Journal of Machine Learning Research 22 (2021) 1-36 Submitted 3/21; Published 6/21

# Mode-wise Tensor Decompositions: Multi-dimensional Generalizations of CUR Decompositions

**HanQin Cai** HQCAI@MATH.UCLA.EDU
*Department of Mathematics*
*University of California, Los Angeles*
*Los Angeles, CA 90095, USA*

**Keaton Hamm** KEATON.HAMM@UTA.EDU
*Department of Mathematics*
*University of Texas at Arlington*
*Arlington, TX 76019, USA*

**Longxiu Huang** HUANGL3@MATH.UCLA.EDU
*Department of Mathematics*
*University of California, Los Angeles*
*Los Angeles, CA 90095, USA*

**Deanna Needell** DEANNA@MATH.UCLA.EDU
*Department of Mathematics*
*University of California, Los Angeles*
*Los Angeles, CA 90095, USA*

**Editor:** Michael Mahoney

## Abstract

Low rank tensor approximation is a fundamental tool in modern machine learning and data science. In this paper, we study the characterization, perturbation analysis, and an efficient sampling strategy for two primary tensor CUR approximations, namely Chidori and Fiber CUR. We characterize exact tensor CUR decompositions for low multilinear rank tensors. We also present theoretical error bounds of the tensor CUR approximations when (adversarial or Gaussian) noise appears. Moreover, we show that low cost uniform sampling is sufficient for tensor CUR approximations if the tensor has an incoherent structure. Empirical performance evaluations, with both synthetic and real-world datasets, establish the speed advantage of the tensor CUR approximations over other state-of-the-art low multilinear rank tensor approximations.

**Keywords:** tensor decomposition, low-rank tensor approximation, CUR decomposition, randomized linear algebra, hyperspectral image compression

## 1. Introduction

A tensor is a multi-dimensional array of numbers, and is the higher-order generalization of vectors and matrices; thus tensors can express more complex intrinsic structures of higher-order data. In various data-rich domains such as computer vision, recommendation systems, medical imaging, data mining, and multi-class learning consisting of multi-modal and multi-relational data, tensors have emerged as a powerful paradigm for managing the data deluge.

Indeed, data is often more naturally represented as a tensor than a vector or matrix; for instance hyperspectral images result in 3–mode tensors, and color videos can be represented as 4–mode tensors. The tensor structure of such data can carry with it more information; for instance, spatial information is kept along the spectral or time slices in these applications. Thus, tensors are not merely a convenient tool for analysis, but are rather a fundamental structure exhibited by data in many domains.

As in the matrix setting, but to an even greater extent, an important tool for handling tensor data efficiently is tensor decomposition, in which the tensor data is represented by a few succinctly representable components. For example, tensor decompositions are used in computer vision (Vasilescu and Terzopoulos, 2002; Yan et al., 2006) to enable the extraction of patterns that generalize well across common modes of variation, whereas in bioinformatics (Yener et al., 2008; Omberg et al., 2009), tensor decomposition has proven useful for the understanding of cellular states and biological processes. Similar to matrices, tensor decompositions can be utilized for compression and low-rank approximation (Mahoney et al., 2008; Zhang et al., 2015; Du et al., 2016; Li et al., 2021). However, there are a greater variety of natural low-rank tensor decompositions than for matrices, in part because the notion of rank for higher order tensors is not unique.

Tensor decompositions have been widely studied both in theory and application for some time (e.g., (Hitchcock, 1927; Kolda and Bader, 2009; Zare et al., 2018)). Examples of different tensor decompositions include the CANDECOMP/PARAFAC (CP) decomposition (Hitchcock, 1927), Tucker decomposition (Tucker, 1966), Hierarchical-Tucker (H-Tucker) decomposition (Grasedyck, 2010), Tensor Train (TT) decomposition (Oseledets, 2011), and Tensor Singular Value Decomposition (t-SVD)(Kilmer et al., 2013). However, most current tensor decomposition schemes require one to first unfold the tensor along a given mode into a matrix, implement a matrix decomposition method, and then relate the result to a tensor form via mode multiplication. Such algorithms are matricial in nature, and fail to properly utilize the tensor structure of the data. Moreover, tensors are often vast in size, so matrix-based algorithms on tensors often have severe computational complexity, whereas algorithms that are fundamentally tensorial in nature will have drastically lower complexity. Let us be concrete here. Suppose a tensor has order $n$ and each dimension is $d$; most of the algorithms to compute the CP decomposition are iterative. For example, the ALS algorithm with line search has a complexity of order $\mathcal{O}(2^n r d^n + n r^3)$ (Phan et al., 2013) where $r$ is the CP–rank. Similarly, if the multilinear rank of the tensor is $(r, \ldots, r)$, computing the HOSVD by computing the compact SVD of the matrices obtained by unfolding the tensor would result in complexity of order $\mathcal{O}(r d^n)$. To accelerate both CP and Tucker decomposition computations, many works have applied different randomized techniques. For instance, there are several sketching algorithms for CP decompositions (Battaglino et al., 2018; Song et al., 2019; Erichson et al., 2020; Gittens et al., 2020; Cheng et al., 2016) based on the sketching techniques for low-rank matrix approximation (Woodruff, 2014), and these techniques have been shown to greatly improve computational efficiency compared to the original ALS algorithm. For Tucker decomposition, several randomized algorithms (Ahmadi-Asl et al., 2021; Che et al., 2021) based on random projection have been developed to accelerate HOSVD and HOOI.

The purpose of this work is to explore tensor-based methods for low-rank tensor decompositions. In particular, we present two flexible tensor decompositions inspired by matrix

CUR decompositions, which utilize small core subtensors to reconstruct the whole. Moreover, our algorithms for forming tensor CUR decompositions do not require unfolding the tensor into a matrix. The uniform sampling based tensor decompositions we discuss subsequently have computational complexity $\mathcal{O}(r^2 n \log^2(d))$, which is a dramatic improvement over those discussed above.

Matrix CUR decompositions (Hamm and Huang, 2020b), which are sometimes called pseudoskeleton decompositions (Goreĭnov et al., 1995; Chiu and Demanet, 2013), utilize the fact that any matrix $X \in \mathbb{R}^{m \times n}$ with $\operatorname{rank}(X) = r$ can be perfectly reconstructed via $X = CU^{\dagger}R$ by choosing submatrices $C = X(:, J)$, $R = X(I, :)$, and $U = X(I, J)$ such that $\operatorname{rank}(U) = r$. Consequently, CUR decompositions only require some of the entries of $X$ governed by the selected column and row submatrices to recover the whole. This observation makes CUR decompositions extremely attractive for large-scale low-rank matrix approximation problems.

It is natural to consider extending the CUR decomposition to tensors in a way that fundamentally utilizes the tensor structure. We will see subsequently that there is no single canonical extension of matrix CUR decompositions to the tensor setting; however, we propose that a *natural* tensor CUR decomposition must be one that selects subsets of each mode, and which is not based on unfolding operations on the tensor.

### 1.1 Contributions

In this paper, our main contributions can be summarized as follows:

1. We first provide some new characterizations for CUR decompositions for tensors and show by example that the characterization for tensors is different from that for matrices. In particular, we show that exact CUR decompositions of low multilinear rank tensors are equivalent to the multilinear rank of the core subtensor chosen being the same as that of the data tensor.

2. Real data tensors rarely exhibit exactly low-rank structure, but they can be modeled as a low multilinear rank tensor plus noise. We undertake a novel perturbation analysis for low multilinear rank tensor CUR approximations, and prove error bounds for the two primary tensor CUR decompositions discussed here. These bounds are qualitative and are given in terms of submatrices of singular vectors of unfoldings of the tensor, and represent the first approximation bounds for these decompositions. We additionally provide some specialized bounds for the case of low multilinear rank tensors perturbed by random Gaussian noise. Our methods and analysis affirmatively answer a question of Mahoney et al. (2008) regarding finding tensor CUR decompositions that preserve multilinear structure of the tensor.

3. When the underlying low multilinear rank tensor has an incoherence property and is perturbed by noise, we show that uniformly sampling indices along each mode yields a low multilinear rank approximation to the tensor with good error bounds.

4. We give guarantees on random sampling procedures of the indices for each mode that ensure that an exact CUR decomposition of a low multilinear rank tensor holds with high probability.

5. We illustrate the effectiveness of the various decompositions proposed here on synthetic tensor datasets and real hyperspectral imaging datasets.

### 1.2 Organization

The rest of the paper is laid out as follows. Section 1.3 contains a comparison of our work with prior art in both the low-rank tensor approximation and tensor CUR literature. Section 2 contains the notations and descriptions of low multilinear rank tensor decompositions necessary for the subsequent analysis. Section 3 contains the statements of the main results of the paper, with Section 3.1 containing the characterization theorems for exact tensor CUR decompositions, Section 3.2 containing the statements of the main approximation bounds for CUR approximations of arbitrary tensors, and Section 3.3 giving error bounds for CUR decompositions obtained via randomly sampling indices. Section 4 contains the proofs of the characterization theorems (Theorems 2 and 3), Section 5 contains proofs of the approximation bounds, and Section 6 contains proofs related to random sampling of core subtensors to either achieve an exact decomposition of low multilinear rank tensors or achieve good approximation of arbitrary tensors. Experiments on synthetic and real hyperspectral data are contained in Section 7, and the paper ends with some concluding remarks and future directions in Section 8.

### 1.3 Prior Art

CUR decompositions for matrices have a long history; a similar matrix form expressed via Schur decompositions goes back at least to Penrose (1956); more recently, they were studied as pseudoskeleton decompositions (Goreĭnov et al., 1995, 1997). Within the randomized linear algebra, theoretical computer science, and machine learning literature, they have been studied beginning with the work of Drineas, Kannan, and Mahoney (Drineas et al., 2006, 2008; Mahoney and Drineas, 2009; Chiu and Demanet, 2013; Sorensen and Embree, 2016; Voronin and Martinsson, 2017). For a more thorough historical discussion, see (Hamm and Huang, 2020b).

The first extension of CUR decompositions to tensors (Mahoney et al., 2008) involved a single-mode unfolding of 3–mode tensors. Accuracy of the tensor-CUR decomposition was transferred from existing guarantees for matrix CUR decompositions, and gave additive error guarantees (containing a factor of $\varepsilon\|\mathcal{A}\|_F^2$). Later, Caiafa and Cichocki (2010) proposed a different variant of tensor CUR that accounts for all modes, which they termed Types 1 and 2 Fiber Sampling Tensor Decompositions. In this work, we dub these with more descriptive monikers Fiber and Chidori CUR decompositions (Section 3). The decompositions we discuss later are generalizations of those of Caiafa and Cichoki, as their work considers tensors with multilinear rank $(r, \dots, r)$ and identically sized index sets ($I_i$ described in Section 3).

In (Hamm and Huang, 2020b), there are several equivalent characterizations for CUR decompositions in the matrix case. We find that there are characterizations for CUR decompositions in the tensor case (Theorems 2 and 3 here). Interestingly, the tensor characterization has some significant differences from that of matrices (see Example 1).

There are very few approximation bounds for any tensor CUR decompositions. Mahoney et al. (2008) provide some basic additive error approximation bounds for a tensor CUR

decomposition made from only taking slices along a single mode. In contrast, the bounds given in this paper (Theorems 5, 7, and 10) are the first general bounds for both Fiber and Chidori CUR decompositions which subsample along all modes. Our analysis positively answers the question of Mahoney et al. by giving sampling methods which are able to truly account for the multilinear structure of the data tensor in a nontrivial way.

A tensor CUR decomposition based on the t-SVD was explored by Wang et al. (2017), and the authors give relative error guarantees for a specified oversampling rate along each mode. Their decomposition is fundamentally different from ours as it uses t-SVD techniques which use block circulant matrices related to the original data tensor. Consequently, their decomposition is much more costly than those described here. Additionally, their factor tensors are typically larger than the constituent parts of the decompositions described here.

For matrices, uniform random sampling is known to provide good CUR approximations under incoherence assumptions, e.g., (Chiu and Demanet, 2013). Theorem 7 extends this analysis to tensors of arbitrary number of modes. Additionally, there are several standard randomized sampling procedures known to give exact matrix CUR decompositions with high probability (Hamm and Huang, 2020a); Theorem 13 provides a sample extension of these results for tensors.

The next section contains further comparison of tensor CUR decompositions with other standard low-rank tensor decompositions such as the HOSVD.

## 2. Preliminaries and Notation

Tensors, matrices, vectors, and scalars are denoted in different typeface for clarity below. In the sequel, calligraphic capital letters are used for tensors, capital letters are used for matrices, lower boldface letters for vectors, and regular letters for scalars. The set of the first $d$ natural numbers is denoted by $[d] := \{1, \cdots, d\}$. We include here some basic notions relating to tensors, and refer the reader to, e.g., (Kolda and Bader, 2009) for a more thorough introduction.

A tensor is a multidimensional array whose dimension is called the *order* (or also *mode*). The space of real tensors of order $n$ and size $d_1 \times \cdots \times d_n$ is denoted as $\mathbb{R}^{d_1 \times \cdots \times d_n}$. The elements of a tensor $\mathcal{X} \in \mathbb{R}^{d_1 \times \cdots \times d_n}$ are denoted by $\mathcal{X}_{i_1, \cdots, i_n}$.

An $n$–mode tensor $\mathcal{X}$ can be matricized, or reshaped into a matrix, in $n$ ways by unfolding it along each of the $n$ modes. The mode-$k$ matricization/unfolding of tensor $\mathcal{X} \in \mathbb{R}^{d_1 \times \cdots \times d_n}$ is the matrix denoted by $\mathcal{X}_{(k)} \in \mathbb{R}^{d_k \times \prod_{j \neq k} d_j}$ whose columns are composed of all the vectors obtained from $\mathcal{X}$ by fixing all indices except for the $k$-th dimension. The mapping $\mathcal{X} \mapsto \mathcal{X}_{(k)}$ is called the mode-$k$ unfolding operator.

Given $\mathcal{X} \in \mathbb{R}^{d_1 \times \cdots \times d_n}$, the norm $\|\mathcal{X}\|_F$ is defined via

$$\|\mathcal{X}\|_F = \left( \sum_{i_1, \cdots, i_n} \mathcal{X}^2_{i_1, \cdots, i_n} \right)^{\frac{1}{2}}.$$

There are various product operations related to tensors; the ones that will be utilized in this paper are the following.

- Outer product: Let $\mathbf{a_1} \in \mathbb{R}^{d_1}, \cdots, \mathbf{a_n} \in \mathbb{R}^{d_n}$. The outer product among these $n$ vectors is a tensor $\mathcal{X} \in \mathbb{R}^{d_1 \times \cdots \times d_n}$ defined as:

$$\mathcal{X} = \mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_n, \quad \mathcal{X}_{i_1,\cdots,i_n} = \prod_{k=1}^{n} \mathbf{a}_k(i_k).$$

  The tensor $\mathcal{X} \in \mathbb{R}^{d_1 \times \cdots \times d_n}$ is of rank one if it can be written as the outer product of $n$ vectors.

- Kronecker product of matrices: The Kronecker product of $A \in \mathbb{R}^{I \times J}$ and $B \in \mathbb{R}^{K \times L}$ is denoted by $A \otimes B$. The result is a matrix of size $(KI) \times (JL)$ defined by

$$A \otimes B = \begin{bmatrix} A_{11}B & A_{12}B & \cdots & A_{1J}B \\ A_{21}B & A_{22}B & \cdots & A_{2J}B \\ \vdots & \vdots & \ddots & \vdots \\ A_{I1}B & A_{I2}B & \cdots & A_{IJ}B \end{bmatrix}.$$

- Mode-$k$ product: Let $\mathcal{X} \in \mathbb{R}^{d_1 \times \cdots \times d_n}$ and $A \in \mathbb{R}^{J \times d_k}$, the multiplication between $\mathcal{X}$ on its $k$-th mode with $A$ is denoted as $\mathcal{Y} = \mathcal{X} \times_k A$ with

$$\mathcal{Y}_{i_1,\cdots,i_{k-1},j,i_{k+1},\cdots,i_n} = \sum_{s=1}^{d_k} \mathcal{X}_{i_1,\cdots,i_{k-1},s,i_{k+1},\cdots,i_n} A_{j,s}.$$

  Note this can be written as a matrix product by noting that $\mathcal{Y}_{(k)} = A\mathcal{X}_{(k)}$. If we have multiple tensor matrix product from different modes, we use the notation $\mathcal{X} \times_{i=t}^{s} A_i$ to denote the product $\mathcal{X} \times_t A_t \times_{t+1} \cdots \times_s A_s$.

