE 1. Randomized Tucker references according to four categories.

Algorithms	Ref
Random Projection	[22]–[27]
Sampling	[28]–[37]
Count-Sketch	[38]–[40]
Random least-squares	[38]

Following the above categories, in this article, we systematically review a variety of randomized algorithms for decomposing large-scale tensors in the Tucker and the HOSVD formats.

The remainder of this article is structured as follows. In Section II, we introduce preliminary concepts and notations used throughout the paper. In Section III, the randomized algorithms for low-rank matrix approximation are presented. In Section IV, the Tucker decomposition and the HOSVD are briefly introduced. The randomized variants of the deterministic Tucker and HOSVD algorithms are presented in Section V. An application of the randomized Tucker algorithms is presented for fast computation of Canonical Polyadic Decomposition (CPD) in Section VI. Further challenges and open problems are discussed in Section VII. Extensive simulations on synthetic and real data tensors are provided in Section VIII to compare the performance of some of the presented randomized algorithms. Conclusions are given in Section IX.

II. PRELIMINARY CONCEPTS AND NOTATIONS

In this section, we present some concepts and notations used throughout the paper.

¹It is also called MultiLinear SVD (MLSVD).

Tensors,² matrices and vectors are respectively denoted by underlined bold upper case letters, e.g., $\underline{\mathbf{X}}$, bold upper case letters, e.g., \mathbf{X} and bold lower case letters, e.g., \mathbf{x} . The number of indices or modes of a tensor is called order. Fibers are first order tensors produced by fixing all modes except one, while slices are second order tensors generated by fixing all modes except two of them.

The symbols \otimes , \dagger , T denote the Kronecker product, the Moore-Penrose pseudoinverse and the transpose of matrices, respectively. The Frobenius norm of tensors and spectral norm of matrices are respectively denoted by $\|\cdot\|_F$ and $\|\cdot\|_2$. In the case of vectors, $\|\cdot\|_2$ denotes the Euclidean norm.

Tensors can be reshaped to matrices and this is called unfolding, matricization or flattening, and the corresponding generated matrices are called unfolding or flattening matrices. In the context of tensor computations, unfolding matrices are of interest because of their applications in computing different types of tensor decompositions where low-rank approximations of these matrices are required in each step of the underlying algorithms. For example, it is known that the factor matrices of the HOSVD are computed via the left singular vectors of the unfolding matrices [21]. The HOSVD is a direct generalization of the classical SVD of matrices to tensors where instead of rank, multilinear rank or Tucker rank is defined which is a N -tuple containing the rank of unfolding matrices along different modes. There are several types of the unfolding matrices including n -unfolding and mode- n unfolding [41]. The first one is used in computation of the Tucker decomposition and the HOSVD.

Given an N th-order tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, then the n -unfolding matrix of the tensor $\underline{\mathbf{X}}$ is denoted by $\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times I_1 \dots I_{n-1} I_{n+1} \dots I_N}$, whose components are defined as follows

$$\mathbf{X}_{(n)}(i_n, j) = \underline{\mathbf{X}}(i_1, i_2, \dots, i_N),$$

where

$$j = 1 + \sum_{k=1, k \neq n}^N (i_k - 1) J_k, \quad J_k = \prod_{m=1, m \neq n}^{k-1} I_m.$$

Although the unfolding operator can be defined in a general setting (see [42]), however throughout the paper, we use the n -unfolding.

Tensors and matrices can be multiplied in different modes which have the same dimensions (sizes). This is called tensor-matrix multiplication along mode n or n -mode (matrix) product of a tensor by a matrix and is a generalization of the matrix-matrix multiplication. To be precise, let $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ and $\mathbf{B} \in \mathbb{R}^{J \times I_n}$, then the tensor $\underline{\mathbf{X}}$ and the matrix \mathbf{B} can be multiplied along mode n as $\underline{\mathbf{X}} \times_n \mathbf{B} \in \mathbb{R}^{I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N}$ which is defined as follows

$$(\underline{\mathbf{X}} \times_n \mathbf{B})_{i_1, \dots, i_{n-1}, j, i_{n+1}, \dots, i_N} = \sum_{i_n=1}^{I_n} x_{i_1, i_2, \dots, i_N} b_{j, i_n},$$

²Tensors are also called multidimensional arrays or multi-ways data. From this perspective, vectors and matrices are first order and second order tensors, respectively.

for $j = 1, 2, \dots, J$. Assume that tensor $\underline{\mathbf{X}}$ and two matrices \mathbf{A}, \mathbf{B} are of conforming sizes, then we have

$$\underline{\mathbf{X}}_{\times n} \mathbf{B}_{\times n} \mathbf{A} = \underline{\mathbf{X}}_{\times n} \mathbf{AB}. \quad (1)$$

If a tensor is multiplied with several matrices along all its modes except one, then its n -unfolding can be computed as follows

$$\begin{aligned} \underline{\mathbf{Y}} &= \underline{\mathbf{X}}_{\times 1} \mathbf{Q}^{(1)} \times_2 \mathbf{Q}^{(2)} \dots \times_N \mathbf{Q}^{(N)} \Leftrightarrow \\ \mathbf{X}_{(n)} &= \mathbf{Q}^{(n)} \mathbf{S}_{(n)} \left(\mathbf{Q}^{(N)} \otimes \dots \otimes \mathbf{Q}^{(n+1)} \otimes \mathbf{Q}^{(n-1)} \otimes \dots \otimes \mathbf{Q}^{(1)} \right)^T. \end{aligned} \quad (2)$$

III. RANDOMIZED LOW-RANK MATRIX ALGORITHMS

Randomized algorithms are utilized for finding low-rank approximation of large-scale data matrices in reasonable amounts of time and memory. They have found many applications in the area of machine learning and data mining, where we encounter large-scale matrices with more than hundreds of millions nonzero entries. Clearly, deterministic approaches are time-consuming in processing such data matrices and it is required to reduce the computational cost and also memory requirement of algorithms in order to process the data matrices more efficiently. Randomized algorithms can be used in such scenarios by finding a low approximation of unfolding matrices $\mathbf{X} = \mathbf{X}_{(n)}$, $n = 1, 2, \dots, N$ using the following formula

$$\begin{aligned} \mathbf{X} &\cong \left(\mathbf{Q}^{(1)} \mathbf{Q}^{(1)T} \right) \mathbf{X} \left(\mathbf{Q}^{(2)} \mathbf{Q}^{(2)T} \right), \\ &= \mathbf{X}_{\times 1} \mathbf{Q}^{(1)} \mathbf{Q}^{(1)T} \times_2 \mathbf{Q}^{(2)} \mathbf{Q}^{(2)T}, \end{aligned} \quad (3)$$

where $\mathbf{Q}^{(1)} \in \mathbb{R}^{I_n \times \hat{R}_n}$, $\mathbf{Q}^{(2)} \in \mathbb{R}^{\left(\prod_{i \neq n} I_i \right) \times \tilde{R}_n}$ ($\hat{R}_n, \tilde{R}_n < \text{rank}(\mathbf{X})$) are approximate orthonormal bases for the column and row spaces of the matrix \mathbf{X} .

For an unfolding matrix $\mathbf{X} = \mathbf{X}_{(n)}$, the dimension of the second mode is usually much larger than the first one, i.e., $\prod_{i \neq n} I_i \gg I_n$, and because of this, it is often enough to make reduction in the second mode and just consider $\mathbf{Q} = \mathbf{Q}^{(1)}$.

The main steps of the randomized algorithms for low-rank matrix approximation are

- **Reduction.** Replacing an extremely large-scale matrix with a new one of smaller size compared to the original one capturing either the column or the row space of the original data matrix as much as possible. This step can be considered as a preprocessing step which makes reduction in the data.
- **SVD computation.** Applying classical deterministic algorithms, e.g., truncated SVD to a reduced data matrix and finding its low-rank approximation.³

³In [43], randomization is also used in the second step. This is considered as a two-step randomized algorithm for Nyström kernel matrix approximation. At the first step, authors use a sub-sampling approach after which they apply randomized SVD instead of deterministic SVD.

- **Recovering.** Recovering the SVD of the original data matrix from the SVD of the compressed one.

The first step can be done by following techniques

- Random Projection
- Sampling
- Count-Sketch.

A. RANDOM PROJECTION

The main idea of this technique is approximating the column (row) space of a given matrix by few new columns (rows) each of which is a linear combination of the columns (rows) of the original data matrix, where the coefficients of the linear combinations come from some probability distributions such as Gaussian, Bernoulli or uniform distribution.⁴

This procedure is performed by multiplying a given matrix by a random matrix from the right-hand or left-hand side. It has been shown that this preserves the Euclidean distances among columns or rows approximately⁵ [44], [45]. Let $\mathbf{X} \in \mathbb{R}^{I \times J}$ be a given data matrix, and R be a target rank. In the first stage of the random projection approach, we generate a random matrix $\Omega = [\omega_1, \omega_2, \dots, \omega_R] \in \mathbb{R}^{J \times R}$ whose components are independent and identically distributed (i.i.d) and taken from some probability distributions, e.g., Gaussian distribution, and then multiply the matrix \mathbf{X} with the columns ω_r as follows

$$\mathbf{y}_r = \mathbf{X} \omega_r \in \mathbb{R}^I, \quad r = 1, 2, \dots, R. \quad (4)$$

The compact form of relation (4) is $\mathbf{Y} = \mathbf{X} \Omega$, where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_R] \in \mathbb{R}^{I \times R}$. The matrix \mathbf{Y} has a smaller size than the matrix \mathbf{X} and is an approximation for the range of \mathbf{X} . Note that the columns of the matrix \mathbf{Y} are generally not orthogonal and in order to compute an orthogonal projection onto the range of \mathbf{Y} , it is necessary to generate a new orthonormal basis of \mathbf{Y} . This can be done by computing the economic QR decomposition of the matrix \mathbf{Y} as $\mathbf{Y} = \mathbf{Q} \mathbf{R}$ and using the matrix $\mathbf{Q} \in \mathbb{R}^{I \times R}$ in subsequent computations. Here, we have

$$\mathbf{X} \cong \mathbf{Q} \mathbf{Q}^T \mathbf{X}, \quad (5)$$

and having computed the SVD of the matrix $\mathbf{B} = \mathbf{Q}^T \mathbf{X} \in \mathbb{R}^{R \times J}$, as $\mathbf{B} = \mathbf{U} \mathbf{S} \mathbf{V}^T$, the SVD of the matrix \mathbf{X} is recovered as follows

$$\mathbf{X} \cong \bar{\mathbf{U}} \mathbf{S} \mathbf{V}^T, \quad (6)$$

where $\bar{\mathbf{U}} = \mathbf{Q} \mathbf{U}$. Note that the size of \mathbf{B} is much smaller than \mathbf{X} and it is much easier to be handled.

Also, if we do not need the SVD of the original data matrix \mathbf{X} and just a low-rank approximation of it is required, this can be easily achieved by the following approximation

$$\mathbf{X} \cong \mathbf{Q} \mathbf{B}, \quad \mathbf{B} = \mathbf{Q}^T \mathbf{X} \in \mathbb{R}^{R \times J}. \quad (7)$$

Low-rank matrix approximation (7) is known as QB approximation of the matrix \mathbf{X} .

⁴It is mentioned in [46] that the difference among different random matrices is negligible. We have also confirmed this in our simulations.

⁵Depending on which side the random matrix is multiplied.

Remark 1 (Oversampling and Power Iteration Methods):

The oversampling and power iteration techniques can be used to improve the solution accuracy of the randomized algorithms. In the oversampling technique, additional random vectors ($R + p$ random vectors instead of R random vectors) are used to better capture the range of a matrix. Also, the power iteration technique is used when the decay rate of the singular values are not very fast. This technique replaces a matrix \mathbf{X} by $(\mathbf{X}\mathbf{X}^T)^q \mathbf{X}$ (q is a nonnegative integer number) and randomized algorithms are applied to this new matrix. Considering the SVD, $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, we have $(\mathbf{X}\mathbf{X}^T)^q \mathbf{X} = \mathbf{U}\mathbf{S}^{2q+1}\mathbf{V}^T$, and it is seen that the left and right singular vectors of the new matrix are the same as those of the matrix \mathbf{X} , but the singular values of the later have a faster decay rate. This can improve the solution accuracy obtained by the randomized algorithms. Selection of the oversampling and the power iteration parameters depend on factors such as *the matrix dimensions, the singular spectrum decay and the random test matrices used in random projection*. However, if the Gaussian random matrices are used, it was experimentally confirmed that oversampling $p = 5, 10$ and power iteration $q = 1, 2$ are enough to achieve a good approximation [46].

The most computationally expensive and time-consuming part of the random projection technique is multiplying an original data matrix with random matrices. This costs $\mathcal{O}(IJR)$ flops. In order to accelerate this step, structured random matrices such as sparse random matrices [47], [48] and subsampled random Fourier transform (or SRFT) [49], [50], subsampled Hadamard transforms, sequence of Givens rotations [50], [51] can be used which are established techniques and used extensively in the literature [52]. For example, the SRFT reduces the computation cost to $\mathcal{O}(IJ \log(R))$. These computations can be further reduced by exploiting compact random matrices [25], [27]. Randomized algorithms reduce the asymptotic complexity of deterministic algorithms for computation of the SVD from $\mathcal{O}(IJR)$ to $\mathcal{O}(IJ \log(R) + (I + J)R^2)$ [46].

The basic form of randomized SVD (BRSVD) algorithm equipped with the oversampling and the power iteration strategies is outlined in Algorithm 1 [46].

For stability issues, one should avoid computing expression $(\mathbf{X}\mathbf{X}^T)^q \mathbf{X}$, explicitly and it should be computed sequentially using economic QR decomposition or LU decomposition [46].

It has been proved that for oversampling parameter $P \geq 2$, power iteration q and target rank R , the accuracy of approximation (5) is

$$\mathbb{E} \left(\left\| \mathbf{X} - \mathbf{Q}\mathbf{Q}^T \mathbf{X} \right\|_2 \right) \leq \left(1 + \sqrt{\frac{R}{P-1}} + \frac{e\sqrt{R+P}}{P} \sqrt{\min(I, J) - k} \right)^{1/(2q+1)} \sigma_{k+1}$$

where \mathbb{E} denotes the mathematical expectation with the Gaussian random matrices and σ_{R+1} is the $(R + 1)$ -th largest singular values of the matrix \mathbf{X} [53]. So, a nearly optimal

Algorithm 1 Basic Randomized SVD Algorithm With Oversampling and Power Iteration

Input : A data matrix $\mathbf{X} \in \mathbb{R}^{I \times J}$, target rank R , oversampling parameter p and power iteration parameter q

Output: SVD factor matrices

$$\mathbf{U} \in \mathbb{R}^{I \times R}, \mathbf{S} \in \mathbb{R}^{R \times R}, \mathbf{V} \in \mathbb{R}^{R \times J}$$

- 1 Generate a random matrix $\mathbf{\Omega} \in \mathbb{R}^{J \times (p+R)}$ with prescribed probability distribution
- 2 Form $\mathbf{Y} = (\mathbf{X}\mathbf{X}^T)^q \mathbf{X}\mathbf{\Omega}$
- 3 Compute QR decomposition $\mathbf{Y} = \mathbf{Q}\mathbf{R}$
- 4 Compute $\mathbf{B} = \mathbf{Q}^T \mathbf{X}$
- 5 Compute the full SVD of the matrix $\mathbf{B} = \bar{\mathbf{U}}\bar{\mathbf{S}}\bar{\mathbf{V}}^T$
- 6 $\tilde{\mathbf{U}} = \mathbf{Q}\bar{\mathbf{U}}$
- 7 $\mathbf{U} = \tilde{\mathbf{U}}(:, 1 : R)$, $\mathbf{S} = \bar{\mathbf{S}}(1 : R, 1 : R)$, $\mathbf{V} = \bar{\mathbf{V}}(:, 1 : R)$

low-rank approximation can be achieved using the randomization framework.

If two dimensions of an unfolding matrix are large, then it is possible to perform reduction on both sides. This can be performed by two random matrices. The structure of two-sided randomized algorithm is outlined in Algorithm 2 [46]. Both Algorithms 1 and 2 are randomized multi-pass algorithms because they compute multiplication with the original data matrix \mathbf{X} in Line 4. These algorithms can be modified to become single-pass, e.g., by considering in Algorithm 1

$$\mathbf{B} \cong (\mathbf{\Omega}_2 \mathbf{Q})^\dagger \mathbf{W}, \quad \mathbf{W} = \mathbf{\Omega}_2 \mathbf{X}, \quad \mathbf{\Omega}_2 \in \mathbb{R}^{R \times I}, \quad (8)$$

and in Algorithm 2

$$\mathbf{B} \cong \left(\mathbf{\Omega}_2 \mathbf{Q}^{(1)} \right)^\dagger \mathbf{W} \left(\mathbf{Q}^{(2)T} \mathbf{\Omega}_1 \right)^\dagger, \quad \mathbf{W} = \mathbf{\Omega}_2 \mathbf{X} \mathbf{\Omega}_1. \quad (9)$$

The benefit of these approaches is that they avoid computation of terms $\mathbf{Q}^T \mathbf{X}$ and $\mathbf{Q}_1^T \mathbf{X} \mathbf{Q}_2$ which may be computationally expensive, especially when the data matrix is stored out-of-cores where the cost of communication may exceed our main computations. Instead, in formulations (8) and (9), the original data matrix \mathbf{X} is sketched by the random projection technique and the corresponding matrix \mathbf{B} is obtained by solving some well-conditioned overdetermined linear least-squares problems [54]. On the other hand, random matrix multiplication can be performed relatively fast by employing structured random matrices. We should note that this strategy passes the original data matrix \mathbf{X} only once because all sketching procedures can be done in the first pass over the raw data. Other types of single-pass techniques can be found in [46], [54]–[58].

