Journal of Machine Learning Research 21 (2020) 1-36 Submitted 5/19; Revised 3/20; Published 3/20

# Union of Low-Rank Tensor Spaces: Clustering and Completion

**Morteza Ashraphijuo** MA3429@COLUMBIA.EDU
*Columbia University*
*New York, NY 10027, USA*

**Xiaodong Wang** XW2008@COLUMBIA.EDU
*Columbia University*
*New York, NY 10027, USA*

**Editor:** Sanjiv Kumar

## Abstract

We consider the problem of clustering and completing a set of tensors with missing data that are drawn from a union of low-rank tensor spaces. In the clustering problem, given a partially sampled tensor data that is composed of a number of subtensors, each chosen from one of a certain number of unknown tensor spaces, we need to group the subtensors that belong to the same tensor space. We provide a geometrical analysis on the sampling pattern and subsequently derive the sampling rate that guarantees the correct clustering under some assumptions with high probability. Moreover, we investigate the fundamental conditions for finite/unique completability for the union of tensor spaces completion problem. Both deterministic and probabilistic conditions on the sampling pattern to ensure finite/unique completability are obtained. For both the clustering and completion problems, our tensor analysis provides significantly better bound than the bound given by the matrix analysis applied to any unfolding of the tensor data.

**Keywords:** Low-rank tensor completion, canonical polyadic (CP) decomposition, union of tensor spaces, clustering tensor spaces, finite completability, unique completability.

## 1. Introduction

Identifying the geometrical properties and relationships of datasets is essential for many data processing tasks, such as completion and denoising. In many applications, we need to analyze a collection of datasets like images, text documents, etc. Assuming that these are two-dimensional (i.e., two dimensions are enough to represent all features and relationships captured in the data), then to model such data structures, we can simply consider a matrix $\mathbf{U} \in \mathbb{R}^{n_1 \times n_2}$ whose columns are chosen from one of $K$ unknown two-dimensional subspaces. The problem of subspace clustering aims to cluster the columns of this matrix into $K$ groups such that the columns in each group belong to the same subspace. Subspace clustering is an important pre-processing step of data analysis when the data lies in a union of subspaces and is well studied (Elhamifar and Vidal, 2009; Liu et al., 2013; Elhamifar and Vidal, 2013). The problem is much more challenging with missing data, i.e., when the matrix $\mathbf{U}$ is incomplete, which is an important problem in subspace learning for real-world scenarios and has been treated broadly (Balzano et al., 2012; Pimentel-Alarcón et al., 2014; Yang et al., 2015). In particular, the information-theoretic bounds on clustering the union of subspaces is investigated in (Pimentel-Alarcón et al., 2016a, 2017; Pimentel-Alarcón and Nowak, 2016; Ashraphijuo and Wang, 2019a), based on the geometrical analysis for matrix completion problems (Pimentel-Alarcón et al., 2016b; Ashraphijuo and Wang, 2019b).

The theoretical analyses for the clustering problem in the literature focus on two-dimensional data. On the other hand, in many applications we need to deal with data that are represented with multiple dimensions in order to capture the correlations across different attributes. Even though a naive way of treating tensors is to "collapse" a tensor into a matrix and then apply matrix analysis, such an approach can not fully exploit the multi-way correlation of the data. In contrast, tensor analysis is capable of taking full advantage of these correlations. Therefore, the problem of clustering a union of low-rank tensor spaces with missing data is a more fundamental problem, which will be studied in this paper. Note that two of the main results of this paper are characterizing the minimum number of samples (i.e., minimum sampling rate) required for clustering a union of tensors with missing data and completing a union of tensors. The low-rank tensor clustering and tensor completion problems with missing data are both important topics in the broad area of machine learning. In particular, the low-rank subspace clustering has many applications in various fields including image processing (Hong et al., 2006), recommender systems (Rennie and Srebro, 2005; Pezeshkpour et al., 2018), gene expressions (Gan et al., 2007), etc. The low-rank tensor completion problem plays a vital role in multilinear data analysis against outliers, gross corruptions and missing values and its diverse applications (Kressner et al., 2014; Zhang et al., 2015). Particularly, low-rank tensor completion finds applications in 3D image reconstruction (Sauve et al., 1999), video inpainting (Patwardhan et al., 2007), hyperspectral data recovery (Li and Li, 2010), higher-order web link analysis (Kolda et al., 2005), etc. In this paper, we provide a fundamental theoretical analysis for the two important mentioned problems. One of the main ideas behind our analysis for the tensor space clustering problem is to take advantage of the condition on the sampling rate that guarantees the unique completability for the tensor completion problem. Then, we use the unique completability property on each of the tensors to correctly identify whether a tensor space fits in that tensor. Moreover, algorithms such as the one developed in (Ashraphijuo et al., 2019) can be developed for achieving such theoretical bounds and the analysis in this paper can be used for rank estimation as studied in (Ashraphijuo et al., 2017).

Since we study the data structures that are partially sampled, another important related problem is the low-rank data retrieval problem and it has many applications in different areas including compressed sensing (Lim and Comon, 2010; Sidiropoulos and Kyrillidis, 2012; Gandy et al., 2011), network coding (Harvey et al., 2005), image processing (Candès et al., 2013; Ji et al., 2010) and data mining (Eldén, 2007) and some literature reviews on this problem can be found in (Candès and Recht, 2009; Candès and Tao, 2010; Cai et al., 2010). Hence, we also study the union of tensor spaces retrieval problem and derive the fundamental conditions on the geometry of the sampling pattern for finite/unique completability based on a polynomial analysis. Then, given the characterized deterministic conditions on the sampling pattern, we provide a probabilistic analysis to obtain a bound on the sampling rate to ensure that these deterministic conditions hold true with high probability.

The remainder of this paper is organized as follows. In Section 2, some preliminaries are presented. In Section 3, we state the two problems treated in this paper, i.e., tensor space clustering with missing data and union of tensor spaces completion, and provide the matrix analysis approach to these problems. Tensor analysis for clustering tensor spaces is presented in Section 4. Then, in Sections 5 and 6, we provide deterministic and probabilistic analyses on the finite/unique completability for union of tensor spaces completion, respectively. Finally, Section 7 concludes the paper.

## 2. Preliminaries

Assume that $\mathcal{U} \in \mathbb{R}^{n_1 \times n_2 \times \ldots n_{d-1} \times n_d}$ is a $d$-way tensor. Throughout this paper, we use the CP rank as the rank of a tensor, which is defined as the minimum number $r$ such that there exist $\mathbf{a}_i^l \in \mathbb{R}^{n_i}$ for $1 \leq i \leq d$ and $1 \leq l \leq r$ and

$$\mathcal{U} = \sum_{l=1}^{r} \mathbf{a}_1^l \otimes \mathbf{a}_2^l \otimes \ldots \otimes \mathbf{a}_d^l, \tag{1}$$

or equivalently,

$$\mathcal{U}(x_1, x_2, \ldots, x_d) = \sum_{l=1}^{r} \mathbf{a}_1^l(x_1)\mathbf{a}_2^l(x_2) \ldots \mathbf{a}_d^l(x_d), \tag{2}$$

where $\otimes$ denotes the tensor product (outer product) and $\mathcal{U}(x_1, x_2, \ldots, x_d)$ denotes the entry of tensor $\mathcal{U}$ with coordinate $\vec{x} = (x_1, x_2, \ldots, x_d)$ and $\mathbf{a}_i^l(x_i)$ denotes the $x_i$-th entry of vector $\mathbf{a}_i^l$. Note that $\mathbf{a}_1^l \otimes \mathbf{a}_2^l \otimes \ldots \otimes \mathbf{a}_d^l \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is a rank-1 tensor, $l = 1, 2, \ldots, r$.

For notational convenience, define $N_{-i} \triangleq \frac{\Pi_{j=1}^d n_j}{n_i}$ (i.e., the product of all but the size of $i$-th dimension), $N_i \triangleq \left(\Pi_{j=1}^i n_j\right)$ (i.e., the product of the sizes of the first $i$ dimensions), $\bar{N}_i \triangleq \left(\Pi_{j=i+1}^d n_j\right)$ (i.e., the product of the sizes of the last $(d-i)$ dimensions). Also, define $x^+ \triangleq \max\{0, x\}$.

**Definition 1** *Define the matrix $\widetilde{\mathbf{U}}_{(i)} \in \mathbb{R}^{N_i \times \bar{N}_i}$ as the $i$-th unfolding of tensor $\mathcal{U}$, such that $\mathcal{U}(\vec{x}) = \widetilde{\mathbf{U}}_{(i)}(\widetilde{M}_i(x_1, \ldots, x_i),\ \bar{M}_i(x_{i+1}, \ldots, x_d))$, where $\widetilde{M}_i : (x_1, \ldots, x_i) \rightarrow \{1, 2, \ldots, N_i\}$ and $\bar{M}_i : (x_{i+1}, \ldots, x_d) \rightarrow \{1, 2, \ldots, \bar{N}_i\}$ are two bijective mappings. These mappings are basically the simple and well-known vectorization mappings.*

**Definition 2** *Define the matrix $\mathbf{U}_{(i)} \in \mathbb{R}^{n_i \times N_{-i}}$ as the $i$-th matricization of tensor $\mathcal{U}$, such that $\mathcal{U}(\vec{x}) = \mathbf{U}_{(i)}(x_i, M_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d))$, where $M_i : (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d) \rightarrow \{1, 2, \ldots, N_{-i}\}$ is a bijective mapping, which is another vectorization mapping.*

**Definition 3** *We call a column $\mathbf{u} \in \mathbb{R}^{N_i}$ an "$i$-th unfolded column," $i = 1, \ldots, d-1$, if there exist $\mathbf{u}_s \in \mathbb{R}^{n_s}$ for $s = 1, \ldots, i$, such that $\mathbf{u} = vec(\mathbf{u}_1 \otimes \ldots \otimes \mathbf{u}_i)$, where $vec(\cdot)$ uses the same bijective mappings that is used for the $i$-th unfolding, i.e., $\widetilde{M}_i : (x_1, \ldots, x_i) \rightarrow \{1, 2, \ldots, N_i\}$, and hence $\mathbf{u}(\widetilde{M}_i((x_1, \ldots, x_i)) = \mathbf{u}_1(x_1) \ldots \mathbf{u}_i(x_i)$. We call a $(d-1)$-th unfolded column as a "structured column."*

In the remainder of this section, we provide some fundamental properties of tensor and rank that will be used later.

**Lemma 4** *The matrix rank of any unfolding or matricization of $\mathcal{U}$ is upper bounded by its CP-rank $r$. In other words, the CP-rank of a tensor is always greater than or equal to the matrix rank of any unfolding or matricization.*

**Proof** In order to show rank $\left(\widetilde{\mathbf{U}}_{(i)}\right) \leq r$, it suffices to show that there exist $\mathbf{b}_1^l \in \mathbb{R}^{N_i}$ and $\mathbf{b}_2^l \in \mathbb{R}^{\bar{N}_i}$ for $1 \leq l \leq r$ such that

$$\widetilde{\mathbf{U}}_{(i)} = \sum_{l=1}^{r} \mathbf{b}_1^l \otimes \mathbf{b}_2^l. \tag{3}$$

Recall the CP decomposition in (1). Then, we define $\mathcal{A}_1^l = \mathbf{a}_1^l \otimes \ldots \otimes \mathbf{a}_i^l$ and $\mathcal{A}_2^l = \mathbf{a}_{i+1}^l \otimes \ldots \otimes \mathbf{a}_d^l$ for $1 \leq l \leq l$ and let $\mathbf{b}_1^l$ and $\mathbf{b}_2^l$ denote the vectorizations of $\mathcal{A}_1^l$ and $\mathcal{A}_2^l$ with the bijective mappings $\widetilde{M}_i : (x_1, \ldots, x_i) \rightarrow \{1, 2, \ldots, N_i\}$ and $\bar{M}_i : (x_{i+1}, \ldots, x_d) \rightarrow \{1, 2, \ldots, \bar{N}_i\}$ of the unfolding $\widetilde{\mathbf{U}}_{(i)}$, respectively. Hence, there exist $\mathbf{b}_1^l \in \mathbb{R}^{N_i}$ and $\mathbf{b}_2^l \in \mathbb{R}^{\bar{N}_i}$ for $1 \leq l \leq r$ such that (3) holds. Similarly, for the matricization $\mathbf{U}_{(i)}$, we can write $\mathbf{U}_{(i)} = \sum_{l=1}^{r} \mathbf{c}_1^l \otimes \mathbf{c}_2^l$, where $\mathbf{c}_1^l = \mathbf{a}_i^l \in \mathbb{R}^{n_i}$ and $\mathbf{c}_2^l \in \mathbb{R}^{N_{-i}}$ is the vectorization of $\mathbf{a}_1^l \otimes \ldots \otimes \mathbf{a}_{i-1}^l \otimes \mathbf{a}_{i+1}^l \otimes \ldots \otimes \mathbf{a}_d^l$ with the bijective mapping $M_i : (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d) \rightarrow \{1, 2, \ldots, N_{-i}\}$ of the matricization $\mathbf{U}_{(i)}$. ■