For the reader's convenience, we also summarize the notation in Table 1.

### 2.1 Tensor Rank

The notion of *rank* for tensors is more complicated than it is for matrices. Indeed, rank is non-uniquely defined. In this work, we will primarily utilize the *multilinear rank* (also called *Tucker rank*) of tensors (Hitchcock, 1928). The multilinear rank of $\mathcal{X}$ is a tuple $\mathbf{r} = (r_1, \cdots, r_n) \in \mathbb{N}^n$, where $r_k = \text{rank}(\mathcal{X}_{(k)})$.

Multilinear rank is relatively easily computed, but is not necessarily the most natural notion of rank. Indeed, the CP rank of a tensor (Hitchcock, 1927, 1928) is the smallest integer $r$ for which $\mathcal{X}$ can be written as the sum or rank-1 tensors. That is, we may write

$$\mathcal{X} = \sum_{i=1}^{r} \lambda_i \, \mathbf{a}_1^{(i)} \otimes \cdots \otimes \mathbf{a}_n^{(i)} \tag{1}$$

for some $\{\lambda_i\} \subseteq \mathbb{R}$ and $\mathbf{a}_k^{(i)} \in \mathbb{R}^{d_k}$. Regard that the notion of rank-1 tensors is unambiguous.

Table 1: Table of Notation.

| Notation | Definition |
|---|---|
| $\mathcal{X}$ | tensor |
| $X$ | matrix |
| $\mathbf{x}$ | vector |
| $x$ | scalar |
| $\mathcal{X} \times_k X$ | tensor $\mathcal{X}$ times matrix $X$ along the $k$-th mode |
| $X \otimes Y$ | Kronecker product of the matrices $X$ and $Y$ |
| $\mathbf{x} \otimes \mathbf{y}$ | outer product of vectors $\mathbf{x}$ and $\mathbf{y}$ |
| $\mathcal{X}_{(k)}$ | mode-$k$ unfolding of $\mathcal{X}$ |
| $X(I, :)$ | row submatrix of $X$ with row indices $I$ |
| $X(:, J)$ | column submatrix of $X$ with column indices $J$ |
| $\mathcal{X}(I_1, \cdots, I_n)$ | subtensor of $\mathcal{X}$ with indices $I_k$ at mode $k$ |
| $[d] = \{1, \cdots, d\}$ | the set of the first $d$ natural numbers |
| $\mathbf{r} = (r_1, \cdots, r_n)$ | multilinear rank |
| $\|\cdot\|_2$ | $\ell_2$ norm for vector, spectral norm for matrix |
| $\|\cdot\|_F$ | Frobenius norm |
| $(\cdot)^\top$ | transpose |
| $(\cdot)^\dagger$ | Moore–Penrose pseudoinverse |

### 2.2 Tensor Decompositions

Tensor decompositions are powerful tools for extracting meaningful, latent structures in heterogeneous, multidimensional data (see, e.g., (Kolda and Bader, 2009)). Similar to the matrix case, there are a wide array of tensor decompositions, and one may select a different one based on the task at hand. For instance, CP decompositions (of the form (1)) are typically the most compact representation of a tensor, but the substantial drawback is that they are NP–hard to compute (Kruskal, 1989). On the other hand, Tucker decompositions and Higher-order Singular Value Decompositions (HOSVD) are natural extensions of the matrix SVD, and thus useful for describing features of the data. Matrix CUR decompositions give factorizations in terms of actual column and row submatrices, and can be cheap to form by random sampling. CUR decompositions are known to provide interpretable representations of data in contrast to the SVD, for example (Mahoney and Drineas, 2009). Similarly, tensor analogues of CUR decompositions represent a tensor via subtubes and fibers of it. We will discuss this further in the following subsection.

The Tucker decomposition was proposed by Tucker (1966) and further developed in (Kroonenberg and De Leeuw, 1980; De Lathauwer et al., 2000a). A special case of Tucker decompositions is called the Higher-order SVD (HOSVD): given an $n$-order tensor $\mathcal{X}$, its HOSVD is defined as the modewise product of a core tensor $\mathcal{T} \in \mathbb{R}^{r_1 \times \cdots \times r_n}$ with $n$ factor matrices $W_k \in \mathbb{R}^{d_k \times r_k}$ (whose columns are orthonormal) along each mode such that

$$\mathcal{X} = \mathcal{T} \times_1 W_1 \times_2 \cdots \times_n W_n =: [\![\mathcal{T}; W_1, \cdots, W_n]\!],$$

where $r_k = \operatorname{rank}(\mathcal{X}_{(k)})$. If we unfold $\mathcal{X}$ along its $k$-th mode, we have

$$\mathcal{X}_{(k)} = W_k \mathcal{T}_{(k)} (W_1 \otimes \cdots \otimes W_{k-1} \otimes W_{k+1} \otimes \cdots \otimes W_n)^\top.$$

The HOSVD can be computed as follows:

1. Unfold $\mathcal{X}$ along mode $k$ to get matrix $\mathcal{X}_{(k)}$;
2. Compute the compact SVD of $\mathcal{X}_{(k)} = W_k \Sigma_k V_k^\top$;
3. $\mathcal{T} = \mathcal{X} \times_1 W_1^\top \times_2 \cdots \times_n W_n^\top$.

For a more comprehensive introduction to tensor decompositions, readers are referred to (Acar and Yener, 2008; Kolda and Bader, 2009; Sidiropoulos et al., 2017).

In the statements below, $a \gtrsim b$ means that $a \geq cb$ for some absolute constant $c > 0$.

### 2.3 Matrix CUR Decompositions

To better compare the matrix and tensor case, we first discuss the characterization of CUR decompositions for matrices obtained in (Hamm and Huang, 2020b).

**Theorem 1 ((Hamm and Huang, 2020b))** *Let $A \in \mathbb{R}^{m \times n}$ and $I \subseteq [m]$, $J \subseteq [n]$. Let $C = A(:, J)$, $U = A(I, J)$, and $R = A(I, :)$. Then the following are equivalent:*

*(i)* $\operatorname{rank}(U) = \operatorname{rank}(A)$,

*(ii)* $A = CU^\dagger R$,

*(iii)* $A = CC^\dagger AR^\dagger R$,

*(iv)* $A^\dagger = R^\dagger U C^\dagger$,

*(v)* $\operatorname{rank}(C) = \operatorname{rank}(R) = \operatorname{rank}(A)$,

*(vi) Suppose columns and rows are rearranged so that $A = \begin{bmatrix} U & B \\ D & E \end{bmatrix}$, and the generalized Schur complement of $A$ with respect to $U$ is defined by $A/U := E - DU^\dagger B$. Then $A/U = 0$.*

*Moreover, if any of the equivalent conditions above hold, then $U^\dagger = C^\dagger A R^\dagger$.*

We note that equivalent condition $(vi)$ in Theorem 1 is not proven in (Hamm and Huang, 2020b), but can readily be deduced using basic methods and Corollary 19.1 of (Matsaglia and PH Styan, 1974).

Notice that if the matrix is treated as a two-way tensor, the matrix CUR decomposition as in $(ii)$ can be written in the form:

$$A = CU^\dagger R = U \times_1 CU^\dagger \times_2 R^\top (U^\top)^\dagger.$$

CUR decompositions provide a representation of data in terms of other data, hence allowing for more ease of interpretation of results. They have been used to practical effect in exploratory data analysis related to natural language processing (Mahoney and Drineas, 2009), subspace clustering (Aldroubi et al., 2018, 2019), and basic science (Yip et al., 2014; Yang et al., 2015). Additionally, CUR is often used as a fast approximation to the SVD

(Drineas et al., 2006, 2008; Boutsidis and Woodruff, 2017; Voronin and Martinsson, 2017) (see also the more general survey of randomized methods (Halko et al., 2011)). Recently, CUR decompositions have been used to accelerate algorithms for Robust PCA Cai et al. (2021a,b). See (Hamm and Huang, 2020b) for a more thorough survey of CUR decompositions and their utility.

In Machine Learning, the Nyström method (CUR decompositions in which the same columns and rows are selected to approximate symmetric positive semi-definite matrices) is heavily utilized to estimate large kernel matrices (Williams and Seeger, 2001; Gittens and Mahoney, 2016; Bertozzi and Flenner, 2016).

For multidimensional data, (Mahoney et al., 2008) proposed tensor-CUR decompositions which only takes advantage of linear but not multilinear structure in data tensors. The authors therein raise the following open problem.

**Question 1 ((Mahoney et al., 2008, Remark 3.3))** *Can one choose slabs and/or fibers to preserve some nontrivial multilinear tensor structure in the original tensor?*

We address this question in this work, and show that the answer is affirmative. Our methods are inspired by matrix CUR techniques, but are determined to account for the tensorial nature of data beyond simple unfolding based methods.

## 3. Main Results

Our main results are broken into three main themes: characterizations of modewise decompositions of tensors inspired by Theorem 1, perturbation analysis for the variants of tensor decompositions proposed here, and upper bounds for low multilinear rank tensor approximations via randomized sampling as well as randomized sampling guarantees for exact reconstruction of low multilinear rank tensors.

Now let us be concrete as to the type of tensor CUR decompositions we will consider here. The first is the most direct analogue of the matrix CUR decomposition in which a core subtensor is selected which we call $\mathcal{R}$, and the other factors are chosen by extruding $\mathcal{R}$ along each mode to produce various subtensors. That is, given indices $I_i \subseteq [d_i]$, we set $\mathcal{R} = \mathcal{A}(I_1, \ldots, I_n)$, $C_i = \mathcal{A}_{(i)}(:, \otimes_{j\neq i} I_j) := (\mathcal{A}(I_1, \cdots, I_{i-1}, :, I_{i+1}, \cdots, I_n))_{(i)}$, and $U_i = C_i(I_i, :)$. This decomposition is illustrated in Figure 1, and we call it the *Chidori CUR decomposition*[1].

The second, more general tensor CUR decomposition discussed here, we call the Fiber CUR decomposition. In this case, to form $C_i$, one is allowed to choose $J_i \subseteq [\prod_{j\neq i} d_j]$ without reference to $I_i$. Thus, the $C_i$ are formed from mode-$i$ fibers which may or may not interact with the core subtensor $\mathcal{R}$. Fiber CUR decompositions are illustrated in Figure 2.

### 3.1 Characterization Theorems

First, we characterize Chidori CUR decompositions and compare with the matrix CUR decomposition of Theorem 1.

1. Chidori joints are used in contemporary Japanese woodworking and are based on an old toy design. The joints bear a striking resemblance to Figure 1.

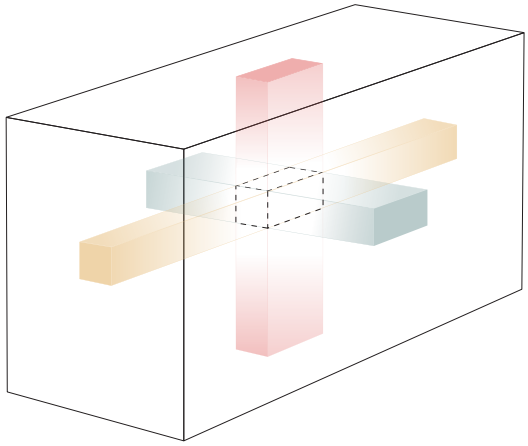

Figure 1: Illustration of Chidori CUR decomposition à la Theorem 2 of a 3-mode tensor in the case when the indices $I_i$ are each an interval and $J_i = \otimes_{j \neq i} I_j$. The matrix $C_1$ is obtained by unfolding the red subtensor along mode 1, $C_2$ by unfolding the green subtensor along mode 2, and $C_3$ by unfolding the yellow subtensor along mode 3. The dotted line shows the boundaries of $\mathcal{R}$. In this case $U_i = \mathcal{R}_{(i)}$ for all $i$.

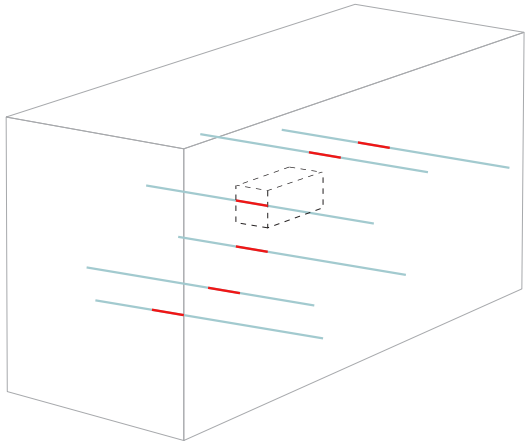

Figure 2: Illustration of the Fiber CUR Decomposition of Theorem 3 in which $J_i$ is not necessarily related to $I_i$. The lines correspond to rows of $C_2$, and red indices within correspond to rows of $U_2$. Note that the lines may (but do not have to) pass through the core subtensor $\mathcal{R}$ outlined by dotted lines. Fibers used to form $C_1$ and $C_3$ are not shown for clarity.

**Theorem 2** *Let $\mathcal{A} \in \mathbb{R}^{d_1 \times \cdots \times d_n}$ with multilinear rank $(r_1, \ldots, r_n)$. Let $I_i \subseteq [d_i]$. Set $\mathcal{R} = \mathcal{A}(I_1, \cdots, I_n)$, $C_i = \mathcal{A}_{(i)}(:, \otimes_{j \neq i} I_j)$, and $U_i = C_i(I_i, :)$. Then the following are equivalent:*

*(i)* $\operatorname{rank}(U_i) = r_i$,

*(ii)* $\mathcal{A} = \mathcal{R} \times_1 (C_1 U_1^\dagger) \times_2 \cdots \times_n (C_n U_n^\dagger)$,

*(iii) the multilinear rank of $\mathcal{R}$ is $(r_1, \cdots, r_n)$,*

*(iv)* $\operatorname{rank}(\mathcal{A}_{(i)}(I_i,:)) = r_i$ *for all* $i = 1, \cdots, n$.

*Moreover, if any of the equivalent statements above hold, then* $\mathcal{A} = \mathcal{A} \times_{i=1}^{n} (C_i C_i^{\dagger})$.

Note that for the indices chosen in Theorem 2, we have $U_i = \mathcal{R}_{(i)}$. Interestingly enough, unlike the matrix case, the projection based decomposition $\mathcal{A} = \mathcal{A} \times_{i=1}^{n} (C_i C_i^{\dagger})$ is not equivalent to the other parts of Theorem 2 as is shown in the following example.

**Example 1** *Let* $\mathcal{A} \in \mathbb{R}^{3\times 3\times 2}$ *with the following frontal slices:*

$$A_1 = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 3 & 8 & 5 \end{bmatrix}, \; A_2 = \begin{bmatrix} 2 & 5 & 3 \\ 4 & 10 & 6 \\ 3 & 7 & 4 \end{bmatrix}.$$

*Then the multilinear rank of* $\mathcal{A}$ *is* $(2,2,2)$. *Set* $I_1 = \{1,2\}$, $I_2 = \{1,2\}$ *and* $I_3 = \{1,2\}$. *Then the frontal slices of* $\mathcal{R}$ *are*

$$R_1 = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \; R_2 = \begin{bmatrix} 2 & 5 \\ 4 & 10 \end{bmatrix}.$$

*The multilinear rank of* $\mathcal{R}$ *is* $(1,2,2)$. *Thus the Chidori CUR decomposition does not hold, i.e.,* $\mathcal{A} \neq \mathcal{R} \times_1 (C_1 U_1^{\dagger}) \times_2 (C_2 U_2^{\dagger}) \times_3 (C_3 U_3^{\dagger})$. *However,* $\operatorname{rank}(C_i) = 2$ *for* $i = 1,2,3$ *which implies that* $\mathcal{A} = \mathcal{A} \times_{i=1}^{3} (C_i C_i^{\dagger})$.

Next we find that, in contrast to the matrix case, the indices $\{I_i\}$ and $\{J_i\}$ do not necessarily have to be correlated (see Figure 2). In the case that the indices $J_i$ are independent from $I_i$, we have the following characterization of the Fiber CUR decomposition.