A main drawback of the random projection algorithms described so far is that they need an estimation of the matrix rank which may be a difficult assumption. To resolve this difficulty, randomized *rank-revealing algorithms*, or equivalently *randomized fixed-precision algorithms* have been developed in [59], [60], where for a given approximation error

Algorithm 2 Two-Sided Randomized SVD

Input : A data matrix $\mathbf{X} \in \mathbb{R}^{I \times J}$, target rank R
Output: SVD factor matrices
 $\mathbf{U} \in \mathbb{R}^{I \times R}$, $\mathbf{S} \in \mathbb{R}^{R \times R}$, $\mathbf{V} \in \mathbb{R}^{R \times J}$

- 1 Draw prescribed random matrices
 $\mathbf{\Omega}_1 \in \mathbb{R}^{J \times R}$, $\mathbf{\Omega}_2 \in \mathbb{R}^{I \times R}$
- 2 Compute $\mathbf{Y}_1 = \mathbf{X}\mathbf{\Omega}_1$ and $\mathbf{Y}_2 = \mathbf{X}^T \mathbf{\Omega}_2$
- 3 Compute economic QR decompositions
 $\mathbf{Y}_1 = \mathbf{Q}^{(1)}\mathbf{R}_1$, $\mathbf{Y}_2 = \mathbf{Q}^{(2)}\mathbf{R}_2$
- 4 Compute $\mathbf{B} = \mathbf{Q}^{(1)T} \mathbf{X} \mathbf{Q}^{(2)}$
- 5 Compute the SVD of matrix $\mathbf{B} = \bar{\mathbf{U}} \bar{\mathbf{S}} \bar{\mathbf{V}}^T$
- 6 Set $\tilde{\mathbf{U}} = \mathbf{Q}^{(1)} \bar{\mathbf{U}}$ and $\tilde{\mathbf{V}} = \mathbf{Q}^{(2)} \bar{\mathbf{V}}$

bound, the rank and corresponding low-rank approximation can be computed automatically.

B. SAMPLING TECHNIQUES

The other type of randomized algorithms for finding low-rank approximations of matrices is based on selecting a part of its columns or rows randomly. The previous approach does not choose individual columns of the matrix \mathbf{X} , but instead it finds a set of linear combinations of columns (rows) capturing the column (row) space of the matrix \mathbf{X} with high probability. Sampling technique is an alternative approach to the random projection where instead of multiplying the original data matrix with random matrices, some columns or rows of it are selected and the original data matrix is compressed in this manner.⁶ The procedure of column or row selection can be performed based on different kinds of probability distributions such as uniform distribution, leverage scores [62], length squared [63], [64] and also with/without replacement.⁷ The accuracies of results highly depend on the probability distribution used in the sampling procedure and either *additive* or *relative* approximation error can be achieved.

Let $\mathbf{X} \in \mathbb{R}^{I \times J}$ be a given data matrix, and \mathbf{C} be a matrix containing the selected columns, then we have two kinds of approximation as⁸ [68]

- **Relative approximation**⁹

$$\|\mathbf{X} - \mathbf{C}\mathbf{C}^\dagger \mathbf{X}\|_F \leq (1 + \epsilon) \|\mathbf{X} - \mathbf{X}_R\|_F, \quad (10)$$

- **Additive approximation**

$$\|\mathbf{X} - \mathbf{C}\mathbf{C}^\dagger \mathbf{X}\|_F^2 \leq \|\mathbf{X} - \mathbf{X}_R\|_F^2 + \epsilon \|\mathbf{X}_R\|_F^2.$$

where \mathbf{X}_R is the best rank R -approximation of the matrix \mathbf{X} . Clearly, randomized algorithms with relative-error

⁶Component selection is another sampling approach where some components (not necessarily fibers) of the original data matrix are selected using some probability distributions [61]. This is also known as sparsification procedure.

⁷It has been reported that for a fixed distribution, and when the columns are uniformly correlated, the uniform sampling without replacement works quite well in practice [65]–[67].

⁸The same concepts are used for spectral norm $\|\cdot\|_2$.

⁹It is also called multiplicative error.

guarantees are of much more interest than the additive-error ones. The number of selected columns should be large enough to capture the range of a matrix and satisfy the relation (10).

The best existing random sampling algorithms use the leverage score probability distribution [69] while their computation is expensive because of the computation of the SVD. In [69], a fast and computationally efficient approach has been proposed for computing the leverage scores. In some applications, we need interpretable low-rank matrix approximations, and columns should be taken from the original data matrix. Here, in the first step, the leverage scores of an underlying data matrix can be computed by the randomized SVD and then the columns of the original data matrix are selected based on the leverage score probability distribution. Note that exploiting random uniform sampling of columns (fibers) provides relatively fast computation. However, this method works well only if a data matrix has low coherence [68] otherwise they do not provide a reliable compression. An alternative approach is to spread out the information of the data matrix or uniformize it by a prior random projection. This preprocessing step allows applying the uniform sampling [28].

The sampling low-rank matrix approximation is outlined in Algorithm 3. It picks R columns of the matrix \mathbf{X} and constructs the matrix \mathbf{Y} [28], where in Line 3 of Algorithm 3, the selected columns are scaled [28]. This scaling procedure provides an unbiased estimator for the matrix-matrix multiplication [68].

Algorithm 3 Sampling Algorithm for Low-Rank Matrix Approximation

Input : A data matrix $\mathbf{X} \in \mathbb{R}^{I \times J}$, $R \in \mathbb{Z}^+$ such that
 $1 \leq R \leq J$, $\{p_j\}_{j=1}^J$, $\sum_{j=1}^J p_j = 1$ and $p_j \geq 0$

Output: Low-rank matrix approximation
 $\mathbf{X} \cong \mathbf{Q}\mathbf{B}$, $\mathbf{Q} \in \mathbb{R}^{I \times R}$, $\mathbf{B} \in \mathbb{R}^{R \times J}$

- 1 **for** $r = 1, 2, \dots, R$ **do**
- 2 Pick $j_r \in \{1, 2, \dots, J\}$ with $\Pr(j_r = \alpha) = p_\alpha$
 $\alpha = 1, 2, \dots, J$
- 3 Set $\tilde{\mathbf{Y}}(:, j_r) = \frac{\mathbf{X}(:, j_r)}{\sqrt{R p_{j_r}}}$
- 4 **end**
- 5 Compute $\tilde{\mathbf{Y}} = \mathbf{Q}\mathbf{R}$, $\mathbf{B} = \mathbf{Q}^T \mathbf{X}$

C. COUNT-SKETCH

The random projection approach needs expensive multiplication of an original data matrix with random matrices. Count-sketch uses an alternative trick to generate a linear combinations of the columns without multiplication with random matrices. Count-sketch was originally introduced in the context of data stream [70] and has been used to speed-up the matrix-matrix computations [71]–[73]. Similar to the random projection and the sampling approaches, the

count-sketch technique aims at capturing the range of a given matrix. It consists of the following three steps:

- Hashing procedure,
- Grouping together the columns with the same hash numbers,
- Signing the columns and summing each group as a representative column.

Assume that we want to compute a rank- R approximation of a matrix. In the hashing procedure, all columns of the data matrix are labeled with numbers $1, 2, \dots, R$ uniformly, i.e., with probability $1/R$.

In the second step, the columns with the same label are grouped together and after signing the columns,¹⁰ the columns in each group are summed up as representative columns.

Clearly, the cost of the first step is negligible, and the main operation in the second step is summing the columns (with integer coefficients ± 1). Compared to multiplying the original data matrix by a random matrix, this technique can significantly reduce the running time.

D. RANDOMIZED ALGORITHMS FOR SOLVING LEAST-SQUARES PROBLEMS

In this section, we briefly describe the randomized algorithms for solving large-scale least-squares problems. In the procedure of computation of the Canonical Polyadic Decomposition (CPD) [74], [75] and the Tucker decomposition, we often need to solve least-squares problems with tall and skinny coefficient matrices, i.e., the number of rows is much more than the number of columns. For example, the following two matrices

$$\bigodot_{i=1}^N \mathbf{Q}^{(n)} \in \mathbb{R}^{\left(\prod_{n=1}^N I_n\right) \times R_n}, \quad (11)$$

$$\bigotimes_{n=1}^N \mathbf{Q}^{(n)} \in \mathbb{R}^{\left(\prod_{n=1}^N I_n\right) \times \left(\prod_{n=1}^N R_n\right)}, \quad (12)$$

have tall and skinny structures where $\mathbf{Q}^{(n)} \in \mathbb{R}^{I_n \times R_n}$, $I_n \gg R_n$, and arise in the procedure of computation of the CPD and the Tucker decomposition [76].

Consider the following least-squares problem

$$\mathbf{x} = \arg \min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2, \quad (13)$$

where the matrix \mathbf{A} is either (11) or (12), the right-hand side is $\mathbf{b} \in \mathbb{R}^{\prod_{n=1}^N I_n}$ and the vector $\mathbf{x} \in \mathbb{R}^{R_n}$ (or $\mathbf{x} \in \mathbb{R}^{\prod_{n=1}^N R_n}$) is the unknown vector needs to be determined. The randomized or sketching techniques replace the least-squares problem (13) by

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{TAx} - \mathbf{Tb}\|_2, \quad (14)$$

¹⁰By signing we mean multiplying the columns by ± 1 uniformly. This is also called binary Rademacher variable.

where

$$\mathbf{T} : \mathbb{R}^{\prod_{n=1}^N I_n} \rightarrow \mathbb{R}^L, \quad L \leq \prod_{n=1}^N I_n,$$

is a map which reduces the dimensionality of the matrix \mathbf{A} and the right-hand side \mathbf{b} , see Figure 2 for a graphical illustration. The transformation \mathbf{T} can be the random projection, the sampling or the count-sketch technique discussed in Section III. It is expected to find a solution $\tilde{\mathbf{x}}$ of the least-squares problem (14) which satisfies

$$\|\mathbf{Ax} - \mathbf{b}\|_2 \leq (1 + \varepsilon) \|\mathbf{Ax}^* - \mathbf{b}\|_2,$$

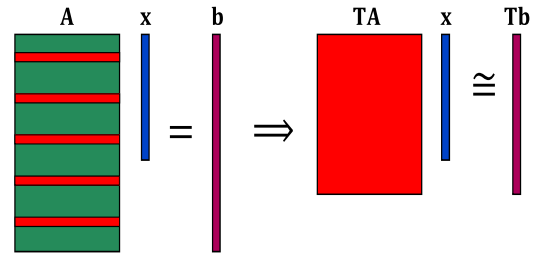


FIGURE 2. Illustration of randomized row sampling technique for solving an overdetermined least-squares problem.

with a high probability where \mathbf{x}^* is the optimal solution of least-squares problem (13) and $\varepsilon > 0$ is a given tolerance [55].

In next section, we introduce the Tucker decomposition and the HOSVD along with algorithms for their computations.

IV. TUCKER DECOMPOSITION AND HIGHER ORDER SINGULAR VALUE DECOMPOSITION (HOSVD)

Let $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, then the Tucker decomposition of the tensor $\underline{\mathbf{X}}$ admits the following model [1]–[3]

$$\underline{\mathbf{X}} \cong \underline{\mathbf{S}} \times_1 \mathbf{Q}^{(1)} \times_2 \mathbf{Q}^{(2)} \dots \times_N \mathbf{Q}^{(N)}, \quad (15)$$

where $\underline{\mathbf{S}} \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_N}$ is a core tensor and $\mathbf{Q}^{(n)} \in \mathbb{R}^{I_n \times R_n}$, $R_n \leq I_n$, $n = 1, 2, \dots, N$ are factor matrices. A shorthand notation for the Tucker decomposition is

$$\underline{\mathbf{X}} \cong \left[\left[\underline{\mathbf{S}}; \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \dots, \mathbf{Q}^{(N)} \right] \right].$$

HOSVD or equivalently Multilinear SVD (MLSVD) is a constrained Tucker decomposition [21] that ensures the orthogonality of factor matrices and all-orthogonality of the core tensor.¹¹ For a graphical illustration of the HOSVD for a 3rd and a 4th-order tensors, see Figure 3.

Unlike the SVD, the core tensor $\underline{\mathbf{S}}$, in general, is not diagonal or even not nonnegative, but it has *pseudo-diagonal* property, which means that the Frobenius norm of slices in each mode is non-increasing as the index is increased [21]. Also, another interpretation of this concept is that the

¹¹A tensor is called all-orthogonal if all its slices in each mode are mutually orthogonal [21].

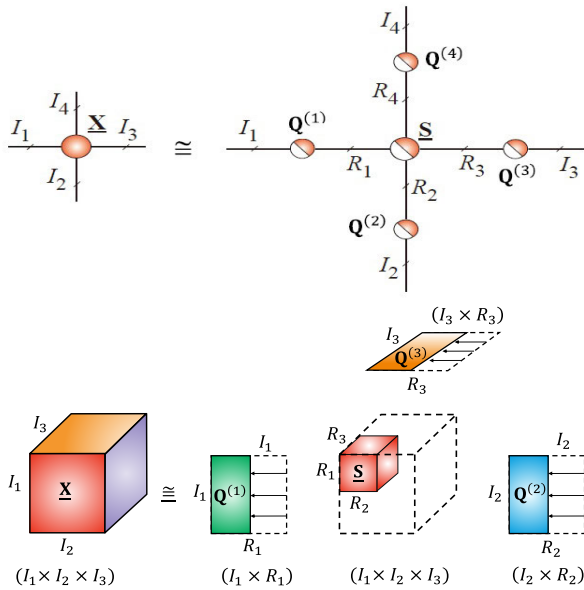


FIGURE 3. Graphical illustrations of the truncated HOSVD, upper is for a fourth order tensors and bottom is for a third order tensors.

intensity (absolute value) of the components of the core tensor \underline{S} is mainly concentrated on one of the corners (position $(1, 1, \dots, 1)$) and it decreases as the components move away from this corner.

Due to the pseudo-diagonality property of the HOSVD which plays a role similar to the singular values in the SVD, a truncated version of the HOSVD is used in practice. To this end, first the orthogonal matrices $\mathbf{Q}^{(n)} \in \mathbb{R}^{I_n \times R_n}$, $n = 1, 2, \dots, N$ in (15) are computed through the truncated SVD (tSVD) or the randomized SVD (rSVD) algorithms for low-rank approximation of the unfolding matrices as

$$\mathbf{X}_{(n)} \cong \mathbf{Q}^{(n)} \mathbf{S}_n \mathbf{V}^{(n)T} \in \mathbb{R}^{I_n \times I_1 \cdots I_{n-1} I_{n+1} \cdots I_N}, \quad (16)$$

where $\mathbf{V}^{(n)} \in \mathbb{R}^{I_1 \cdots I_{n-1} I_{n+1} \cdots I_N \times R_n}$, $\mathbf{S}_n \in \mathbb{R}^{R_n \times R_n}$ and R_n is the numerical rank of the unfolding matrix $\mathbf{X}_{(n)}$. Then, due to the orthogonality of the factor matrices $\mathbf{Q}^{(n)} \in \mathbb{R}^{R_n \times I_n}$, $n = 1, 2, \dots, N$, the core tensor can be computed as

$$\underline{S} = \underline{X} \times_1 \mathbf{Q}^{(1)T} \times_2 \mathbf{Q}^{(2)T} \cdots \times_N \mathbf{Q}^{(N)T} \in \mathbb{R}^{R_1 \times R_2 \times \cdots \times R_N}. \quad (17)$$

Since, we only need the left singular vectors in (16), the eigenvalue decomposition (EVD) of the Gramian matrix $\mathbf{X}_{(n)} \mathbf{X}_{(n)}^T = \mathbf{Q}^{(n)} \Sigma^2 \mathbf{Q}^{(n)T}$ can also be used in our computations. The Crank-Nicholson-like algorithm [77], [78], rank revealing QR (LU) decompositions [79], [80] or Alternating Least Squares (ALS) type techniques [81], [82] are other possible technique for computation of the factor matrices.

The N -tuple (R_1, R_2, \dots, R_N) is called multilinear rank or Tucker rank of tensor \underline{X} .

The truncated HOSVD does not provide the best multilinear rank approximation in the least-squares sense while a

quasi-best approximation can be achieved [21] as follows

$$\|\underline{X} - \underline{S} \times_1 \mathbf{Q}^{(1)} \times_2 \cdots \times_N \mathbf{Q}^{(N)}\|_F \leq \sqrt{N} \|\underline{X} - \underline{X}_{\text{Best}}\|_F,$$

where $\underline{X}_{\text{best}}$ is the best multilinear rank approximation of the tensor \underline{X} .