**Lemma 5** *The CP-rank of a tensor $\mathcal{U}$ is equal to the* **minimum number of structured columns** *that span all columns of $\widetilde{\mathbf{U}}_{(d-1)}$.*

**Proof** Recall from Lemma 4 that there exist $\mathbf{b}_1^l \in \mathbb{R}^{N_i}$ and $\mathbf{b}_2^l \in \mathbb{R}^{\bar{N}_i}$ for $1 \leq l \leq r$ such that (3) holds. Define $\mathbf{B}_1 = [\mathbf{b}_1^1|\mathbf{b}_1^2|\ldots|\mathbf{b}_1^r] \in \mathbb{R}^{N_i \times r}$ and $\mathbf{B}_2 = [\mathbf{b}_2^1|\mathbf{b}_2^2|\ldots|\mathbf{b}_2^r]^\top \in \mathbb{R}^{r \times \bar{N}_i}$. Then, (3) can be rewritten as $\widetilde{\mathbf{U}}_{(i)} = \mathbf{B}_1 \mathbf{B}_2$ and therefore, there exist $r$ columns such that each one is an $i$-th unfolded column (columns of $\mathbf{B}_1$) and each column of the $i$-th unfolding of $\mathcal{U}$ can be written as a linear combination of the mentioned $r$ columns. Moreover, the columns of $\mathbf{B}_2^\top$ have a similar structure, i.e., they are $(d-i)$-th unfolded columns. Note that in the case of $i = d-1$, the columns of $\mathbf{B}_1$ are "structured columns" and the columns of $\mathbf{B}_2^\top$ do no have any particular structure, i.e., $\mathbf{B}_2$ is an arbitrary matrix. *Therefore, rank($\mathcal{U}$) $= r$ means that $r$ is the minimum number of structured columns that span all columns of $\widetilde{\mathbf{U}}_{(d-1)}$.* ■

**Remark 6** *rank($\mathcal{U}$) $= r$ concludes that there exists a set $\mathcal{S}$ consisting of $r$ structured columns whose column span (denoted by $\mathcal{T}$) includes any column of $\widetilde{\mathbf{U}}_{(d-1)}$. In other words, the column span of these $r$ structured columns, i.e., $\mathcal{T}$, is an unfolded tensor space of rank $r$. In fact, the image of the bijective mapping $(\widetilde{M}_{d-1}, \bar{M}_{d-1})$ with the domain of space of all $d$-way tensors generated by this unfolded tensor space is $\mathcal{T}$.*

## 3. Problem Statements, Main Results And Matrix Approaches

### 3.1. Problem Statements

Assume that $\mathcal{U}_k \in \mathbb{R}^{n_1 \times n_2 \times \ldots n_{d-1} \times c_k}$ is a $d$-way tensor, $k = 1, 2, \ldots, K$. Define $\mathcal{U} \in \mathbb{R}^{n_1 \times n_2 \times \ldots, n_d}$ as the concatenation of the mentioned tensors along the $d$-th dimension (not in any specific order), where $n_d = \sum_{k=1}^{K} c_k$. Let $r_k$ denote the rank of $\mathcal{U}_k$, $k = 1, 2, \ldots, K$. We assume that $\mathcal{U}$ is (or $\mathcal{U}_k$'s are) randomly sampled, i.e., each entry of $\mathcal{U}$ is independently sampled with probability $0 < p < 1$. Let $\Omega$ be an $n_1 \times n_2 \times \ldots \times n_d$ binary tensor of sampling pattern such that $\Omega(\vec{x}) = 1$ if $\mathcal{U}(\vec{x})$ is sampled and $\Omega(\vec{x}) = 0$ otherwise. Let $\mathcal{U}_\Omega$ denote the incomplete tensor consisting of only the sampled entries of $\mathcal{U}$. We assume that the rank values $r_1, \dots, r_K$ and the sampled tensor $\mathcal{U}_\Omega$ are given. Furthermore, we assume that $\mathcal{U}_k$ is generated generically from the corresponding tensor space (of rank $r_k$), i.e., each entry of the vector along the $d$-th dimension in CP-decomposition of $\mathcal{U}_k$ is drawn independently according to a continuous uniform distribution with respect to the Lebesgue measure on $\mathbb{R}$, $k = 1, \dots, K$. Denote $C_k = c_1 + \cdots + c_k$ for $k = 1, \dots, K$. Throughout the paper, we assume the genericty of the data and that the entries of the tensor are sampled uniformly and independently with some probability $p$.

For the sake of notational simplicity, instead of working with tensors, we look into the image of the bijective mapping $(\widetilde{M}_{d-1}, \bar{M}_{d-1})$ with the domain $\mathcal{U}$, i.e., $\widetilde{\mathbf{U}}_{(d-1)}$. Specifically, consider the set $\mathcal{S}_k$ consisting of $r_k$ structured columns (i.e., $(d-1)$-th unfolded columns) that are chosen generically, $k = 1, \dots, K$. In other words, each of them is obtained through vectorization of the outer product of $d-1$ vectors in $\mathbb{R}^{n_i}$, for $1 \leq i \leq d-1$, and each entry of any of these $d-1$ vectors is drawn independently according to a continuous uniform distribution with respect to the Lebesgue measure on $\mathbb{R}$. Moreover, let $\mathcal{T}_k$ denote the column span of $\mathcal{S}_k$. Then, consider a matrix $\widetilde{\mathbf{U}}^k_{(d-1)} \in \mathbb{R}^{N_{d-1} \times c_k}$ such that $c_k \geq r_k$ and the columns of $\widetilde{\mathbf{U}}^k_{(d-1)}$ are drawn generically from the column span of $\mathcal{S}_k$. Hence, the folded $d$-way tensor corresponding to $\widetilde{\mathbf{U}}^k_{(d-1)}$ (denoted by $\mathcal{U}_k$) is of CP-rank $r_k$ and the column span of $\mathcal{S}_k$ provides an unfolded tensor space of rank $r_i$, $i = 1, \dots, k$.
Summary of notations:

| Symbol | Definition |
|---|---|
| $K$ | The number of tensors. |
| $\mathcal{U}_k$ | The $k$-th tensor, which belongs to $\mathbb{R}^{n_1 \times n_2 \times \dots n_{d-1} \times c_k}$. |
| $\Omega$ | The binary sampling tensor, which belongs to $\{0,1\}^{n_1 \times n_2 \times \dots n_{d-1} \times n_d}$. |
| $C_k$ | $C_k = c_1 + \cdots + c_k$ for $k = 1, \dots, K$. |
| $N_{-i}$ | $\frac{\Pi_{j=1}^d n_j}{n_i}$. |
| $N_i$ | $\Pi_{j=1}^i n_j$. |
| $\bar{N}_i$ | $\Pi_{j=i+1}^d n_j$. |
| $\tilde{\mathbf{U}}_{(i)}$ | The $i$-th unfolding of tensor $\mathcal{U}$, which belongs to $\mathbb{R}^{N_i \times \bar{N}_i}$. |
| $\mathbf{U}_{(i)}$ | The $i$-th matricization of tensor $\mathcal{U}$, which belongs to $\mathbb{R}^{n_i \times N_{-i}}$. |
| $\mathcal{T}_k$ | The tensor space that each column of $\widetilde{\mathbf{U}}^k_{(d-1)}$ is chosen generically from. |
| $\mathcal{S}_k$ | The set of structured columns that are a basis for the tensor space $\mathcal{T}_k$, i.e., $\mathcal{S}_k = \{\mathbf{u}_1, \dots \mathbf{u}_{r_k}\}$. |

### 3.1.1. TENSOR SPACE CLUSTERING WITH MISSING DATA

In Section 4, we are interested in clustering the subtensors of $\mathcal{U}$ of size $\mathbb{R}^{n_1 \times n_2 \times \dots n_{d-1} \times 1}$ in $K$ groups such that members of each group belong to the same $d$-dimensional source of size $n_1 \times n_2 \times \dots \times n_{d-1} \times c_k$ and rank $r_k$ with high probability. This is equivalent to clustering the columns of $\widetilde{\mathbf{U}}_{(d-1)}$ such that their corresponding tensors (folded versions) belong to the same tensor space.

Assuming that $\mathcal{I}_k$ denotes a set of $c_k$ columns generically chosen from $\mathcal{T}_k$, i.e., the column span of the unfolded tensor basis $\mathcal{S}_k$, $k = 1, \dots, K$, and we call $\{\mathcal{I}_1, \dots, \mathcal{I}_K\}$ the $K$ unknown sources. In fact, the $(d-1)$-th unfolding $\widetilde{\mathbf{U}}^k_{(d-1)} \in \mathbb{R}^{N_{d-1} \times c_k}$ of tensor $\mathcal{U}_k$ consists of the $c_k$ columns in $\mathcal{I}_k$. As mentioned earlier, in the clustering problem, we assume that $\sum_{k=1}^K c_k$ structured columns of

$\mathcal{I}_k$'s are randomly sampled while we do not know the source of each structured column. Then, we are interested in correctly clustering these sampled structured columns with high probability, i.e., correctly identifying the source $\mathcal{I}_k$ that each structured column is chosen from with high probability.

#### 3.1.2. UNION OF TENSOR SPACES COMPLETION

For this problem, we assume that: (i) the sets $\mathcal{S}_k$ consisting of $r_k$ structured columns are such that $\mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq \ldots \subseteq \mathcal{S}_K$; (ii) the subtensors are already correctly clustered. Hence, in this setting we can assume that the first $c_1$ columns of $\widetilde{\mathbf{U}}_{(d-1)}$ are $\widetilde{\mathbf{U}}^1_{(d-1)}$, the next $c_2$ columns of $\widetilde{\mathbf{U}}_{(d-1)}$ are $\widetilde{\mathbf{U}}^2_{(d-1)}$, and so on, i.e., $\widetilde{\mathbf{U}}_{(d-1)} = [\widetilde{\mathbf{U}}^1_{(d-1)}|\ldots|\widetilde{\mathbf{U}}^K_{(d-1)}]$.

In Section 5, we are interested in obtaining the deterministic conditions on the sampling pattern $\Omega$ such that there are finite number of completions of $\mathcal{U}$ that have the mentioned data structure and satisfy all rank constraints. Then, in Section 6, we will further obtain a lower bound on the sampling rate that ensures the similar results with high probability.

### 3.2. Summary of Results and Main Steps of Analysis

An executive summary of the steps in our analysis for each section is as follows:

(i) Clustering results for union of tensors (Section 4) - Minimum uniform sampling rate $p$ to correctly cluster the tensors:

| |
|---|
| (1) Required sampling rate for unique completability of a single tensor $\Longleftrightarrow$ Lemma 11. |
| (2) A condition that ensures infinite completability of a single tensor (will be used in the proof of the main theorem to contradicts the unique completability) $\Longleftrightarrow$ Lemma 16. |
| (1) and (2) $\Longleftrightarrow$ Characterize the required sampling rate for correctly clustering the union of tensors with a certain probability (Theorem 17). |

(ii) Deterministic analysis for completion of union of tensors (Section 5) - Necessary and sufficient conditions on the sampling pattern $\Omega$:

| |
|---|
| Completion Problem $\Longleftrightarrow$ Solving polynomial equations in terms of the entries of canonical decomposition of the union of tensors given in Definition 23 $\Longleftrightarrow$ Lemma 31. |
| (1) Finite completability of $\mathcal{U}$ given $\mathbf{A}_i^k$'s $\Longleftrightarrow$ Assumption 1. |
| (2) Finite completability of $\mathbf{A}_i^k$'s $\Longleftrightarrow$ Existence of $(n_1 + \cdots + n_{d-1})r_K - \sum_{k=1}^{K} r_k(r_k - r_{k-1}) - r_K(d-2)$ algebraically independent polynomials in $\mathcal{P}(\Omega)$ (Lemma 33) $\Longleftrightarrow$ Relationship between $\Omega$ (via the constraint tensor $\breve{\Omega}$) and the maximum number of algebraically independent polynomials (Lemma 34). |
| (1) and (2) $\Longleftrightarrow$ Geometric patterns for finite completability of $\mathcal{U}$ (Theorem 36). |

(iii) Probabilistic analysis for completion of union of tensors (Section 6) - Minimum uniform sampling rate $p$ to ensure finite/unique completability:

| |
|---|
| (1) Connecting conditions on the number of samples and geometry of $\Omega$ (Lemma 41) |
| (1) and Theorem 36 (geometry of $\breve{\Omega}$) $\Longrightarrow$ Probabilistic condition for finite completability (Theorem 42). |

### 3.3. Matrix Analysis Approach To The Clustering Problem

One naive way of treating the above tensor space clustering problem is to ignore the "unfolded structure" of the columns in $\mathcal{I}_k$ and simply apply the subspace clustering results for matrices.