**Theorem 3** *Let* $\mathcal{A} \in \mathbb{R}^{d_1 \times \cdots \times d_n}$ *with multilinear rank* $(r_1, \ldots, r_n)$. *Let* $I_i \subseteq [d_i]$ *and* $J_i \subseteq [\prod_{j \neq i} d_j]$. *Set* $\mathcal{R} = \mathcal{A}(I_1, \cdots, I_n)$, $C_i = \mathcal{A}_{(i)}(:, J_i)$ *and* $U_i = C_i(I_i, :)$. *Then the following statements are equivalent*

*(i)* $\operatorname{rank}(U_i) = r_i$,

*(ii)* $\mathcal{A} = \mathcal{R} \times_1 (C_1 U_1^{\dagger}) \times_2 \cdots \times_n (C_n U_n^{\dagger})$,

*(iii)* $\operatorname{rank}(C_i) = r_i$ *for all* $i$ *and the multilinear rank of* $\mathcal{R} = (r_1, \cdots, r_n)$.

*Moreover, if any of the equivalent statements above hold, then* $\mathcal{A} = \mathcal{A} \times_{i=1}^{n} (C_i C_i^{\dagger})$.

The implications $(i) \Longrightarrow (ii)$ in Theorems 2 and 3 are essentially contained in (Caiafa and Cichocki, 2010), though they consider tensors with constant multilinear rank $(r, \ldots, r)$, and force the condition $|I_i| = |I_j|$ for all $i$ and $j$. All other directions of these characterization theorems are new to the best of our knowledge, and it is of interest that the characterizations of each type of tensor CUR decomposition presented here (Fiber and Chidori) are different from the matrix case (e.g., the projection based version is no longer equivalent as Example 1 shows).

**Remark 4** *Mahoney et al. (2008) consider tensor CUR decompositions for order-3 tensors of the form* $\mathcal{R} \times_3 (C_3 U_3)$ *(after translating to our notation). One can show that this is essentially a particular case of the Chidori CUR decomposition in which* $C_1 = U_1 = \mathcal{A}_{(1)}$ *and* $C_2 = U_2 = \mathcal{A}_{(2)}$, *although the matrix* $U_3$ *undergoes some additional scaling in their algorithm not present here.*

### 3.2 Perturbation Bounds

The above characterizations are interesting from a mathematical viewpoint, but here we turn to more practical purposes and undertake an analysis of how well the decompositions mentioned above perform as low multilinear rank approximations to arbitrary tensors. To state our results, we consider the additive noise model that we observe $\widetilde{\mathcal{A}} = \mathcal{A} + \mathcal{E}$, where $\mathcal{A}$ has low multilinear rank $(r_1, \ldots, r_n)$ with $r_i < d_i$ for all $i$, and $\mathcal{E}$ is an arbitrary noise tensor. Our main concern is to address the following.

**Question 2** *If we choose the subtensors of $\widetilde{\mathcal{A}}$ in the manner of Section 3.1, how do the Chidori and Fiber CUR approximations of $\widetilde{\mathcal{A}}$ as suggested by Theorems 2 and 3 relate to Chidori and Fiber CUR decompositions of $\mathcal{A}$?*

To understand the results answering this question, we first set some notation. In what follows, tildes represent subtensors or submatrices of $\widetilde{\mathcal{A}}$, the letter $\mathcal{R}$ corresponds to core subtensors of the appropriate tensor, $C$ corresponds to either fibers or subtubes depending on if we are discussing Fiber or Chidori CUR, and $U$ corresponds to submatrices of $C$. If $\mathcal{R}$, $C$, or $U$ appears without a tilde, it is a subtensor/matrix of $\mathcal{A}$. In particular, consider

$$\begin{gathered} \widetilde{\mathcal{R}} = \mathcal{A}(I_1, \cdots, I_n), \qquad \mathcal{E}_{\mathcal{R}} = \mathcal{E}(I_1, \cdots, I_n), \\ \widetilde{C}_i = \widetilde{\mathcal{A}}_{(i)}(:, J_i), \qquad E_{J_i} = \mathcal{E}_{(i)}(:, J_i), \qquad \widetilde{U}_i = \widetilde{C}_i(I_i, :), \qquad E_{I_i,J_i} = E_{J_i}(I_i, :) \end{gathered} \tag{2}$$

for some index sets $I_i \subseteq [d_i]$ and $J_i \subseteq [\prod_{j \neq i} d_j]$, and we write

$$\widetilde{\mathcal{R}} = \mathcal{R} + \mathcal{E}_{\mathcal{R}}, \qquad \widetilde{C}_i = C_i + E_{J_i}, \qquad \widetilde{U}_i = U_i + E_{I_i,J_i}, \tag{3}$$

where $\mathcal{R} = \mathcal{A}(I_1, \ldots, I_n) \in \mathbb{R}^{|I_1| \times \cdots \times |I_n|}$, $C_i = \mathcal{A}_i(:, J_i) \in \mathbb{R}^{d_i \times |J_i|}$ and $U_i = C_i(I_i, :) \in \mathbb{R}^{|I_i| \times |J_i|}$. We also consider enforcing the rank on the submatrices $\widetilde{U}_i$ formed from the Chidori and Fiber CUR decompositions. Here, $\widetilde{U}_{i,r_i}$ is the best rank $r_i$ approximation of $\widetilde{U}_i$, and $\widetilde{U}_{i,r_i}^{\dagger}$ is its Moore–Penrose pseudoinverse. With these notations, our goal is to estimate the error

$$\mathcal{A} - \mathcal{A}_{\text{app}} := \mathcal{A} - \widetilde{\mathcal{R}} \times_{i=1}^{n} (\widetilde{C}_i \widetilde{U}_{i,r_i}^{\dagger}). \tag{4}$$

To measure accuracy, we consider the Frobenius norm of the difference of $\mathcal{A}$ and its low multilinear rank approximation $\mathcal{A}_{\text{app}}$.

**Theorem 5** *Let $\widetilde{\mathcal{A}} = \mathcal{A} + \mathcal{E}$, where the multilinear rank of $\mathcal{A}$ is $(r_1, \cdots, r_n)$ and the compact SVD of $\mathcal{A}_{(i)}$ is $\mathcal{A}_{(i)} = W_i \Sigma_i V_i^{\top}$. Let $I_i \subseteq [d_i]$ and $J_i \subseteq [\prod_{j \neq i} d_j]$. Invoke the notations of (2)–(4), and suppose that $\sigma_{r_i}(U_i) > 8 \left\| E_{I_i,J_i} \right\|_2$ for all $i$. Then,*

$$\begin{aligned} \|\mathcal{A} - \mathcal{A}_{\text{app}}\|_F &\leq \frac{9^n}{4^n} \left( \prod_{i=1}^{n} \left\| W_{i,I_i}^{\dagger} \right\|_2 \right) \|\mathcal{E}_{\mathcal{R}}\|_F \\ &+ \sum_{j=1}^{n} \frac{9^{n-j}}{4^{n-j}} \left\| \mathcal{R}_{(j)} \right\|_2 \left( \prod_{i \neq j} \left\| W_{i,I_i}^{\dagger} \right\|_2 \right) \left( 5 \left\| U_j^{\dagger} \right\|_2 \left\| W_{j,I_j}^{\dagger} \right\|_2 \left\| E_{I_j,J_j} \right\|_F + 2 \left\| U_j^{\dagger} \right\|_2 \left\| E_{J_j} \right\|_F \right), \end{aligned}$$

*where $W_{i,I_i} := W_i(I_i, :)$.*

**Remark 6** *The bounds in Theorem 5 contain the exponential terms* $\left(\frac{9}{4}\right)^n$*, which is unavoidable, but the base number* $\frac{9}{4}$ *is not necessarily sharp. For more details see Equation* (7)*.*

The error bounds in Theorem 5 are qualitative in that they are given in terms of the Frobenius norms of subtensors of the noise tensor $\mathcal{E}$. These bounds can be applied generally, but can also be applied to give error estimates for approximating the HOSVD by low multilinear tensor approximation; this can be achieved by setting $\mathcal{A}$ to be the HOSVD approximation of a certain multilinear rank of $\widetilde{\mathcal{A}}$. Note also that $\|\mathcal{R}_{(j)}\|_2 \leq \|\mathcal{R}\|_F$ for all $j$ by basic norm inequalities, so if desired, this can be taken out of the summation for simpler error bounds.

The drawback of the above bounds is that they include norms of subtensors of $\mathcal{A}$, which can be difficult to estimate in general. In the case that the indices $J_i$ depend on $I_i$, we can achieve more specialized bounds that can be expressed in terms of submatrices of singular vectors of the unfoldings of $\mathcal{A}$ as in the following theorem.

**Theorem 7** *Invoke the notations of Theorem 5 and* (2)–(4)*, and let* $J_i = \otimes_{j\neq i} I_j$*. Suppose that* $\sigma_{r_i}(U_i) > 8\|E_{I_i,J_i}\|_2$ *for every* $1 \leq i \leq n$*, and let the compact SVD of* $\mathcal{A}_{(i)}$ *be* $W_i\Sigma_i V_i^\top$*. Then*

$$\|\mathcal{A} - \mathcal{A}_{\text{app}}\|_F \leq \frac{9^n}{4^n}\left(\prod_{i=1}^{n}\left\|W_{i,I_i}^\dagger\right\|_2\right)\|\mathcal{E}_{\mathcal{R}}\|_F$$
$$+\sum_{j=1}^{n}\frac{9^{n-j}}{4^{n-j}}\left\|\mathcal{R}_{(j)}\right\|_2\left(\prod_{i\neq j}\left\|W_{i,I_i}^\dagger\right\|_2^2\right)\left\|\mathcal{A}_{(j)}^\dagger\right\|_2\left\|W_{j,I_j}^\dagger\right\|_2\left(5\left\|W_{j,I_j}^\dagger\right\|_2\left\|E_{I_j,J_j}\right\|_F + 2\left\|E_{J_j}\right\|_F\right).$$

**Remark 8** *Notice that the bounds in Theorem 5 and 7 are in terms of the submatrices of singular vectors of unfoldings of the tensor. These bounds are quite general, but they illustrate how one ought to sample* $I_i$ *and* $J_i$ *to ensure good bounds. Theorems 10 and 11 give more user friendly versions of these bounds for the concrete case when a low multilinear rank tensor* $\mathcal{A}$ *is perturbed by noise.*

*In the matrix case, there are error bounds for CUR decompositions in terms of the optimal low rank approximations of the given matrix. It is interesting, yet challenging, to derive upper bounds for tensor CUR decompositions in terms of the optimal low multilinear rank approximations of the given tensor.*

The norms of pseudoinverses of submatrices of singular vectors can vary extensively depending on the sampling method. In particular, if maximum volume sampling is used (Civril and Magdon-Ismail, 2009; Goreinov et al., 2010; Mikhalev and Oseledets, 2018), then one can give generic bounds on these terms (see (Hamm and Huang, 2021) for examples of upper bounds for maximum volume sampling for matrix CUR decompositions). However, maximum volume sampling is often intractable in practice. However, if the matrices $\mathcal{A}_{(i)}$ have good incoherence, then uniform sampling yields submatrices $W_{i,J_i}^\dagger$ with small norm (cf. (Tropp, 2011, Lemma 3.4)).

### 3.3 Error Bounds for Random Sampling Core Subtensors

Randomized sampling for column and row submatrices has been shown to be an effective method for low-rank approximation, and can provide a fast and reliable estimation of the SVD of a matrix (Frieze et al., 2004; Drineas et al., 2006; Rudelson and Vershynin, 2007; Wang and Zhang, 2013). First, let us state the formation of both Chidori and Fiber CUR decompositions via random sampling in algorithmic form here.

**Algorithm 1** Randomized Chidori CUR Decomposition

**Input:** $\mathcal{A} \in \mathbb{R}^{d_1 \times \cdots \times d_n}$, sample sizes $t_i$, and probability distributions $\{p^{(i)}\}$ over $[d_i]$, $i = 1, \ldots, n$.
**for** $i = 1 : n$ **do**
  Sample $t_i$ indices from $[d_i]$ without replacement from $\{p^{(i)}\}$, and denote the index set $I_i$
  $J_i = \otimes_{j \neq i} I_i$
  $C_i = \mathcal{A}_{(i)}(:, J_i)$
  $U_i = C_i(I_i, :)$
**end for**
$\mathcal{R} = \mathcal{A}(I_1, \ldots, I_n)$
**Output:** $\mathcal{R}, C_i, U_i$ such that $\mathcal{A} \approx \mathcal{R} \times_{i=1}^{n} (C_i U_i^{\dagger})$.

**Algorithm 2** Randomized Fiber CUR Decomposition

**Input:** $\mathcal{A} \in \mathbb{R}^{d_1 \times \cdots \times d_n}$, sample sizes $t_i, s_i$, and probability distributions $\{p^{(i)}\}, \{q^{(i)}\}$ over $[d_i]$ and $[\prod_{j \neq i} d_j]$, respectively, $i = 1, \ldots, n$.
**for** $i = 1 : n$ **do**
  Sample $t_i$ indices from $[d_i]$ without replacement from $\{p^{(i)}\}$, and denote the index set $I_i$
  Sample $s_i$ indices from $[\prod_{j \neq i} d_j]$ without replacement from $\{q^{(i)}\}$, and denote the index set $J_i$
  $C_i = \mathcal{A}_{(i)}(:, J_i)$
  $U_i = C_i(I_i, :)$
**end for**
$\mathcal{R} = \mathcal{A}(I_1, \ldots, I_n)$
**Output:** $\mathcal{R}, C_i, U_i$ such that $\mathcal{A} \approx \mathcal{R} \times_{i=1}^{n} (C_i U_i^{\dagger})$.

For general matrices, sampling columns with or without replacement from sophisticated distributions such as leverage scores is known to provide quality submatrices that represent the column space of the data faithfully (Mahoney and Drineas, 2009), but these distributions come at the cost of being expensive to compute. On the other hand, uniform sampling of columns is cheap, but is not always reliable (for instance on extremely sparse matrices). Nonetheless, it is well understood that uniformly sampling column submatrices is both cheap and effective when the initial matrix is incoherent, e.g., (Talwalkar and Rostamizadeh, 2010; Chiu and Demanet, 2013). Here we extend these ideas to tensors – a task that first requires defining what tensor incoherence even is.

**Definition 9 ((Xia et al., 2021))** *Let $W \in \mathbb{R}^{d\times r}$ have orthonormal columns. Its coherence is defined as*

$$\mu(W) = \frac{d}{r}\max_{1\le i\le d}\|W(i,:)\|_2^2.$$

*For a tensor $\mathcal{A}\in\mathbb{R}^{d_1\times\cdots\times d_n}$ such that $\mathcal{A}_{(j)} = W_j\Sigma_j V_j^\top$ is its compact singular value decomposition with $\sigma_{j1},\cdots,\sigma_{jr_j}$ on the diagonal of $\Sigma_j$, we define its coherence as*

$$\mu(\mathcal{A}) := \max\left\{\mu(W_1),\cdots,\mu(W_n)\right\},$$

*and define*

$$\sigma_{\min}(\mathcal{A}) = \min\left\{\sigma_{1r_1},\cdots,\sigma_{nr_n}\right\},$$
$$\sigma_{\max}(\mathcal{A}) = \max\left\{\sigma_{11},\cdots,\sigma_{n1}\right\}.$$

From the above definition, a tensor's incoherence is defined to be the maximal incoherence of all its unfoldings. With this definition in hand, we may state our main result on low multilinear rank tensor approximation via uniform sampling.

Note that in Theorem 10 below, the incoherence of the tensor impacts how large the index sets $I_i$ must be to guarantee a good approximation; this is similar to existing matrix constructions (Candes and Romberg, 2007; Tropp, 2011; Chiu and Demanet, 2013), and one should expect a (hopefully mild) oversampling factor for good approximation bounds in random sampling methods.