Substituting (17) in (15), we have

$$\underline{X} \cong \underline{X} \times_1 \mathbf{Q}^{(1)} \mathbf{Q}^{(1)T} \times_2 \mathbf{Q}^{(2)} \mathbf{Q}^{(2)T} \cdots \times_N \mathbf{Q}^{(N)} \mathbf{Q}^{(N)T}. \quad (18)$$

Note that if the factor matrices $\mathbf{Q}^{(n)}$, $n = 1, 2, \dots, N$ are not orthogonal and just have full column rank, then (17) and (18) are replaced by¹²

$$\underline{S} \cong \underline{X} \times_1 \mathbf{Q}^{(1)\dagger} \times_2 \mathbf{Q}^{(2)\dagger} \cdots \times_N \mathbf{Q}^{(N)\dagger}, \quad (19)$$

and

$$\underline{X} \cong \underline{X} \times_1 \mathbf{Q}^{(1)} \mathbf{Q}^{(1)\dagger} \times_2 \mathbf{Q}^{(2)} \mathbf{Q}^{(2)\dagger} \cdots \times_N \mathbf{Q}^{(N)} \mathbf{Q}^{(N)\dagger}, \quad (20)$$

where \dagger is the Moore-Penrose pseudoinverse operator. It is worth mentioning that if the dimension of a particular mode, say mode n , is relatively small, then we can ignore reduction in the mentioned mode or, equivalently, ignore its corresponding orthogonal matrix $\mathbf{Q}^{(n)}$ and remove it from (18) and (20). In view of (18) and (20), it is seen that to find an approximate HOSVD or an approximate Tucker decomposition, good approximations for the range of unfolding matrices are required. More precisely, the main problem is how to find a good approximation for the range of unfolding matrices $\mathbf{X}_{(n)}$, i.e., $\mathbf{Q}^{(n)}$. This problem can be formally formulated as follows:

Given a data tensor $\underline{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and tolerance ϵ , the problem is finding orthogonal matrices $\mathbf{Q}^{(n)} \in \mathbb{R}^{I_n \times R_n}$, satisfying

$$\|\underline{X} - \underline{X} \times_1 \mathbf{Q}^{(1)} \mathbf{Q}^{(1)T} \cdots \times_N \mathbf{Q}^{(N)} \mathbf{Q}^{(N)T}\|_F \leq \sqrt{N} \epsilon, \quad (21)$$

and in the case when $\mathbf{Q}^{(n)}$ are not orthogonal

$$\|\underline{X} - \underline{X} \times_1 \mathbf{Q}^{(1)} \mathbf{Q}^{(1)\dagger} \cdots \times_N \mathbf{Q}^{(N)} \mathbf{Q}^{(N)\dagger}\|_F \leq \sqrt{N} \epsilon. \quad (22)$$

We will return to these problems later in Section V.

A. SEQUENTIALLY TRUNCATED HOSVD (STHOSVD) ALGORITHM

In formulation (16), low-rank approximations of the unfolding matrices $\mathbf{X}_{(n)}$ are required where all unfolding matrices have the same number of elements as the original data tensor \underline{X} . A more efficient algorithm with a less computational complexity for computing the HOSVD is *Sequentially Truncated HOSVD* (STHOSVD) algorithm [83]. Instead of dealing with all unfolding matrices (all with the same number of elements as the original data tensor \underline{X}) and finding their low-rank matrix approximations, this algorithm reduces the size of underlying unfolding matrices sequentially. More precisely, at each iteration of the algorithm, a new unfolding matrix which approximates a specific factor matrix has a

¹²If \mathbf{Q} is full-rank then $\mathbf{Q}^\dagger \mathbf{Q} = \mathbf{I}$, where \mathbf{I} is the identity matrix of conforming dimension.

smaller size than the previous ones. This results in significant speed-up with comparable accuracy or sometimes even better accuracy.¹³ This procedure is described in Algorithm 4 [83]. It is worth mentioning that in Lines 4 of Algorithm 4, an n -mode (matrix) product of the tensor $\underline{\mathbf{S}}$ by the matrix $\mathbf{Q}^{(n)T}$ is required. This can be performed in an equivalent way. To be more precise, let $[\mathbf{Q}^{(n)}, \Lambda^{(n)}, \mathbf{V}^{(n)}]$ be the factor matrices of the truncated SVD of the unfolding matrix $\mathbf{S}_{(n)}$. If $\underline{\mathbf{Y}} \leftarrow \underline{\mathbf{S}} \times_n \mathbf{Q}^{(n)T}$ then it is easy to show that $\mathbf{Y}_{(n)} = \Lambda^{(n)} \mathbf{V}^{(n)T}$. This means that in Algorithm 4, in Line 3, if we compute all factor matrices $[\mathbf{Q}^{(n)}, \Lambda^{(n)}, \mathbf{V}^{(n)}]$, then Line 4 can be replaced by $\mathbf{S}_{(n)} = \Lambda^{(n)} \mathbf{V}^{(n)T}$.

Another advantage of the STHOSVD compared to the THOSVD is that in the STHOSVD algorithm, the core tensor is automatically computed by the algorithm in the last step while in the THOSVD algorithm, all factor matrices are first computed and then the core tensor is computed through formulation (17) or (19). This may lead to the intermediate data explosion phenomenon [15]. Note that Algorithm 4, starts from the first factor matrix and other factor matrices are computed in ascending order, i.e. $p = [1, 2, \dots, N]$, but other orderings are also possible. It turns out that the ordering used within the STHOSVD algorithm affects the approximation accuracy and speed-up [83].

Algorithm 4 Sequentially Truncated HOSVD (STHOSVD) Algorithm

Input : A data tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ and a tolerance ϵ ;
Output: Approximative HOSVD of the tensor $\underline{\mathbf{X}}$
 $\underline{\mathbf{X}} \cong [\underline{\mathbf{S}}; \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \dots, \mathbf{Q}^{(N)}]$ and multilinear rank (R_1, R_2, \dots, R_N)

- 1 Set $\underline{\mathbf{S}} = \underline{\mathbf{X}}$
- 2 **for** $n = 1, 2, \dots, N$ **do**
- 3 $[\mathbf{Q}^{(n)}, \sim, \sim] = \text{truncated-svd}(\mathbf{S}_{(n)}, \frac{\epsilon}{N})$
- 4 $\underline{\mathbf{S}} = \underline{\mathbf{S}} \times_n \mathbf{Q}^{(n)T}$
- 5 **end**

B. HIGHER ORDER ORTHOGONAL ITERATION (HOOI) ALGORITHM

Unlike *Eckart-Young* property for matrices [84], neither the truncated HOSVD (THOSVD) nor the sequentially truncated HOSVD (STHOSVD) provides the best multilinear rank approximation for higher order tensors. Several algorithms have been developed for the computation of the best multilinear rank approximation among which we can mention Newton-Grassmann algorithm [85], Riemannian trust-region algorithm [86], Higher-Order Orthogonal Iteration (HOOI) algorithm¹⁴ [87] etc. HOOI algorithm is the simplest kind of

¹³Although there is no theoretical evidence establishing better accuracy of SHOSVD compared to the HOSVD, numerical simulations highly support this. However, there are some examples for which the HOSVD provides better accuracy compared to the SHOSVD [83].

¹⁴This algorithm is also known as Tucker-ALS algorithm.

these algorithms which is based on the idea of ALS technique where at each iteration of the algorithm, all factor matrices are fixed except one and it is updated by solving a least-squares problem. To be more precise, consider the Tucker decomposition (15) and apply the n -unfolding on both sides of (15), as

$$\mathbf{X}_{(n)} = \mathbf{Q}^{(n)} \mathbf{S}_{(n)} \left(\mathbf{Q}^{(N)} \otimes \dots \otimes \mathbf{Q}^{(n+1)} \otimes \mathbf{Q}^{(n-1)} \otimes \dots \otimes \mathbf{Q}^{(1)} \right)^T,$$

for $n = 1, 2, \dots, N$. Assume that the core tensor $\underline{\mathbf{S}}$ and all but one factor matrices $\{\mathbf{Q}^{(p)}\}$, $p \neq n$, are known. The unknown factor matrix $\mathbf{Q}^{(n)}$ can be computed by solving the following least-squares problem¹⁵

$$\mathbf{Q}^{(n)} = \arg \min_{\mathbf{Q} \in \mathbb{R}^{I_n \times R_n}} \left\| \mathbf{A}^{(n)} \mathbf{Q}^T - \mathbf{X}_{(n)}^T \right\|_F^2, \quad (23)$$

where

$$\mathbf{A}^{(n)} = \left(\bigotimes_{p=N, p \neq n}^1 \mathbf{Q}^{(p)} \right) \mathbf{S}_{(n)}^T.$$

It can be shown that the solution to the least-squares problem (23) is equivalent to finding R_n leading left singular vectors of $\mathbf{Z}_{(n)}$ [21] where

$$\mathbf{Z} = \underline{\mathbf{X}} \times_{p \neq n} \left\{ \mathbf{Q}^{(p)} \right\}^T. \quad (24)$$

This procedure is repeated for all factor matrices $\mathbf{Q}^{(n)} \in \mathbb{R}^{I_n \times R_n}$, $n = 1, 2, \dots, N$, sequentially and in an usual Gauss-Seidel manner. After computing all factor matrices, the core tensor $\underline{\mathbf{S}}$ is updated in the following form

$$\underline{\mathbf{S}} = \underline{\mathbf{X}} \times_1 \mathbf{Q}^{(1)T} \dots \times_N \mathbf{Q}^{(N)T}. \quad (25)$$

Also the computation of the core tensor $\underline{\mathbf{S}}$ in (25), can be equivalently computed by solving the following least-squares problem

$$\underline{\mathbf{S}} = \arg \min_{\underline{\mathbf{S}} \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_N}} \left\| \mathbf{B} \mathbf{s}(\cdot) - \mathbf{x}(\cdot) \right\|_2^2, \quad (26)$$

where

$$\mathbf{B} = \bigotimes_{n=N}^1 \mathbf{Q}^{(n)},$$

and $\mathbf{x}(\cdot)$ and $\mathbf{s}(\cdot)$ are the vectorization of the tensors $\underline{\mathbf{X}}$ and $\underline{\mathbf{S}}$, respectively. This procedure is outlined in Algorithm 5 [87].

V. RANDOMIZED TUCKER DECOMPOSITION AND RANDOMIZED HOSVD

It can be seen that the most computationally expensive part of the HOSVD or the STHOSVD is low-rank matrix approximation of the unfolding matrices. In randomized Tucker algorithms or randomized HOSVD, the truncated (or economic) SVD is replaced by the randomized low-rank matrix algorithms. In this section, we discuss a variety of randomized algorithms for computation of the Tucker decomposition and the HOSVD.

¹⁵In [90], the l_1 -norm form of least-squares problem (23) is solved.

Algorithm 5 Higher-Order Orthogonal Iteration (HOOI) Algorithm

Input : A data tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, and a multilinear rank (R_1, R_2, \dots, R_N)

Output: Approximative Tucker representation of the tensor $\underline{\mathbf{X}}$ as $\underline{\mathbf{X}} \cong [\underline{\mathbf{S}}; \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \dots, \mathbf{Q}^{(N)}]$

- 1 Initialize factor matrices $\mathbf{Q}^{(2)}, \dots, \mathbf{Q}^{(N)}$ with the Truncated HOSVD or random matrices
- 2 **while** A stopping criterion is not satisfied **do**
- 3 **for** $n = 1, 2, \dots, N$ **do**
- 4 $\underline{\mathbf{Z}} = \underline{\mathbf{X}}_{\times p \neq n} \{\mathbf{Q}^{(p)T}\};$
- 5 $\mathbf{Q}^{(n)}$ is constructed from R_n leading left singular vectors of $\underline{\mathbf{Z}}_{(n)}$
- 6 **end**
- 7 $\underline{\mathbf{S}} = \underline{\mathbf{Z}}_{\times 1} \mathbf{Q}^{(1)T} \times_2 \mathbf{Q}^{(2)T} \dots \times_N \mathbf{Q}^{(N)T}$
- 8 **end**

A. RANDOMIZED RANDOM PROJECTION TUCKER DECOMPOSITION

Exploiting the idea of random projection, the randomized variants of the HOSVD and the HOOI algorithms can be straightforwardly developed. These algorithms are outlined in Algorithms 6 and 7 [24], [26].

In Algorithms 6 and 7, the unfolding matrix is not necessary to be computed explicitly. In [88], instead of $\underline{\mathbf{W}} = \underline{\mathbf{Z}}_{(n)} \Omega^{(n)}$, it is suggested to perform $\underline{\mathbf{Z}}_{\times p \neq n} \{\Omega_p^{(n)}\}$ where $\Omega_p^{(n)} \in \mathbb{R}^{I_n \times R_n}$, $p = 1, 2, \dots, N$ are random matrices and then compute the unfolding matrix $\mathbf{W}_{(n)}$. In view of (2), the later is equivalent to the multiplication of the unfolding matrix $\mathbf{W}_{(n)}$ to the Kronecker product of the random matrices.¹⁶ This is a memory efficient approach it needs

$\mathcal{O}(\sum_{p \neq n} I_p R_p)$, while the former needs $\mathcal{O}(\prod_{p \neq n} I_p R_n)$. Also a memory efficient variant of the power iteration expression $(\underline{\mathbf{Z}}_{(n)} \mathbf{Z}_{(n)}^T)^q \underline{\mathbf{Z}}_{(n)} \Omega^{(n)}$ in the tensor format is discussed in [89].

Both Algorithms 6 and 7 are nonadaptive because they require an estimation of the multilinear rank (R_1, R_2, \dots, R_N) which may be a difficult assumption in real applications. To overcome this difficulty, it is possible to apply fixed precision randomized algorithms [59], [60] within Algorithms 6 and 7 instead of standard (basic) randomized Algorithm 1. This idea has been utilized in [23].

Also, a different strategy has been used in [22] for adaptively computing the multilinear rank of tensors. This idea basically relies on solving the Problems (21) and (22) numerically and finding orthonormal matrices $\mathbf{Q}^{(n)}$. Using the following identity

$$\|\underline{\mathbf{X}}\|_F^2 = \|\underline{\mathbf{X}}_{\times n} (\mathbf{Q}^{(n)} \mathbf{Q}^{(n)T})\|_F^2 + \|\underline{\mathbf{X}}_{\times n} (\mathbf{I}_{I_n} - \mathbf{Q}^{(n)} \mathbf{Q}^{(n)T})\|_F^2,$$

¹⁶In [22], [25], the Khatri-Rao product of random matrices is used.

it can be shown that [83]

$$\begin{aligned} \|\underline{\mathbf{X}} - \underline{\mathbf{X}}_{\times 1} (\mathbf{Q}^{(1)} \mathbf{Q}^{(1)T}) \times_2 \dots \times_N (\mathbf{Q}^{(N)} \mathbf{Q}^{(N)T})\|_F^2 \\ \leq \sum_{n=1}^N \|\underline{\mathbf{X}} - \underline{\mathbf{X}}_{\times n} (\mathbf{Q}^{(n)} \mathbf{Q}^{(n)T})\|_F^2, \end{aligned}$$

for $n = 1, 2, \dots, N$.

This implies that finding orthonormal factor matrices is equivalent to finding the orthonormal matrices $\mathbf{Q}^{(n)} \in \mathbb{R}^{I_n \times R_n}$ that satisfy

$$\begin{aligned} \|\underline{\mathbf{X}} - \underline{\mathbf{X}}_{\times n} (\mathbf{Q}^{(n)} \mathbf{Q}^{(n)T})\|_F^2 \\ = \|\underline{\mathbf{X}}_{\times n} (\mathbf{I}_{I_n} - \mathbf{Q}^{(n)} \mathbf{Q}^{(n)T})\|_F^2 \leq \frac{\varepsilon}{N}. \end{aligned} \quad (27)$$

The unfolding form of (27) is

$$\|(\mathbf{I}_{I_n} - \mathbf{Q}^{(n)} \mathbf{Q}^{(n)T}) \mathbf{X}_{(n)}\|_F^2 \leq \frac{\varepsilon}{N}, \quad (28)$$

and as a result, to find the factor matrix $\mathbf{Q}^{(n)}$, it is sufficient to compute an orthonormal matrix which captures the range of the unfolding matrix $\mathbf{X}_{(n)}$. Here, we can benefit from the efficient and optimized randomized algorithms for matrices. In [22], the sub-problem (27) which is in tensor form is solved.

Algorithm 6 Random Projection HOSVD (RP-HOSVD) Algorithm

Input : A data tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, and a multilinear rank (R_1, R_2, \dots, R_N)

Output: Approximative HOSVD of the tensor $\underline{\mathbf{X}}$ as $\underline{\mathbf{X}} \cong [\underline{\mathbf{S}}; \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \dots, \mathbf{Q}^{(N)}]$

- 1 $\underline{\mathbf{Z}} = \underline{\mathbf{X}}$
- 2 **for** $n = 1, 2, \dots, N$ **do**
- 3 Compute $\mathbf{W}^{(n)} = \underline{\mathbf{Z}}_{(n)} \Omega^{(n)}$, where $\Omega^{(n)}$ is an $(\prod_{k \neq n} I_k) \times R_n$ random Gaussian matrix
- 4 Compute $\mathbf{Q}^{(n)}$ as an orthonormal basis of $\mathbf{W}^{(n)}$ by using, e.g., the economic QR decomposition;
- 5 **end**
- 6 Compute the core tensor as $\underline{\mathbf{S}} \equiv \underline{\mathbf{X}}_{\times 1} \mathbf{Q}^{(1)T} \times_2 \mathbf{Q}^{(2)T} \dots \times_N \mathbf{Q}^{(N)T}$

Similar to the HOSVD and the HOOI algorithms, a randomized variant of the STHOSVD algorithm can also be developed. Algorithm 8 is a random projection STHOSVD algorithm in which Algorithm 1 is exploited for computing low-rank approximations of the n -unfolding matrices [23]. The adaptive randomized STHOSVD is also proposed in [23] wherein the randomized rank-revealing algorithms [59], [60] are exploited instead of the Algorithm 1. All discussed algorithms so far do not preserve the structure of the underlying data tensor such as sparsity or nonnegativity. A structural preserving variant of Algorithm 8 is developed in [23] where the core tensor takes elements from the original data matrix.