First, we restate the main result of (Pimentel-Alarcón and Nowak, 2016) (i.e., Theorems 1 and 3 in (Pimentel-Alarcón and Nowak, 2016)). Consider a similar problem as described in Section 3.1 for two-dimensional data, or matrix. Let $\mathcal{T}_1, \dots, \mathcal{T}_K$ be subspaces independently chosen from $\text{Gr}(n_1, r)$ (set of all $r$-dimensional subspaces of the $n_1$-dimensional space), and source $\mathcal{I}_k$ includes $c_k$ columns generically chosen from $\mathcal{T}_k$, $k = 1, \dots, K$. The matrix $\mathbf{U} \in \mathbb{R}^{n_1 \times n_2}$ is such that $n_2 = \sum_{k=1}^{K} c_k$ and includes all columns of $\mathcal{I}_k$ for $k = 1, \dots, K$. $\mathbf{\Omega}$ is an $n_1 \times n_2$ binary sampling matrix.

**Lemma 7** (Pimentel-Alarcón and Nowak, 2016) *Assume that the subspaces $\mathcal{T}_1, \dots, \mathcal{T}_K$ are independently chosen from Gr$(n_1, r)$, $r \leq \frac{n_1}{6}$ and $c_k \geq (r+1)(n_1 - r + 1)$, $k = 1, \dots, K$. Moreover suppose that each column of $\mathbf{U}_{\mathbf{\Omega}}$ includes at least $l$ sampled entries such that*

$$l > \max\{12 \log \left( \frac{n_1(r+1)}{\epsilon} \right) + 12, 2r\}. \tag{4}$$

*Let $\bar{\mathcal{T}}$ denote an $r$-dimensional subspace that fits exactly $\bar{c}$ columns of $\mathbf{U}_{\mathbf{\Omega}}$ (i.e., $\bar{c}$ is the maximum number of columns of $\mathbf{U}_{\mathbf{\Omega}}$ that can be covered by $\bar{\mathcal{T}}$) and assume that $\bar{c} \geq (r+1)(n_1 - r + 1)$. Then, with probability at least $1 - K\epsilon$, the following statement holds: All the $\bar{c}$ columns of $\mathbf{U}_{\mathbf{\Omega}}$ covered by $\bar{\mathcal{T}}$ belong to one source $\mathcal{I}_{k_0}$ for some $1 \leq k_0 \leq K$ and the rest of the columns of $\mathbf{U}_{\mathbf{\Omega}}$ do not belong to $\mathcal{I}_{k_0}$ and moreover, $\bar{c} = c_{k_0}$ and $\bar{\mathcal{T}} = \mathcal{T}_{k_0}$.*

**Remark 8** *The above lemma provides the conditions for clustering the columns that belong to one particular subspace with high probability.*

For the tensor space clustering problem in Section 3.1, assume that $r_1 = \cdots = r_K = r$ and the tensor spaces are chosen independently. Then, we can simply apply Lemma 7 to $\widetilde{\mathbf{U}}_{(d-1)}$ to cluster one of the tensor spaces if

$$c_k \geq (r_{\max} + 1)(N_{-d} - r_{\max} + 1), \qquad k = 1, \dots, K, \tag{5}$$

and

$$l > \max\{12 \log \left( \frac{N_{-d}(r_{\max} + 1)}{\epsilon} \right) + 12, 2r_{\max}\}. \tag{6}$$

Note that the assumptions such as $r_1 = \cdots = r_K = r$, the tensor spaces are chosen independently, and (5) are very strong and we are interested in clustering the tensor spaces without requiring such assumptions.

**Remark 9** *The reason that the above naive method is not very efficient and requires a very strong assumption* (5) *is that we completely ignored the "unfolded structure" of the columns of* $\widetilde{\mathbf{U}}_{(d-1)}$ *and effectively relaxed the tensor space clustering problem to a matrix clustering problem, i.e., relaxing the structure of CP decomposition to the simple structure of matrix decomposition.*

### 3.4. Matrix Analysis Approach To The Completion Problem

A naive approach is again to ignore the unfolding structure of the columns of $\widetilde{\mathbf{U}}_{(d-1)}$ and treat it as a union of low-rank matrices completion problem. For notational clarity, denote $\mathbf{V} \triangleq \widetilde{\mathbf{U}}_{(d-1)} \in \mathbb{R}^{N_{-d}\times(c_1+\cdots+c_K)}$, $\mathbf{V}_k \triangleq \widetilde{\mathbf{U}}^k_{(d-1)} \in \mathbb{R}^{N_{-d}\times(c_1+\cdots+c_k)}$. Note that the CP-rank of a tensor is an upper bound on the rank of any unfoldings of that tensor and therefore, we can simply conclude that $\text{rank}\,(\mathbf{V}_k) \leq r_k$, $k = 1,\dots,K$. Morover, the tensor space structure for the completion problem ($\mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq \dots \subseteq \mathcal{S}_K$) simply results that $\text{span}\,(\mathbf{V}_1) \subseteq \text{span}\,(\mathbf{V}_2) \subseteq \dots \subseteq \text{span}\,(\mathbf{V}_K)$. Now, let us consider the special case when $\text{rank}\,(\mathbf{V}_k) = r_k$, $k = 1,\dots,K$ and we know that given $\text{span}\,(\mathbf{V}_1) \subseteq \text{span}\,(\mathbf{V}_2) \subseteq \dots \subseteq \text{span}\,(\mathbf{V}_K)$, the following lemma found in (Ashraphijuo and Wang, 2018) gives a lower bound on the number of samples such that $\mathbf{V}$ is uniquely completable.

**Lemma 10** *Assume that* $r_k \leq N_{-d}$ *and* $c_k \geq (r_k - r_{k-1} + 1)(N_{-d} - r_k)$ *for* $k = 1,\dots,K$, *and each column of* $\mathbf{\Omega}$ *(i.e.,* $\widetilde{\mathbf{\Omega}}_{(d-1)}$ *for* $\mathcal{U}$*) includes at least* $l$ *nonzero entries where*

$$l > \max\left\{9\,\log\left(\frac{N_{-d}}{\epsilon}\right) + 3\,\log\left(\frac{\max_{1\leq k\leq K}\{r_k - r_{k-1}\}2K}{\epsilon}\right) + 6, 2r_K\right\}. \tag{7}$$

*Then, with probability at least* $1-\epsilon$, $\mathbf{V}$ *is uniquely completable.*

Recall that $N_{-d} = n_1 n_2 \dots n_{d-1}$ and hence the condition $c_k \geq (r_k - r_{k-1} + 1)(N_{-d} - r_k)$ implies that a very large number of subtensors are needed, similar to the clustering case [cf. Eq. (5)].

## 4. Tensor Space Clustering: Main Results

### 4.1. Useful Results on Tensor Completion

The tensor completion problem is to recover the missing entries of a randomly sampled tensor given the rank of the original tensor. In this subsection, we provide some results on the CP-rank tensor completion, which are instrumental to solving the tensor space clustering problem.

#### Unique Completability of Tensors

In (Ashraphijuo and Wang, 2017), the conditions on the sampling pattern and sampling probability are given to ensure finite/unique completability of the sampled tensor of the given rank, where finite/unique completability means the number of possible completions of the given rank constraint is finite/one. Consider a $d$-way tensor $\mathcal{U} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ drawn generically from the manifold of tensors of the same size and of CP-rank $r$. The following lemma is taken from (Ashraphijuo and Wang, 2017).

**Lemma 11** *Assume that* $d > 2$, $(\min_{1\leq i\leq d-1} n_i) > 200$, $n_d \geq (r+2)(\sum_{i=1}^{d-1} n_i)$, $r \leq \frac{\min_{1\leq i\leq d-1} n_i}{6}$. *Assume that each column of* $\widetilde{\mathbf{U}}_{(d-1)}$ *includes at least* $l$ *nonzero entries, where*

$$l > \max\left\{27\ \log\left(\frac{2\max_{1\leq i\leq d-1} n_i}{\epsilon}\right) + 9\ \log\left(\frac{8r(d-1)}{\epsilon}\right) + 18, 6r\right\}. \tag{8}$$

*Then, with probability at least* $1-\epsilon$, *there exist only one completion of the sampled tensor* $\mathcal{U}$ *with CP rank* $r$, *which is the original sampled tensor.*

**Tensor Completion With One Missing Entry**

Here, we consider the tensor completion problem with only one missing entry. This will be useful in proving Lemma 16.

**Lemma 12** *Let* $\mathcal{U}$ *be a rank-*$(r-1)$ *tensor and* $\mathcal{U}(\vec{x}) = y$ *be an entry of this tensor. Assume that changing the value of entry* $\mathcal{U}(\vec{x})$ *from* $y$ *to* $y'$ *results in* $\mathcal{U}'$, *which is a rank-*$r$ *tensor. Then, there are infinitely many scalars* $y''$ *such that changing the value of entry* $\mathcal{U}(\vec{x})$ *from* $y$ *to* $y''$ *results in a rank-*$r$ *tensor.*

**Proof** First, we claim that changing the value of only one entry of a tensor can increase the rank of the tensor by at most one. Recall that the rank of a tensor is the minimum number of structured columns whose column span includes all columns of $\widetilde{\mathbf{U}}_{(d-1)}$. Assume that only one entry of the tensor is changed and consider the column of $\widetilde{\mathbf{U}}_{(d-1)}$ where the changed entry resides. Before changing this entry, the mentioned column was covered by the column span of $r-1$ structured columns. To show the earlier claim, we need to show that there exists a structured column such that together with the previous $r-1$ structured columns, they span the modified column of $\widetilde{\mathbf{U}}_{(d-1)}$. Note that a column $\mathbf{u} \in \mathbb{R}^{N_{d-1}}$ with only a single nonzero entry $\mathcal{U}(\widetilde{M}_{d-1}(x_1,\dots,x_{d-1})) = u \neq 0$, is a structured column, as we can rewrite $\mathbf{u}$ as the vectorization of the outer product of $[\underbrace{0,\dots,0}_{x_i-1},1,0,\dots,0]^\top \in \mathbb{R}^{n_i}$ for $1 \leq i \leq d-2$ and $[\underbrace{0,\dots,0}_{x_{d-1}-1},u,0,\dots,0]^\top \in \mathbb{R}^{n_{d-1}}$. As a result, by adding a structured column with a single nonzero entry in the corresponding location of the modified entry (with the value of the nonzero entry equal to the amount of modification in that entry) to the mentioned $r-1$ structured columns, we obtain an unfolded tensor space of at most rank $r$ and therefore, our earlier claim is proved. As a consequence, changing the value of only one entry of a tensor results in changing the rank of the tensor by at most one. Hence, for any scalar $y''$, changing the value of entry $\mathcal{U}(\vec{x})$ from $y$ to $y''$ results in either a rank-$(r-2)$, rank-$(r-1)$ or rank-$r$ tensor. Note that rank-$(r-2)$ is not possible since otherwise by changing the value of $\mathcal{U}(\vec{x})$ from $y'$ to $y''$ the rank is changed by 2.

Now given a rank-$(r-1)$ tensor, the CP decomposition

$$\mathcal{U}(\vec{x}) = \sum_{l=1}^{r-1} \mathbf{a}_1^l(x_1)\mathbf{a}_2^l(x_2)\dots\mathbf{a}_d^l(x_d), \tag{9}$$

results in a system of $|\Omega|$ polynomial equations $\mathbf{p}$ in terms of the variables $\mathbf{a}_i^l$, $i = 1,\dots,d$, $l = 1,\dots,r-1$. By assumption, when the value of the entry $\mathcal{U}(\vec{x})$ is changed from $y$ to $y'$, the rank is changed from $r-1$ to $r$. This means that when $\mathcal{U}(\vec{x}) = y$, $\mathbf{p}$ is feasible; and when

$\mathcal{U}(\vec{x}) = y'$, $\mathbf{p}$ is not feasible. On the other hand, since changing $\mathcal{U}(\vec{x})$ to any $y''$ results in either rank-$(r-1)$ or rank-$r$ tensor, we conclude that if changing $\mathcal{U}(\vec{x})$ to $y'$ leads to infeasible $\mathbf{p}$, then rank is $r$. The rest of the proof then follows from Lemma 15. ■

Next, we prove Lemma 15 which was used in the above lemma. We first need the following definition.

**Definition 13** *A closed set is a set that contains all its limit points, i.e., for any sequence of points* $\lim_{t\to\infty} a_t = a_0$ *such that* $a_t$ *belongs to the mentioned set for* $t \geq 1$, *we conclude that* $a_0$ *belongs to this set as well.*

The feasible region of a system of equations of $n$ variables is the set of all $n$-dimensional real valued vectors that are solutions to the system.

**Lemma 14** *Consider an arbitrary system of polynomial equations* $\mathbf{p} = (p_1, \ldots, p_m) = \mathbf{0}$ *in terms of* $n$ *variables. The feasible region of* $\mathbf{p}$ *in* $\mathbb{R}^n$ *is a closed set.*

**Proof** First, we show that the feasible region of any single polynomial equation $p_i = 0$, is a closed set in $\mathbb{R}^n$. Let $\mathcal{F}_i$ denote this feasible region, $i = 1, \ldots, m$. Consider a sequence of points $\{\vec{x}_t \in \mathbb{R}^n | t = 1, 2, \ldots\}$ such that $\vec{x}_t \in \mathcal{F}_i$ and $\lim_{t\to\infty} \vec{x}_t = \vec{x}_0$. The assumption $\vec{x}_t \in \mathcal{F}_i$ simply results that $p_i(\vec{x}_t) = 0$. Note that $p_i$ is continuous with respect to the vector of variables $\vec{x} = (x_1, \ldots, x_n)$ and therefore, $\lim_{t\to\infty} p_i(\vec{x}_t) = 0$. As a result, $p_i(\vec{x}_0) = 0$ and hence, $\vec{x}_0 \in \mathcal{F}_i$. Hence, $\mathcal{F}_i$ is closed.