**Theorem 10** *Let $\mathcal{A}\in\mathbb{R}^{d_1\times\cdots\times d_n}$ with multilinear rank $(r_1,\cdots,r_n)$ and $\mathcal{A}_{(i)} = W_i\Sigma_i V_i^\top$ be the compact singular value decomposition of $\mathcal{A}_{(i)}$. Suppose that $\widetilde{\mathcal{A}} = \mathcal{A}+\mathcal{E}$. Suppose that $\mathcal{A}_{\text{app}} = \widetilde{\mathcal{R}}\times_{i=1}^n\left(\widetilde{C}_i\widetilde{U}_{i,r_i}^\dagger\right)$ is formed via the Chidori CUR decomposition (Algorithm 1) with input $\widetilde{\mathcal{A}}$, uniform probabilities, and $|I_i|\ge\gamma_i\mu(W_i)r_i$ for some $\gamma_i>0$. If $\delta\in[0,1)$ such that $(1-\delta)^{\frac{n}{2}}\sqrt{\frac{\prod_{i=1}^n|I_i|}{\prod_{i=1}^n d_i}}\sigma_{\min}(\mathcal{A})\ge 8\|E_{I_i,J_i}\|_2$, then*

$$\begin{aligned}\|\mathcal{A}-\mathcal{A}_{\text{app}}\|_F \le{}& \frac{9^n\sqrt{\prod_{i=1}^n d_i}}{4^n(1-\delta)^{\frac{n}{2}}\sqrt{\prod_{i=1}^n|I_i|}}\|\mathcal{E}_{\mathcal{R}}\|_F\\ &+\frac{9^n}{4^{n-1}}\frac{\sigma_{\max}(\mathcal{A})}{\sigma_{\min}(\mathcal{A})}\left(\frac{1+\eta}{1-\delta}\right)^{\frac{n}{2}}\sqrt{\prod_{i=1}^n\frac{d_i}{(1-\delta)|I_i|}}\|\mathcal{E}_{\mathcal{R}}\|_F\\ &+\frac{2\sigma_{\max}(\mathcal{A})}{\sigma_{\min}(\mathcal{A})}\left(\frac{1+\eta}{1-\delta}\right)^{\frac{n}{2}}\sum_{j=1}^n\frac{9^{n-j}}{4^{n-j}}\sqrt{\prod_{i\ne j}\frac{d_i}{(1-\delta)|I_i|}}\left\|E_{J_j}\right\|_F\end{aligned}\tag{5}$$

*with probability at least $1-\sum_{i=1}^n r_i\left(\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\gamma_i}+\left(\frac{e^{\eta}}{(1+\eta)^{1+\eta}}\right)^{\gamma_i}\right)$ for every $\eta\ge 0$.*

Let us provide a concrete choice of parameters for illustration of the bounds above: we assume a simple case where $d_i = d$ and $r_i = r$ for all $i$, we set $\delta=\eta=0.5$ and $\gamma = 10\log(d)$, and consider $\mathcal{E}$ to have i.i.d. Gaussian entries. The result is the following.

**Theorem 11** *Let $\mathcal{A} \in \mathbb{R}^{d\times\cdots\times d}$ be an $n$-mode tensor with multilinear rank $(r,\cdots,r)$ and $\mathcal{A}_{(i)} = W_i\Sigma_i V_i^\top$ be the compact SVD of $\mathcal{A}_{(i)}$. Suppose that $\widetilde{\mathcal{A}} = \mathcal{A} + \mathcal{E}$ where $\mathcal{E}$ is random tensor with i.i.d. Gaussian distributed entries with mean 0 and variance $\sigma$ i.e., $\mathcal{E}_{i_1,\cdots,i_n} \sim \mathcal{N}(0,\sigma)$. Suppose $I_i \subseteq [d]$ with $|I_1| = \cdots = |I_n| = \ell \geq 10\mu(\mathcal{A})r\log(d)$, and $\mathcal{A}_{\text{app}} = \widetilde{\mathcal{R}} \times_{i=1}^n \left(\widetilde{C}_i\widetilde{U}_{i,r_i}^\dagger\right)$ is formed via the Chidori CUR decomposition (Algorithm 1) with input $\widetilde{\mathcal{A}}$ and uniform probabilities. Let $p$ be in the open interval $(1, 1+\frac{\log(2)}{\log(\ell)})$. If $\ell^p\sigma_{\min}^2(\mathcal{A}) \geq 2^{n+8}(\log(\ell^{n-1}+d))d^n\sigma^2$, then*

$$\|\mathcal{A} - \mathcal{A}_{\text{app}}\|_F \leq \left(\frac{3^{2n}}{2^{\frac{3}{2}n-1}} + \frac{\sigma_{\max}(\mathcal{A})}{\sigma_{\min}(\mathcal{A})}\cdot\frac{3^{\frac{5n}{2}}}{2^{\frac{3n}{2}-4}}\right)\ell^{\frac{1-p}{2}}\sqrt{\log(\ell^{n-1}+d)}d^{\frac{n}{2}}\sigma$$

*with probability at least $1 - \frac{2rn}{d} - \frac{2n}{(\ell^{n-1}+d)^{2\ell^{1-p}-1}}$.*

**Remark 12** *The conclusion of Theorem 11 holds unchanged for Rademacher random noise with each entry of $\mathcal{E}$ being $\pm\sigma$ with equal probability. This is due to the use of (Tropp, 2012, Theorem 1.5) in the proof, which holds for both types of noise (see Section 3.3 for more detail).*

To conclude our results section, we show an example of how one can guarantee an exact decomposition of a low multilinear rank tensor with high probability via random sampling of indices. The theorem is stated in terms of so-called column length sampling of the unfolded versions of the tensor. Given a matrix $A$, we define probability distributions over its columns and rows, respectively, via

$$p_j(A) := \frac{\|A(j,)\|_2^2}{\|A\|_F^2}, \qquad q_j(A) := \frac{\|A(:,j)\|_2^2}{\|A\|_F^2}.$$

Subsequently, for a tensor $\mathcal{A}$, we set

$$p_j^{(i)} := p_j(\mathcal{A}_{(i)}), \quad q_j^{(i)} := q_j(\mathcal{A}_{(i)}).$$

In the following theorems, $c > 0$ is an absolute constant, and comes from the analysis of Rudelson and Vershynin (2007).

**Theorem 13** *Let $\mathcal{A} \in \mathbb{R}^{d_1\times\cdots\times d_n}$ with multilinear rank $(r_1,\cdots,r_n)$. Let $0 < \varepsilon_i < \kappa(\mathcal{A}_{(i)})^{-1}$. Let $I_i \subseteq [d_i]$, $J_i \subseteq \left[\prod_{j\neq i} d_j\right]$ satisfy*

$$|I_i| \gtrsim \left(\frac{r_i\log(d_i)}{\varepsilon_i^4}\right)\log\left(\frac{r_i\log(d_i)}{\varepsilon_i^4}\right), \qquad |J_i| \gtrsim \left(\frac{r_i\log(\prod_{j\neq i}d_j)}{\varepsilon_i^4}\right)\log\left(\frac{r_i\log(\prod_{j\neq i}d_j)}{\varepsilon_i^4}\right).$$

*and let $\mathcal{R}, C_i, U_i$ be obtained from the Fiber CUR decomposition (Algorithm 2) with probabilities $p_j^{(i)}$ and $q_j^{(i)}$. Then with probability at least*

$$\prod_{i=1}^n\left(1-\frac{2}{d_i^c}\right)\left(1-\frac{2}{\left(\prod_{j\neq i}d_j\right)^c}\right),$$

*we have $\operatorname{rank}(U_i) = r_i$, hence $\mathcal{A} = \mathcal{R}\times_{i=1}^n (C_iU_i^\dagger)$.*

**Theorem 14** *Let $\mathcal{A} \in \mathbb{R}^{d_1 \times \cdots \times d_n}$ with multilinear rank $(r_1, \cdots, r_n)$. Let $0 < \varepsilon_i < \kappa(\mathcal{A}_{(i)})^{-1}$. Let $I_i \subseteq [d_i]$ satisfy $|I_i| \gtrsim \left(\frac{r_i \log(d_i)}{\varepsilon_i^4}\right) \log\left(\frac{r_i \log(d_i)}{\varepsilon^4}\right)$. Let $\mathcal{R}, C_i, U_i$ be obtained via the Chidori CUR decomposition (Algorithm 1) with probabilities $p_j^{(i)}$. Then with probability at least $\prod_{i=1}^n \left(1 - \frac{2}{d_i^c}\right)$, we have $\mathcal{A} = \mathcal{R} \times_{i=1}^n (C_i U_i^\dagger)$.*

More results of a similar nature to Theorems 13 and 14 can be obtained by extending the analysis done here to other sampling probabilities as was done in (Hamm and Huang, 2020b), but we only state a sample of what one can obtain here for simplicity. One can readily state analogues of the above results for uniform sampling, or small perturbations of the distributions given above mimicking Theorem 5.1, and Corollaries 5.2, 5.4, and 5.5 of Hamm and Huang (2020b), which would yield exact Fiber or Chidori CUR decompositions with high probability via uniform and leverage score sampling (here one would need to utilize the leverage score distributions for each of the unfoldings of $\mathcal{A}$, which would be computationally intensive).

### 3.4 Computational Complexity

Let $\mathcal{A} \in \mathbb{R}^{d_1 \times \cdots \times d_n}$ with multilinear rank $(r_1, \cdots, r_n)$. The complexity of randomized HOSVD decomposition is $\mathcal{O}(r \prod_{i=1}^n d_i)$ with $r = \min_i\{r_i\}$. Whereas, computing the tensor-CUR approximation only involves the computations of the pseudoinverse of $U_i$.

1. (Chidori CUR) If we sample $I_i$ uniformly with $|I_i| = \mathcal{O}(r_i \log(d_i))$ and set $J_i = \otimes_{j \neq i} I_j$, then the complexity of computing the pseudoinverse of $U_i$ is $\mathcal{O}(r_i \prod_{j=1}^n (r_j \log(d_j)))$.

2. (Fiber CUR) If we sample $I_i$ and $J_i$ uniformly, then the size of $I_i$ and $J_i$ should be $\mathcal{O}(r_i \log(d_i))$ and $\mathcal{O}\left(r_i \log(\prod_{j \neq i} d_j)\right)$. Thus the complexity of computing the pseudoinverse of $U_i$ is $\mathcal{O}\left(r_i^2 \log(d_i)\left(\sum_{j \neq i} \log(d_j)\right)\right)$.

**Conversion of tensor CUR to HOSVD:** The low multilinear rank approximations discussed here are in terms of modewise products of a core subtensor of the data tensor and matrices formed by subsampling unfolded versions of the tensor, but if a user wishes to obtain an approximation in Tucker format, it is easily done via Algorithm 3. Note that converting from HOSVD to a tensor CUR decomposition is not as straightforward.

**Algorithm 3** Efficient conversion from CUR to HOSVD

1: **Input:** $\mathcal{R}, C_i, U_i$: CUR decomposition of the tensor $\mathcal{A}$.
2: $[Q_i, R_i] = \mathrm{qr}(C_i U_i^\dagger)$ for $i = 1, \cdots, n$
3: $\mathcal{T}_1 = \mathcal{R} \times_1 R_1 \times_2 \cdots \times_n R_n$
4: Compute HOSVD of $\mathcal{T}_1$ to find $\mathcal{T}_1 = \mathcal{T} \times_1 V_1 \times_2 \cdots \times_n V_n$
5: **Output:** $[\![\mathcal{T}; Q_1 V_1, \cdots, Q_n V_n]\!]$: HOSVD decomposition of $\mathcal{A}$.

## 4. Proofs of Characterization Theorems

Here we provide the proofs of the characterization theorems in Section 3.1. We do so in reverse order to the presentation due to the fact that Theorem 3 is the more general statement, and thus implies some of the statements of Theorem 2.

### 4.1 Proof of Theorem 3

$(i) \implies (ii)$: A special case of this was proved by Caiafa and Cichocki (2010), but we give a complete and simplified proof here.

Since $\text{rank}\,(U_1) = r_1$, we have $\mathcal{A}_{(1)} = C_1 U_1^\dagger (\mathcal{R}_1)_{(1)}$ with $\mathcal{R}_1 \in \mathbb{R}^{|I_1| \times d_2 \times \cdots \times d_n}$ defined to be $\mathcal{A}(I_1, :, \cdots)$, then we have $\mathcal{A} = \mathcal{R}_1 \times_1 (C_1 U_1^\dagger)$. Similarly, we have

$$\mathcal{A}_{(2)} = C_2 U_2^\dagger (\mathcal{A}(:, I_2, :, \cdots))_{(2)},$$

which implies that

$$(\mathcal{R}_1)_{(2)} = (\mathcal{A}(I_1, :, \cdots))_{(2)} \quad = \quad C_2 U_2^\dagger (\mathcal{A}(I_1, I_2, :, \cdots))_{(2)}.$$

Now set $\mathcal{R}_2 = \mathcal{A}(I_1, I_2, :, \cdots)$. Thus, $(\mathcal{R}_1)_{(2)} = C_2 U_2^\dagger (\mathcal{R}_2)_{(2)}$ and $\mathcal{R}_2 = \mathcal{R}_1 \times_2 (C_2 U_2^\dagger)$. Continuing in this manner, we have $\mathcal{R}_{n-1} = \mathcal{R}_n \times_n (C_n U_n^\dagger)$. Thus we have

$$\mathcal{A} = \mathcal{R}_1 \times_1 (C_1 U_1^\dagger) = \cdots = \mathcal{R} \times_1 (C_1 U_1^\dagger) \times_2 \cdots \times_n (C_n U_n^\dagger). \tag{6}$$

$(ii) \implies (iii)$: Since $\mathcal{R}$ and $C_i$ only contain part of $\mathcal{A}$, we have multilinear rank of $\mathcal{R} \leq$ multilinear rank of $\mathcal{A} = (r_1, \cdots, r_n)$ and $\text{rank}\,(C_i) \leq \text{rank}\,(\mathcal{A}_{(i)}) = r_i$. In addition, $\mathcal{A} = \mathcal{R} \times_1 (C_1 U_1^\dagger) \times_2 \cdots \times_n (C_n U_n^\dagger)$, so $r_i = \text{rank}\,(\mathcal{A}_{(i)}) \leq \min\{\text{rank}\,(\mathcal{R}_{(i)}), \text{rank}\,(C_i)\}$. Therefore, $(iii)$ holds.

$(iii) \implies (i)$ Notice that since $U_i$ is a submatrix of $C_i$, we have $\text{rank}\,(U_i) \leq r_i$. Since $\text{rank}\,(\mathcal{R}_{(i)}) = r_i$ and $\mathcal{R}_{(i)}$ is a submatrix of $(\mathcal{A}(\cdots, I_i, \cdots))_{(i)}$, we have

$$\text{rank}\,((\mathcal{A}(\cdots, I_i, \cdots))_{(i)}) = r_i.$$

By Theorem 1, we see that $\text{rank}\,(U_i) = r_i$, completing the proof. ■

Now we turn to the proof of Theorem 2, which handles the case when the column indices $J_i$ of $C_i$ are chosen to be $J_i = \otimes_{j \neq i} I_j$ i.e., $C_i = (\mathcal{A}(I_1, \cdots, I_{i-1}, :, I_{i+1}, \cdots, I_n))_{(i)}$.

### 4.2 Proof of Theorem 2

Evidently, the equivalence $(i) \iff (ii)$ follows as a special case of Theorem 3. Similarly, we have $(ii) \implies (iii)$.

To see $(iii) \implies (ii)$, note that since $\mathcal{A}$ has multilinear rank $(r_1, \cdots, r_n)$, $\mathcal{A}$ has HOSVD decomposition $\mathcal{A} = \mathcal{T} \times_1 W_1 \times_2 \cdots \times_n W_n$ such that $\mathcal{T} \in \mathbb{R}^{r_1 \times \cdots \times r_n}$ satisfies $\text{rank}\,(\mathcal{T}_{(i)}) = r_i$ and $W_i \in \mathbb{R}^{d_i \times r_i}$ has rank $r_i$. By the condition $\text{rank}\,(\mathcal{A}_{(i)}(I_i, :)) = r_i$, we have that $\text{rank}\left(W_i(I_i, :)\mathcal{T}_{(i)} \left(\otimes_{j \neq i} W_j^\top\right)\right) = r_i$. Thus $\text{rank}\,(W_i(I_i, :)) = r_i$. Notice that

$$C_i = \mathcal{A}_{(i)}(: \otimes_{j \neq i} I_j) = W_i \mathcal{T}_{(i)} \left(\otimes_{j \neq i} (W_j(I_j, :))\right)^\top.$$

Sylvester's rank inequality implies that

$$\begin{aligned}\operatorname{rank}(C_i) &= \operatorname{rank}\left(W_i \mathcal{T}_{(i)} \left(\otimes_{j\neq i}(W_j(I_j,:))\right)^\top\right)\\ &\geq \operatorname{rank}\left(W_i\mathcal{T}_{(i)}\right) + \operatorname{rank}\left(\otimes_{j\neq i}(W_j(I_j,:))\right) - \prod_{j\neq i} r_j\\ &= r_i + \prod_{j\neq i} r_j - \prod_{j\neq i} r_j = r_i.\end{aligned}$$

The proof that $\mathcal{R}$ has multilinear rank $(r_1,\cdots,r_n)$ is similar, which yields $(iv) \Longrightarrow (iii)$. The condition on the rank of $C_i$ and $\mathcal{R}$ imply that $\mathcal{A} = \mathcal{R} \times_{i=1}^n (C_i U_i^\dagger)$ by Theorem 3.