Algorithm 7 Random Projection HOOI (RP-HOOI) Algorithm

Input : A data tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, and a multilinear rank (R_1, R_2, \dots, R_N)

Output: Approximative HOSVD of the tensor $\underline{\mathbf{X}}$ as $\underline{\mathbf{X}} \cong [\underline{\mathbf{S}}; \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \dots, \mathbf{Q}^{(N)}]$

- 1 Initialize factor matrices $\mathbf{Q}^{(n)} \in \mathbb{R}^{I_n \times R_n}$ as factor matrices of the HOSVD or random Gaussian matrices;
- 2 **while** A stopping criterion is not satisfied **do**
- 3 **for** $n = 1, 2, \dots, N$ **do**
- 4 $\underline{\mathbf{S}} = \underline{\mathbf{X}} \times_{p \neq n} \{\mathbf{Q}^{(p)T}\}$
- 5 Compute $\mathbf{W}^{(n)} = \mathbf{S}_{(n)} \boldsymbol{\Omega}^{(n)}$ where $\boldsymbol{\Omega}^{(n)} \in \mathbb{R}^{p \times R_n}$ is a random matrix drawn from Gaussian distribution
- 6 Compute $\mathbf{Q}^{(n)} \in \mathbb{R}^{I_n \times R_n}$ as an orthonormal basis of $\mathbf{W}^{(n)}$, e.g., by using QR decomposition
- 7 **end**
- 8 $\underline{\mathbf{S}} = \underline{\mathbf{S}} \times_N \mathbf{Q}^{(N)T}$
- 9 **end**

Algorithm 8 Randomized Sequentially Truncated HOSVD (R-STHOSVD) Algorithm

Input : A data tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, and a multilinear rank (R_1, R_2, \dots, R_N)

Output: Approximative Tucker representation of the tensor $\underline{\mathbf{X}}$ as $\underline{\mathbf{X}} \cong [\underline{\mathbf{S}}; \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \dots, \mathbf{Q}^{(N)}]$

- 1 $\mathbf{S} = \mathbf{X}_{(n)}$
- 2 **for** $n = 1, 2, \dots, N$ **do**
- 3 $[\mathbf{Q}^{(n)}, \sim, \sim] = \text{Apply Algorithm 1 to the } n\text{-unfolding matrix } \mathbf{S}_{(n)} \text{ with target rank } R_n$
- 4 $\underline{\mathbf{S}} = \underline{\mathbf{S}} \times_n \mathbf{Q}^{(n)T}$
- 5 **end**

The algorithms and techniques discussed so far need to pass the original data tensor $\underline{\mathbf{X}}$ multiple times. This restricts their applicability for the data tensors stored out-of-cores because of high communication cost. Due to this issue, we need to develop so-called randomized Pass-Efficient algorithms where the original data tensor should be passed as less as possible. For example, one possible option is using Algorithm (2) or pass-efficient algorithms proposed in [46], [55]–[58], for low-rank approximation of n -unfolding matrices.

An elegant pass-efficient algorithm is developed [91] and is described in Algorithm 9. It first captures the important actions of the tensor $\underline{\mathbf{X}}$ in each mode (Lines 1-2), i.e., the range of unfolding matrices, and also their corresponding interactions (Line 3), i.e., the compressed core tensor generated by multiplying the original data tensor with random matrices in different modes. These are performed by multiplying the unfolding matrices by random matrices and at the same time multiplying an original tensor by random matrices

along its different modes. The structure of this procedure is described in Lines 1 and 3 of Algorithm 9 where

$$\boldsymbol{\Omega}_n \in \mathbb{R}^{\left(\prod_{i \neq n} I_i\right) \times K_n}, \quad \tilde{\boldsymbol{\Omega}}_n \in \mathbb{R}^{S_n \times I_n},$$

and $R_n \leq K_n \leq S_n$, $n = 1, 2, \dots, N$. Here, (R_1, R_2, \dots, R_N) is a target multilinear rank. Note that in Algorithm 9, $\mathbf{Q}^{(n)} \in \mathbb{R}^{I_n \times K_n}$, $\underline{\mathbf{H}} \in \mathbb{R}^{S_1 \times S_2 \times \dots \times S_N}$, and in order to obtain a truncated HOSVD, the obtained core tensor $\underline{\mathbf{S}}$ and its corresponding factor matrices $\mathbf{Q}^{(N)}$ are truncated to multilinear rank (R_1, R_2, \dots, R_N) .

Algorithm 9 Randomized Pass-Efficient Randomized Algorithm for the Tucker Decomposition (R-PET)

Input : A data tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, sketching parameters $K, S, S > K$ and a predefined multilinear rank (R_1, R_2, \dots, R_N)

Output: Approximative Tucker representation of the tensor $\underline{\mathbf{X}}$ as $\underline{\mathbf{X}} \cong [\underline{\mathbf{S}}; \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \dots, \mathbf{Q}^{(N)}]$

- 1 Compute factor sketches, $\mathbf{Y}_n = \mathbf{X}_{(n)} \boldsymbol{\Omega}_n$, $n = 1, 2, \dots, N$
- 2 Recover factor matrices, $[\mathbf{Q}^{(n)}, \sim] = \mathbf{QR}(\mathbf{Y}_n)$
- 3 Compute core sketch $\underline{\mathbf{H}} = \underline{\mathbf{X}} \times_1 \tilde{\boldsymbol{\Omega}}_1 \times \dots \times_N \tilde{\boldsymbol{\Omega}}_N$
- 4 Recover core tensor $\underline{\mathbf{S}} = \underline{\mathbf{H}} \times_1 (\tilde{\boldsymbol{\Omega}}_1 \mathbf{Q}^{(1)})^\dagger \times_2 \dots \times_N (\tilde{\boldsymbol{\Omega}}_N \mathbf{Q}^{(N)})^\dagger$
- 5 Truncate the core tensor $\underline{\mathbf{S}}$ and the factor matrices $\mathbf{Q}^{(n)}$, $n = 1, 2, \dots, N$ with multilinear rank (R_1, R_2, \dots, R_N)

B. RANDOMIZED SAMPLING TUCKER DECOMPOSITION

In randomized sampling Tucker decomposition, each factor matrix $\mathbf{Q}^{(n)}$ is computed by sampling the columns of the corresponding unfolding matrix $\mathbf{X}_{(n)}$ or equivalently sampling the fibers¹⁷ of the original data tensor $\underline{\mathbf{X}}$. The structure of this approach is outlined in Algorithm 10 [94]. After computing an approximation for the range of unfolding matrices, they can be used to find a Tucker approximation (18), although they can also be orthogonalized by economic QR decomposition to be used for computation of the HOSVD. Note that in the R-ST algorithm, the factor matrices preserve the structure of the original data tensor (nonnegativity, sparsity, smoothness) but not necessarily the core tensor. As we mentioned in Section III-B, the procedure of column or row selection can be performed based on different kinds of probability distributions such as uniform or other distributions and also with or without replacement. Moreover, the algorithms naturally provide either additive or relative errors. Sampling based on the leverage scores are proposed in [32], [35], [36].

Cross-approximation¹⁸ [95]–[97], can be considered as a sampling technique with a difference that the sampling pro-

¹⁷Algorithms based on sampling slices are proposed in [34], [92], [93], but those in [92], [93] are not related to the Tucker decomposition.

¹⁸It is also called (pseudo)-skeleton decomposition or CUR decomposition.

Algorithm 10 Randomized Sampling Tucker Approximation (R-ST)

Input : A data tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, and a predefined multilinear rank (R_1, R_2, \dots, R_N)

Output: Approximative Tucker representation of the tensor $\underline{\mathbf{X}}$ as $\underline{\mathbf{X}} \cong [\underline{\mathbf{S}}; \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \dots, \mathbf{Q}^{(N)}]$

- 1 **for** $n = 1, 2, \dots, N$ **do**
- 2 Sample some columns of $\mathbf{X}_{(n)}$ based on a probability distribution and store them in the factor matrix $\mathbf{Q}^{(n)} \in \mathbb{R}^{I_n \times R_n}$
- 3 **end**
- 4 Compute the core tensor $\underline{\mathbf{S}} = \underline{\mathbf{X}} \times_1 \mathbf{Q}_1^\dagger \times_2 \mathbf{Q}_2^\dagger \dots \times_N \mathbf{Q}_N^\dagger$

cedure is performed heuristically instead of randomly. Three known heuristic algorithms are:

- Maxvol-based low-rank matrix approximation [98], [99]
- Cross2D matrix approximation [100], [101]
- Discrete Empirical Interpolatory Method (DEIM) [103], [104]
- Pivoted QR decomposition [105]

It is known that the quality of the cross approximation quite depends on the module of the determinant of the intersection matrix¹⁹ which is called *matrix volume*. More precisely, a set of columns and rows should be selected with an intersection matrix whose volume is as much as possible. Clearly, this is an NP hard problem because we need to check the volume of all possible intersection matrices produced by different selection of columns and rows. However, heuristic algorithms exist for computing suboptimal solutions.

Several generalizations of the matrix cross-approximation to the low Tucker rank approximation are proposed in [29], [31]. The proposed algorithms in [31] basically apply the Cross2D algorithm [100], [101] to the unfolding matrices sequentially where the long rows of the unfolding matrices can be treated as slice matrices. Here again a cross approximation of these slices are computed and this idea totally reduces the computational complexity of the cross-approximation. Two algorithms proposed in [29] are based on the fiber selection and they do not apply cross-approximation to unfolding matrices. The first algorithm is analytic while second one is iterative (adaptive). The analytic one is of less practical interest because it requires quite large number of fibers, but the adaptive one is more efficient and is a straightforward generalization of cross-approximation of matrices to tensors. A similar approach is proposed in [102] in the sense of number of selected fibers in each mode.

The fibers can also be sampled using the pivoted QR decomposition applied to the unfolding matrices $\mathbf{X}_{(n)}$. Inspired by the matrix case [105], *Higher Order Interpolatory Decomposition (HOID)* is proposed in [32]. Given an N th-order tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, the pivoted QR factorization

is applied to the unfolding matrix $\mathbf{X}_{(n)} \in \mathbb{R}^{I_1 \times J}$, ($J = \prod_{i \neq n} I_i$) as

$$\mathbf{X}_{(n)} \Pi = \mathbf{Q} \mathbf{R}, \quad (29)$$

where $\Pi \in \mathbb{R}^{J \times J}$ and $\mathbf{Q} \in \mathbb{R}^{I_n \times I_n}$ are a permutation matrix and an orthogonal matrix, respectively. The pivoted QR decomposition can be computed using strong rank-revealing QR (RRQR) algorithm [106]. The permutation matrix Π and the corresponding orthogonal matrix \mathbf{Q} are partitioned as follows

$$\Pi = [\Pi_1 \ \Pi_2], \quad \mathbf{Q} = [\mathbf{Q}_1 \ \mathbf{Q}_2], \quad (30)$$

where $\Pi_1 \in \mathbb{R}^{J \times K}$, $\Pi_2 \in \mathbb{R}^{J \times (J-K)}$, $\mathbf{Q}_1 \in \mathbb{R}^{I_n \times K}$, $\mathbf{Q}_2 \in \mathbb{R}^{I_n \times (I_n-K)}$. Here Equation (29) can be rewritten as

$$\mathbf{X}_{(n)} [\Pi_1 \ \Pi_2] = [\mathbf{Q}_1 \ \mathbf{Q}_2] \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{0} & \mathbf{R}_{22} \end{bmatrix}, \quad (31)$$

where $\mathbf{R}_{11} \in \mathbb{R}^{K \times K}$, $\mathbf{R}_{12} \in \mathbb{R}^{K \times (I_n-K)}$, $\mathbf{R}_{22} \in \mathbb{R}^{(I_n-K) \times (J-K)}$ and $\mathbf{0} \in \mathbb{R}^{(I_n-K) \times K}$ with all entries equal to zero. From (31), by straightforward computations, we have

$$\begin{aligned} \mathbf{Q}^{(n)} &\equiv \mathbf{X}_{(n)} \Pi_1 = \mathbf{Q}_1 \mathbf{R}_{11}, \\ \mathbf{X}_{(n)} \Pi_2 &= \mathbf{Q}_1 \mathbf{R}_{12} + \mathbf{Q}_2 \mathbf{R}_{22} \cong \mathbf{Q}_1 \mathbf{R}_{12}, \end{aligned}$$

if $\|\mathbf{R}_{22}\|_2$ is small enough. It can be seen that

$$\mathbf{X}_{(n)} \cong [\mathbf{Q}_1 \mathbf{R}_{11}, \ \mathbf{Q}_1 \mathbf{R}_{12}] \Pi^T, \quad (32)$$

and substituting $\mathbf{Q}_1 = \mathbf{Q}^{(n)} \mathbf{R}_{11}^{-1}$ in (32), we have

$$\mathbf{X}_{(n)} \cong \mathbf{Q}^{(n)} \mathbf{F}^T, \quad \mathbf{F}^T = [\mathbf{I} \ \mathbf{R}_{11}^{-1} \mathbf{R}_{12}] \Pi^T, \quad (33)$$

which is a low-rank matrix approximation of the matrix $\mathbf{X}_{(n)}$. The matrix $\mathbf{Q}^{(n)}$ which is a full-rank matrix²⁰ can be used as an approximation for the basis of the range of $\mathbf{X}_{(n)}$. It is worth mentioning that $\mathbf{Q}^{(n)} = \mathbf{X}_{(n)}(:, \mathbf{p})$ where \mathbf{p} is the indices of the selected columns and as a result it may not be necessarily orthonormal. The HOID algorithm applies the earlier procedure to all unfolding matrices $\mathbf{X}_{(n)}$ and computes the basis matrices $\mathbf{Q}^{(n)}$. Afterwards, the core tensor $\underline{\mathbf{S}}$ can be computed by

$$\underline{\mathbf{S}} = \underline{\mathbf{X}} \times_1 \mathbf{Q}^{(1)\dagger} \times_2 \mathbf{Q}^{(2)\dagger} \dots \times_N \mathbf{Q}^{(N)\dagger}.$$

A randomized variant of this algorithm can be developed by replacing randomized QR decomposition instead of the QR decomposition. To this end, we should first make reduction in the second mode of the unfolding matrices using random projection as $\mathbf{Y} = \mathbf{X}_{(n)} \Omega$ where $\Omega \in \mathbb{R}^{\prod_{i \neq n} I_i \times R}$ is a random matrix after which the economic QR decomposition of the matrix \mathbf{Y} is computed as $\mathbf{Y} = \mathbf{Q} \mathbf{R}$. The orthonormal matrix \mathbf{Q} is an approximate orthonormal basis for the range of the matrix $\mathbf{X}_{(n)}$ which can be used in the above procedure. This procedure is summarized in Algorithm 11. In situation that

¹⁹By an intersection matrix, we mean a matrix which produced by the intersection of sampled columns and rows.

²⁰Because it is the multiplication of a nonsingular matrix and an orthogonal matrix.

the singular values of the matrix $\mathbf{X}_{(n)}$ do not decay very fast, the orthonormal basis may not be accurate and power iteration technique is required (see Remark 1).

If the dimension of the first mode is also large, it is suggested in [107] to first make reduction in the first mode of the matrix using random projection as $\mathbf{Y} = \Omega \mathbf{X}_{(n)}$ where $\Omega \in \mathbb{R}^{(R+P) \times I_n}$ is a random matrix (P is the oversampling parameter), after which the procedure described above is applied to the matrix \mathbf{Y} for column selection. It is shown that the selected indices of columns for the matrix \mathbf{Y} can be used for the original data matrix \mathbf{X} . More precisely, if $\mathbf{Y} \cong \mathbf{Y}(:, \mathbf{p}) \mathbf{F}^T$ then the indices \mathbf{p} corresponding to the selected columns can be used for the matrix $\mathbf{X}_{(n)}$, i.e., $\mathbf{X}_{(n)} \cong \mathbf{X}_{(n)}(:, \mathbf{p}) \mathbf{F}^T$.

Remark 2: The procedure of column selection can be performed using the leverage scores sampling or the DEIM algorithm. These techniques were studied in [32].

Algorithm 11 Randomized Higher Order Interpolatory Decomposition (HOID) (R-HOID) Algorithm

Input : A data tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, and a multilinear rank (R_1, R_2, \dots, R_N)

Output: Approximative Tucker representation of the tensor $\underline{\mathbf{X}}$ as $[[\underline{\mathbf{S}}, \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \dots, \mathbf{Q}^{(N)}]]$

```

1 for  $n = 1, 2, \dots, N$  do
2   Compute a randomized interpolatory decomposition
   of unfolding  $\mathbf{X}_{(n)} \cong \mathbf{Q}^{(n)} \mathbf{F}_n^T$ , where  $\mathbf{Q}^{(n)} \in \mathbb{R}^{I_n \times R_n}$ ,
   are columns of  $\mathbf{X}_{(n)}$ 
3 end
4 Compute the core tensor  $\underline{\mathbf{S}} \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_N}$  as

```

$$\underline{\mathbf{S}} \equiv \underline{\mathbf{X}} \times_1 \mathbf{Q}^{(1)\dagger} \times_2 \dots \times_N \mathbf{Q}^{(N)\dagger}$$

Other sampling techniques for the HOSVD and the HOOI algorithms can be found in [33], [34], [37]. The proposed algorithms in [33] are based on the sparsification idea where in the first step, a sparse tensor is generated from the original data tensor. Then, it is used for subsequent computations. The idea of the stochastic gradient descent is used in [37], where at each iteration only a subtensor of the original tensor is considered. The randomized algorithms proposed in [34] are based on sampling fibers instead of components.

C. RANDOMIZED LEAST-SQUARES TUCKER DECOMPOSITION

At each iteration of the randomized Algorithm 7, several multiplications with random matrices and also contraction of the core tensor with factor matrices are required which may be expensive if the algorithm needs many iterations to converge. Motivated by this fact, in [38], a randomized least-squares HOOI algorithm is proposed. This algorithm is summarized in Algorithm 12 in which instead of exploiting updating rules 4 and 7 in Algorithm 5, the equivalent least-squares problems (23) and (26) are solved, respectively. To be precise, instead of performing Lines 4-5 and 7 of Algorithm 5,

the least-squares problem (23) is solved by random least-squares algorithms. Please note that the coefficient matrices in least-squares problems (23) and (26) are not required to be computed explicitly. Also, the TensorSketch [72] is used in [38] to solve the underlying least-squares problem.