Define $\mathcal{F} = \cap_{1\leq i\leq m} \mathcal{F}_i$, which denotes the feasible region of $\mathbf{p}$ in $\mathbb{R}^n$. Note that the intersection of several closed sets in $\mathbb{R}^n$ is a closed set in $\mathbb{R}^n$ (11.1.5 Closed Set Properties in (Bartle and Sherbert, 2011)). Hence, $\mathcal{F}$ is a closed set in $\mathbb{R}^n$. ■

The following lemma is used in the proof of Lemma 12.

**Lemma 15** *Consider an arbitrary system of polynomial equations* $\mathbf{p} = (p_1, \ldots, p_{m-1}, p_m - c) = \mathbf{0}$ *in terms of* $n$ *variables, where* $c \in \mathbb{R}$ *is a constant. Assume that there exist* $y, y' \in \mathbb{R}$ *such that the feasible region of* $\mathbf{p}$ *is non-empty for* $c = y$ *and empty for* $c = y'$. *Then, there exist infinitely many real scalars* $y''$ *such that the feasible region of* $\mathbf{p}$ *is empty for* $c = y''$.

**Proof**

It suffices to show that for any neighborhood $\mathcal{N}(y') = [y' - \epsilon, y' + \epsilon]$ (for any $\epsilon > 0$) there exists a scalar $y'' \in \mathcal{N}(y')$ such that $y'' \neq y'$ and $c = y''$ results that the feasible region of $\mathbf{p}$ is empty, since by considering smaller and smaller neighborhoods, e.g., $\epsilon_n = \frac{1}{n}$, we obtain infinitely many scalars $y''$, such that the feasible region of $\mathbf{p}$ is empty for $c = y''$.

Now, by contradiction, assume that there exist a neighborhood $\mathcal{N}(y') = [y' - \epsilon, y' + \epsilon]$ (for some $\epsilon > 0$) such that for any scalar $y'' \in \mathcal{N}(y')$ that $y'' \neq y'$, $c = y''$ results that the feasible region of $\mathbf{p}$ is non-empty.

Let $\mathcal{F}' \subseteq \mathbb{R}^n$ denote the feasible region of the system of polynomial equations $\mathbf{p}' = (p_1, \ldots, p_{m-1}) = \mathbf{0}$. Note that $c = y''$ results that the feasible region of $\mathbf{p}$ is non-empty if and only if there exists a vector $\vec{x''} \in \mathcal{F}'$ such that $p_m(\vec{x''}) = y''$. In other words, $y''$ belongs to the image of $p_m$ with $\mathcal{F}'$

as the domain ($p_m(\mathcal{F}')$) if and only if $c = y''$ results that the feasible region of $\mathbf{p}$ is non-empty. Consequently, for any $y'' \in \mathcal{N}(y')$ and $y'' \neq y'$, there exist at least one vector $\vec{x}'' \in \mathcal{F}'$ such that $p_m(\vec{x}'') = y''$.

As a result of the fact that $y' - \epsilon$ and $y' + \epsilon$ belong to the image of $p_m$ with $\mathcal{F}'$ as the domain, there exist $\vec{x}_1'' \in \mathcal{F}'$ and $\vec{x}_2'' \in \mathcal{F}'$ such that $p_m(\vec{x}_1'') = y' - \epsilon$ and $p_m(\vec{x}_2'') = y' + \epsilon$. Let $l$ denote the one dimensional line that connects $\vec{x}_1''$ to $\vec{x}_2''$ in $\mathbb{R}^n$ and $l'$ denote the corresponding segment on this line with endpoints $\vec{x}_1''$ and $\vec{x}_2''$.

According to Lemma 14, $\mathcal{F}'$ is a closed set. On the other hand, it is easily verified that any one dimensional segment in $\mathbb{R}^n$ that includes its endpoints is a closed set. Hence, $l'$ is a closed set and therefore, $\mathcal{F}'' = \mathcal{F}' \cap l'$ is a closed set (Bartle and Sherbert, 2011). Note that $\mathcal{F}''$ denotes all the points on the segment that connects $\vec{x}_1''$ to $\vec{x}_2''$, i.e., $l'$, that also belong to $\mathcal{F}'$. Due to the continuity of $p_m$, we have a sequence of points $\vec{x}_t$ in $\mathcal{F}''$ that $\lim_{t\to\infty} p_m(\vec{x}_t) = y'$. Since, $\mathcal{F}''$ is a closed set, it includes $\vec{x}_0$, the limit of the sequence and therefore, $p_m(\vec{x}_0) = y'$. This is a contradiction and the proof is complete. ■

### Infinite Completability of Tensors

The following lemma provides a condition under which there exist infinitely many rank-$r$ completions of a sampled tensor.

**Lemma 16** *Let $\mathcal{U}_\Omega$ be a sampled tensor. Assume that there exist a rank-$(r-i)$ completion (with $0 < i < r$) and a rank-$r$ completion. Then, there exist infinitely many rank-$r$ completions of $\mathcal{U}_\Omega$.*

**Proof** As mentioned in the proof of Lemma 12, changing the value of only one entry of a tensor results in changing the rank of the tensor by at most one. Let $\mathcal{U}_1$ and $\mathcal{U}_2$ denote the rank-$(r-i)$ and rank-$r$ completions, respectively. $\mathcal{U}_1$ and $\mathcal{U}_2$ are the same over the sampled entries, i.e., $(\mathcal{U}_1)_\Omega = (\mathcal{U}_2)_\Omega$, and their difference is only over some of the non-sampled entries. We change the values of non-sampled entries of $\mathcal{U}_1$ *one by one* to the values of the corresponding non-sampled entries of $\mathcal{U}_2$, which will eventually result in $\mathcal{U}_2$ if we continue this for all non-sampled entries. While performing this simple process, we simply increase the rank from $r-1$ to $r$ at some step by changing a non-sampled entry. This is because at the beginning the rank of the tensor is $r - i \leq r - 1$ and at the end the rank is $r$ and also at each step the rank changes by at most one.

Hence, there exists a rank-$(r-1)$ completion $\mathcal{U}_3$ of the sampled tensor $\mathcal{U}_\Omega$ such that changing the value of an entry $\mathcal{U}_3(\vec{x})$ from $y$ to $y'$ increases the rank to $r$ for some scalars $y$ and $y'$. Hence, according to Lemma 12, there exists infinitely many rank-$r$ completions of $\mathcal{U}_\Omega$. ■

### 4.2. Conditions on Tensor Space Clustering

**Assumptions for Theorem 17** : *For the tensor space clustering problem discussed in Section 3.1 without loss of generality, assume that $r_1 \leq r_2 \leq \ldots \leq r_K$, $(\min_{1\leq i\leq d-1} n_i) > 200$ and denote $r_{\max} = \max_{1\leq k\leq K} r_k = r_K$. Assume further that $r_{\max} \leq \frac{\min_{1\leq i\leq d-1} n_i}{6}$,*

$$c_k \geq K(r_{\max} + 2)(\sum_{i=1}^{d-1} n_i), \qquad \textit{for } k = 1, \ldots, K, \tag{10}$$

*and also, each column of $\widetilde{\mathbf{U}}_{(d-1)}$ includes at least $l$ sampled entries where*

$$l > \max\left\{27 \ \log\left(\frac{2\max_{1\leq i\leq d-1} n_i}{\epsilon}\right) + 9 \ \log\left(\frac{8r_{\max}(d-1)}{\epsilon}\right) + 18, 6r_{\max}\right\}. \qquad (11)$$

*Let $\bar{\mathcal{S}}$ be a set of $r_1$ structured columns and $\bar{\mathcal{T}}$ denote the unfolded tensor space generated by its columns span that fits exactly $\bar{c}$ columns of $\widetilde{\mathbf{U}}_{(d-1)}$ (i.e., $\bar{c}$ is the maximum number of columns of $\widetilde{\mathbf{U}}_{(d-1)}$ that can be covered by $\bar{\mathcal{S}}$) and assume that $\bar{c} \geq K(r_{\max}+2)(\sum_{i=1}^{d-1} n_i)$.*

**Theorem 17** *Assume that all the above conditions described in "Assumptions for Theorem 17" hold true. Then, with probability at least $1-\epsilon$ the following statement holds: All the $\bar{c}$ columns of $\widetilde{\mathbf{U}}_{(d-1)}$ covered by $\bar{\mathcal{S}}$ belong to one source $\mathcal{I}_{k_0}$ for some $1 \leq k_0 \leq K$ such that $r_{k_0} = r_1$ (if $r_1 < r_2$ then $k_0 = 1$ and otherwise there are more options for $k_0$) and the rest of the columns of $\widetilde{\mathbf{U}}_{(d-1)}$ do not belong to $\mathcal{I}_{k_0}$ and moreover, $\bar{c} = c_{k_0}$ and $\bar{\mathcal{T}} = \mathcal{T}_{k_0}$.*

**Proof** According to the pigeonhole principle, at least $\lceil \frac{\bar{c}}{K} \rceil \geq (r_{\max}+2)(\sum_{i=1}^{d-1} n_i)$ columns of the $\bar{c}$ covered columns by $\bar{\mathcal{S}}$ are chosen from one source $\mathcal{I}_{k_0}$. Note that due to the assumption $r_{\max} \geq r_{k_0}$, we have $(r_{\max}+2)(\sum_{i=1}^{d-1} n_i) \geq (r_{k_0}+2)(\sum_{i=1}^{d-1} n_i)$ and hence, there are at least $(r_{k_0}+2)(\sum_{i=1}^{d-1} n_i)$ columns covered by $\bar{\mathcal{S}}$ that are chosen from one source $\mathcal{I}_{k_0}$. Then, according to Lemma 11, there exists a unique rank-$r_{k_0}$ completion of the tensor corresponding to the mentioned $(r_{k_0}+2)(\sum_{i=1}^{d-1} n_i)$ columns with probability at least $1-\epsilon$. Hence, assuming that the mentioned unique completability holds, it suffices to show the mentioned claims in the statement of the theorem hold with probability one.

First, we show that $r_{k_0} = r_1$. By contradiction, assume otherwise that $r_1 < r_{k_0}$. Recall that $\bar{\mathcal{T}}$ is an $r_1$-dimensional tensor space that fits the mentioned $(r_{k_0}+2)(\sum_{i=1}^{d-1} n_i)$ columns and hence, there exists a rank-$r_1$ completion of the tensor corresponding to these columns. Moreover, note that the original data gives a rank-$r_{k_0}$ completion of the tensor corresponding to these columns. Hence, according to Lemma 16, there exist infinitely many rank-$r_{k_0}$ completions of the tensor corresponding to these columns, which contradicts the earlier uniqueness assumption. As a result, we have $r_{k_0} = r_1$ with probability one. Now that $r_{k_0} = r_1$, according to the uniqueness of rank-$r_{k_0}$ completion assumption, and due to the fact that both subspaces $\bar{\mathcal{T}}$ and $\mathcal{T}_{k_0}$ are $r_1$-dimensional (since $r_{k_0} = r_1$), we simply conclude $\bar{\mathcal{T}} = \mathcal{T}_{k_0}$. Consequently, $\bar{\mathcal{T}}$ covers all $c_{k_0}$ columns of $\widetilde{\mathbf{U}}_{(d-1)}$ that belong to $\mathcal{I}_{k_0}$. In order to complete the proof, it suffices to show that $\bar{c} = c_{k_0}$, i.e., $\bar{\mathcal{T}}$ does not cover any other column of $\widetilde{\mathbf{U}}_{(d-1)}$ that belongs to other sources $\mathcal{I}_k$ for $k \neq k_0$, with probability one. Since we have $r_{k_0} = r_1 = \min\{r_1, r_2, \ldots, r_K\}$, we conclude that $\mathcal{T}_k \not\subseteq \mathcal{T}_{k_0}$ for $k \neq k_0$ and therefore, any column chosen from sources other than $\mathcal{I}_{k_0}$ does not belong to $\mathcal{T}_{k_0}$ with probability one. Note that this statement is not valid if $r_{k_0} \neq \min\{r_1, r_2, \ldots, r_K\}$ and this is why we cluster the tensor space with the lowest dimension for now.