To see $(i) \Longrightarrow (iv)$, notice that $U_i = C_i(I_i,:) = \mathcal{A}_{(i)}(I_i, \otimes_{j\neq i} I_j)$ which is a submatrix of $\mathcal{A}_{(i)}(I_i,:)$. Since $\operatorname{rank}(U_i) = r_i$, we have $\operatorname{rank}(\mathcal{A}_{(i)}(I_i,:)) \geq \operatorname{rank}(U_i) = r_i$. Additionally, $\operatorname{rank}(\mathcal{A}_{(i)}) = r_i$, so $\operatorname{rank}(\mathcal{A}_{(i)}(I_i,:)) = r_i$. Thus, $(iv)$ holds. ■

## 5. Perturbation Analysis

Real data tensors are rarely exactly low rank, but can be modeled as low multilinear rank data + noise. In this section, we give proofs for the perturbation analysis expoused in Section 3.2. Our primary task is to consider a tensor of the form:

$$\widetilde{\mathcal{A}} = \mathcal{A} + \mathcal{E},$$

where $\mathcal{A} \in \mathbb{R}^{d_1\times\cdots\times d_n}$ has low multilinear rank $(r_1,\cdots,r_n)$ with $r_i < d_i$, and $\mathcal{E} \in \mathbb{R}^{d_1\times\cdots\times d_n}$ is an arbitrary noise tensor.

### 5.1 Ingredients for the Proof of Theorem 5

The main ingredients for proving Theorem 5 is the following.

**Theorem 15** *Let $\widetilde{\mathcal{A}} = \mathcal{A} + \mathcal{E} \in \mathbb{R}^{d_1\times\cdots\times d_n}$, where the multilinear rank of $\mathcal{A}$ is $(r_1,\ldots,r_n)$. Let $I_i \subseteq [d_i]$ and $J_i \subseteq [\prod_{j\neq i} d_j]$. Invoke the notations of (2)–(4), and suppose that $\sigma_{r_i}(U_i) > 8\|E_{I_i,J_i}\|_2$ for all $i$. Then,*

$$\begin{aligned}\|\mathcal{A} - \mathcal{A}_{\text{app}}\|_F &\leq \frac{9^n}{4^n}\left(\prod_{i=1}^n \left\|C_i U_i^\dagger\right\|_2\right)\|\mathcal{E}_{\mathcal{R}}\|_F\\ &+ \sum_{j=1}^n \frac{9^{n-j}}{4^{n-j}}\|\mathcal{R}_{(j)}\|_2\left(\prod_{i\neq j}\left\|C_i U_i^\dagger\right\|_2\right)\left(5\left\|U_j^\dagger\right\|_2\left\|C_j U_j^\dagger\right\|_2\|E_{I_j,J_j}\|_F + 2\left\|U_j^\dagger\right\|_2\|E_{J_j}\|_F\right).\end{aligned}$$

**Proposition 16 ((Hamm and Huang, 2021, Proposition 3.1))** *Suppose that $A \in \mathbb{R}^{m\times n}$ has rank $r$ and its compact SVD is $A = W_A \Sigma_A V_A^\top$. Let $C$, $U$, and $R$ be the submatrices of $A$ (with selected row and column indices $I$ and $J$, respectively) such that $A = CU^\dagger R$. Then*

$$\|CU^\dagger\|_2 = \|W_{A,I}^\dagger\|_2,$$

*where $W_{A,I} = W_A(I,:)$.*

### 5.2 Proof of Theorem 5

Combining Theorem 15 with Proposition 16 directly yields the conclusion of Theorem 5. ■

### 5.3 Proof of Theorem 7

Note that when the column index $J_i$ of $C_i$ can be written as $J_i = \otimes_{j\neq i} I_j$, we have $U_i = \mathcal{R}_{(i)}$. If $\mathcal{A}$ has HOSVD $\mathcal{A} = \mathcal{T} \times_1 W_1 \times_2 \cdots \times_n W_n$, then $U_i = W_{i,I_i}\mathcal{T}_{(i)}\left(\otimes_{j\neq i} W_{j,I_j}\right)^\top$. Thus $\left\|U_i^\dagger\right\|_2 \leq \left\|\mathcal{T}_{(i)}^\dagger\right\|_2 \prod_{j=1}^n \left\|W_{j,I_j}^\dagger\right\|_2 = \left\|\mathcal{A}_{(i)}^\dagger\right\|_2 \prod_{j=1}^n \left\|W_{j,I_j}^\dagger\right\|_2$. Combining this bound with Theorem 5 yields the required bound. ■

### 5.4 Ingredients for the proof of Theorem 15

To prove Theorem 15, we first need some preliminary observations.

**Lemma 17** *Let $A$ and $B$ be two rank $r$ matrices, then*

$$\left\|AA^\dagger - BB^\dagger\right\|_2 \leq \frac{\|A-B\|_2}{\sigma_r(A)}, \quad \left\|A^\dagger A - B^\dagger B\right\|_2 \leq \frac{\|A-B\|_2}{\sigma_r(A)},$$

*and*

$$\left\|AA^\dagger - BB^\dagger\right\|_F \leq \frac{\sqrt{2}\|A-B\|_F}{\sigma_r(A)}, \quad \left\|A^\dagger A - B^\dagger B\right\|_F \leq \frac{\sqrt{2}\|A-B\|_F}{\sigma_r(A)}.$$

**Proof** Let $A = W\Sigma V^\top$ be the compact SVD of $A$. Then $AA^\dagger = W\Sigma V^\top V\Sigma^{-1}W^\top = WW^\top$. Similarly, $A^\dagger A = VV^\top$. Then, we achieve the claims by applying (Cai et al., 2019, Lemma 6). ■

Prior to stating a corollary of this lemma, note that $E_{I,J} = E(I,J)$ in the matrix case. Likewise, note that $\widetilde{U}_r$ is the best rank–$r$ approximation of $\widetilde{U}$, and $\widetilde{U}_r^\dagger$ is the Moore–Penrose pseudoinverse of $\widetilde{U}_r$.

**Corollary 18** *Suppose that $A \in \mathbb{R}^{m\times n}$ has rank $r$ with compact SVD $A = W\Sigma V^*$ and $\widetilde{A} = A + E$. Let $\widetilde{C} = \widetilde{A}(:,J)$ and $\widetilde{U} = \widetilde{A}(I,J)$ with $I \subseteq [m]$ and $J \subseteq [n]$. If $\sigma_r(U) > 4\left\|E_{I,J}\right\|_2$, then*

$$\left\|\widetilde{C}\widetilde{U}_r^\dagger - CU^\dagger\right\| \leq \frac{\left\|U^\dagger\right\|_2\left(\left\|CU^\dagger\right\|_2\left\|E_{I,J}\right\| + \|E_J\|\right)}{1 - 4\left\|U^\dagger\right\|_2\left\|E_{I,J}\right\|_2} + 2\sqrt{2}\left\|CU^\dagger\right\|_2\left\|U^\dagger\right\|_2\left\|E_{I,J}\right\|.$$

*Here $\|\cdot\|$ can be $\|\cdot\|_2$ or $\|\cdot\|_F$.*

**Proof** Since $\sigma_r(U) > 4\|E_{I,J}\|_2$, we have $\sigma_r(\widetilde{U}) \geq \sigma_r(U) - \|E_{I,J}\|_2 > 3\|E_{I,J}\|_2 \geq 0$. Thus $\operatorname{rank}(\widetilde{U}_r) = r$. Notice that $C = CU^\dagger U$. Then

$$\|\widetilde{C}\widetilde{U}_r^\dagger - CU^\dagger\| \leq \|C\widetilde{U}_r^\dagger - CU^\dagger\| + \|E_J\widetilde{U}_r^\dagger\|$$

$$
\begin{aligned}
&= \|CU^\dagger U\widetilde{U}_r^\dagger - CU^\dagger UU^\dagger\| + \|E_J\widetilde{U}_r^\dagger\| \\
&\le \|CU^\dagger\|_2\|(\widetilde{U} - E_{I,J})\widetilde{U}_r^\dagger - UU^\dagger\| + \|E_J\widetilde{U}_r^\dagger\| \\
&\le \|CU^\dagger\|_2\|\widetilde{U}\widetilde{U}_r^\dagger - UU^\dagger\| + \|CU^\dagger\|_2\|\widetilde{U}_r^\dagger\|_2\,\|E_{I,J}\| + \|\widetilde{U}_r^\dagger\|_2\|E_J\| \\
&\le \sqrt{2}\|CU^\dagger\|_2\|\widetilde{U}_r - U\|\|U^\dagger\| + \|CU^\dagger\|_2\|\widetilde{U}_r^\dagger\|_2\,\|E_{I,J}\| + \|\widetilde{U}_r^\dagger\|_2\|E_J\| \\
&\le 2\sqrt{2}\left\|CU^\dagger\right\|_2\left\|U^\dagger\right\|_2\|E_{I,J}\| + \frac{\left\|U^\dagger\right\|_2\left(\left\|CU^\dagger\right\|_2\|E_{I,J}\| + \|E_J\|\right)}{1 - 4\left\|U^\dagger\right\|_2\left\|E_{I,J}\right\|_2},
\end{aligned}
$$

where the penultimate inequality uses Lemma 17. We also used the facts that $\widetilde{U}\widetilde{U}_r^\dagger = \widetilde{U}_r\widetilde{U}_r^\dagger$ (similar to (Hamm and Huang, 2021)), and that $\|\widetilde{U}_r - U\| \le \|\widetilde{U}_r - \widetilde{U}\| + \|\widetilde{U} - U\| \le 2\|E_{I,J}\|$. ■

**Lemma 19 ((Hamm and Huang, 2021, Lemma 8.3))** *Suppose that $A \in \mathbb{R}^{m\times n}$ has rank $r$ with compact SVD $A = W\Sigma V^\top$ and $\widetilde{A} = A + E$. Let $\widetilde{C} = \widetilde{A}(:, J)$ and $\widetilde{U} = \widetilde{A}(I, J)$ with $I \subseteq [m]$ and $J \subseteq [n]$. If $\sigma_r(U) > 4\left\|E_{I,J}\right\|_2$, then*

$$
\left\|\widetilde{C}\widetilde{U}_r^\dagger\right\| \le \left\|CU^\dagger\right\|\left(1 + \frac{\left\|U^\dagger\right\|_2\left\|E_{I,J}\right\|_2}{1 - 4\left\|U^\dagger\right\|_2\left\|E_{I,J}\right\|_2}\right) + \frac{\left\|U^\dagger\right\|_2\|E_J\|}{1 - 4\left\|U^\dagger\right\|_2\left\|E_{I,J}\right\|_2}.
$$

Now, we are in the stage to prove Theorem 15.

### 5.5 Proof of Theorem 15

Notice that $\sigma_{r_i}(U_i) > 8\|E_{I_i,J_i}\|_2 \ge 0$ implies that $\operatorname{rank}(U_i) = r_i$, thus $\mathcal{A} = \mathcal{R} \times_{i=1}^n \left(C_iU_i^\dagger\right)$ by Theorem 3. Therefore,

$$
\begin{aligned}
&\|\mathcal{A} - \mathcal{A}_{\text{app}}\|_F \\
&= \left\|\mathcal{R} \times_{i=1}^n \left(C_iU_i^\dagger\right) - \widetilde{\mathcal{R}} \times_{i=1}^n \left(\widetilde{C}_i\widetilde{U}_{i,r_i}^\dagger\right)\right\|_F \\
&\le \left\|(\mathcal{R} - \widetilde{\mathcal{R}}) \times_{i=1}^n (\widetilde{C}_i\widetilde{U}_{i,r_i}^\dagger)\right\|_F + \sum_{j=1}^n \left\|\mathcal{R} \times_{i=1}^{j-1} C_iU_i^\dagger \times (C_jU_j^\dagger - \widetilde{C}_j\widetilde{U}_{j,r_j}^\dagger) \times_{i=j+1}^n \widetilde{C}_i\widetilde{U}_{i,r_i}^\dagger\right\|_F \\
&\le \|\mathcal{E}_{\mathcal{R}}\|_F \prod_{i=1}^n \left\|\widetilde{C}_i\widetilde{U}_{i,r}^\dagger\right\|_2 + \sum_{j=1}^n \|\mathcal{R}_{(j)}\|_2 \left(\prod_{i=1}^{j-1}\left\|C_iU_i^\dagger\right\|_2\right)\left\|C_jU_j^\dagger - \widetilde{C}_j\widetilde{U}_{j,r_j}^\dagger\right\|_F \left(\prod_{i=j+1}^n \left\|\widetilde{C}_i\widetilde{U}_{i,r_i}^\dagger\right\|_2\right).
\end{aligned}
$$

By Lemma 19, we have

$$
\begin{aligned}
\left\|\widetilde{C}_i\widetilde{U}_{i,r}^\dagger\right\|_2 &\le \left\|C_iU_i^\dagger\right\|_2\left(1 + \frac{\|U_i^\dagger\|_2\left\|E_{I_i,J_i}\right\|_2}{1 - 4\|U_i^\dagger\|_2\left\|E_{I_i,J_i}\right\|_2}\right) + \frac{\|U_i^\dagger\|_2\|E_{J_i}\|_2}{1 - 4\|U_i^\dagger\|_2\|E_{I_i,J_i}\|_2} \\
&\le \frac{5}{4}\left\|C_iU_i^\dagger\right\|_2 + \frac{1}{4} \\
&\le \frac{9}{4}\left\|C_iU_i^\dagger\right\|_2.
\end{aligned} \tag{7}
$$

By Corollary 18,

$$\begin{aligned}
&\left\|C_j U_j^\dagger - \widetilde{C}_j \widetilde{U}_{j,r_j}^\dagger\right\|_F \\
&\leq \frac{\left\|U_j^\dagger\right\|_2 \left(\left\|C_j U_j^\dagger\right\|_2 \left\|E_{I_j,J_j}\right\|_F + \left\|E_{J_j}\right\|_F\right)}{1 - 4\left\|U_j^\dagger\right\|_2 \left\|E_{I_j,J_j}\right\|_2} + 2\sqrt{2}\left\|C_j U_j^\dagger\right\|_2 \left\|U_j^\dagger\right\|_2 \left\|E_{I_j,J_j}\right\|_F \\
&\leq 2\left\|U_j^\dagger\right\|_2 \left(\left\|C_j U_j^\dagger\right\|_2 \left\|E_{I_j,J_j}\right\|_F + \left\|E_{J_j}\right\|_F\right) + 2\sqrt{2}\left\|C_j U_j^\dagger\right\|_2 \left\|U_j^\dagger\right\|_2 \left\|E_{I_j,J_j}\right\|_F \\
&\leq 5\left\|U_j^\dagger\right\|_2 \left\|C_j U_j^\dagger\right\|_2 \left\|E_{I_j,J_j}\right\|_F + 2\left\|U_j^\dagger\right\|_2 \left\|E_{J_j}\right\|_F.
\end{aligned} \tag{8}$$

Combining (7) and (8), we get

$$\begin{aligned}
\|\mathcal{A} - \mathcal{A}_{\text{app}}\|_F &\leq \frac{9^n}{4^n}\left(\prod_{i=1}^n \left\|C_i U_i^\dagger\right\|_2\right)\|\mathcal{E}_{\mathcal{R}}\|_F \\
&+ \sum_{j=1}^n \frac{9^{n-j}}{4^{n-j}}\|\mathcal{R}_{(j)}\|_2 \left(\prod_{i\neq j}\left\|C_i U_i^\dagger\right\|_2\right)\left(5\left\|U_j^\dagger\right\|_2 \left\|C_j U_j^\dagger\right\|_2 \left\|E_{I_j,J_j}\right\|_F + 2\left\|U_j^\dagger\right\|_2 \left\|E_{J_j}\right\|_F\right).
\end{aligned}$$

This concludes the proof. ■

## 6. Random Sampling Methods

The main purpose of this section is to prove the results of Section 3.3. From Theorem 3 and 5, we see that the existence of a tensor-CUR decomposition and the error for a tensor-CUR approximation both depend crucially on the chosen indices $I_i$ and $J_i$. Here, we study randomized sampling procedures for selecting these indices to determine how the sampling affects the quality of the tensor approximation, and when it can guarantee an exact decomposition when $\mathcal{A}$ has low multilinear rank.