Algorithm 12 Randomized Least-Squares Higher Order Orthogonal Iteration (HOOI) (R-LSHOOI) Algorithm

Input : A data tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, and a multilinear rank (R_1, R_2, \dots, R_N) ;

Output: Approximative Tucker representation of the tensor $\underline{\mathbf{X}}$ as $\underline{\mathbf{X}} \cong [[\underline{\mathbf{S}}, \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \dots, \mathbf{Q}^{(N)}]]$

```

1 Initialize factor matrices  $\mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \dots, \mathbf{Q}^{(N)}$  and core
  tensor  $\underline{\mathbf{S}}$  with i.i.d uniformly random entries
2 while A stopping criterion is not satisfied do
3   for  $n = 1, 2, \dots, N$  do
4     Compute  $\mathbf{Q}^{(n)}$  by solving least-squares problem
     (23) via a sketching technique for
      $n = 1, 2, \dots, N$ ;
5   end
6   Find the core tensor  $\underline{\mathbf{S}}$  by solving least-squares
   problem (26) via sketching technique;
7 end

```

D. RANDOMIZED COUNT-SKETCH TUCKER DECOMPOSITION

The count-sketch technique discussed in Subsection III-C, can be straightforwardly generalized to tensors. Motivated by the paper [108] in which the underlying least-squares problems in the CP-ALS²¹ algorithm [76] are solved via the sampling techniques, in [38], the count-sketch technique is used to solve the least-squares problems (23) and (26). More precisely, the reduction map \mathbf{T} in the least-squares problem (14), is the count-sketch operator. Because of a special structure of the coefficient matrix of least-squares problems (23) and (26), sketching procedure can be performed very fast by employing FFT [38].

The count-sketch technique is used in [109] for tensor interpolative decomposition [110] and also in [111] for the CP decomposition. The concept of Higher-order Count Sketch is developed in [39] for higher order tensors to fully exploit the multidimensional structure of the data tensor.

Another possible direction is using the count-sketch technique for low-rank matrix approximations of the unfolding matrices which are required for tensor decomposition. Here, the factor matrices are approximated by computing low-rank approximation of the unfolding matrices using the count-sketch technique and then the core tensor is computed via (19).

²¹ALS means Alternating Least-Squares [76].

VI. APPLICATION OF THE RANDOMIZED HOSVD IN FAST CANONICAL POLYADIC DECOMPOSITION (CPD)

The idea of making a prior reduction in a raw data tensor into the Tucker format and then decomposing its core tensor in the CPD format, was first proposed by Bro in his PhD thesis [112]. This technique is applicable only if the rank of the tensor does not exceed its dimensions. Motivated by this difficulty, a prior reduction in the Tensor Train format [113] was recently suggested in [114]. In this section, we focus on the former and discuss how randomization can be exploited within the tensor decomposition procedure.

The main procedure of this technique is as follows

- Dimensionally reduction of a given data tensor using the Tucker decomposition as a preprocessing step,
- Computing the CPD of the compressed core tensor of the Tucker decomposition,
- Recovering the factor matrices of the CPD of the original data tensor from the factor matrices of the CPD of the core tensor.

Assuming that the data tensor has a low multilinear rank approximation, the first step mentioned above can be done through the randomized HOSVD algorithms. This makes the decomposition procedure much faster compared with the deterministic ones. This idea is used in [115].

Assume that the Tucker representation of a data tensor $\underline{\mathbf{X}}$ is computed as $[[\underline{\mathbf{S}}; \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \dots, \mathbf{Q}^{(N)}]]$ using one of the randomized HOSVD or randomized Tucker decomposition algorithms discussed in Section V. Let $\underline{\mathbf{S}} \cong [[\tilde{\mathbf{A}}^{(1)}, \tilde{\mathbf{A}}^{(2)}, \dots, \tilde{\mathbf{A}}^{(N)}]]$ be the CPD of the compressed core tensor $\underline{\mathbf{S}}$. Then, the CPD of the original tensor $\underline{\mathbf{X}}$ can be recovered from the CPD of the compressed tensor $\underline{\mathbf{S}}$. More precisely, the CPD of the original tensor $\underline{\mathbf{X}}$ is

$$\underline{\mathbf{X}} \cong [[\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)}]],$$

where $\mathbf{A}^{(n)} = \mathbf{Q}^{(n)} \tilde{\mathbf{A}}^{(n)}$, $n = 1, 2, \dots, N$.

Remark 3: The idea of a prior reduction of a data tensor into the TT format for computation of the CPD was suggested in [114]. Randomized variants of these algorithms can also be developed.

VII. DISCUSSION ON FURTHER CHALLENGES

Recently, several scalable algorithms were proposed in [116]–[118] for computation of the HOSVD. These algorithms inherently do not have a randomized structure and basically they exploit the idea of *on-the-fly-computation* and *parallel row-wise update rule roles* to avoid the *intermediate data explosion* problem. Combining these algorithms with randomized techniques is a potential topic needs to be investigated. For example, this issue was recently studied in [119]. Block Term Decomposition (BTD) [120]–[122] is a generalization of the CP decomposition and the Tucker decomposition and has found many applications such as blind source separation [123], feature extraction [124], electroencephalogram (EEG) analysis [125] etc. Proposing randomized algorithms for this tensor decomposition is a potential research topic needs to be investigated.

Motivated by some applications in Cyber-Physical-Social Systems and Internet of Things (IOT), distributed algorithms for computation of the HOSVD were recently developed in [126]–[130]. The proposed algorithms mainly distribute the unfolding matrices among several machines and integrate their low-rank matrix approximations to find the HOSVD approximation of the original data tensor. These algorithm can be further improved by exploiting randomized algorithms.

VIII. SIMULATIONS

In this section, we experimentally examine some of the selected and most efficient randomized algorithms presented in Section V and compare their performance and efficiency. All numerical simulations were performed on a cluster with 508Gb RAM and 48 CPUs 1.2GHz. Although, the sampling procedure can be done based on different types of probability distributions, but we used uniform distribution in our computations. Note that in the case of randomized sampling algorithms, we have used sampling without replacement as it works better than sampling with replacement in machine learning applications [65], [66]. We experienced the same results. We have set oversampling $P = 10$ and power iteration $q = 2$ for all random projection algorithms (except the noiseless case where we have not used any oversampling or power iteration parameter). For the R-PET algorithm (Algorithm 9), we have used $S = 2K + 1$ in all experiments where for each example, a specific K was used and for the STHOSVD algorithm, we always considered the ascending order $p = [1, 2, \dots, N]$. The relative error is used as a criterion to assess the performance of the randomized algorithms for the Tucker approximation and is defined as follows

$$E = \frac{\|\underline{\mathbf{X}} - \hat{\underline{\mathbf{X}}}\|_F}{\|\underline{\mathbf{X}}\|_F}, \quad (34)$$

where $\underline{\mathbf{X}}$ is a given data tensor and $\hat{\underline{\mathbf{X}}}$ is its Tucker approximation, respectively.

We also consider the case that the tensor $\underline{\mathbf{X}}$ is corrupted by a noise term as follows

$$\underline{\mathbf{Y}} = \underline{\mathbf{X}} + \gamma \underline{\mathbf{N}}, \quad (35)$$

where $\underline{\mathbf{N}}$ is a Gaussian random tensor²² and γ is a parameter to change the noise level. The Signal-to-Noise Ratio (SNR) measure is used to evaluate the noise level and is defined as follows

$$SNR [dB] = 10 \log \left(\frac{\|\underline{\mathbf{X}}\|_F^2}{\|\gamma \underline{\mathbf{N}}\|_F^2} \right).$$

For the noisy case, the performance of an algorithm is evaluated in terms of fit which is defined as follows

$$Fit = 1 - E,$$

where E was defined by (34).

²²A tensor whose elements are independent and identically distributed (i.i.d) and taken from normal distribution with zero mean and variance 1.

A. TEST 1 – SYNTHETIC DATA TENSORS

The synthetic data tensors used in our experiments are: random tensors with low multilinear rank (noiseless and noisy cases), random sparse tensors, Hilbert tensors and function based tensors.

1) RANDOM LOW MULTILINEAR RANK TENSORS

In this experiment, we consider a 3rd order tensor of size $1000 \times 1000 \times 1000$ with multilinear rank $(20, 40, 30)$. We first randomly generate a Gaussian core tensor $\mathbf{S} \in \mathbb{R}^{20 \times 40 \times 30}$ and Gaussian factor matrices $\mathbf{Q}^{(1)} \in \mathbb{R}^{1000 \times 20}$, $\mathbf{Q}^{(2)} \in \mathbb{R}^{1000 \times 40}$, $\mathbf{Q}^{(3)} \in \mathbb{R}^{1000 \times 30}$. Then, the following tensor

$$\underline{\mathbf{X}} = \llbracket \mathbf{S}; \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \mathbf{Q}^{(3)} \rrbracket, \quad (36)$$

is generated and some of the deterministic and randomized algorithms are applied on it. The running time and corresponding relative error of algorithms are averaged over 50 Monte Carlo simulations. To show the acceleration of randomized algorithms against the deterministic counterparts, we first report the running time of the THOSVD and the STHOSVD algorithms and their randomized variants. The results for the noiseless data tensor are displayed in Figure 4 and the results for the noisy case (20 dB) are displayed in Figure 5. For computing the leading singular vectors of unfolding matrices, we used both the truncated SVD (MATLAB function svds) and the EVD (MATLAB function eigs) and compared their executive times. In displayed figures, the notations “svds” and “eigs” mean that the underlying algorithms exploit the truncated SVD and the EVD for computing the factor matrices, respectively. Our results show that the EVD is much faster than using the svds for computing the factor matrices and their randomized versions make the algorithms much faster. The performance comparison of deterministic and randomized algorithms for the noiseless case are reported in Table 2. It is seen that randomized algorithms provide approximately the same results as the deterministic ones but with less computational time. For the noisy data tensor with 20 dB, the fit of the THOSVD, the STHOSVD algorithms and their randomized variants

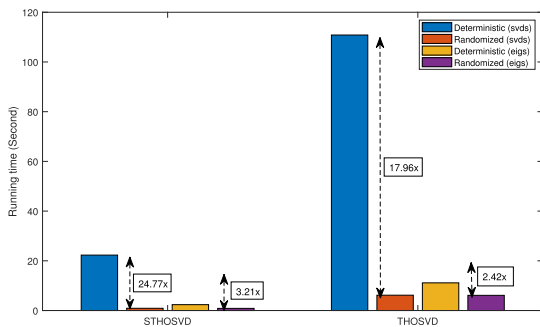


FIGURE 4. Running time comparison of deterministic and randomized variants of the THOSVD and the STHOSVD algorithms (svds and eigs versions) for a noiseless random data tensor of size $1000 \times 1000 \times 1000$ (10^9 entries) and multilinear rank $(20, 40, 30)$.

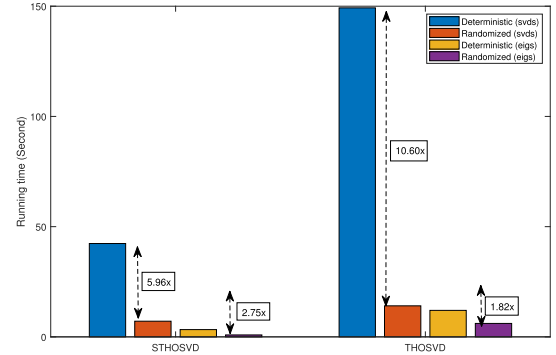


FIGURE 5. Running time comparison of deterministic and randomized variants of the THOSVD and the STHOSVD algorithms (svds and eigs versions) for a noisy random data tensor of size $1000 \times 1000 \times 1000$ and multilinear rank $(20, 40, 30)$ with SNR=20 dB.

were almost the same and equal to 90% but the randomized algorithms were much faster as seen in Figure 5. For the noiseless data tensor, we did not use any power iteration or oversampling parameter. For the noisy case, we set power iteration $q = 2$ and oversampling $P = 10$, and this is why for a data tensor with the same size and multilinear rank, more acceleration is achieved using randomized algorithms for the noiseless data tensor compared to the noisy data tensor.

We also compared the performance of different randomized algorithms for the noiseless data tensor in Figure 6 (also roughly the same results were obtained for the noisy data tensor and we skip them). The results show that the R-STHOSVD algorithm (Algorithm 8) and the randomized sampling Tucker (R-ST) algorithm (Algorithm 10) are the fastest algorithms for computation of the HOSVD or the Tucker decomposition. Note the R-ST algorithm based on uniform sampling is also relatively fast but when the noise level is relatively high or the singular values of unfolding matrices do not decay very fast, its accuracy is degraded due to additive error bounds. In such situations, the leverage-score

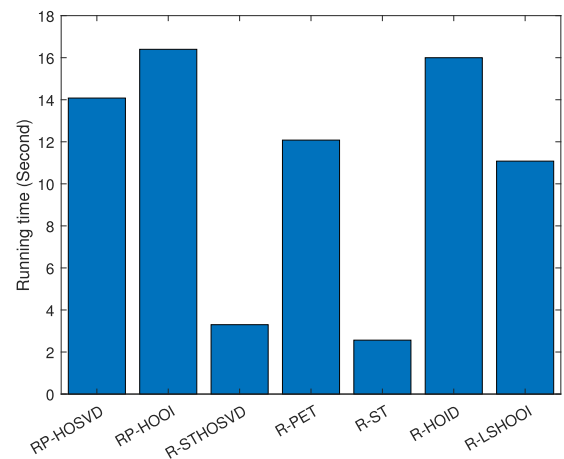


FIGURE 6. Running time comparison of various randomized algorithms for a noiseless random data tensor of size $1000 \times 1000 \times 1000$ and multilinear rank $(20, 40, 30)$.

TABLE 2. Relative error comparison of different deterministic and randomized algorithms for computing low multilinear rank approximation of data tensors (36), (37), (38) and (39).

Data	HOOI	STHOSVD	HOSVD	RP-HOSVD	RP-HOOI	R-STHOSVD	R-PET	R-ST	R-HOID	R-LSHOOI
Data tensor (36)	1.0303e-13	1.5103e-13	1.5203e-13	6.0334e-13	3.0012e-13	6.0334e-13	1.5620e-12	1.2812e-11	4.7531e-13	6.2167e-13
Data tensor (37)	1.0031e-8	1.6095e-8	1.6124e-8	3.6206e-8	1.0031e-8	3.1637e-8	3.4651e-7	4.1674e-6	6.2213e-8	3.4352e-8
Sparse tensor (38)	2.7732e-6	4.8484e-6	9.5498e-6	9.5498e-6	2.7732e-6	7.5750e-6	7.5750e-5	7.5750e-4	7.5750e-6	4.7531e-6
Hilbert tensor (39)	1.0044e-10	4.9682e-10	8.4250e-10	1.1041e-10	1.2122e-10	1.1478e-10	4.2234e-9	6.3753e-8	6.9878e-10	4.7531e-10

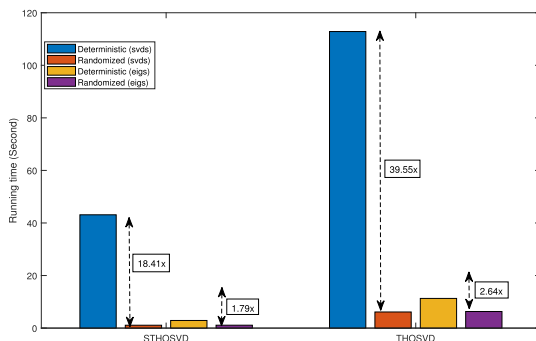
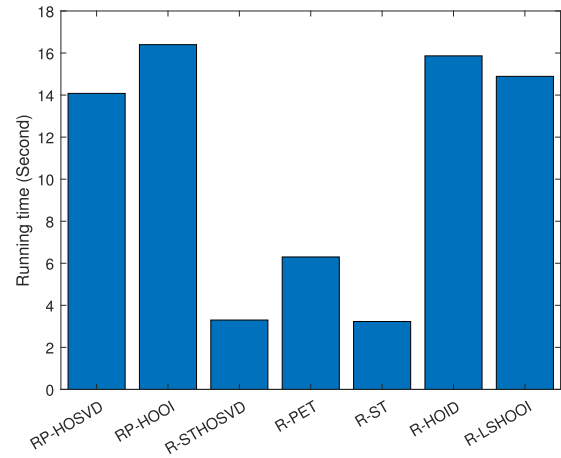
probability distribution can provide more accurate solutions with relative error bound guarantee but requiring higher computational cost.

2) FUNCTION BASED TENSORS

Consider a 3rd order tensor of size $1000 \times 1000 \times 1000$ whose components are generated as follows

$$\underline{\mathbf{X}}(i, j, k) = \frac{1}{\sqrt[5]{i^5 + j^5 + k^5}}. \quad (37)$$

It is shown in [131]–[137] that such a data tensor admits a very low multilinear rank approximation since the singular values of the corresponding unfolding matrices decay very fast, (for more related function based tensors, see [131]–[137] and the references therein). We set the multilinear rank as (30, 30, 30) in our computations and apply deterministic and randomized algorithms on the mentioned data tensor. The running time of the THOSVD, the STHOSVD and their randomized variants are reported in Figure 7. Also, the performance comparison of deterministic and randomized algorithms are reported in Table 2. Here again the simulation results indicate the superiority of randomized algorithms over the deterministic ones. To compare the running time of different randomized algorithms, see Figure 8. The results show that the R-STHOSVD and the R-ST algorithms are the most efficient randomized algorithms for the low multilinear rank approximation of the data tensor (37).

**FIGURE 7.** Running time comparison of deterministic and randomized variants of the THOSVD and the STHOSVD algorithms (svds and eigs versions) for the function based tensor (37) of size $1000 \times 1000 \times 1000$ and multilinear rank (30, 30, 30).**FIGURE 8.** Running time comparison of various randomized algorithms for the function based tensor (37) of size $1000 \times 1000 \times 1000$ and multilinear rank (30, 30, 30).