Now, by contradiction, assume that a column $\widetilde{\mathbf{u}}_{\Omega_{(d-1)}}$ of $\widetilde{\mathbf{U}}_{\Omega_{(d-1)}}$ is chosen from $\mathcal{I}_{k_1}$ for some $k_1 \neq k_0$, and it can be covered by $\bar{\mathcal{S}}$. Consider $r_1$ random columns of $\widetilde{\mathbf{U}}_{\Omega_{(d-1)}}$ that belong to $\mathcal{I}_{k_0}$ and denote it by $\widetilde{\mathbf{U}}^0_{\Omega_{(d-1)}}$ and let $\widetilde{\mathbf{U}}^0_{(d-1)}$ be the unique rank-$r_1$ completion of $\widetilde{\mathbf{U}}^0_{\Omega_{(d-1)}}$. Define $\widetilde{\mathbf{U}}^1_{\Omega'_{(d-1)}} = [\widetilde{\mathbf{U}}^0_{(d-1)} | \widetilde{\mathbf{u}}_{\Omega_{(d-1)}}] \in \mathbb{R}^{N_{-d} \times (r_1+1)}$ (only the last column of $\widetilde{\mathbf{U}}^1_{\Omega'_{(d-1)}}$ is incomplete). Note that (11) ensures that $\widetilde{\mathbf{u}}_{\Omega_{(d-1)}}$ includes at least $r_1+1$ sampled entries and therefore, $\widetilde{\mathbf{U}}^1_{\Omega'_{(d-1)}}$ includes an $(r_1+1) \times (r_1+1)$ submatrix such that the $r_1+1$ of the sampled entries of $\widetilde{\mathbf{u}}_{\Omega_{(d-1)}}$ are included

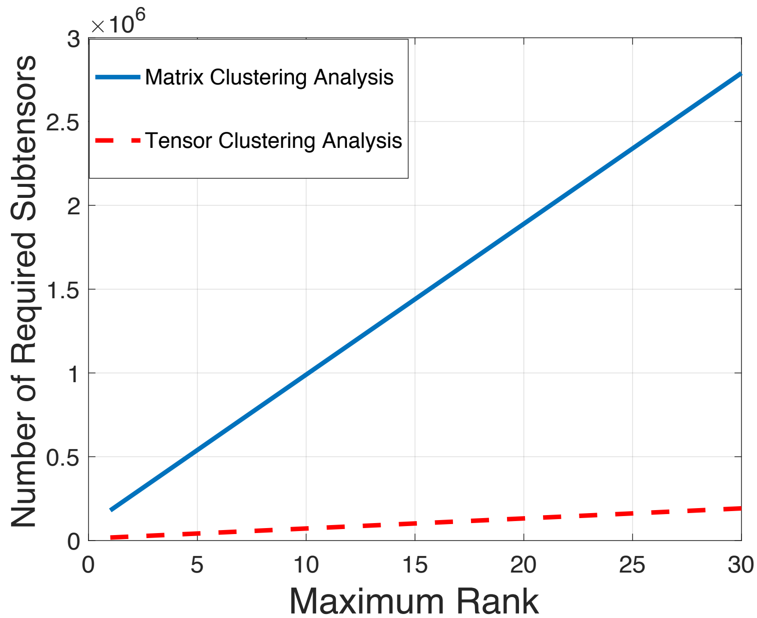


Figure 1: The required number of subtensors $c_k$ for correct clustering.

and denote such submatrix by $\widetilde{\mathbf{U}}^{1'}_{(d-1)}$. Note that $\widetilde{\mathbf{u}}_{(d-1)}$ is chosen generically from the column span of $\mathcal{T}_k$ and we know that $\mathcal{T}_k \not\subseteq \mathcal{T}_{k_0}$. Therefore, the matrix rank of $\widetilde{\mathbf{U}}^{1'}_{(d-1)} \in \mathbb{R}^{(r_1+1)\times(r_1+1)}$ is $r_1+1$. Hence, any completion of $\widetilde{\mathbf{U}}^{1}_{\Omega'_{(d-1)}}$ has matrix rank of at least $r_1+1$ and as a result, any completion of the tensor corresponding to the unfolding $\widetilde{\mathbf{U}}^{1}_{\Omega'_{(d-1)}}$ has CP-rank of at least $r_1+1$. Therefore, $\bar{\mathcal{S}}$ cannot fit $\widetilde{\mathbf{u}}_{\Omega_{(d-1)}}$ with probability one since $\bar{\mathcal{S}}$ is an $r_1$-dimensional tensor space, and the proof is complete due to this contradiction. ∎

**Remark 18** *The required number of subtensors chosen from each source in the matrix analysis approach given by* (5) *is* $\mathcal{O}(r_{\max} n_1 \dots n_{d-1})$. *However, this number reduces to* $\mathcal{O}(r_{\max}(n_1+\dots+n_{d-1}))$ *in the tensor approach according to Theorem 17. This huge improvement is a consequence of taking advantage of tensor analysis as opposed to matrix analysis. Moreover, Theorem 17 does not require the assumptions of* $r_1 = r_2 = \dots = r_K$ *and independently chosen tensor spaces.*

**Example 1** *Consider an example in which* $d = 3$, $K = 10$, $n_1 = n_2 = 300$, $\epsilon = 0.1$. *We compare the number of subtensors required for correctly clustering using the matrix analysis and our proposed tensor analysis in Figure 1, given by* (5) *and* (10)*, respectively. Note that the bound obtained by matrix analysis is valid only if* $r_1 = r_2 = \dots = r_{10}$*, which is not the case for the proposed tensor analysis. Then, we compare the required number of samples per column of the* $(d-1)$*-th unfolding for correct clustering using the matrix analysis and our proposed tensor analysis, given by* (6) *and* (11)*, respectively, in Figure 2. It is clearly seen from the two figures that the proposed tensor analysis substantially reduces the number of subtensors and samples needed for correct clustering.*

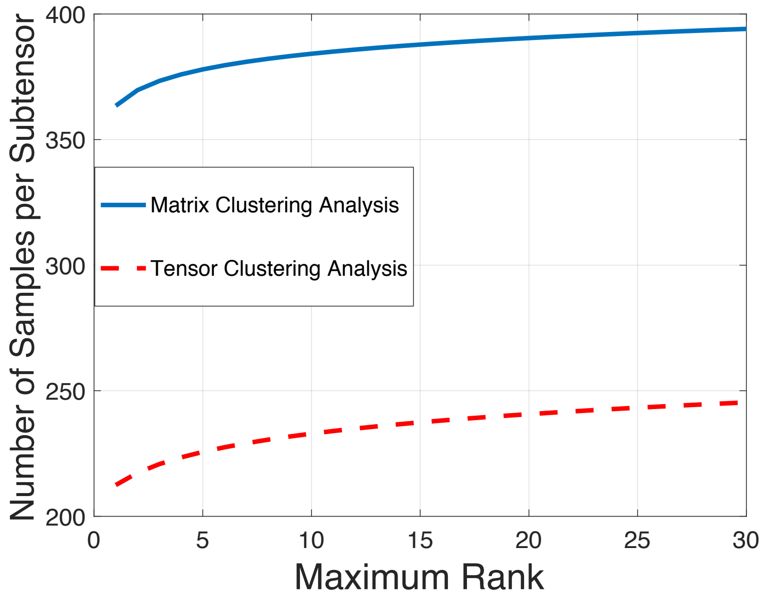


Figure 2: The required number of samples per subtensor for correct clustering.

According to Theorem 17, we identify all columns of $\widetilde{\mathbf{U}}_{(d-1)}$ chosen from a source (generated by a set of $r_1$ structured columns). Then, we can simply exclude the subtensors (corresponding to the identified columns) from the sampled tensor. Then, the problem reduces to the similar problem with $K-1$ tensor spaces of ranks $r_2 \leq \ldots \leq r_K$. Hence, the same analysis as in Theorem 17 is applicable again. In particular, let $\bar{\mathcal{S}}_1, \ldots, \bar{\mathcal{S}}_{K'}$ (for some $1 \leq K' < K$) denote different sets of $r_1, \ldots, r_{K'}$ structured columns that fit exactly $\bar{c}_1, \ldots, \bar{c}_{K'}$ columns of $\widetilde{\mathbf{U}}_{(d-1)}$, respectively, and assume that $\bar{c}_k \geq K(r_{\max}+2)(\sum_{i=1}^{d-1} n_i)$, $k = 1, \ldots, K'$. Let $\bar{\mathcal{T}}_1, \ldots, \bar{\mathcal{T}}_{K'}$ denote the column span of $\bar{\mathcal{S}}_1, \ldots, \bar{\mathcal{S}}_{K'}$, respectively. Moreover, assume that there exist $K(r_{\max}+2)(\sum_{i=1}^{d-1} n_i)$ columns of $\widetilde{\mathbf{U}}_{(d-1)}$ covered by $\bar{\mathcal{S}}_k$ that cannot be covered by any of $\bar{\mathcal{S}}_1, \ldots, \bar{\mathcal{S}}_{k-1}$, $k = 1, \ldots, K'$. Then, using Theorem 17, we have $\bar{\mathcal{T}}_1 = \mathcal{T}_{k_1}$ and $\bar{c}_1 = c_{k_1}$ with probability at least $1-\epsilon$. Hence, we can exclude all the $c_{k_1}$ the subtensors (corresponding to the identified $c_{k_1}$ columns) from the sampled tensor and the identified tensor space $\mathcal{T}_{k_1}$. Therefore, similarly, we can apply Theorem 17 again. We consider the scenario of correct identification of the subtensors corresponding to the first source, which holds true with probability at least $1-\epsilon$ and continue clustering the rest of the subtensors.

Given the previous correct clustering, we can cluster the subtensors of the next tensor space correctly with probability at least $1-\epsilon$. This can be done because due to the assumption, after excluding the columns of $\widetilde{\mathbf{U}}_{(d-1)}$ of the first source, there still exist $K(r_{\max}+2)(\sum_{i=1}^{d-1} n_i)$ columns covered by $\bar{\mathcal{S}}_2$ that cannot be covered by $\bar{\mathcal{S}}_1$. Hence, we apply Theorem 17 again and therefore, with probability at least $(1-\epsilon)^2$ the following statement holds: All the $\bar{c}_k$ columns of $\widetilde{\mathbf{U}}_{(d-1)}$ covered by $\bar{\mathcal{S}}_k$ belong to one source $\mathcal{I}_{k'}$ such that $r_{k'} = r_k$ and the rest of the columns of $\widetilde{\mathbf{U}}_{(d-1)}$ do not belong to $\mathcal{I}_{k'}$ and moreover, $\bar{c}_k = c_{k'}$ and $\bar{\mathcal{T}}_k = \mathcal{T}_{k'}$, $k = 1, 2$. By simply repeating this procedure, we conclude the following corollary.

**Assumptions for Corollary 19** : *Without loss of generality, assume that* $r_1 \leq r_2 \leq \ldots \leq r_K$, $(\min_{1\leq i\leq d-1} n_i) > 200$ *and denote* $r_{\max} = \max_{1\leq k\leq K} r_k = r_K$. *Assume further that* $r_{\max} \leq \frac{\min_{1\leq i\leq d-1} n_i}{6}$, (10) *holds, and also, each column of* $\widetilde{\mathbf{U}}_{(d-1)}$ *includes at least* $l$ *sampled* *entries such that* (11) *holds. Let* $\bar{\mathcal{S}}_1,\dots,\bar{\mathcal{S}}_{K'}$ *(for some* $1 \leq K' < K$*) denote different sets of* $r_1,\dots,r_{K'}$ *structured columns that fit exactly* $\bar{c}_1,\dots,\bar{c}_{K'}$ *columns of* $\widetilde{\mathbf{U}}_{(d-1)}$*, respectively, and assume that* $\bar{c}_k \geq K(r_{\max}+2)(\sum_{i=1}^{d-1} n_i)$*,* $k=1,\dots,K'$*. Let* $\bar{\mathcal{T}}_1,\dots,\bar{\mathcal{T}}_{K'}$ *denote the column span of* $\bar{\mathcal{S}}_1,\dots,\bar{\mathcal{S}}_{K'}$*, respectively. Moreover, assume that there exist* $K(r_{\max}+2)(\sum_{i=1}^{d-1} n_i)$ *columns of* $\widetilde{\mathbf{U}}_{(d-1)}$ *covered by* $\bar{\mathcal{S}}_k$ *that cannot be covered by any of* $\bar{\mathcal{S}}_1,\dots,\bar{\mathcal{S}}_{k-1}$*,* $k=1,\dots,K'$*.*

**Corollary 19** *Assume that all the above conditions described in "Assumptions for Corollary 19" hold true. Then, with probability at least* $(1-\epsilon)^{K'}$ *the following statement holds: All the* $\bar{c}_k$ *columns of* $\widetilde{\mathbf{U}}_{(d-1)}$ *covered by* $\bar{\mathcal{S}}_k$ *belong to one source* $\mathcal{I}_{k'}$ *such that* $r_{k'}=r_k$ *and the rest of the columns of* $\widetilde{\mathbf{U}}_{(d-1)}$ *do not belong to* $\mathcal{I}_{k'}$ *and moreover,* $\bar{c}_k = c_{k'}$ *and* $\bar{\mathcal{T}}_k = \mathcal{T}_{k'}$*,* $k=1,\dots,K'$*.*

## 5. Union of Tensor Spaces Completion: Deterministic Analysis

### 5.1. Canonical Decomposition

Recall that for the problem of union of tensor spaces completion, we have $\widetilde{\mathbf{U}}_{(d-1)} = [\widetilde{\mathbf{U}}^1_{(d-1)}|\widetilde{\mathbf{U}}^2_{(d-1)}|\dots|\widetilde{\mathbf{U}}^K_{(d-1)}]$ and each column of $\widetilde{\mathbf{U}}^k_{(d-1)}$ is chosen generically from a tensor space $\mathcal{T}_k$, $k=1,\dots,K$. Also, $\mathcal{S}_k$ denotes the set of structured columns that are a basis for the tensor space $\mathcal{T}_k$, i.e., $\mathcal{T}_k$ is the column span of the structured columns in $\mathcal{S}_k$, $k=1,\dots,K$. Moreover, this union of tensor spaces is such that the $r_k$ structured columns of $\mathcal{S}_k$ can be denoted by $\mathbf{u}_1,\dots \mathbf{u}_{r_k}$, i.e., $\mathcal{S}_k = \{\mathbf{u}_1,\dots \mathbf{u}_{r_k}\}$, $k=1,\dots,K$, and as a consequence we have $\mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq \dots \subseteq \mathcal{S}_K$.