### 6.1 Guaranteeing Exact Decompositions

We begin by answering the question: when does random sampling guarantee an exact tensor-CUR decomposition for low multilinear rank $\mathcal{A}$. While this is not overly important in practice, we are able to use this analysis to provide approximation guarantees in the general case. First, we review a related sampling result for matrix CUR decompositions.

**Lemma 20 ((Hamm and Huang, 2020a, Theorem 4.1))** *Let $A \in \mathbb{R}^{m\times n}$ have rank $r$. Let $0 < \varepsilon < \kappa(A)^{-1}$. Let $d_1 \in [m]$, $d_2 \in [n]$ satisfy*

$$d_1 \gtrsim \left(\frac{r\log(m)}{\varepsilon^4}\right)\log\left(\frac{r\log(m)}{\varepsilon^4}\right), \qquad d_2 \gtrsim \left(\frac{r\log(n)}{\varepsilon^4}\right)\log\left(\frac{r\log(n)}{\varepsilon^4}\right).$$

*Choose $I \subseteq [m]$ by sampling $d_1$ rows of $A$ independently without replacement according to probabilities $q_j^{(i)}$ and choose $J \subseteq [n]$ by sampling $d_2$ columns of $A$ independently with*

*replacement according to* $p_j^{(i)}$. *Set* $R = A(I,:)$, $C = A(:,J)$, *and* $U = A(I,J)$. *Then with probability at least* $(1-\frac{2}{n^c})(1-\frac{2}{m^c})$,

$$\operatorname{rank}(U) = r.$$

**Proof** [Proof of Theorem 13] Lemma 20 applied to each $C_i, U_i$ implies condition $(i)$ of Theorem 3, which implies that $\mathcal{A} = \mathcal{R} \times_1 (C_1 U_1^\dagger) \times_2 \cdots \times_n (C_n U_n^\dagger)$ as required. ■

Recall that according to Theorem 2, if $J_i = \otimes_{j\neq i} I_i$, and $\operatorname{rank}(\mathcal{A}_i(I_i,:)) = r_i$ for all $i$, then $\mathcal{A} = \mathcal{R} \times_{i=1}^n (C_i U_i^\dagger)$. Theorem 14 shows a sampling scheme only requiring choosing indices $I_i$ to achieve an exact Chidori CUR decomposition with high probability.

**Proof** [Proof of Theorem 14] This theorem is derived in the same way as Theorem 13 but applying Theorem 2 instead of Theorem 3. ■

### 6.2 Ingredients for the Proof of Theorem 10

We now turn toward the proof of the main result (Theorem 10) regarding approximation bounds of tensor-CUR approximations via randomized sampling. Before proving Theorem 10, we state an fundamental result of Tropp about uniform random sampling of matrices with low incoherence .

**Lemma 21 ((Tropp, 2011, Lemma 3.4))** *Suppose that* $W \in \mathbb{R}^{d\times r}$ *has orthonormal columns. If* $I \subseteq [d]$ *with* $|I| \geq \gamma\mu(W)r$ *for some* $\gamma > 0$ *is chosen by sampling* $[d]$ *uniformly without replacement, then*

$$\mathbb{P}\left(\left\|W(I,:)^\dagger\right\|_2 \leq \sqrt{\frac{d}{(1-\delta)|I|}}\right) \geq 1 - r\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^\gamma, \quad \text{for all} \quad \delta \in [0,1),$$

$$\mathbb{P}\left(\|W(I,:)\|_2 \leq \sqrt{\frac{(1+\eta)|I|}{d}}\right) \geq 1 - r\left(\frac{e^{\eta}}{(1+\eta)^{1+\eta}}\right)^\gamma, \quad \text{for all} \quad \eta \geq 0.$$

Lemma 21 easily results in the following corollary for uniformly sampling indices of a tensor.

**Corollary 22** *Let* $\mathcal{A} \in \mathbb{R}^{d_1\times\cdots\times d_n}$ *with multilinear rank* $(r_1,\cdots,r_n)$ *and let* $\mathcal{A}_{(i)} = W_i\Sigma_i V_i^\top$ *be* $\mathcal{A}_{(i)}$*'s compact SVD. If* $I_i \subseteq [d_i]$ *with* $|I_i| \geq \gamma_i\mu(W_i)r_i$ *for some* $\gamma_i > 0$ *are chosen by sampling* $[d_i]$ *uniformly without replacement, then with probability at least* $1 - \sum_{i=1}^n \left(r_i\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\gamma_i} + r_i\left(\frac{e^{\eta}}{(1+\eta)^{1+\eta}}\right)^{\gamma_i}\right)$

$$\|W_i(I_i,:)^\dagger\|_2 \leq \sqrt{\frac{d_i}{(1-\delta)|I_i|}}, \qquad \|W_i(I_i,:)\|_2 \leq \sqrt{\frac{(1+\eta)|I_i|}{d_i}} \tag{9}$$

*for all* $i = 1,\cdots,n$ *and all* $\delta \in [0,1)$ *and* $\eta \geq 0$.

### 6.3 Proof of Theorem 10

The proof follows essentially by combining Theorem 7 and Corollary 22. Suppose that the HOSVD of $\mathcal{A}$ is $\mathcal{A} = \mathcal{T} \times_1 W_1 \times_2 \cdots \times_n W_n$. Then

$$\begin{aligned}\|\mathcal{R}_{(i)}\|_2 &= \|W_{i,I_i}\mathcal{T}_{(i)}\left(\otimes_{j\neq i}W_{j,I_j}\right)^\top\|_2 \\ &\leq \|\mathcal{T}_{(i)}\|_2 \prod_{j=1}^n \|W_{j,I_j}\|_2 \\ &= \sigma_{i1}\prod_{j=1}^n \|W_{j,I_j}\|_2 \leq \sigma_{\max}(\mathcal{A})\prod_{j=1}^n \|W_{j,I_j}\|_2.\end{aligned}$$

According to Corollary 22, with probability at least

$$1 - \sum_{i=1}^n \left( r_i \left( \frac{e^{-\delta}}{(1-\delta)^{1-\delta}} \right)^{\gamma_i} + r_i \left( \frac{e^{\eta}}{(1+\eta)^{1+\eta}} \right)^{\gamma_i} \right)$$

we have

$$\|W_i(I_i,:)^\dagger\|_2 \leq \sqrt{\frac{d_i}{(1-\delta)|I_i|}}, \qquad \|W_i(I_i,:)\|_2 \leq \sqrt{\frac{(1+\eta)|I_i|}{d_i}}.$$

Therefore,

$$\|\mathcal{R}_{(i)}\|_2 \leq \sqrt{\frac{(1+\eta)^n \prod_{j=1}^n |I_i|}{\prod_{j=1}^n d_j}}\sigma_{\max}(\mathcal{A})$$

and

$$\begin{aligned}\left\|U_i^\dagger\right\|_2 = \left\|\mathcal{R}_{(i)}^\dagger\right\| &= \left\|\left(W_{i,I_i}\mathcal{T}_{(i)}\left(\otimes_{j\neq i}W_{j,I_j}\right)^\top\right)^\dagger\right\|_2 \\ &\leq \|\mathcal{T}_i^\dagger\|_2 \prod_{j=1}^n \|W_{j,I_j}^\dagger\|_2 \\ &\leq \sqrt{\frac{\prod_{j=1}^n d_j}{(1-\delta)^n \prod_{j=1}^n |I_j|}}\left\|\mathcal{A}_{(i)}^\dagger\right\|_2 \\ &= \sqrt{\frac{\prod_{j=1}^n d_j}{(1-\delta)^n \prod_{j=1}^n |I_j|}}\frac{1}{\sigma_{ir_i}} \\ &\leq \frac{1}{\sigma_{\min}(\mathcal{A})}\sqrt{\frac{\prod_{j=1}^n d_j}{(1-\delta)^n \prod_{j=1}^n |I_j|}}.\end{aligned}$$

Additionally,

$$\sqrt{\frac{(1-\delta)^n \prod_{i=1}^n |I_i|}{\prod_{i=1}^n d_i}}\sigma_{\min}(\mathcal{A}) \geq 8\|E_{I_i,J_i}\|_2,$$

thus $\sigma_{r_i}(U_i) \geq 8\|E_{I_i,J_i}\|$ holds given the condition that $\|W_i(I_i,:)^\dagger\|_2 \leq \sqrt{\frac{d_i}{(1-\delta)|I_i|}}$. Combing this with Theorem 7, with probability at least

$$1 - \sum_{i=1}^n \left( r_i \left( \frac{e^{-\delta}}{(1-\delta)^{1-\delta}} \right)^{\gamma_i} + r_i \left( \frac{e^{\eta}}{(1+\eta)^{1+\eta}} \right)^{\gamma_i} \right),$$

we have

$$
\begin{aligned}
&\|\mathcal{A}-\mathcal{A}_{\text{app}}\|_F \\
\le& \frac{9^n}{4^n}\left(\prod_{i=1}^n \left\|W_{i,I_i}^\dagger\right\|_2\right)\|\mathcal{E}_{\mathcal{R}}\|_F \\
&+\sum_{j=1}^n \frac{9^{n-j}}{4^{n-j}}\left\|\mathcal{R}_{(j)}\right\|_2\left(\prod_{i\neq j}\left\|W_{i,I_i}^\dagger\right\|_2^2\right)\left\|W_{j,I_j}^\dagger\right\|_2\left\|\mathcal{A}_{(j)}^\dagger\right\|_2\left(5\left\|W_{j,I_j}^\dagger\right\|_2\left\|E_{I_j,J_j}\right\|_F+2\left\|E_{J_j}\right\|_F\right) \\
\le& \frac{9^n}{4^n}\sqrt{\frac{\prod_{j=1}^n d_j}{(1-\delta)^n\prod_{j=1}^n|I_j|}}\|\mathcal{E}_{\mathcal{R}}\|_F \\
&+\sum_{j=1}^n \frac{\sigma_{\max}(\mathcal{A})9^{n-j}}{4^{n-j}}\sqrt{\frac{(1+\eta)^n\prod_{j=1}^n|I_i|}{\prod_{j=1}^n d_j}}\frac{\prod_{i\neq j}d_i}{(1-\delta)^{n-1}\prod_{i\neq j}|I_i|}\sqrt{\frac{d_j}{(1-\delta)|I_j|}}\left\|\mathcal{A}_{(j)}^\dagger\right\|_2 \\
&\cdot\left(5\sqrt{\frac{d_j}{(1-\delta)|I_j|}}\left\|E_{I_j,J_j}\right\|_F+2\left\|E_{J_j}\right\|_F\right)\Bigg) \\
\le& \frac{9^n}{4^n}\sqrt{\frac{\prod_{j=1}^n d_j}{(1-\delta)^n\prod_{j=1}^n|I_j|}}\|\mathcal{E}_{\mathcal{R}}\|_F \\
&+\frac{\sigma_{\max}(\mathcal{A})}{\sigma_{\min}(\mathcal{A})}\sqrt{\frac{(1+\eta)^n}{(1-\delta)^n}}\sum_{j=1}^n\frac{9^{n-j}}{4^{n-j}}\sqrt{\frac{\prod_{i\neq j}d_i}{(1-\delta)^{n-1}\prod_{i\neq j}|I_i|}}\left(5\sqrt{\frac{d_j}{(1-\delta)|I_j|}}\left\|E_{I_j,J_j}\right\|_F+2\left\|E_{J_j}\right\|_F\right) \\
\le& \frac{9^n\sqrt{\prod_{i=1}^n d_i}}{4^n(1-\delta)^{\frac{n}{2}}\sqrt{\prod_{i=1}^n|I_i|}}\|\mathcal{E}_{\mathcal{R}}\|_F+\frac{9^n}{4^{n-1}}\frac{\sigma_{\max}(\mathcal{A})}{\sigma_{\min}(\mathcal{A})}\left(\frac{1+\eta}{1-\delta}\right)^{\frac{n}{2}}\sqrt{\prod_{i=1}^n\frac{d_i}{(1-\delta)|I_i|}}\|\mathcal{E}_{\mathcal{R}}\|_F \\
&+\frac{2\sigma_{\max}(\mathcal{A})}{\sigma_{\min}(\mathcal{A})}\left(\frac{1+\eta}{1-\delta}\right)^{\frac{n}{2}}\sum_{j=1}^n\frac{9^{n-j}}{4^{n-j}}\sqrt{\prod_{i\neq j}\frac{d_i}{(1-\delta)|I_i|}}\left\|E_{J_j}\right\|_F.
\end{aligned}
$$

The last equation holds because of $\|E_{I_i,J_i}\|_F = \|\mathcal{E}_{\mathcal{R}}\|_F$ for $i=1,\cdots,n$ and $\sum_{j=1}^n \frac{9^{n-j}}{4^{n-j}} \le \frac{4}{5}\cdot\frac{9^n}{4^n}$. This completes the proof. ■

### 6.4 Proof of Theorem 11

To prove this corollary, we need to use (Tropp, 2012, Theorem 1.5). For the reader's convenience, we state the theorem here:

**Theorem 23 ((Tropp, 2012, Theorem 1.5))** *Consider a finite sequence $\{B_k\}$ of fixed matrices with dimension $m_1 \times m_2$, and let $\{\xi_k\}$ be a finite sequence of independent standard normal variables. Define the variance parameter*

$$\phi^2 := \max\left\{\left\|\sum_k B_k B_k^\top\right\|_2, \left\|\sum_k B_k^\top B_k\right\|_2\right\}.$$

*Then, for all* $t \geq 0$,

$$\left\| \sum_k \xi_k B_k \right\|_2 \leq t$$

*with probability at least* $1 - (m_1 + m_2)e^{-\frac{t^2}{2\phi^2}}$.