3) RANDOM SPARSE TENSORS

Consider the following data tensor

$$\underline{\mathbf{X}} = \sum_{i=1}^{10} \frac{\gamma}{i^2} \mathbf{x}_i \circ \mathbf{y}_i \circ \mathbf{z}_i + \sum_{i=11}^{200} \frac{1}{i^2} \mathbf{x}_i \circ \mathbf{y}_i \circ \mathbf{z}_i, \quad (38)$$

where \mathbf{x}_i , \mathbf{y}_i and \mathbf{z}_i are sparse vectors with only 0.05 sparsity (percent non-zero numbers). We set $I_1 = I_2 = I_3 = 1000$ and $\gamma = 1000$ and generate a 3rd-order tensor $\underline{\mathbf{X}}$. The parameters used in this simulation are the same as previous experiments and we use a multilinear rank $R = (15, 15, 15)$. The running time of the THOSVD and the STHOSVD algorithms and their randomized forms are reported in Figure 9. Also, the performance comparison of deterministic and randomized algorithms are reported in Table 2. Figure 9 and Table 2, show that the randomized algorithms can significantly speed-up the decomposition procedure with achieving roughly the same accuracy as the deterministic algorithms. In Figure 10, we make a comparison among different randomized algorithms. The results show that the R-STHOSVD and the R-ST algorithms are the most promising approaches for the low multilinear approximation of the data tensor (38). It is also seen that the randomized algorithms can achieve almost the same accuracy as the deterministic ones but much faster.

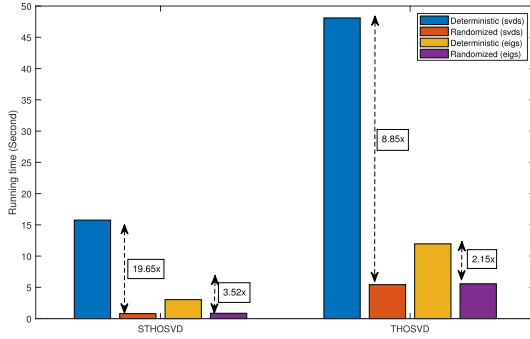


FIGURE 9. Running time comparison of deterministic and randomized variants of the THOSVD and the STHOSVD algorithms (svds and eigs versions) for the sparse tensor (38) of size $1000 \times 1000 \times 1000$ and multilinear rank (15, 15, 15).

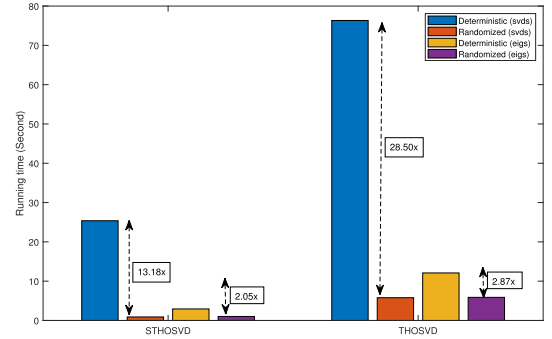


FIGURE 11. Running time comparison of deterministic and randomized variants of the THOSVD and the STHOSVD algorithms (svds and eigs versions) for the Hilbert tensor of size $1000 \times 1000 \times 1000$ and multilinear rank (20, 20, 20).

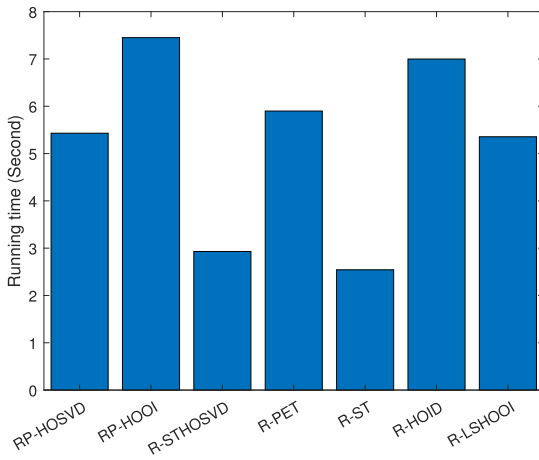


FIGURE 10. Running time comparison of various randomized algorithms for the sparse tensor (38) of size $1000 \times 1000 \times 1000$ and multilinear rank (15, 15, 15).

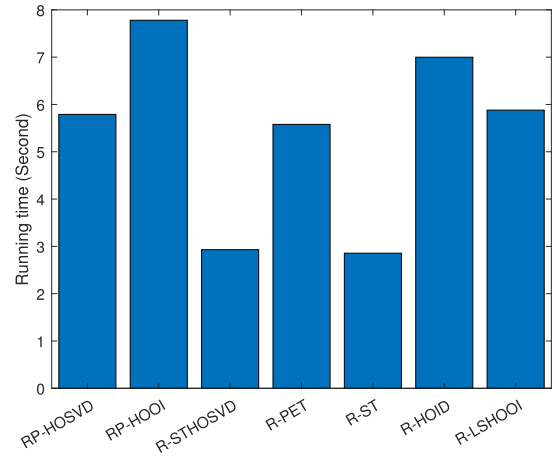


FIGURE 12. Running time comparison of various randomized algorithms for the Hilbert tensor of size $1000 \times 1000 \times 1000$ and multilinear rank (20, 20, 20).

4) HILBERT TENSORS

An N th order Hilbert tensor is defined as follows

$$\underline{\mathbf{X}}(i_1, i_2, \dots, i_N) = \frac{1}{i_1 + i_2 + \dots + i_N - N + 1}, \quad (39)$$

which is a natural generalization of Hilbert matrices (second order tensors). We generate a 3th-order Hilbert tensor of size $1000 \times 1000 \times 1000$ and use approximate multilinear rank (20, 20, 20) within the randomized algorithms. The running time of the THOSVD and the STHOSVD algorithms and their randomized variants are reported in Figure 11. Also, the performance comparison of deterministic and randomized algorithms are reported in Table 2. Once again the superiority of randomized algorithms over the deterministic ones is visible. In Figure 12, we make a comparison among different randomized algorithms. The results show that the R-STHOSVD and the R-ST algorithms are the best and most efficient algorithms for the low multilinear rank approximation of the data tensor (39). It is seen that the randomized algorithms can achieve almost the same accuracy as the deterministic ones but much faster.

B. TEST 2 – REAL DATA TENSORS

1) COIL-100 DATA SET

Columbia object image library COIL-100 is a dataset which consists of 7200 color images (100 objects under 72 different rotations) [138]. Each image is of size $128 \times 128 \times 3$ and the whole dataset naturally can be represented as a 5th-order tensor of size $128 \times 128 \times 3 \times 72 \times 100$. The first 5 images and some random rotations are depicted in Figure 13. We reshape this tensor to a 3rd-order tensor of size $1024 \times 1152 \times 300$ and apply THOSVD and STHOSVD algorithms and their randomized variants on this 3rd order data tensor. The speedup achieved by the R-STHOSVD and the R-THOSVD for different uniform multilinear ranks (R, R, R) , $R = 5, 10, 15, \dots, 30$ are shown in Figure 14. Also, the reconstructed images using the STHOSVD and R-STHOSVD for multilinear rank (450, 450, 250) are displayed in Figure 16. The results indicate that the randomized algorithms provide roughly the same performance as the deterministic algorithms but in much less running times.

The running time comparison of various randomized algorithms for multilinear rank (20, 20, 20) are reported



FIGURE 13. The first 8 images of the COIL-100 dataset for three random rotations.

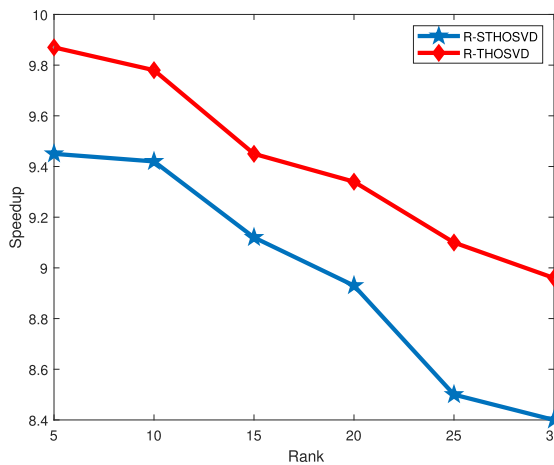


FIGURE 14. The running time speed-up of randomized STHOSVD and randomized THOSVD for decomposing COIL-100 for selected multilinear rank (R, R, R) .

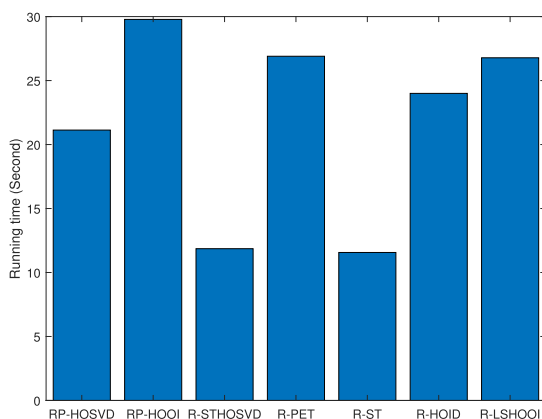


FIGURE 15. The running time comparison of various randomized algorithms for decomposing COIL100 data set and multilinear rank $(450, 450, 250)$.

in Figure 15. Our results again show the effectiveness of the R-STHOSVD for decomposing large-scale data tensors.



FIGURE 16. The reconstructed images using randomized and deterministic STHOSVD algorithms for the first 8 images.



FIGURE 17. Visualization of the original frame number 50 of the grayscale video.

2) NELL-2 DATA TENSOR (LARGE-SCALE SPARSE TENSORS)

In this experiment, we show the applicability and effectiveness of the randomized algorithms for handling large-scale sparse data tensors. To this end, we considered the *NELL-2* [139] which is accessible at the formidable repository of open sparse tensors and tool (FROSTT) database [140]. It is a 3rd order tensor of size $12092 \times 9184 \times 28818$ with 76879419 number of nonzero elements. The *NELL-2* dataset was tested on machine learning algorithms which exploit relationship between variables [139]. It was generated at Chicago university as a part of “Read the Web” project.

For this sparse data tensor, all the deterministic algorithms introduced in the paper were not able to decompose the mentioned data tensors in the Tucker format due to high memory and computational complexity issues while the randomized STHOSVD was able to decompose it. The running

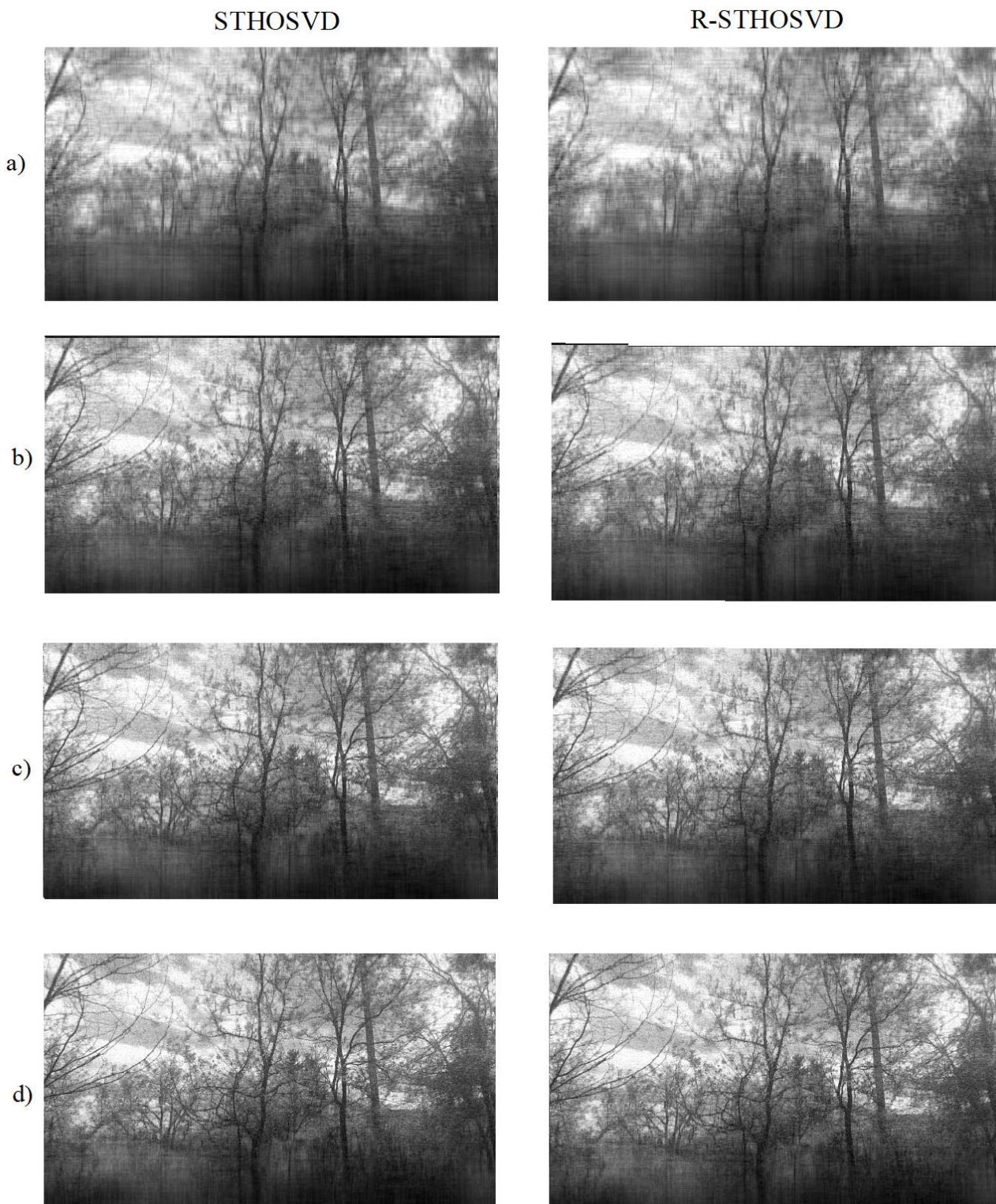


FIGURE 18. Visualization of the reconstructed frame number 50 using deterministic and randomized STHOSVD algorithms for different multilinear ranks, a) Multilinear rank (50, 50, 50), b) Multilinear rank (100, 100, 100), c) Multilinear rank (150, 150, 150), d) Multilinear rank (200, 200, 200).

time and relative-error achieved by the RSTHOSVD algorithm for the uniform multilinear rank (R, R, R) , $R = 100, 125, 150 \dots, 200$ are reported in Table 3.

C. VIDEO DATASETS

In this experiment we compress a large-scale grayscale video in the Tucker format. The data tensor was firstly used to

TABLE 3. The running time (in second) and relative errors of the randomized STHOSVD Algorithm for NELL-2 dataset.

Target Multilinear Rank	STHOSVD		R-STHOSVD	
	Running time	Relative error	Running time	Relative error
(50,50,50)	X	X	212.56	0.6248
(75,75,75)	X	X	442.72	0.5720
(100,100,100)	X	X	853.25	0.5255
(125,125,125)	X	X	1347.94	0.4944
(150,150,150)	X	X	1412.56	0.4695
(175,175,175)	X	X	1514.43	0.4484
(200,200,200)	X	X	2009.51	0.4283

TABLE 4. The relative errors of deterministic and randomized algorithms for the grayscale video.

Target multilinear rank	Relative error				Running time (Second)				Compression ratio
	HOSVD	R-HOSVD	STHOSVD	R-STHOSVD	HOSVD	R-HOSVD	STHOSVD	R-STHOSVD	
(10,10,10)	0.3226	0.3346	0.3210	0.3247	122.90	17.42	38.42	3.40	1.1393e-05
(20,20,20)	0.3030	0.3178	0.3020	0.3069	140.11	19.98	41.20	3.82	2.4062e-05
(30,30,30)	0.2913	0.3069	0.2901	0.2975	133.59	21.29	36.09	4.17	3.9282e-05
(40,40,40)	0.2827	0.2989	0.2814	0.2894	136.55	24.60	42.04	5.59	5.8327e-05
(50,50,50)	0.2761	0.2937	0.2742	0.2833	133.60	29.81	43.49	6.85	8.2475e-05
(60,60,60)	0.2699	0.2879	0.2681	0.2772	158.87	34.81	48.91	7.90	1.1300e-04
(70,70,70)	0.2647	0.2835	0.2625	0.2728	159.74	38.46	49.35	8.77	1.5118e-04
(80,80,80)	0.2600	0.2779	0.2575	0.2678	157.28	39.18	50.22	10.51	1.9828e-04
(90,90,90)	0.2553	0.2740	0.2528	0.2634	156.29	44.98	50.34	10.26	2.5559e-04
(100,100,100)	0.2509	0.2697	0.2485	0.2597	158.16	47.69	52.07	13.15	3.2437e-04
(110,110,110)	0.2471	0.2656	0.2445	0.2559	159.13	53.78	50.61	12.94	4.0591e-04
(120,120,120)	0.2432	0.2618	0.2407	0.2522	156.14	54.72	51.77	15.01	5.0148e-04
(130,130,130)	0.2397	0.2583	0.2371	0.2486	157.94	58.61	53.65	16.41	6.1235e-04
(140,140,140)	0.2364	0.2544	0.2337	0.2451	159.34	56.99	53.01	17.66	7.3981e-04
(150,150,150)	0.2333	0.2511	0.2305	0.2421	163.63	65.13	55.07	17.65	8.8511e-04
(160,160,160)	0.2301	0.2475	0.2273	0.2389	160.37	68.05	54.04	20.17	0.0010
(170,170,170)	0.2270	0.2443	0.2243	0.2359	161.34	70.54	54.93	21.15	0.0012
(180,180,180)	0.2239	0.2408	0.2213	0.2331	159.83	74.41	56.63	22.73	0.0014
(190,190,190)	0.2209	0.2378	0.2184	0.2301	161.92	72.89	57.45	23.21	0.0017
(200,200,200)	0.2181	0.2346	0.2156	0.2271	162.20	82.63	57.64	25.59	0.0019

classify the frames in [38]. The original data tensor is a 4rd order tensor of size $2200 \times 1080 \times 3 \times 1980$ consisting of 1980 RGB frames. For our task, we consider its grayscale form which is of size $2200 \times 1080 \times 1980$. The frame number 50 is displayed in Figure 17. We applied the THOSVD and the STHOSVD algorithms on the mentioned grayscale video for different uniform multilinear rank (R, R, R) , $R = 10, 20, \dots, 200$. The relative error and running time comparison are reported in Table 4. Also, the reconstructed images using randomized and deterministic STHOSVD algorithms are visualized in Figure 18. In both scenarios, it is seen that the randomized algorithms achieve roughly the same relative error as the deterministic algorithms but much faster.