Recall that $\mathbf{u}_l = \text{vec}(\mathbf{a}^l_1 \otimes \dots \otimes \mathbf{a}^l_{d-1})$, $\mathbf{a}^l_i \in \mathbb{R}^{n_i}$ for $l=1,\dots,r_K$, where $\text{vec}(\cdot)$ uses the same bijective mappings that is used for the $(d-1)$-th unfolding, i.e., $\bar{M}_{d-1} : (x_1,\dots,x_{d-1}) \rightarrow \{1,2,\dots,N_{d-1}\}$. Denote $\mathbf{A}^k_i = [\mathbf{a}^{r_{k-1}+1}_i|\mathbf{a}^{r_{k-1}+2}_i|\dots|\mathbf{a}^{r_k}_i] \in \mathbb{R}^{n_i \times (r_k - r_{k-1})}$ for $i=1,\dots,d$ and $k=1,\dots,K$. Note that the decomposition in (1) is not unique (in fact there are infinitely many different decompositions). Next, we will impose certain structure on $\mathbf{a}^l_i$ such that the decomposition becomes unique and we will denote such unique decomposition by $\bar{\mathbf{a}}^l_i$, $i=1,\dots,d$ and $l=1,\dots,r_K$, which will be called the canonical decomposition.

The canonical structure features two properties: Property I and Property II. Property I given in Lemma 20, imposes structures on $\bar{\mathbf{a}}^l_d$, $l=1,\dots,r_K$, and is a consequence of the rank constraints and the assumption $\mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq \dots \subseteq \mathcal{S}_K$. Property II given in Lemma 24, imposes structures on $\bar{\mathbf{a}}^l_i$, $i=1,\dots,d-1$ and $l=1,\dots,r_K$, and is a consequence of the assumption that $\mathcal{S}_k$ is generated generically from its structured columns. This canonical structure is such that among all possible CP-decompositions of $\mathcal{U}$, exactly one of them satisfies the proposed canonical structure.

**Lemma 20** *There exist* $\bar{\mathbf{a}}^l_d \in \mathbb{R}^{n_d}$ *for* $l=1,\dots,r_K$ *such that* $\mathcal{U} = \sum_{l=1}^{r_K} \mathbf{a}^l_1 \otimes \dots \otimes \mathbf{a}^l_{d-1} \otimes \bar{\mathbf{a}}^l_d$ *and for any* $k=1,\dots,K$*,* $x=1,\dots,(c_1+\dots+c_{k-1})$ *and* $l=r_{k-1}+1,\dots,r_k$ *we have* $\bar{\mathbf{a}}^l_d(x)=0$*. In other words,* $\bar{\mathbf{A}}^k_d(1:(c_1+\dots+c_{k-1}), 1:r_k-r_{k-1}) = \mathbf{0}_{(c_1+\dots+c_{k-1})\times(r_k-r_{k-1})}$*. We call such property in the CP-decomposition of* $\mathcal{U}$ *as Property I.*

**Proof** Note that since each column of $\widetilde{\mathbf{U}}^k_{(d-1)}$ is chosen from $\mathcal{T}_k$, there exist $\mathbf{B}^k \in \mathbb{R}^{r_k \times c_k}$ such that $\widetilde{\mathbf{U}}^k_{(d-1)} = [\mathbf{u}_1|\dots|\mathbf{u}_{r_k}]\mathbf{B}^k$, $k=1,\dots,K$. Recall that $\widetilde{\mathbf{U}}_{(d-1)} = [\widetilde{\mathbf{U}}^1_{(d-1)}|\dots|\widetilde{\mathbf{U}}^K_{(d-1)}]$. Therefore, by considering the union of the mentioned decompositions of $\widetilde{\mathbf{U}}^k_{(d-1)}$ for $k=1,\dots,K$, we conclude

$$\widetilde{\mathbf{U}}_{(d-1)} = [\mathbf{u}_1|\dots|\mathbf{u}_{r_K}] \underbrace{[\mathbf{C}^1|\dots|\mathbf{C}^K]}_{[\bar{\mathbf{a}}_d^1|\dots|\bar{\mathbf{a}}_d^{r_K}]^\top}, \tag{12}$$

where $\mathbf{C}^k = [\mathbf{B}^{k^\top}|\mathbf{0}_{c_k \times (r_K - r_k)}]^\top \in \mathbb{R}^{r_K \times c_k}$ and $[\bar{\mathbf{a}}_d^1|\dots|\bar{\mathbf{a}}_d^{r_K}]^\top = [\mathbf{C}^1|\dots|\mathbf{C}^K]$. Hence, for any $k = 1,\dots,K$ and $l = r_{k-1}+1,\dots,r_k$ we have $\bar{\mathbf{a}}_d^l(x) = 0$ if $x \leq c_1 + \cdots + c_{k-1}$. Since $\mathbf{u}_l = \text{vec}(\mathbf{a}_1^l \otimes \dots \otimes \mathbf{a}_{d-1}^l)$ for $l = 1,\dots,r_K$, (12) can be written as

$$\mathcal{U} = \sum_{l=1}^{r_K} \mathbf{a}_1^l \otimes \dots \otimes \mathbf{a}_{d-1}^l \otimes \bar{\mathbf{a}}_d^l, \tag{13}$$

and hence, the proof is complete. ∎

**Remark 21** *Note that according to the described structure, $\bar{\mathbf{A}}_d^1$ has no pattern and all its entries are unknown variables.*

**Example 2** *Consider an example where $d = 3$, $K = 2$, $r_1 = 2$, $r_2 = 4$, $n_1 = 5$, $n_2 = 4$, $c_1 = 3$ and $c_2 = 3$ (recall that $n_3 = c_1 + c_2$). Here we show the mentioned structure for this example.*

*By Property I, we have*

$\bar{\mathbf{A}}_3^1 =$

| $\bar{\mathbf{a}}_3^1$ | $\bar{\mathbf{a}}_3^2$ |
|---|---|
| $\bar{\mathbf{A}}_3^1(1,1)$ | $\bar{\mathbf{A}}_3^1(1,2)$ |
| $\bar{\mathbf{A}}_3^1(2,1)$ | $\bar{\mathbf{A}}_3^1(2,2)$ |
| $\bar{\mathbf{A}}_3^1(3,1)$ | $\bar{\mathbf{A}}_3^1(3,2)$ |
| $\bar{\mathbf{A}}_3^1(4,1)$ | $\bar{\mathbf{A}}_3^1(4,2)$ |
| $\bar{\mathbf{A}}_3^1(5,1)$ | $\bar{\mathbf{A}}_3^1(5,2)$ |
| $\bar{\mathbf{A}}_3^1(6,1)$ | $\bar{\mathbf{A}}_3^1(6,2)$ |

, $\bar{\mathbf{A}}_3^2 =$

| $\bar{\mathbf{a}}_3^3$ | $\bar{\mathbf{a}}_3^4$ |
|---|---|
| 0 | 0 |
| 0 | 0 |
| 0 | 0 |
| $\bar{\mathbf{A}}_3^2(4,1)$ | $\bar{\mathbf{A}}_3^2(4,2)$ |
| $\bar{\mathbf{A}}_3^2(5,1)$ | $\bar{\mathbf{A}}_3^2(5,2)$ |
| $\bar{\mathbf{A}}_3^2(6,1)$ | $\bar{\mathbf{A}}_3^2(6,2)$ |

.

**Definition 22** *Consider an arbitrary number $i_0 \in \{1,\dots,d-1\}$. Property II for a CP-decomposition of $\mathcal{U}$ is defined as follows:*

*(i) All entries of the first row of $\bar{\mathbf{A}}_i^k$ are equal to one for $i \in \{1,\dots,d-1\} \backslash i_0$ and $1 \leq k \leq K$.*

*(ii)* $\bar{\mathbf{A}}_{i_0}^1(1:r_1, 1:r_1) = \mathbf{I}_{r_1}$, $\bar{\mathbf{A}}_{i_0}^2(1:r_2, 1:(r_2-r_1)) = [\mathbf{0}_{(r_2-r_1)\times r_1}|\mathbf{I}_{(r_2-r_1)}]^\top$, $\dots$, $\bar{\mathbf{A}}_{i_0}^K(1:r_K, 1:(r_K - r_{K-1})) = [\mathbf{0}_{(r_K-r_{K-1})\times r_{K-1}}|\mathbf{I}_{(r_K-r_{K-1})}]^\top$.

**Example 2. (continued)** *By Property II, for $i_0 = 1$, $\bar{\mathbf{A}}_1^1$ includes an $r_1 \times r_1$ identity matrix on the top. $\bar{\mathbf{A}}_1^2$ includes an $r_1 \times (r_2 - r_1)$ zero matrix on the top and then an $(r_2 - r_1) \times (r_2 - r_1)$ identity matrix below it. And for $i = 2$, $\bar{\mathbf{A}}_2^1$ and $\bar{\mathbf{A}}_2^2$ only include a row of all ones on the top.*

$\bar{\mathbf{A}}_1^1 =$

| $\bar{\mathbf{a}}_1^1$ | $\bar{\mathbf{a}}_1^2$ |
|---|---|
| 1 | 0 |
| 0 | 1 |
| $\bar{\mathbf{A}}_1^1(3,1)$ | $\bar{\mathbf{A}}_1^1(3,2)$ |
| $\bar{\mathbf{A}}_1^1(4,1)$ | $\bar{\mathbf{A}}_1^1(4,2)$ |
| $\bar{\mathbf{A}}_1^1(5,1)$ | $\bar{\mathbf{A}}_1^1(5,2)$ |

, $\bar{\mathbf{A}}_1^2 =$

| $\bar{\mathbf{a}}_1^3$ | $\bar{\mathbf{a}}_1^4$ |
|---|---|
| 0 | 0 |
| 0 | 0 |
| 1 | 0 |
| 0 | 1 |
| $\bar{\mathbf{A}}_1^2(5,1)$ | $\bar{\mathbf{A}}_1^2(5,2)$ |

,

*and*

$$\bar{\mathbf{A}}_2^1 = \begin{array}{|c|c|}\hline \bar{\mathbf{a}}_2^1 & \bar{\mathbf{a}}_2^2 \\ \hline\hline 1 & 1 \\ \hline \bar{\mathbf{A}}_2^1(2,1) & \bar{\mathbf{A}}_2^1(2,2) \\ \hline \bar{\mathbf{A}}_2^1(3,1) & \bar{\mathbf{A}}_2^1(3,2) \\ \hline \bar{\mathbf{A}}_2^1(4,1) & \bar{\mathbf{A}}_2^1(4,2) \\ \hline \end{array}, \bar{\mathbf{A}}_2^2 = \begin{array}{|c|c|}\hline \bar{\mathbf{a}}_2^3 & \bar{\mathbf{a}}_2^4 \\ \hline\hline 1 & 1 \\ \hline \bar{\mathbf{A}}_2^2(2,1) & \bar{\mathbf{A}}_2^2(2,2) \\ \hline \bar{\mathbf{A}}_2^2(3,1) & \bar{\mathbf{A}}_2^2(3,2) \\ \hline \bar{\mathbf{A}}_2^2(4,1) & \bar{\mathbf{A}}_2^2(4,2) \\ \hline \end{array}.$$

**Definition 23** *A CP-decomposition of $\mathcal{U}$ is called a canonical decomposition if and only if both Property I and Property II hold.*

**Example 3** *To illustrate the canonical decomposition of $\mathcal{U}$, we have provided the known entries or the canonical pattern for the components of the decomposition for $1 \leq l \leq r_1$ and $r_1 + 1 \leq l \leq r_2$ (i.e., $\bar{\mathbf{A}}_i^1$ and $\bar{\mathbf{A}}_i^2$) as the following.*