Now, let's prove Theorem 11 with the case $d \leq \ell^{n-1}$. Since $d_1 = \cdots = d_n = d$, $r_1 = \cdots = r_n$, $|I_1| = \cdots = |I_n| = \ell$, $\delta = \eta = \frac{1}{2}$, and $\gamma = 10r\log(d)$, (5) of Theorem 10 can be restated as

$$\begin{aligned}
&\|\mathcal{A} - \mathcal{A}_{\text{app}}\|_F \\
&\leq \frac{9^n\sqrt{\prod_{i=1}^n d_i}}{4^n(1-\delta)^{\frac{n}{2}}\sqrt{\prod_{i=1}^n |I_i|}}\|\mathcal{E}_{\mathcal{R}}\|_F + \frac{9^n}{4^{n-1}}\frac{\sigma_{\max}(\mathcal{A})}{\sigma_{\min}(\mathcal{A})}\left(\frac{1+\eta}{1-\delta}\right)^{\frac{n}{2}}\sqrt{\prod_{i=1}^n \frac{d_i}{(1-\delta)|I_i|}}\|\mathcal{E}_{\mathcal{R}}\|_F \\
&\quad + \frac{2\sigma_{\max}(\mathcal{A})}{\sigma_{\min}(\mathcal{A})}\left(\frac{1+\eta}{1-\delta}\right)^{\frac{n}{2}}\sum_{j=1}^n \frac{9^{n-j}}{4^{n-j}}\sqrt{\prod_{i\neq j} \frac{d_i}{(1-\delta)|I_i|}}\left\|E_{J_j}\right\|_F \\
&= \left(\frac{9\sqrt{2d}}{4\sqrt{\ell}}\right)^n \|\mathcal{E}_{\mathcal{R}}\|_F + \frac{9^n}{4^{n-1}}\cdot\frac{\sigma_{\max}(\mathcal{A})}{\sigma_{\min}(\mathcal{A})}\cdot 3^{\frac{n}{2}}\cdot\sqrt{\frac{(2d)^n}{\ell^n}}\|\mathcal{E}_{\mathcal{R}}\|_F \\
&\quad + \frac{2\sigma_{\max}(\mathcal{A})}{\sigma_{\min}(\mathcal{A})}\cdot 3^{\frac{n}{2}}\cdot\sqrt{\frac{(2d)^{n-1}}{\ell^{n-1}}}\cdot\sum_{j=1}^n \frac{9^{n-j}}{4^{n-j}}\|E_{J_j}\|_F,
\end{aligned}$$

with probability at least

$$\begin{aligned}
&1 - \sum_{i=1}^n \left(r_i\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\gamma_i} + r_i\left(\frac{e^{\eta}}{(1+\eta)^{1+\eta}}\right)^{\gamma_i}\right) \\
&= 1 - nr\left(\left(\frac{e^{-\frac{1}{2}}}{(\frac{1}{2})^{\frac{1}{2}}}\right)^{10\log(d)} + \left(\frac{e^{\frac{1}{2}}}{(\frac{3}{2})^{\frac{3}{2}}}\right)^{10\log(d)}\right) \\
&\geq 1 - nr\left(\frac{1}{d^{3/2}} + \frac{1}{d^{1.08}}\right) \geq 1 - \frac{2nr}{d}.
\end{aligned}$$

Additionally, $\mathcal{E}_{i_1,\cdots,i_n}$ are i.i.d. and $\mathcal{E}_{i_1,\cdots,i_n} \sim \mathcal{N}(0,\sigma)$. Thus, $E_{I_i,J_i}$ can be rewritten in the form $\sum_{k,s}\xi_{k,s}B_{k,s}$ where $B_{k,s} \in \mathbb{R}^{|I_i|\times|J_i|}$ is a matrix where the only nonzero entry is $\sigma$ with index $(k,s)$ and $\xi_{k,s} \sim \mathcal{N}(0,1)$. Then $\phi^2 = \max\left\{\|\sum_{k,s} B_{k,s}^\top B_{k,s}\|_2, \|\sum_{k,s} B_{ks}, B_{k,s}^\top\|_2\right\} = \sigma^2\ell^{n-1}$. By applying Theorem 23, we have with probability at least $1 - (\ell^{n-1}+d)^{1-2\ell^{1-p}}$

$$\|E_{I_i,J_i}\|_2 \leq 2\sigma\sqrt{\ell^{n-p}\log(\ell^{n-1}+d)} \quad \text{and} \quad \|E_{I_i,J_i}\|_F \leq 2\sigma\sqrt{\ell^{n+1-p}\log(\ell^{n-1}+d)},$$

where $1 < p < 1 + \frac{\log(2)}{\log(\ell)}$. Similarly, we have with probability at least $1 - (\ell^{n-1}+d)^{1-2\ell^{1-p}}$

$$\|E_{J_i}\|_2 \leq 2\sigma\sqrt{\ell^{n-p}\log(\ell^{n-1}+d)} \quad \text{and} \quad \|E_{J_i}\|_F \leq 2\sigma\sqrt{d\ell^{n-p}\log(\ell^{n-1}+d)}.$$

By using the union bound of all these probabilities, we find that

$$\begin{aligned}
&\|\mathcal{A}-\mathcal{A}_{\text{app}}\|_F \\
\leq &\left(\frac{9\sqrt{2d}}{4\sqrt{\ell}}\right)^n \cdot 2\sigma\sqrt{\ell^{n+1-p}\log(\ell^{n-1}+d)} \\
&+\frac{9^n}{4^{n-1}}\cdot\frac{\sigma_{\max}(\mathcal{A})}{\sigma_{\min}(\mathcal{A})}\cdot 3^{\frac{n}{2}}\cdot\sqrt{\frac{(2d)^n}{\ell^n}}\cdot 2\sigma\sqrt{\ell^{n+1-p}\log(\ell^{n-1}+d)} \\
&+\frac{2\sigma_{\max}(\mathcal{A})}{\sigma_{\min}(\mathcal{A})}\cdot 3^{\frac{n}{2}}\sum_{j=1}^{n}\frac{9^{n-j}}{4^{n-j}}\sqrt{\frac{(2d)^{n-1}}{\ell^{n-1}}}\cdot 2\sigma\sqrt{d\ell^{n-p}\log(\ell^{n-1}+d)} \\
\leq &\left(\frac{3^{2n}}{2^{\frac{3}{2}n-1}}+\frac{\sigma_{\max}(\mathcal{A})}{\sigma_{\min}(\mathcal{A})}\cdot\frac{3^{\frac{5n}{2}}}{2^{\frac{3n}{2}-4}}\right)\ell^{\frac{1-p}{2}}\sqrt{\log(\ell^{n-1}+d)}d^{\frac{n}{2}}\sigma
\end{aligned}$$

with probability at least $1-\frac{2rn}{d}-\frac{2n}{(\ell^{n-1}+d)^{2\ell^{1-p}-1}}$.

The proof of the case when $d>\ell^{n-1}$ follows from a slight modification of the same argument and the fact that $\phi^2=\sigma^2 d$. ■

## 7. Numerical Experiments

In this section, we evaluate the empirical performance of the proposed tensor CUR decompositions against other state-of-the-art low multilinear rank tensor approximation methods: HOSVD from (De Lathauwer et al., 2000b; Tucker, 1966), sequentially truncated HOSVD (st-HOSVD) from (Vannieuwenhoven et al., 2012), and higher-order orthogonal iteration (HOOI) from (De Lathauwer et al., 2000b; Kroonenberg and De Leeuw, 1980). All tests are conducted on a Ubuntu workstation equipped with Intel i9-9940X CPU and 128GB DDR4 RAM, and executed from Matlab R2020a. We use the implementations of HOSVD, st-HOSVD and HOOI from `tensor_toolbox v3.1`[2]. The codes for Fiber and Chidori CUR decompositions are available online at `https://github.com/caesarcai/Modewise_Tensor_Decomp`.

To implement the tensor CUR decompositions, we set $C_i$ to have size $d_i\times 2r_i\log(\prod_{j\neq i}d_j)$ and $U_i$ to be $r_i\log(d_i)\times 2r_i\log(\prod_{j\neq i}d_j)$ for the Fiber CUR decomposition, and set $C_i$ to have size $d_i\times\prod_{j\neq i}r_j\log(d_j)$ and $U_i$ to be $r_i\log(d_i)\times\prod_{j\neq i}r_j\log(d_j)$ for the Chidori CUR decomposition. The indices used to determine $\mathcal{R}$, $C_i$, and $U_i$ are sampled uniformly at random, and the sampling size comes from the sampling guarantee results of Section 3.3.

### 7.1 Synthetic Dataset

We generate a 3-mode tensor $\mathcal{X}\in\mathbb{R}^{d\times d\times d}$ with rank $(r,r,r)$ by

$$\mathcal{X}:=\mathcal{T}\times_1 G_1\times_2 G_2\times_3 G_3,$$

where the core tensor $\mathcal{T}\in\mathbb{R}^{r\times r\times r}$ and matrices $G_1,G_2,G_3\in\mathbb{R}^{d\times r}$ are random tensor/-matrices with i.i.d. Gaussian distributed entries ($\sim\mathcal{N}(0,1)$). In addition, we generate the

2. Website: `https://gitlab.com/tensors/tensor_toolbox/-/releases/v3.1`.

additive i.i.d. Gaussian noise $\mathcal{E} \in \mathbb{R}^{d\times d\times d}$ with some given variance $\sigma$ so that we have the noisy observation

$$\widetilde{\mathcal{X}} := \mathcal{X} + \mathcal{E}.$$

We evaluate the robustness and computational efficiency of all tested methods under four different noise levels (i.e., $\sigma = 10^{-1}, 10^{-4}, 10^{-7}$, and 0). For all methods, we denote the output of multilinear rank $(r, r, r)$ tensor approximation to be $\mathcal{X}_r$, and the relative approximation error is

$$\text{err} := \frac{\|\mathcal{X}_r - \mathcal{X}\|_F}{\|\mathcal{X}\|_F}.$$

The test results are averaged over 50 trials and summarized in Figure 3. One can see that both variations of the proposed tensor CUR decomposition are substantially faster than all other state-of-the-art methods. In particularly, Fiber CUR achieves over 150× speedup when $d$ is large. In the noiseless case (i.e., $\sigma = 0$), all methods, including the proposed ones, approximate the low multilinear rank tensor with the same accuracy. However, when additive noise appears, the proposed methods have slightly worse but still good approximation accuracy. We provide more detailed and enlarged runtime plots for only the tensor CUR decompositions in Figure 4. As discussed in Section 3.4, we verify that both tensor CUR decompositions have much lower computational complexities than the start-of-the-art, with Chidori CUR being slightly slower than Fiber CUR.

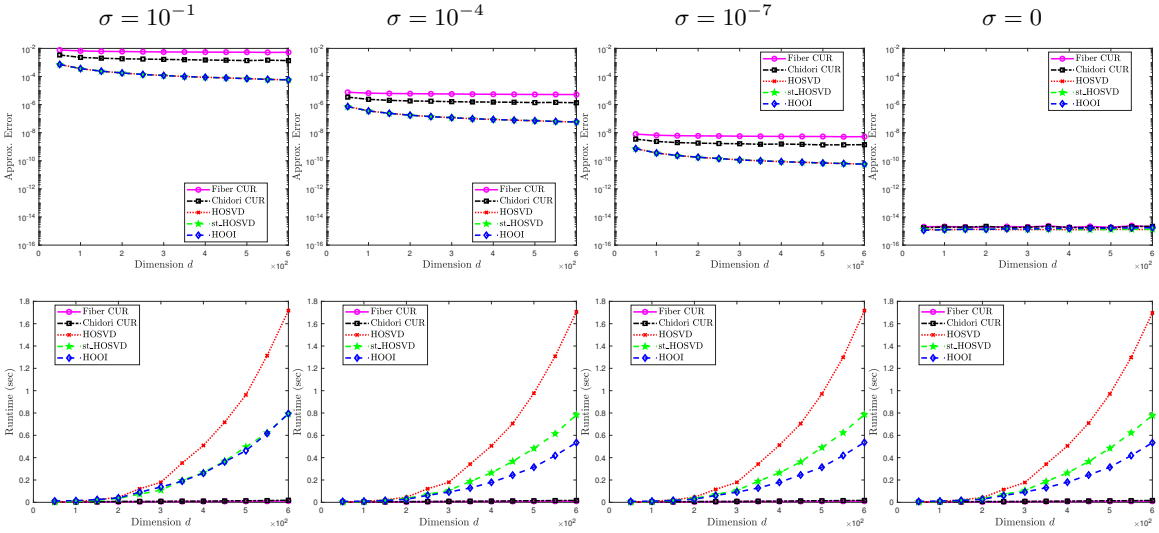


Figure 3: Comparison of low rank tensor approximation methods under different noise levels $\sigma$. Rank $(5, 5, 5)$ is used in all tests and $d$ varies from 50 to 600. **Top row:** relative approximation errors *vs.* tensor dimensions. **Bottom row:** runtime *vs.* tensor dimensions.

**Remark 24** *In Figure 4, the time complexities, with respect to the problem dimension $d$, of both tensor CUR decompositions appear to be $\mathcal{O}(d)$ instead of the claimed $\mathcal{O}(n\log^2(d))$ and $\mathcal{O}(\log^n(d))$ where $n$ is the number of modes of the tensor (see Section 3.4). This is due to the inefficiency of the subtensor/submatrix extraction in current tensor toolbox, even when the extracted data is partially contiguous. If we exclude the time for subtensor/submatrix*

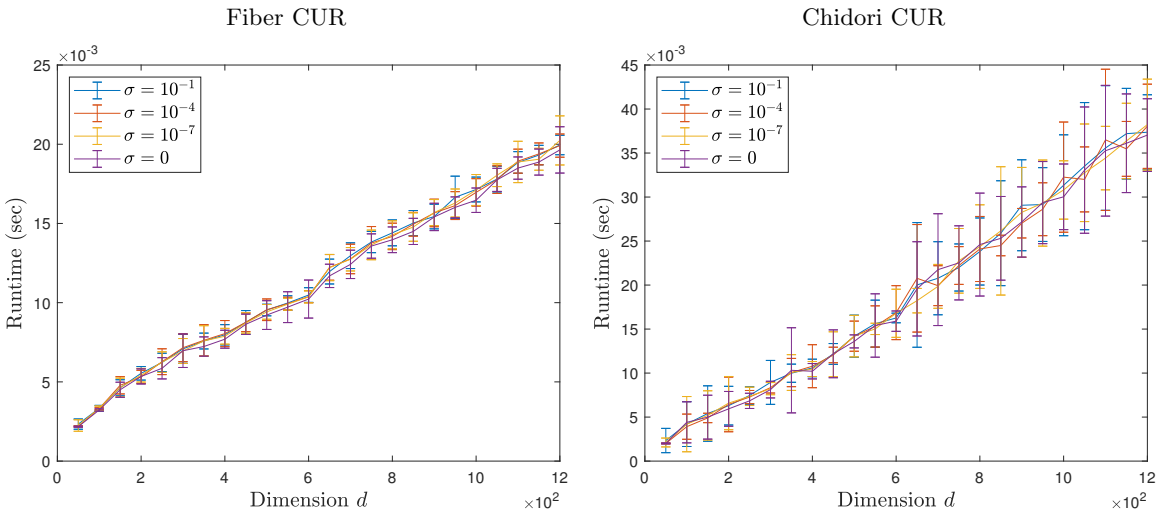


Figure 4: Runtime with variance bar *vs.* tensor dimensions for Fiber and Chidori CUR under different noise levels $\sigma$. Rank $(5, 5, 5)$ is used for all tests and $d$ varies from 50 to 1200.

*extraction, then the runtime plots match the theoretical complexities. The empirical time performance of the tensor CUR decompositions will be further improved with future releases of the tensor toolbox when the subtensor/submatrix extraction becomes more efficient.*

### 7.2 Hyperspectral Image Compression

Hyperspectral image compression is a standard real-world benchmark for low multilinear rank tensor approximations (Mahoney et al., 2008; Zhang et al., 2015; Du et al., 2016; Li et al., 2021). We consider the use of the Fiber and Chidori CUR decompositions for hyperspectral image compression on three benchmark datasets from (Foster et al., 2004): Ribeira, Braga, and Ruivaes[3]. The runtime and approximation performance of the tensor CUR decompositions are compared against the other state-of-the-art methods. Approximation performance is evaluated by signal-to-noise ratio (SNR) defined by

$$\mathrm{SNR}_{\mathrm{dB}} = 10 \log \left( \frac{\|\mathcal{X}\|_F^2}{\|\mathcal{X} - \mathcal{X}_r\|_F^2} \right),$$

where $\mathcal{X}$ is the color image corresponding to the original hyperspectral data and $\mathcal{X}_r$ corresponds to the compressed data. The experimental results, along with the size and rank information of the dataset, are summarized in Table 2. In particular, we determine the ranks based on the original hyperspectral images' HOSVD where we try our best to balance between compression efficiency and quality. One can see both tensor CUR decompositions yield drastically improved speed performance with Fiber CUR being the fastest of the two variants. On the other hand, Chidori CUR achieves superior approximation results, and Fiber CUR is still very competitive. These datasets are in the form of tall tensors with

3. The datasets can be found online at `https://personalpages.manchester.ac.uk/staff/d.h.foster/Hyperspectral_images_of_natural_scenes_04.html`.