IX. CONCLUSION

In this article, we reviewed and extended a variety of state-of-the-art randomized algorithms for computing the Tucker decomposition and the Higher Order SVD (HOSVD). We studied both single-pass and multi-pass randomized algorithms and also random projection and sampling techniques for computing the Tucker decomposition and the HOSVD. Simulations were conducted on synthetic and real datasets to compare the performance and running time of the deterministic and the randomized algorithms with particular focus on

showing the superiority of randomized algorithms against the deterministic ones.

ACKNOWLEDGMENT

(Salman Ahmadi-Asl, Stanislav Abukhovich, and Maame G. Asante-Mensah contributed equally to this work.) Salman Ahmadi-Asl would like to thank Dr. Varvara Logacheva for her constructive comments and useful discussion on the paper.

REFERENCES

- [1] L. R. Tucker, "The extension of factor analysis to threedimensional matrices," in *Contributions to Mathematical Psychology*, vol. 110119. New York, NY, USA: Holt, Rinehart and Winston (HRW), 1964.
- [2] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, Sep. 1966.
- [3] L. R. Tucker, "Implications of factor analysis of three-way matrices for measurement of change," *Problems Measuring Change*, vol. 15, pp. 122–137, 1963.
- [4] Y.-D. Kim, E. Park, S. Yoo, T. Choi, L. Yang, and D. Shin, "Compression of deep convolutional neural networks for fast and low power mobile applications," 2015, *arXiv:1511.06530*. [Online]. Available: <http://arxiv.org/abs/1511.06530>
- [5] J. Gusak, M. Kholiavchenko, E. Ponomarev, L. Markeeva, I. Oseledets, and A. Cichocki, "MUSCO: Multi-stage compression of neural networks," 2019, *arXiv:1903.09973*. [Online]. Available: <http://arxiv.org/abs/1903.09973>
- [6] B. Savas and L. Eldén, "Handwritten digit classification using higher order singular value decomposition," *Pattern Recognit.*, vol. 40, no. 3, pp. 993–1003, Mar. 2007.

- [7] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2002, pp. 447–460.
- [8] D. Rafailidis and P. Daras, "The TFC model: Tensor factorization and tag clustering for item recommendation in social tagging systems," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 43, no. 3, pp. 673–688, May 2013.
- [9] A. Nanopoulos, D. Rafailidis, P. Symeonidis, and Y. Manolopoulos, "MusicBox: Personalized music recommendation based on cubic analysis of social tags," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 2, pp. 407–412, Feb. 2010.
- [10] J.-T. Sun, H.-J. Zeng, H. Liu, Y. Lu, and Z. Chen, "CubeSVD: A novel approach to personalized Web search," in *Proc. 14th Int. Conf. World Wide Web (WWW)*, 2005, pp. 382–390.
- [11] P. Symeonidis, "ClustHOSVD: Item recommendation by combining semantically enhanced tag clustering with tensor HOSVD," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 46, no. 9, pp. 1240–1251, Sep. 2016.
- [12] L. De Lathauwer and J. Vandewalle, "Dimensionality reduction in higher-order signal processing and rank- (R_1, R_2, \dots, R_N) reduction in multilinear algebra," *Linear Algebra Appl.*, vol. 391, pp. 31–55, Nov. 2004.
- [13] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.
- [14] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017.
- [15] T. G. Kolda and J. Sun, "Scalable tensor decompositions for multi-aspect data mining," in *Proc. 8th IEEE Int. Conf. Data Mining*, Dec. 2008, pp. 363–372.
- [16] R. Badeau and R. Boyer, "Fast multilinear singular value decomposition for structured tensors," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1008–1021, Jan. 2008.
- [17] A. R. Zhang, Y. Luo, G. Raskutti, and M. Yuan, "ISLET: Fast and optimal low-rank tensor regression via importance sketching," *SIAM J. Math. Data Sci.*, vol. 2, no. 2, pp. 444–479, Jan. 2020.
- [18] Y. Feng and G. Zhou, "Orthogonal random projection for tensor completion," *IET Comput. Vis.*, vol. 14, no. 5, pp. 233–240, Aug. 2020.
- [19] S. Ahmadi-Asl, A. Cichocki, A. H. Phan, M. G. Asante-Mensah, M. M. Ghazani, T. Tanaka, and I. Oseledets, "Randomized algorithms for fast computation of low rank tensor ring model," *Mach. Learn., Sci. Technol.*, vol. 2, no. 1, pp. 305–340, 2020.
- [20] A. Kolbeinsson, J. Kossaifi, Y. Panagakis, A. Bulat, A. Anandkumar, I. Tzoulaki, and P. Matthews, "Tensor dropout for robust learning," 2019, *arXiv:1902.10758*. [Online]. Available: <http://arxiv.org/abs/1902.10758>
- [21] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [22] M. Che and Y. Wei, "Randomized algorithms for the approximations of Tucker and the tensor train decompositions," *Adv. Comput. Math.*, vol. 45, pp. 1–34, Jul. 2018.
- [23] R. Minster, A. K. Saibaba, and M. E. Kilmer, "Randomized algorithms for low-rank tensor decompositions in the Tucker format," *SIAM J. Math. Data Sci.*, vol. 2, no. 1, pp. 189–215, Jan. 2020.
- [24] G. Zhou, A. Cichocki, and S. Xie, "Decomposition of big tensors with low multilinear rank," 2014, *arXiv:1412.1885*. [Online]. Available: <http://arxiv.org/abs/1412.1885>
- [25] Y. Sun, Y. Guo, J. A. Tropp, and M. Udell, "Tensor random projection for low memory dimension reduction," in *Proc. NIPS Workshop Relational Represent. Learn.*, 2018, pp. 1–15.
- [26] A. S. J. W. Wolf, "Low rank tensor decompositions for high dimensional data approximation, recovery and prediction," Ph.D. dissertation, Dept. Math., Tech. Univ. Berlin, Berlin, Germany, 2019.
- [27] K. Batselier, W. Yu, L. Daniel, and N. Wong, "Computing low-rank approximations of large-scale matrices with the tensor network randomized SVD," *SIAM J. Matrix Anal. Appl.*, vol. 39, no. 3, pp. 1221–1244, Jan. 2018.
- [28] P. Drineas, R. Kannan, and M. W. Mahoney, "Fast Monte Carlo algorithms for matrices II: Computing a low-rank approximation to a matrix," *SIAM J. Comput.*, vol. 36, no. 1, pp. 158–183, Jan. 2006.
- [29] C. F. Caiafa and A. Cichocki, "Generalizing the column-row matrix decomposition to multi-way arrays," *Linear Algebra Appl.*, vol. 433, no. 3, pp. 557–573, Sep. 2010.
- [30] J. Sun, S. Papadimitriou, C.-Y. Lin, N. Cao, S. Liu, and W. Qian, "MultiVis: Content-based social network exploration through multi-way visual analysis," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2009, pp. 1064–1075.
- [31] I. V. Oseledets, D. V. Savostianov, and E. E. Tyrtyshnikov, "Tucker dimensionality reduction of three-dimensional arrays in linear time," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 939–956, Jan. 2008.
- [32] A. K. Saibaba, "HOLD: Higher order interpolatory decomposition for tensors based on Tucker representation," *SIAM J. Matrix Anal. Appl.*, vol. 37, no. 3, pp. 1223–1249, Jan. 2016.
- [33] C. E. Tsourakakis, "MACH: Fast randomized tensor decompositions," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2010, pp. 689–700.
- [34] Z. Song, D. P. Woodruff, and P. Zhong, "Relative error tensor low rank approximation," in *Proc. 30th Annu. ACM-SIAM Symp. Discrete Algorithms*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2019, pp. 2772–2789.
- [35] I. Perros, R. Chen, R. Vuduc, and J. Sun, "Sparse hierarchical Tucker factorization and its application to healthcare," in *Proc. IEEE Int. Conf. Data Mining*, Nov. 2015, pp. 943–948.
- [36] D. Cheng, R. Peng, Y. Liu, and I. Perros, "SPALS: Fast alternating least squares via implicit leverage scores sampling," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 721–729.
- [37] A. Traoré, M. Berar, and A. Rakotomamonjy, "Singleshot: A scalable Tucker tensor decomposition," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 6304–6315.
- [38] O. A. Malik and S. Becker, "Low-rank Tucker decomposition of large tensors using tensorsketch," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 10117–10127.
- [39] Y. Shi and A. Anandkumar, "Higher-order count sketch: Dimensionality reduction that retains efficient tensor operations," 2019, *arXiv:1901.11261*. [Online]. Available: <http://arxiv.org/abs/1901.11261>
- [40] Y. Shi, "Efficient tensor operations via compression and parallel computation," Ph.D. dissertation, UC Irvine, Irvine, CA, USA, 2019.
- [41] A. Cichocki, N. Lee, I. Oseledets, A.-H. Phan, Q. Zhao, and D. P. Mandic, "Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions," *Found. Trends Mach. Learn.*, vol. 9, nos. 4–5, pp. 249–429, 2016.
- [42] T. G. Kolda, "Multilinear operators for higher-order decompositions," Sandia Nat. Lab., Albuquerque, NM, USA, Tech. Rep. SAND2006-2081, 2006.
- [43] M. Li, W. Bi, J. T. Kwok, and B.-L. Lu, "Large-scale Nyström kernel matrix approximation using randomized SVD," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 1, pp. 152–164, Jan. 2015.
- [44] D. Ahfock, W. J. Astle, and S. Richardson, "Statistical properties of sketching algorithms," 2017, *arXiv:1706.03665*. [Online]. Available: <http://arxiv.org/abs/1706.03665>
- [45] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," *Contemp. Math.*, vol. 26, nos. 189–206, p. 1, 1984.
- [46] N. Halko, P. G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Rev.*, vol. 53, no. 2, pp. 217–288, Jan. 2011.
- [47] D. Achlioptas, "Database-friendly random projections: Johnson-Lindenstrauss with binary coins," *J. Comput. Syst. Sci.*, vol. 66, no. 4, pp. 671–687, Jun. 2003.
- [48] P. Li, T. J. Hastie, and K. W. Church, "Very sparse random projections," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2006, pp. 287–296.
- [49] F. Woolfe, E. Liberty, V. Rokhlin, and M. Tytgert, "A fast randomized algorithm for the approximation of matrices," *Appl. Comput. Harmon. Anal.*, vol. 25, no. 3, pp. 335–366, Nov. 2008.
- [50] E. Liberty, "Accelerated dense random projections," Ph.D. dissertation, Yale Univ., New Haven, CT, USA, 2009.
- [51] N. Ailon and E. Liberty, "Fast dimension reduction using Rademacher series on dual BCH codes," *Discrete Comput. Geometry*, vol. 42, no. 4, p. 615, Sep. 2008.
- [52] V. Rokhlin and M. Tytgert, "A fast randomized algorithm for overdetermined linear least-squares regression," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 36, pp. 13212–13217, Sep. 2008.
- [53] P. G. Martinsson, "Randomized methods for matrix computations," *Math. Data*, vol. 25, pp. 187–231, Feb. 2019.
- [54] J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher, "Practical sketching algorithms for low-rank matrix approximation," *SIAM J. Matrix Anal. Appl.*, vol. 38, no. 4, pp. 1454–1485, Jan. 2017.

- [55] D. P. Woodruff, "Sketching as a tool for numerical linear algebra," *Found. Trend Theor. Comput. Sci.*, vol. 10, no. 2, pp. 1–157, 2014.
- [56] C. Boutsidis, D. P. Woodruff, and P. Zhong, "Optimal principal component analysis in distributed and streaming models," in *Proc. 48th Annu. ACM Symp. Theory Comput.*, Jun. 2016, pp. 236–249.
- [57] J. Upadhyay, "The price of privacy for low-rank factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 4176–4187.
- [58] C. Bajaj, Y. Wang, and T. Wang, "SketchyCoreSVD: SketchySVD from random subsampling of the data matrix," 2019, *arXiv:1907.13634*. [Online]. Available: <http://arxiv.org/abs/1907.13634>
- [59] P.-G. Martinsson and S. Voronin, "A randomized blocked algorithm for efficiently computing rank-revealing factorizations of matrices," *SIAM J. Sci. Comput.*, vol. 38, no. 5, pp. S485–S507, Jan. 2016.
- [60] W. Yu, Y. Gu, and Y. Li, "Efficient randomized algorithms for the fixed-precision low-rank matrix approximation," *SIAM J. Matrix Anal. Appl.*, vol. 39, no. 3, pp. 1339–1359, Jan. 2018.
- [61] P. Drineas and M. W. Mahoney, "RandNLA: Randomized numerical linear algebra," *Commun. ACM*, vol. 59, no. 6, pp. 80–90, May 2016.
- [62] P. Drineas, M. W. Mahoney, and S. Muthukrishnan, "Relative-error CUR matrix decompositions," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 2, pp. 844–881, Jan. 2008.
- [63] A. Frieze, R. Kannan, and S. Vempala, "Fast Monte-Carlo algorithms for finding low-rank approximations," *J. ACM*, vol. 51, no. 6, pp. 1025–1041, Nov. 2004.
- [64] A. Frieze, R. Kannan, and S. Vempala, "Fast Monte-Carlo algorithms for finding low-rank approximations," in *Proc. 39th Annu. Symp. Found. Comput. Sci. (FOCS)*, 1998, pp. 370–378.
- [65] E. Drinea, P. Drineas, and P. Huggins, "A randomized singular value decomposition algorithm for image processing applications," in *Proc. 8th Panhellenic Conf. Informat.*, 2001, pp. 278–288.
- [66] S. Kumar, M. Mohri, and A. Talwalkar, "Sampling techniques for the Nyström method method," in *Proc. Artif. Intell. Statist.*, 2009, pp. 304–311.
- [67] S. Kumar, M. Mohri, and A. Talwalkar, "Sampling methods for the Nyström method," *J. Mach. Learn. Res.*, vol. 13, pp. 981–1006, Apr. 2012.
- [68] M. W. M. Boyd, "Randomized algorithms for matrices and data," *Found. Trends Mach. Learn.*, vol. 3, no. 2, pp. 123–224, 2010.
- [69] P. Drineas, M. Magdon-Ismail, M. W. Mahoney, and D. P. Woodruff, "Fast approximation of matrix coherence and statistical leverage," *J. Mach. Learn. Res.*, vol. 13, pp. 3475–3506, Dec. 2012.
- [70] M. Charikar, K. Chen, and M. Farach-Colton, "Finding frequent items in data streams," in *Proc. Int. Colloq. Automata, Lang., Program.* Berlin, Germany: Springer, 2002, pp. 693–703.
- [71] N. Pham and R. Pagh, "Fast and scalable polynomial kernels via explicit feature maps," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2013, pp. 239–247.
- [72] R. Pagh, "Compressed matrix multiplication," *ACM Trans. Comput. Theory*, vol. 5, no. 3, pp. 1–17, Aug. 2013.
- [73] K. L. Clarkson and D. P. Woodruff, "Low-rank approximation and regression in input sparsity time," *J. ACM*, vol. 63, no. 6, p. 54, 2017.
- [74] F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," *J. Math. Phys.*, vol. 6, nos. 1–4, pp. 164–189, Apr. 1927.
- [75] F. L. Hitchcock, "Multiple invariants and generalized rank of a P-way matrix or tensor," *J. Math. Phys.*, vol. 7, nos. 1–4, pp. 39–79, Apr. 1928.
- [76] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Aug. 2009.
- [77] A.-H. Phan, A. Cichocki, and P. Tichavský, "On fast algorithms for orthogonal tucker decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 6766–6770.
- [78] F. Camarrone and M. M. Van Hulle, "Fast multiway partial least squares regression," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 2, pp. 433–443, Feb. 2019.
- [79] T. F. Chan, "Rank revealing QR factorizations," *Linear Algebra Appl.*, vols. 88–89, pp. 67–82, Apr. 1987.
- [80] C.-T. Pan, "On the existence and computation of rank-revealing LU factorizations," *Linear Algebra Appl.*, vol. 316, nos. 1–3, pp. 199–222, Sep. 2000.
- [81] C. Xiao, C. Yang, and M. Li, "Efficient alternating least squares algorithms for truncated HOSVD of higher-order tensors," 2020, *arXiv:2004.02583*. [Online]. Available: <http://arxiv.org/abs/2004.02583>
- [82] M. Li, C. Xiao, and C. Yang, "A-tucker: Input-adaptive and matricization-free tucker decomposition for dense tensors on CPUs and GPUs," 2020, *arXiv:2010.10131*. [Online]. Available: <http://arxiv.org/abs/2010.10131>
- [83] N. Vannieuwenhoven, R. Vandebril, and K. Meerbergen, "A new truncation strategy for the higher-order singular value decomposition," *SIAM J. Sci. Comput.*, vol. 34, no. 2, pp. A1027–A1052, Jan. 2012.
- [84] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, Sep. 1936.
- [85] L. Eldén and B. Savas, "A Newton–Grassmann method for computing the best multilinear rank- (R_1, R_2, R_3) approximation of a tensor," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 2, pp. 248–271, 2009.
- [86] M. Ishteva, P.-A. Absil, S. Van Huffel, and L. De Lathauwer, "Best low multilinear rank approximation of higher-order tensors, based on the Riemannian trust-region scheme," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 1, pp. 115–135, Jan. 2011.
- [87] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank- (R_1, R_2, \dots, R_N) approximation of higher-order tensors," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1324–1342, 2000.
- [88] M. Che, Y. Wei, and H. Yan, "Randomized algorithms for the low multilinear rank approximations of tensors," 2019, *arXiv:1908.11031*. [Online]. Available: <http://arxiv.org/abs/1908.11031>
- [89] M. Che, Y. Wei, and H. Yan, "Randomized algorithms for the low multilinear rank approximations of tensors," *SIAM J. Matrix Anal. Appl.*, vol. 41, no. 2, pp. 605–639, 2020.
- [90] D. G. Chachlakis, A. Prater-Bennette, and P. P. Markopoulos, "L1-norm tucker tensor decomposition," *IEEE Access*, vol. 7, pp. 178454–178465, 2019.
- [91] Y. Sun, Y. Guo, C. Luo, J. Tropp, and M. Udell, "Low-rank tucker approximation of a tensor from streaming data," *SIAM J. Math. Data Sci.*, vol. 2, no. 4, pp. 1123–1150, Jan. 2020.
- [92] D. A. Tarzanagh and G. Michailidis, "Fast randomized algorithms for t-product based tensor operations and decompositions with applications to imaging data," *SIAM J. Imag. Sci.*, vol. 11, no. 4, pp. 2629–2664, Jan. 2018.
- [93] M. W. Mahoney, M. Maggioni, and P. Drineas, "Tensor-CUR decompositions for tensor-based data," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 957–987, Jan. 2008.
- [94] P. Drineas and M. W. Mahoney, "A randomized algorithm for a tensor-based generalization of the singular value decomposition," *Linear Algebra Appl.*, vol. 420, nos. 2–3, pp. 553–571, Jan. 2007.
- [95] F. R. Gantmacher, "The theory of matrices," Chelsea Publishing, New York, NY, USA, Tech. Rep., 1959, vol. 2.
- [96] S. A. Goreinov, E. E. Tyrtshnikov, and N. L. Zamarashkin, "A theory of pseudoskeleton approximations," *Linear Algebra Appl.*, vol. 261, nos. 1–3, pp. 1–21, Aug. 1997.
- [97] S. A. Goreinov, N. L. Zamarashkin, and E. E. Tyrtshnikov, "Pseudo-skeleton approximations by matrices of maximal volume," *Math. Notes*, vol. 62, no. 4, pp. 515–519, 1997.
- [98] S. A. Goreinov, I. V. Oseledets, D. V. Savostyanov, E. E. Tyrtshnikov, and N. L. Zamarashkin, "How to find a good submatrix," in *Matrix Methods: Theory, Algorithms and Applications: Dedicated to the Memory of Gene Golub*. Singapore: World Scientific, 2010, pp. 247–256.
- [99] S. A. Goreinov and E. E. Tyrtshnikov, "The maximal-volume concept in approximation by low-rank matrices," *Contemp. Math.*, vol. 280, pp. 47–52, 2001.
- [100] D. Savostyanov, "Polilinear approximation of matrices and integral equations," Ph.D. dissertation, INM RAS, Moscow, Russia, (in Russian), 2006.
- [101] E. Tyrtshnikov, "Incomplete cross approximation in the mosaicskeleton method," *Computing*, vol. 64, no. 4, pp. 367–380, 2000.
- [102] S. Friedland, V. Mehrmann, A. Miedlar, and M. Nkengla, "Fast low rank approximations of matrices and tensors," *Electron. J. Linear Algebra*, vol. 22, pp. 1031–1048, Jan. 2011.
- [103] S. Chaturantabut and D. C. Sorensen, "Discrete empirical interpolation for nonlinear model reduction," in *Proc. 48th IEEE Conf. Decis. Control (CDC) Held Jointly With 28th Chin. Control Conf.*, Dec. 2009, pp. 4316–4321.
- [104] D. C. Sorensen and M. Embree, "A DEIM induced CUR factorization," *SIAM J. Sci. Comput.*, vol. 38, no. 3, pp. A1454–A1482, Jan. 2016.
- [105] G. W. Stewart, "Four algorithms for the efficient computation of truncated pivoted QR approximations to a sparse matrix," *Numerische Math.*, vol. 83, no. 2, pp. 313–323, Aug. 1999.
- [106] M. Gu and S. C. Eisenstat, "Efficient algorithms for computing a strong rank-revealing QR factorization," *SIAM J. Sci. Comput.*, vol. 17, no. 4, pp. 848–869, Jul. 1996.