*For $1 \leq l \leq r_1$:*

$$\bar{\mathbf{A}}_{i_0}^1 = \begin{array}{|c|c|c|c|}\hline \bar{\mathbf{a}}_{i_0}^1 & \bar{\mathbf{a}}_{i_0}^2 & \dots & \bar{\mathbf{a}}_{i_0}^{r_1} \\ \hline\hline 1 & 0 & \dots & 0 \\ \hline 0 & 1 & \dots & 0 \\ \hline \vdots & \vdots & \dots & \vdots \\ \hline 0 & 0 & \dots & 1 \\ \hline \bar{\mathbf{a}}_{i_0}^1(r_1+1) & \bar{\mathbf{a}}_{i_0}^2(r_1+1) & \dots & \bar{\mathbf{a}}_{i_0}^{r_1}(r_1+1) \\ \hline \vdots & \vdots & \dots & \vdots \\ \hline \bar{\mathbf{a}}_{i_0}^1(n_{i_0}) & \bar{\mathbf{a}}_{i_0}^2(n_{i_0}) & \dots & \bar{\mathbf{a}}_{i_0}^{r_1}(n_{i_0}) \\ \hline \end{array},$$

*and for $i \in \{1, \dots, d-1\} \backslash i_0$*

$$\bar{\mathbf{A}}_i^1 = \begin{array}{|c|c|c|}\hline \bar{\mathbf{a}}_i^1 & \dots & \bar{\mathbf{a}}_i^{r_1} \\ \hline\hline 1 & \dots & 1 \\ \hline \bar{\mathbf{a}}_i^1(2) & \dots & \bar{\mathbf{a}}_i^{r_1}(2) \\ \hline \vdots & \dots & \vdots \\ \hline \bar{\mathbf{a}}_i^1(n_i) & \dots & \bar{\mathbf{a}}_i^{r_1}(n_i) \\ \hline \end{array}.$$

*For $r_1 + 1 \leq l \leq r_2$:*

$$\bar{\mathbf{A}}_{i_0}^2 = \begin{array}{|c|c|c|c|}\hline \bar{\mathbf{a}}_{i_0}^{r_1+1} & \bar{\mathbf{a}}_{i_0}^{r_1+2} & \dots & \bar{\mathbf{a}}_{i_0}^{r_2} \\ \hline\hline \mathbf{0}_{r_1} & \mathbf{0}_{r_1} & \dots & \mathbf{0}_{r_1} \\ \hline 1 & 0 & \dots & 0 \\ \hline 0 & 1 & \dots & 0 \\ \hline \vdots & \vdots & \dots & \vdots \\ \hline 0 & 0 & \dots & 1 \\ \hline \bar{\mathbf{a}}_{i_0}^{r_1+1}(r_2+1) & \bar{\mathbf{a}}_{i_0}^{r_1+2}(r_2+1) & \dots & \bar{\mathbf{a}}_{i_0}^{r_2}(r_2+1) \\ \hline \vdots & \vdots & \dots & \vdots \\ \hline \bar{\mathbf{a}}_{i_0}^{r_1+1}(n_{i_0}) & \bar{\mathbf{a}}_{i_0}^{r_1+2}(n_{i_0}) & \dots & \bar{\mathbf{a}}_{i_0}^{r_2}(n_{i_0}) \\ \hline \end{array},$$

*where $\mathbf{0}_{r_1}$ is all-zero vector of size $r_1$ and for $i \in \{1,\dots,d-1\}\backslash\{i_0\}$*

$$\bar{\mathbf{A}}_i^2 = \begin{array}{|c|c|c|} \hline \bar{\mathbf{a}}_i^{r_1+1} & \dots & \bar{\mathbf{a}}_i^{r_2} \\ \hline \hline 1 & \dots & 1 \\ \hline \bar{\mathbf{a}}_i^{r_1+1}(2) & \dots & \bar{\mathbf{a}}_i^{r_2}(2) \\ \hline \vdots & \dots & \vdots \\ \hline \bar{\mathbf{a}}_i^{r_1+1}(n_i) & \dots & \bar{\mathbf{a}}_i^{r_2}(n_i) \\ \hline \end{array} .$$

*Moreover, we have*

$$\bar{\mathbf{A}}_d^1 = \begin{array}{|c|c|c|} \hline \bar{\mathbf{a}}_d^{1} & \dots & \bar{\mathbf{a}}_d^{r_1} \\ \hline \hline \bar{\mathbf{a}}_d^{1}(1) & \dots & \bar{\mathbf{a}}_d^{r_1}(1) \\ \hline \vdots & \dots & \vdots \\ \hline \bar{\mathbf{a}}_d^{1}(n_d) & \dots & \bar{\mathbf{a}}_d^{r_1}(n_d) \\ \hline \end{array}, \ \bar{\mathbf{A}}_d^2 = \begin{array}{|c|c|c|} \hline \bar{\mathbf{a}}_d^{r_1+1} & \dots & \bar{\mathbf{a}}_d^{r_2} \\ \hline \hline \mathbf{0}_{c_1} & \dots & \mathbf{0}_{c_1} \\ \hline \bar{\mathbf{a}}_d^{r_1+1}(c_1+1) & \dots & \bar{\mathbf{a}}_d^{r_2}(c_1+1) \\ \hline \vdots & \dots & \vdots \\ \hline \bar{\mathbf{a}}_d^{r_1+1}(n_d) & \dots & \bar{\mathbf{a}}_d^{r_2}(n_d) \\ \hline \end{array} .$$

*Note that the canonical structure is mainly imposed on $\bar{\mathbf{a}}_{i_0}^l$ and $\bar{\mathbf{a}}_d^l$. And for $i \neq i_0, d$, the canonical structure means $\bar{\mathbf{a}}_i^l(1) = 1$.*

**Lemma 24** *Given that tensor $\mathcal{U}$ is chosen generically from the described union of tensor spaces, with probability one, there exists a unique canonical decomposition of $\mathcal{U}$.*

**Proof** We show our claim by induction on the value of $K$. For $K = 1$ the canonical structure is exactly the canonical structure in (Ashraphijuo and Wang, 2017) and therefore, according to Lemma 3 in (Ashraphijuo and Wang, 2017), there exists a unique canonical decomposition. Now, we assume that the statement holds true for $K = 1$ and we show the statement holds true for $K = 2$ as well (similarly we can show the statement for $K = k' + 1$ given the statement for $K = k'$).

For $K = 2$, recall that $\mathcal{U}_1 \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times c_1}$, $\mathcal{U}_2 \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times c_2}$ and $\mathcal{U} \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times (c_1+c_2)}$ is the concatenation of $\mathcal{U}_1$ and $\mathcal{U}_2$ along the $d$-th dimension. Also, rank$(\mathcal{U}_1) = r_1$, rank$(\mathcal{U}_2) =$ rank$(\mathcal{U}) = r_2$. Then we need to show that the following equation

$$\mathcal{U} = \sum_{l=1}^{r_2} \bar{\mathbf{a}}_1^l \otimes \bar{\mathbf{a}}_2^l \otimes \ldots \otimes \bar{\mathbf{a}}_{d-1}^l \otimes \bar{\mathbf{a}}_d^l, \tag{14}$$

has a unique solution, where $\{\bar{\mathbf{a}}_i^l\}$ satisfy Properties I and II. From (14), we can write

$$\mathcal{U}_1 = \mathcal{U}(:,:,\ldots,:,1:c_1) = \sum_{l=1}^{r_2} \bar{\mathbf{a}}_1^l \otimes \bar{\mathbf{a}}_2^l \otimes \ldots \otimes \bar{\mathbf{a}}_{d-1}^l \otimes \bar{\mathbf{a}}_d^l(1:c_1), \tag{15}$$

and

$$\mathcal{U}_2 = \mathcal{U}(:,:,\ldots,:,c_1+1:c_2) = \sum_{l=1}^{r_2} \bar{\mathbf{a}}_1^l \otimes \bar{\mathbf{a}}_2^l \otimes \ldots \otimes \bar{\mathbf{a}}_{d-1}^l \otimes \bar{\mathbf{a}}_d^l(c_1+1:c_2). \tag{16}$$

By the induction hypothesis, $\mathcal{U}_1$ has a unique canonical decomposition. Also note that by Property I, $\bar{\mathbf{a}}_d^l(x) = 0$ for $r_1 + 1 \leq l \leq r_2$ and $1 \leq x \leq c_1$. Hence, (15) becomes

$$\mathcal{U}_1 = \sum_{l=1}^{r_1} \bar{\mathbf{a}}_1^l \otimes \bar{\mathbf{a}}_2^l \otimes \ldots \otimes \bar{\mathbf{a}}_{d-1}^l \otimes \bar{\mathbf{a}}_d^l(1 : c_1). \tag{17}$$

Note that $\text{rank}(\mathcal{U}_1) = r_1$ and also the canonical structure on the entries that are involved in (17) are exactly the same canonical structure for $K = 1$ and a rank-$r_1$ tensor case. Therefore, according to the induction hypothesis, all the entries involved in (17), i.e., $\bar{\mathbf{a}}_1^l, \ldots, \bar{\mathbf{a}}_{d-1}^l, \bar{\mathbf{a}}_d^l(1 : c_1)$ for $l = 1, \ldots, r_1$, will be determined uniquely. As a results, we can rewrite (16) as

$$\begin{aligned}\mathcal{U}_2 = &\sum_{l=1}^{r_1} \bar{\mathbf{a}}_1^l \otimes \bar{\mathbf{a}}_2^l \otimes \ldots \otimes \bar{\mathbf{a}}_{d-1}^l \otimes \bar{\mathbf{a}}_d^l(c_1 + 1 : c_2) \\ + &\sum_{l=r_1+1}^{r_2} \bar{\mathbf{a}}_1^l \otimes \bar{\mathbf{a}}_2^l \otimes \ldots \otimes \bar{\mathbf{a}}_{d-1}^l \otimes \bar{\mathbf{a}}_d^l(c_1 + 1 : c_2).\end{aligned} \tag{18}$$

In the following, we first show that $\bar{\mathbf{a}}_d^l(c_1 + 1 : c_2)$ in the first term of (18) can be determined uniquely by the genericity assumption; then we show the vectors in the second term of (18) are unique by invoking the induction hypothesis.

For notational simplicity in this proof, assume that $i_0 = 1$. Then, from (18) we have

$$\begin{aligned}\mathcal{U}_2(1 : r_1, 1, \ldots, 1, 1 : c_2) &= \textstyle\sum_{l=1}^{r_2} \bar{\mathbf{a}}_1^l(1 : r_1) \otimes \bar{\mathbf{a}}_2^l(1) \otimes \ldots \otimes \bar{\mathbf{a}}_{d-1}^l(1) \otimes \bar{\mathbf{a}}_d^l(c_1 + 1 : c_1 + c_2) \\ &\overset{(a)}{=} \textstyle\sum_{l=1}^{r_1} \bar{\mathbf{a}}_1^l(1 : r_1) \otimes \bar{\mathbf{a}}_2^l(1) \otimes \ldots \otimes \bar{\mathbf{a}}_{d-1}^l(1) \otimes \bar{\mathbf{a}}_d^l(c_1 + 1 : c_1 + c_2),\end{aligned} \tag{19}$$

where (a) follows from the fact that the first $r_1$ entries of $\bar{\mathbf{a}}_1^l$ is zero for $l = r_1 + 1, \ldots, r_2$ (according to Property II). This linear system of $r_1 c_2$ equations in terms of the entries of $\bar{\mathbf{a}}_d^l(c_1 + 1 : c_1 + c_2)$ for $1 \leq l \leq r_1$ can be solved uniquely due to the genericity assumption. Note that $\bar{\mathbf{a}}_d^l(c_1 + 1 : c_1 + c_2)$ for $1 \leq l \leq r_1$ are the coefficients of the first $r_1$ structured columns. Now, we look at it the rest of the entries of (18), given by

$$\begin{aligned}\mathcal{U}_2(r_1 + 1 : n_1, :, \ldots, :) = &\sum_{l=1}^{r_1} \bar{\mathbf{a}}_1^l(r_1 + 1 : n_1) \otimes \bar{\mathbf{a}}_2^l \otimes \ldots \otimes \bar{\mathbf{a}}_{d-1}^l \otimes \bar{\mathbf{a}}_d^l(c_1 + 1 : c_1 + c_2) \\ + &\sum_{l=r_1+1}^{r_2} \bar{\mathbf{a}}_1^l(r_1 + 1 : n_1) \otimes \bar{\mathbf{a}}_2^l \otimes \ldots \otimes \bar{\mathbf{a}}_{d-1}^l \otimes \bar{\mathbf{a}}_d^l(c_1 + 1 : c_1 + c_2).\end{aligned} \tag{20}$$

Since all vectors in the first line of (20) have been shown unique, we consider

$$\begin{aligned}\mathcal{U}' = \mathcal{U}_2'(r_1 + 1 : n_1, :, \ldots, :) = \\ \textstyle\sum_{l=r_1+1}^{r_2} \bar{\mathbf{a}}_1^l(r_1 + 1 : n_1) \otimes \bar{\mathbf{a}}_2^l \otimes \ldots \otimes \bar{\mathbf{a}}_{d-1}^l \otimes \bar{\mathbf{a}}_d^l(c_1 + 1 : c_1 + c_2).\end{aligned} \tag{21}$$

Note that $\text{rank}(\mathcal{U}') = r_2 - r_1$ and also the canonical structure on the vectors in (21) is exactly the same canonical structure for $K = 1$ and a rank-$(r_2 - r_1)$ tensor case. Therefore, according to the induction hypothesis, all vectors in (21) can be determined uniquely. ∎

### 5.2. Polynomials

**Remark 25** *We consider the unique canonical decomposition for the sampled tensor and using the sampled entries, recovering the original tensor is equivalent with recovering the entries of the components of this unique decomposition. Henceforth, for notational simplicity, we will use $\mathbf{a}_i^l$ instead of $\bar{\mathbf{a}}_i^l$.*

Denote $C_k = c_1 + \cdots + c_k$ for $k = 1, \dots, K$. Then, according to Property I, we have

$$\mathcal{U}_k = \sum_{l=1}^{r_k} \mathbf{a}_1^l \otimes \dots \mathbf{a}_{d-1}^l \otimes \mathbf{a}_d^l(C_{k-1}+1 : C_k), \tag{22}$$

or equivalently

$$\mathcal{U}(x_1, \dots, x_{d-1}, x_d + C_{k-1}) = \mathcal{U}_k(x_1, \dots, x_{d-1}, x_d) = \sum_{l=1}^{r_k} \mathbf{a}_1^l(x_1) \dots \mathbf{a}_{d-1}^l(x_{d-1}) \mathbf{a}_d^l(x_d + C_{k-1}), \tag{23}$$

for $1 \leq x_d \leq c_k$ or equivalently, $C_{k-1} + 1 \leq x_d + C_{k-1} \leq C_k$.