Table 2: Runtime and relative error.

| | | Ribeira | Braga | Ruivaes |
|---|---|---|---|---|
| Size | | $1017 \times 1340 \times 33$ | $1021 \times 1338 \times 33$ | $1017 \times 1338 \times 33$ |
| Rank | | $(60, 60, 7)$ | $(60, 60, 5)$ | $(65, 65, 4)$ |
| Runtime (seconds) | Fiber CUR | **0.29** | **0.26** | **0.31** |
| | Chidori CUR | 0.66 | 0.59 | 0.55 |
| | HOSVD | 1.49 | 1.41 | 1.42 |
| | st_HOSVD | 0.83 | 0.77 | 0.76 |
| | HOOI | 2.29 | 2.67 | 3.30 |
| SNR (dB) | Fiber CUR | 24.14 | 17.93 | 15.53 |
| | Chidori CUR | **24.39** | **18.56** | **15.84** |
| | HOSVD | 22.99 | 17.70 | 15.48 |
| | st_HOSVD | 22.18 | 17.90 | 15.49 |
| | HOOI | 24.33 | 18.00 | 15.61 |

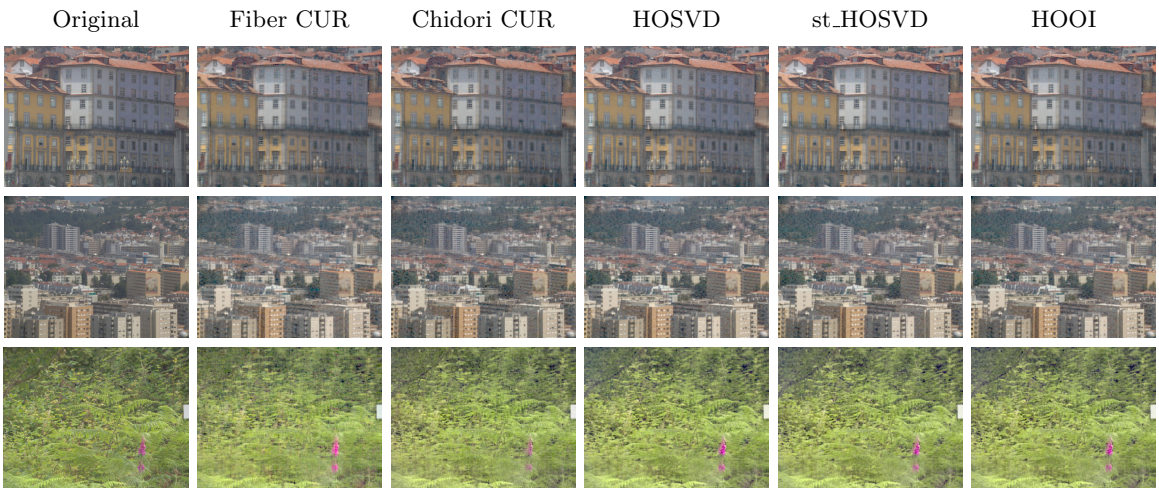


Figure 5: Visual comparison of the original and compressed hyperspectral images. From top to bottom, each row of the images are for the datasets Ribeira, Braga and Ruivaes, respectively.

relatively high ranks, which are conditions typically adverse to CUR methods. Nonetheless, tensor CUR decompositions were able to achieve remarkable advantages in this real-world test.

Finally, we present the visual comparison of the compression results in Figure 5. We find all methods achieve visually good compression regardless of SNR. Once again, the tensor CUR decompositions are able to finish the compression task in a much shorted time without substantially sacrificing quality.

## 8. Conclusions and Future Prospects

While extensions of CUR decompositions to tensors are not entirely new, this work gives some of the first nontrivial bounds for quality of approximation of low multilinear rank tensors via two types of approximations: Fiber and Chidori CUR. This was achieved through considering arbitrary tensors as perturbations of low multilinear rank ones. As additional points of interest, we characterized these decompositions and provided error estimates and exact decomposition guarantees under a simple random sampling scheme on the indices.

We also demonstrated that the Fiber and Chidori CUR decompositions obtained via uniformly randomly sampling incoherent tensors are significantly faster than state-of-the-art low-rank tensor approximations without sacrificing quality of reconstruction on both synthetic and benchmark hyperspectral image data sets.

Given the success of matrix CUR decompositions in a wide range of applications and the ubiquity of tensor data, we expect that tensor CUR decompositions, including those discussed here, will become standard tools for practitioners. In the future, it would be of interest to understand how the idea of reconstructing a tensor from fibers and subtensors of it can be applied to yield fast algorithms for robust decompositions (similar to robust PCA for matrices) and tensor completion.

## Acknowledgments

DN and LH were supported by NSF DMS #2011140 and NSF BIGDATA #1740325. KH was sponsored in part by the Army Research Office under grant number W911NF-20-1-0076. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

We thank Amy Hamm Design for the production of Figures 1 and 2, and Dustin Mixon for suggesting the name Chidori.

## References

Evrim Acar and Bülent Yener. Unsupervised multiway data analysis: A literature survey. *IEEE Transactions on Knowledge and Data Engineering*, 21(1):6–20, 2008.

Salman Ahmadi-Asl, Stanislav Abukhovich, Maame G Asante-Mensah, Andrzej Cichocki, Anh Huy Phan, Tohishisa Tanaka, and Ivan Oseledets. Randomized algorithms for computation of Tucker decomposition and higher order SVD (HOSVD). *IEEE Access*, 9: 28684–28706, 2021.

Akram Aldroubi, Ali Sekmen, Ahmet Bugra Koku, and Ahmet Faruk Çakmak. Similarity matrix framework for data from union of subspaces. *Applied and Computational Harmonic Analysis*, 45(2):425–435, 2018.

Akram Aldroubi, Keaton Hamm, Ahmet Bugra Koku, and Ali Sekmen. CUR decompositions, similarity matrices, and subspace clustering. *Frontiers in Applied Mathematics and Statistics*, 4:65, 2019.

Casey Battaglino, Grey Ballard, and Tamara G Kolda. A practical randomized CP tensor decomposition. *SIAM Journal on Matrix Analysis and Applications*, 39(2):876–901, 2018.

Andrea L Bertozzi and Arjuna Flenner. Diffuse interface models on graphs for classification of high dimensional data. *SIAM Review*, 58(2):293–328, 2016.

Christos Boutsidis and David P Woodruff. Optimal CUR matrix decompositions. *SIAM Journal on Computing*, 46(2):543–589, 2017.

HanQin Cai, Jian-Feng Cai, and Ke Wei. Accelerated alternating projections for robust principal component analysis. *The Journal of Machine Learning Research*, 20(1):685–717, 2019.

HanQin Cai, Keaton Hamm, Longxiu Huang, Jiaqi Li, and Tao Wang. Rapid robust principal component analysis: CUR accelerated inexact low rank estimation. *IEEE Signal Processing Letters*, 28:116–120, 2021a.

HanQin Cai, Keaton Hamm, Longxiu Huang, and Deanna Needell. Robust CUR decomposition: Theory and imaging applications. *arXiv preprint arXiv:2101.05231*, 2021b.

Cesar F Caiafa and Andrzej Cichocki. Generalizing the column–row matrix decomposition to multi-way arrays. *Linear Algebra and its Applications*, 433(3):557–573, 2010.

Emmanuel Candes and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969, 2007.

Maolin Che, Yimin Wei, and Hong Yan. Randomized algorithms for the low multilinear rank approximations of tensors. *Journal of Computational and Applied Mathematics*, 390: 113380, 2021.

Dehua Cheng, Richard Peng, Yan Liu, and Ioakeim Perros. SPALS: Fast alternating least squares via implicit leverage scores sampling. *Advances in Neural Information Processing Systems*, 29:721–729, 2016.

Jiawei Chiu and Laurent Demanet. Sublinear randomized algorithms for skeleton decompositions. *SIAM Journal on Matrix Analysis and Applications*, 34(3):1361–1383, 2013.

Ali Civril and Malik Magdon-Ismail. On selecting a maximum volume sub-matrix of a matrix and related problems. *Theoretical Computer Science*, 410(47-49):4801–4811, 2009.

Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000a.

Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. On the best rank-1 and rank-$(r_1, ..., r_n)$ approximation of higher-order tensors. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1324–1342, 2000b.

Petros Drineas, Ravi Kannan, and Michael W Mahoney. Fast Monte Carlo algorithms for matrices III: Computing a compressed approximate matrix decomposition. *SIAM Journal on Computing*, 36(1):184–206, 2006.

Petros Drineas, Michael W Mahoney, and S Muthukrishnan. Relative-error CUR matrix decompositions. *SIAM Journal on Matrix Analysis and Applications*, 30(2):844–881, 2008.

Bo Du, Mengfei Zhang, Lefei Zhang, Ruimin Hu, and Dacheng Tao. PLTD: Patch-based low-rank tensor decomposition for hyperspectral images. *IEEE Transactions on Multimedia*, 19(1):67–79, 2016.

N Benjamin Erichson, Krithika Manohar, Steven L Brunton, and J Nathan Kutz. Randomized CP tensor decomposition. *Machine Learning: Science and Technology*, 1(2):025012, 2020.

David H Foster, Sérgio MC Nascimento, and Kinjiro Amano. Information limits on neural identification of colored surfaces in natural scenes. *Visual neuroscience*, 21(3):331, 2004.

Alan Frieze, Ravi Kannan, and Santosh Vempala. Fast Monte-Carlo algorithms for finding low-rank approximations. *Journal of the ACM (JACM)*, 51(6):1025–1041, 2004.

Alex Gittens and Michael W Mahoney. Revisiting the Nyström method for improved large-scale machine learning. *The Journal of Machine Learning Research*, 17(1):3977–4041, 2016.

Alex Gittens, Kareem Aggour, and Bülent Yener. Adaptive sketching for fast and convergent Canonical Polyadic decomposition. In *International Conference on Machine Learning*, pages 3566–3575. PMLR, 2020.

Sergei A Goreinov, Ivan V Oseledets, Dimitry V Savostyanov, Eugene E Tyrtyshnikov, and Nikolay L Zamarashkin. How to find a good submatrix. In *Matrix Methods: Theory, Algorithms And Applications: Dedicated to the Memory of Gene Golub*, pages 247–256. World Scientific, 2010.

Sergei A. Goreĭnov, Nikolai L. Zamarashkin, and Eugene E. Tyrtyshnikov. Pseudo-skeleton approximations. *Doklay Akdemii Nauk*, 343(2):151–152, 1995.

Sergei A Goreĭnov, Nikolai Leonidovich Zamarashkin, and Evgenii Evgen'evich Tyrtyshnikov. Pseudo-skeleton approximations by matrices of maximal volume. *Mathematical Notes*, 62(4):515–519, 1997.

Lars Grasedyck. Hierarchical singular value decomposition of tensors. *SIAM Journal on Matrix Analysis and Applications*, 31(4):2029–2054, 2010.

Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.

Keaton Hamm and Longxiu Huang. Stability of sampling for CUR decompositions. *Foundations of Data Science*, 2(2):83–99, 2020a.

Keaton Hamm and Longxiu Huang. Perspectives on CUR decompositions. *Applied and Computational Harmonic Analysis*, 48(3):1088–1099, 2020b.

Keaton Hamm and Longxiu Huang. Perturbations of CUR decompositions. *SIAM Journal on Matrix Analysis and Applications*, 42(1):351–375, 2021.

Frank L Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematical Physics*, 6(1-4):164–189, 1927.

Frank L Hitchcock. Multiple invariants and generalized rank of a $p$-way matrix or tensor. *Journal of Mathematical Physics*, 7(1-4):39–79, 1928.

Misha E Kilmer, Karen Braman, Ning Hao, and Randy C Hoover. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM Journal on Matrix Analysis and Applications*, 34(1):148–172, 2013.

Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

Pieter M Kroonenberg and Jan De Leeuw. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45(1):69–97, 1980.

Joseph B Kruskal. Rank, decomposition, and uniqueness for 3-way and N-way arrays. *Multiway Data Analysis*, pages 7–18, 1989.

Rui Li, Zhibin Pan, Yang Wang, and Ping Wang. The correlation-based Tucker decomposition for hyperspectral image compression. *Neurocomputing*, 419:357–370, 2021.

Michael W Mahoney and Petros Drineas. CUR matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences*, 106(3):697–702, 2009.

Michael W Mahoney, Mauro Maggioni, and Petros Drineas. Tensor-CUR decompositions for tensor-based data. *SIAM Journal on Matrix Analysis and Applications*, 30(3):957–987, 2008.

George Matsaglia and George PH Styan. Equalities and inequalities for ranks of matrices. *Linear and multilinear Algebra*, 2(3):269–292, 1974.

Aleksandr Mikhalev and Ivan V Oseledets. Rectangular maximum-volume submatrices and their applications. *Linear Algebra and its Applications*, 538:187–211, 2018.

Larsson Omberg, Joel R Meyerson, Kayta Kobayashi, Lucy S Drury, John FX Diffley, and Orly Alter. Global effects of DNA replication and DNA replication origin activity on eukaryotic gene expression. *Molecular systems biology*, 5(1):312, 2009.

Ivan V Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33 (5):2295–2317, 2011.

R. Penrose. On best approximate solutions of linear matrix equations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 52(1):17–19, 1956.

Anh-Huy Phan, Petr Tichavskỳ, and Andrzej Cichocki. CANDECOMP/PARAFAC decomposition of high-order tensors through tensor reshaping. *IEEE Transactions on Signal Processing*, 61(19):4847–4860, 2013.

Mark Rudelson and Roman Vershynin. Sampling from large matrices: An approach through geometric functional analysis. *Journal of the ACM (JACM)*, 54(4):21–es, 2007.

Nicholas D Sidiropoulos, Lieven De Lathauwer, Xiao Fu, Kejun Huang, Evangelos E Papalexakis, and Christos Faloutsos. Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65(13):3551–3582, 2017.

Zhao Song, David P Woodruff, and Peilin Zhong. Relative error tensor low rank approximation. In *Proc. ACM-SIAM Symposium on Discrete Algorithms*, pages 2772–2789. SIAM, 2019.

Danny C Sorensen and Mark Embree. A DEIM induced CUR factorization. *SIAM Journal on Scientific Computing*, 38(3):A1454–A1482, 2016.

Ameet Talwalkar and Afshin Rostamizadeh. Matrix coherence and the Nyström method. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, pages 572–579, 2010.

Joel A Tropp. Improved analysis of the subsampled randomized Hadamard transform. *Advances in Adaptive Data Analysis*, 3(01n02):115–126, 2011.

Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434, 2012.

Ledyard R Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

Nick Vannieuwenhoven, Raf Vandebril, and Karl Meerbergen. A new truncation strategy for the higher-order singular value decomposition. *SIAM Journal on Scientific Computing*, 34(2):A1027–A1052, 2012.

M Alex O Vasilescu and Demetri Terzopoulos. Multilinear analysis of image ensembles: Tensorfaces. In *European conference on computer vision*, pages 447–460, 2002.

Sergey Voronin and Per-Gunnar Martinsson. Efficient algorithms for CUR and interpolative matrix decompositions. *Advances in Computational Mathematics*, 43(3):495–516, 2017.

Lele Wang, Kun Xie, Thabo Semong, and Huibin Zhou. Missing data recovery based on tensor-CUR decomposition. *IEEE Access*, 6:532–544, 2017.

Shusen Wang and Zhihua Zhang. Improving CUR matrix decomposition and the Nyström approximation via adaptive sampling. *The Journal of Machine Learning Research*, 14(1): 2729–2769, 2013.

Christopher Williams and Matthias Seeger. Using the Nyström method to speed up kernel machines. In *Proceedings of the 14th annual conference on neural information processing systems*, pages 682–688, 2001.

David Woodruff. Sketching as a tool for numerical linear algebra. *Foundations and Trends in Theoretical Computer Science*, 10(1–2):1–157, 2014.

Dong Xia, Ming Yuan, Cun-Hui Zhang, et al. Statistically optimal and computationally efficient low rank tensor completion from noisy entries. *Annals of Statistics*, 49(1):76–99, 2021.

Shuicheng Yan, Dong Xu, Qiang Yang, Lei Zhang, Xiaoou Tang, and Hong-Jiang Zhang. Multilinear discriminant analysis for face recognition. *IEEE Transactions on Image Processing*, 16(1):212–220, 2006.

Jiyan Yang, Oliver Rubel, Michael W Mahoney, and Benjamin P Bowen. Identifying important ions and positions in mass spectrometry imaging data using CUR matrix decompositions. *Analytical Chemistry*, 87(9):4658–4666, 2015.

Bülent Yener, Evrim Acar, Pheadra Aguis, Kristin Bennett, Scott L Vandenberg, and George E Plopper. Multiway modeling and analysis in stem cell systems biology. *BMC Systems Biology*, 2(1):63, 2008.

Ching-Wa Yip, Michael W Mahoney, Alexander S Szalay, István Csabai, Tamás Budavári, Rosemary FG Wyse, and Laszlo Dobos. Objective identification of informative wavelength regions in galaxy spectra. *The Astronomical Journal*, 147(5):110, 2014.

Ali Zare, Alp Ozdemir, Mark A Iwen, and Selin Aviyente. Extension of PCA to higher order data structures: An introduction to tensors, tensor decompositions, and tensor PCA. *Proceedings of the IEEE*, 106(8):1341–1358, 2018.

Lefei Zhang, Liangpei Zhang, Dacheng Tao, Xin Huang, and Bo Du. Compression of hyperspectral remote sensing images by tensor approach. *Neurocomputing*, 147:358–363, 2015.