- [107] S. Voronin and P.-G. Martinsson, "Efficient algorithms for CUR and interpolative matrix decompositions," *Adv. Comput. Math.*, vol. 43, no. 3, pp. 495–516, 2017.
- [108] C. Battaglini, G. Ballard, and T. G. Kolda, "A practical randomized CP tensor decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 39, no. 2, pp. 876–901, Jan. 2018.
- [109] O. A. Malik and S. Becker, "Fast randomized matrix and tensor interpolative decomposition using CountSketch," 2019, *arXiv:1901.10559*. [Online]. Available: <http://arxiv.org/abs/1901.10559>
- [110] D. J. Biagioni, D. Beylkin, and G. Beylkin, "Randomized interpolative decomposition of separated representations," *J. Comput. Phys.*, vol. 281, pp. 116–134, Jan. 2015.
- [111] Y. Wang, H.-Y. Tung, A. J. Smola, and A. Anandkumar, "Fast and guaranteed tensor decomposition via sketching," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 991–999.
- [112] R. Bro, "Multi-way analysis in the food industry—models, algorithms, and applications," in *Proc. MRI, EPG EMA, ICSLP*, 2000, pp. 1–290.
- [113] I. V. Oseledets, "Tensor-train decomposition," *SIAM J. Sci. Comput.*, vol. 33, no. 5, pp. 2295–2317, Jan. 2011.
- [114] A.-H. Phan, A. Cichocki, I. Oseledets, S. A. Asl, G. Calvi, and D. Mandic, "Tensor networks for latent variable analysis: Higher order canonical polyadic decomposition," 2018, *arXiv:1809.00535*. [Online]. Available: <http://arxiv.org/abs/1809.00535>
- [115] N. B. Erichson, K. Manohar, S. L. Brunton, and J. N. Kutz, "Randomized CP tensor decomposition," 2017, *arXiv:1703.09074*. [Online]. Available: <http://arxiv.org/abs/1703.09074>
- [116] S. Oh, N. Park, S. Lee, and U. Kang, "Scalable tucker factorization for sparse tensors—algorithms and discoveries," in *Proc. IEEE 34th Int. Conf. Data Eng. (ICDE)*, Apr. 2018, pp. 1120–1131.
- [117] S. Oh, N. Park, J.-G. Jang, L. Sael, and U. Kang, "High-performance tucker factorization on heterogeneous platforms," *IEEE Trans. Parallel Distrib. Syst.*, vol. 30, no. 10, pp. 2237–2248, Oct. 2019.
- [118] J. Oh, K. Shin, E. E. Papalexakis, C. Faloutsos, and H. Yu, "S-HOT: Scalable high-order tucker decomposition," in *Proc. 10th ACM Int. Conf. Web Search Data Mining*, Feb. 2017, pp. 761–770.
- [119] J.-G. Jang and U. Kang, "D-tucker: Fast and memory-efficient tucker decomposition for dense tensors," in *Proc. IEEE 36th Int. Conf. Data Eng. (ICDE)*, Apr. 2020, pp. 1850–1853.
- [120] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—Part I: Lemmas for partitioned matrices," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1022–1032, Jan. 2008.
- [121] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—Part II: Definitions and uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1033–1066, Jan. 2008.
- [122] L. De Lathauwer and D. Nion, "Decompositions of a higher-order tensor in block terms—Part III: Alternating least squares algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1067–1083, Jan. 2008.
- [123] L. De Lathauwer, "Block component analysis, a new concept for blind source separation," in *Proc. Int. Conf. Latent Variable Anal. Signal Separat.* Berlin, Germany: Springer, 2012, pp. 1–8.
- [124] L. N. Ribeiro, A. L. F. de Almeida, and V. Zarzoso, "Enhanced block term decomposition for atrial activity extraction in atrial fibrillation ECG," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jul. 2016, pp. 1–5.
- [125] B. Hunyadi, D. Camps, L. Sorber, W. V. Paesschen, M. D. Vos, S. V. Huffel, and L. D. Lathauwer, "Block term decomposition for modelling epileptic seizures," *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 1, p. 139, Dec. 2014.
- [126] X. Wang, L. T. Yang, X. Chen, L. Wang, R. Ranjan, X. Chen, and M. J. Deen, "A multi-order distributed HOSVD with its incremental computing for big services in cyber-physical-social systems," *IEEE Trans. Big Data*, vol. 6, no. 4, pp. 666–678, Dec. 2020.
- [127] X. Wang, W. Wang, L. T. Yang, S. Liao, D. Yin, and M. J. Deen, "A distributed HOSVD method with its incremental computation for big data in cyber-physical-social systems," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 2, pp. 481–492, Jun. 2018.
- [128] X. Wang, L. T. Yang, X. Chen, M. J. Deen, and J. Jin, "Improved multi-order distributed HOSVD with its incremental computing for smart city services," *IEEE Trans. Sustain. Comput.*, early access, Nov. 15, 2018, doi: [10.1109/TSUSC.2018.2881439](https://doi.org/10.1109/TSUSC.2018.2881439).
- [129] X. Wang, L. T. Yang, H. Liu, and M. J. Deen, "A big data-as-a-service framework: State-of-the-art and perspectives," *IEEE Trans. Big Data*, vol. 4, no. 3, pp. 325–340, Sep. 2018.
- [130] X. Wang, L. T. Yang, X. Chen, J.-J. Han, and J. Feng, "A tensor computation and optimization model for cyber-physical-social big data," *IEEE Trans. Sustain. Comput.*, vol. 4, no. 4, pp. 326–339, Oct. 2019.
- [131] G. Beylkin and M. J. Mohlenkamp, "Numerical operator calculus in higher dimensions," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 16, pp. 10246–10251, Aug. 2002.
- [132] G. Beylkin and M. J. Mohlenkamp, "Algorithms for numerical analysis in high dimensions," *SIAM J. Sci. Comput.*, vol. 26, no. 6, pp. 2133–2159, Jan. 2005.
- [133] S. R. Chinnamsetty, M. Espig, B. N. Khoromskij, W. Hackbusch, and H.-J. Flad, "Tensor product approximation with optimal rank in quantum chemistry," *J. Chem. Phys.*, vol. 127, no. 8, Aug. 2007, Art. no. 084110.
- [134] L. Grasedyck, "Existence and computation of low kronecker-rank approximations for large linear systems of tensor product structure," *Computing*, vol. 72, nos. 3–4, pp. 247–265, Jun. 2004.
- [135] B. N. Khoromskij, "On tensor approximation of green iterations For Kohn-Sham equations," *Comput. Vis. Sci.*, vol. 11, nos. 4–6, pp. 259–271, 2008.
- [136] W. Hackbusch, B. N. Khoromskij, and E. E. Tyrtysnikov, "Hierarchical kronecker tensor-product approximations," *J. Numer. Math.*, vol. 13, no. 2, pp. 119–156, Jan. 2005.
- [137] E. E. Tyrtysnikov, "Tensor approximations of matrices generated by asymptotically smooth functions," *Sbornik, Math.*, vol. 194, no. 6, p. 941, 2003.
- [138] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (coil-100)," Columbia Univ., New York, NY, USA, Tech. Rep., 1996.
- [139] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. Hruschka, Jr., and T. Mitchell, "Toward an architecture for never-ending language learning," in *Proc. AAAI*, vol. 5, 2010, p. 3.
- [140] S. Smith, J. W. Choi, J. Li, R. Vuduc, J. Park, X. Liu, and G. Karypis, "FROSTT: The formidable repository of open sparse tensors and tools," Chicago Univ., Chicago, IL, USA, Tech. Rep., 2017.



algebra, tensor computations, and machine learning.



SALMAN AHMADI-ASL received the master's degree in applied mathematics from Tarbiat Modares University, Tehran, Iran, in 2012, and the Ph.D. degree in applied mathematics from the Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran, in 2016. He is currently a Research Scientist with the Center of Computational and Data-Intensive Science and Engineering, Skoltech, Moscow, Russia. His current research interests include randomized algorithms for multilinear

STANISLAV ABUKHOVICH received the master's degree in applied mathematics from Lomonosov Moscow State University, Moscow, Russia, in 2019. He is currently pursuing the Ph.D. degree with the Center of Computational and Data-Intensive Science and Engineering, Skoltech, Moscow. His current research interests include deep learning, tensor decompositions, and randomized algorithms for multilinear algebra.



MAAME G. ASANTE-MENASH (Student Member, IEEE) is currently pursuing the Ph.D. degree with the Center of Computational and Data-Intensive Science and Engineering, Skoltech, Moscow, Russia. His current research interests include deep learning, tensor completion, and randomized algorithms for multilinear algebra.



ANDRZEJ CICHOCKI (Fellow, IEEE) received the M.Sc. (Hons.), Ph.D., and Dr.Sc. (habilitation) degrees in electrical engineering from the Warsaw University of Technology, Warsaw, Poland.

He spent several years at the University of Erlangen-Nurnberg, Erlangen, Germany, as an Alexander-von-Humboldt Research Fellow and a Guest Professor. From 1995 to 2018, he was a Team Leader and the Head of the Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Wako, Japan. He is currently a Professor with the Skolkovo Institute of Science and Technology (Skoltech), Moscow, Russia, and a Visiting/Adjunct Professor with the Tokyo University of Agriculture and Technology (TUAT), Tokyo, Japan; Hangzhou Dianzi University (HDU), Hangzhou, China; Nicolaus Copernicus University (UMK), Torun, Poland; and the Institute of Systems Research (IBS), Polish Academy of Science, Warsaw. He has authored more than 500 peer-reviewed papers and six monographs in English (two of them translated to Chinese). His publications currently report over 37 000 citations according to Google Scholar, with an h-index of 84. He is currently among three most cited Polish computer scientists. His current research interests include deep learning, tensor decompositions, tensor networks for big data analytics, multiway blind source separation, and brain-computer interface and their biomedical applications. He serves/served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNALS PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, and *Journal of Neuroscience Methods*. He was the founding Editor-in-Chief of the *Journal Computational Intelligence and Neuroscience*.



ANH-HUY PHAN (Member, IEEE) received the master's degree from the Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, in 2005, and the Ph.D. degree from the Kyushu Institute of Technology, Kitakyushu, Japan, in 2011. From October 2011 to March 2018, he was a Research Scientist with the Laboratory for Advanced Brain Signal Processing, Brain Science Institute (BSI), RIKEN, Tokyo, Japan, where he was a Visiting Research Scientist with the Toyota Collaboration Center, from April 2012 to April 2015. Since May 2018, he has been an Assistant Professor with the Center for Computational and Data-Intensive Science and Engineering, Skoltech, Moscow, Russia. He is also a Visiting Associate Professor with the Tokyo University of Agriculture and Technology (TUAT), Tokyo. He has authored three monographs. His research interests include multilinear algebra, tensor computation, tensor networks, nonlinear system, blind source separation, and brain-computer interface. He received best paper awards for articles in the *IEEE Signal Processing Magazine* in 2018 and in the International Conference on Neural Information Processing in 2016, and the Outstanding Reviewer Award for maintaining the prestige of the IEEE International Conference on Acoustics, Speech and Signal Processing 2019.



TOSHIHISA TANAKA (Senior Member, IEEE) received the B.E., M.E., and Ph.D. degrees from the Tokyo Institute of Technology, Meguro, Japan, in 1997, 2000, and 2002, respectively.

From October 2002 to March 2004, he was a Research Scientist at the RIKEN Brain Science Institute. In April 2004, he joined the Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, where he is currently an Associate Professor. In 2005, he was a Royal Society Visiting Fellow at the Communications and Signal Processing Group, Imperial College London, U.K. From June 2011 to October 2011, he was a Visiting Faculty Member with the Department of Electrical Engineering, University of Hawaii at Manoa. His research interests include a broad area of signal processing and machine learning, including brain and biomedical signal processing, brain-machine interfaces, and adaptive systems. From 2000 to 2002, he was a JSPS Research Fellow. He is a Co-Editor of *Signal Processing Techniques for Knowledge Extraction and Information Fusion* (with Mandic, New York, NY, USA: Springer, 2008).

Dr. Tanaka is a member of the IEICE, APSIPA, and Society for Neurosciences. Furthermore, he serves as a Member-at-Large and Board of Governors of Asia-Pacific Signal and Information Processing Association (APSIPA). He was the Chair of the Technical Committee on Biomedical Signal Processing, APSIPA. He has served as an Associate Editor and a Guest Editor of special issues in journals, including *Neurocomputing* and *IEICE Transactions on Fundamentals*. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, *Computational Intelligence and Neuroscience* (Hindawi), and *Advances in Data Science and Adaptive Analysis* (World Scientific).



IVAN OSELEDETS graduated from the Moscow Institute of Physics and Technology, Dolgoprudny, Russia, in 2006, and received the Ph.D. and D.Sc. degrees from the Institute of Numerical Mathematics of Russian Academy of Sciences (INM RAS) named under G. Marchuk, Moscow, Russia, in 2007 and 2012, respectively. He is currently an Associate Professor with Skoltech, Moscow, and a Leading Researcher with INM RAS. He has coauthored more than 100 papers. His current research

interests include linear algebra, tensor methods, machine learning, and deep learning. He is a recipient of several awards, including the SIAM Outstanding Paper Prize in 2018 and the Russian President Award for Science and Innovation for young scientists.

...