The reason to represent the size of the last dimension by $C_K$ instead of $n_d$ is due to the fact that the $K$ original tensors are unionized over the last dimension and since in all of the statements and proofs we need the size of each of these tensors we used the notations $c_1, \dots, c_K$ and their sums $C_k = c_1 + \cdots + c_k$. Hence, $n_d$ is basically $C_K$ but as we use $c_k$'s in the statements throughout the paper, we used $C_K$ instead of $n_d$ here as well.

**Remark 26** *It can be seen from* (23)*, any observed entry $\mathcal{U}_k(\vec{x})$ results in an equation that involves one entry of $\mathbf{a}_i^l$, $i = 1, \dots, d$ and $l = 1, \dots, r_k$. Considering the entries of $\mathbf{a}_i^l$'s as variables (right-hand side of* (23)*), each observed entry results in a polynomial in terms of these variables. Moreover, for any observed entry $\mathcal{U}_k(\vec{x})$, the value of $x_i$ specifies the location of the entry of $\mathbf{a}_i^l$ that is involved in the corresponding polynomial, $i = 1, \dots, d$ and $l = 1, \dots, r_k$.*

**Assumption** 1: Each column of $\widetilde{\mathbf{U}}_{(d-1)}^k$ (or equivalently each row of $\mathbf{U}_{(d)}^k$) includes at least $r_k$ observed entries, for $1 \leq k \leq K$.

**Lemma 27** *Given $\mathbf{A}_i^k$'s for $i = 1, \dots, d-1$ and $1 \leq k \leq K$, Assumption* 1 *holds if and only if $\mathbf{A}_d^k$ can be determined uniquely for $1 \leq k \leq K$.*

**Proof** We show that each column of $\widetilde{\mathbf{U}}_{(d-1)}^k$ includes at least $r_k$ observed entries is equivalent to unique solvability of $\mathbf{a}_d^l(C_{k-1}+1 : C_k)$'s for $1 \leq l \leq r_k$, $k \in \{1, \dots, K\}$. This completes the proof since the rest of the entries of $\mathbf{A}_d^k$ for $1 \leq k \leq K$ are zero due to Property I. It can be seen from (12) that each observed entry in the $i$-th column of $\widetilde{\mathbf{U}}_{(d-1)}^k$ results in a degree-1 polynomial (because $[\mathbf{u}_1 | \dots | \mathbf{u}_{r_K}]$ is given) in terms of the $r_k$ entries of the $i$-th row of $[\mathbf{a}_d^1(C_{k-1}+1 : C_k) | \dots | \mathbf{a}_d^{r_k}(C_{k-1}+1 : C_k)] \in \mathbb{R}^{c_k \times r_k}$. Therefore, Assumption 1 is equivalent with assuming that we have $r_k$ degree-1 polynomials in terms of the entries of each row of $[\mathbf{a}_d^1(C_{k-1}+1 : C_k) | \dots | \mathbf{a}_d^{r_k}(C_{k-1}+1 : C_k)]$. Genericity of the coefficients of these polynomials results that with probability one each row of $[\mathbf{a}_d^1(C_{k-1}+1 : C_k) | \dots | \mathbf{a}_d^{r_k}(C_{k-1}+1 : C_k)]$ can be determined uniquely. ∎

**Remark 28** *According to Lemma 27, $\mathbf{A}_d^k$ can be determined uniquely in terms of the entries of $\mathbf{A}_i^k$'s for $i = 1, \dots, d-1$ and $k = 1, \dots, K$ and recall that each observed entry is equivalent to a polynomial in the form of* (23). *Consider all such polynomials excluding those that have been used to obtain $\mathbf{A}_d^k$ ($r_k$ samples in each column of $\widetilde{\mathbf{U}}_{(d-1)}^k$) and denote this set of polynomials in terms of the entries of $\mathbf{A}_i^k$'s for $i = 1, \dots, d-1$ and $k = 1, \dots, K$ by $\mathcal{P}(\Omega)$.*

### 5.3. Constraint Tensor

Recall that $\Omega \in \mathbb{R}^{n_1 \times \dots \times n_d}$ is the binary sampling pattern tensor. Let $\Omega_k \in \mathbb{R}^{n_1 \times \dots \times n_{d-1} \times c_k}$ denote the subtensor of $\Omega$ corresponding to $\mathcal{U}_k$ and $\mathcal{P}(\Omega_k)$ denote those polynomials of $\mathcal{P}(\Omega)$ that correspond to an observed entry of $\mathcal{U}_k$.

In the following, a procedure is described to construct a binary tensor $\breve{\Omega}$ based on $\Omega$, which is used to obtain the conditions for finite and unique completability. First, we construct $\breve{\Omega}_1$ from $\Omega_1$ according to the same procedure described in (Ashraphijuo and Wang, 2017), which is described here again for completeness.

For each subtensor $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times 1}$ of $\mathcal{U}_1$, let $N_\Omega(\mathcal{Y})$ denote the number of sampled entries in $\mathcal{Y}$. Then, since $r_1$ polynomials have been used to obtain $[\mathbf{a}_d^1(1:C_1)|\dots|\mathbf{a}_d^{r_1}(1:C_1)]$ (as $\mathcal{Y}$ is the tensor corresponding to a column of the $(d-1)$-th unfolding of $\mathcal{U}_1$), $\mathcal{Y}$ contributes $N_\Omega(\mathcal{Y}) - r_1$ polynomial equations in terms of the entries of $\mathbf{a}_i^l$'s for $1 \leq i \leq d-1$ among all $N_\Omega(\mathcal{U}_1) - r_1 c_1$ polynomials in $\mathcal{P}(\Omega_1)$.

$\mathcal{U}_1$ includes $c_1$ subtensors that belong to $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times 1}$ (columns of the $(d-1)$-th unfolding of $\mathcal{U}_1$) and let $\mathcal{Y}_t$ for $1 \leq t \leq c_1$ denote these $c_1$ subtensors. Define a binary valued tensor $\breve{\mathcal{Y}}_t \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times y_t}$, where $y_t = N_\Omega(\mathcal{Y}_t) - r_1$ and its entries are described as the following. We can look at $\breve{\mathcal{Y}}_t$ as $y_t$ tensors each belonging to $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times 1}$. We choose $r_1$ of the $N_\Omega(\mathcal{Y}_t)$ observed entries and for each of the mentioned $y_t$ tensors in $\breve{\mathcal{Y}}_t$, and we set the entries corresponding to these $r_1$ entries equal to 1. For each of the other $y_t$ observed entries (excluding the mentioned $r_1$ observed entries that we chose), we pick one of the $y_t$ tensors of $\breve{\mathcal{Y}}_t$ and set its corresponding entry (the same location as that specific observed entry) equal to $1$ and set the rest of the entries equal to $0$. In the case that $y_t = 0$ we simply ignore $\breve{\mathcal{Y}}_t$, i.e., $\breve{\mathcal{Y}}_t = \emptyset$.

By putting together all $c_1$ tensors in dimension $d$, we construct a binary valued tensor $\breve{\Omega}_1 \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times y}$, where $y = \sum_{i=1}^{c_1} y_t = N_\Omega(\mathcal{U}_1) - r_1 c_1$ and denote it by $\breve{\Omega}_1$. Similarly, we construct $\breve{\Omega}_k$ for $k = 2, \dots, K$. Then, we combine all these $\breve{\Omega}_k$'s together along dimension $d$ and call it the **constraint tensor** $\breve{\Omega}$. Observe that each subtensor of $\breve{\Omega}_k$ which belongs to $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times 1}$ includes exactly $r_k + 1$ nonzero entries.

For notational simplicity, define $\breve{\Omega}_{k:K}$ and $\Omega_{k:K}$ as the remaining of the constraint tensor $\breve{\Omega}$ and the sampling pattern $\Omega$ after removing $\breve{\Omega}_1, \dots, \breve{\Omega}_{k-1}$ and $\Omega_1, \dots, \Omega_{k-1}$, respectively. Moreover, in this paper, when we use $\breve{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times t}$ to denote a subtensor of $\breve{\Omega}$, $\breve{\Omega}'_{k:K}$ denotes the subtensor of $\breve{\Omega}'$ that also is a subtensor of $\breve{\Omega}_{k:K}$. Let $m_i(\breve{\Omega}')$ and $m_i(\breve{\Omega}'_{k:K})$ denote the number of nonzero rows of matricizations $\breve{\mathbf{\Omega}}'_{(i)}$ and $\breve{\mathbf{\Omega}}'_{k:K_{(i)}}$, respectively. Also, let $\mathcal{P}(\breve{\Omega}')$ denote the set of polynomials that correspond to nonzero entries of $\breve{\Omega}'$.

**Remark 29** *Note that when the $j$-th row of the $i$-th matricization of $\breve{\Omega}'_{k_0:K}$ has a non-zero entry, it means that the $j$-th row of the $i$-th matricization of $\Omega'_{k_0:K}$ has a non-zero entry and therefore, there exists a sampled entry (corresponding to the non-zero entry) such that its $i$-th coordinate is equal to $j$ and also it belongs to $\mathcal{U}_k$ for some $k \geq k_0$. Therefore, the entries of $j$-th row of $\mathbf{A}_i^{k_0}$ are involved in* *the polynomial corresponding to this sampled entry. Hence, the number of non-zero rows of the $i$-th matricization of $\breve{\Omega}'_{k_0:K}$ is equal to the number of rows of $\mathbf{A}_i^{k_0}$ that are involved in the polynomial in $\mathcal{P}(\breve{\Omega}')$.*

**Remark 30** *Note that each $\breve{\Omega}_k$ is constructed based on the number $r_k$ and the $r_k c_k$ sampled entries from Assumption 1, $k = 1, \ldots, K$. However, if we simply ignore the first $(K-1)$ rank constraints and apply the analysis in (Ashraphijuo and Wang, 2017), the constraint tensor will be constructed based only on the number $r_K$ and the $r_K(c_1 + \cdots + c_K)$ sampled entries from Assumption 1 in (Ashraphijuo and Wang, 2017). In the following example, we construct the constraint tensor based on the above description and compare it with the simple scenario from (Ashraphijuo and Wang, 2017).*

**Example 4** Consider an example where $d = 3$, $K = 2$, $r_1 = 1$, $r_2 = 2$, $n_1 = 4$, $n_2 = 4$, $c_1 = 2$ and $c_2 = 2$ (recall that $n_3 = c_1 + c_2$). Assume that $\Omega(x, y, z) = 1$ if $(x, y, z) \in \mathcal{S}$ and $\Omega(x, y, z) = 0$ otherwise (i.e., $\mathcal{S}$ denotes the set of sampled entries), where

$$\mathcal{S} = \{(1,1,1),(2,3,1),(4,3,1),(1,1,2),(2,4,2),(1,3,3),(3,2,3),(2,2,4),(2,4,4),(3,4,4),(4,1,4)\}.$$

Hence, observed entries $(1,1,1),(2,3,1),(4,3,1)$ belong to $\mathcal{Y}_1$, observed entries $(1,1,2),(2,4,2)$ belong to $\mathcal{Y}_2$, observed entries $(1,3,3),(3,2,3)$ belong to $\mathcal{Y}_3$, and observed entries $(2,2,4),(2,4,4)$, $(3,4,4),(4,1,4)$ belong to $\mathcal{Y}_4$. As a result, $y_1 = 3 - r_1 = 2$, $y_2 = 2 - r_1 = 1$, $y_3 = 2 - r_2 = 0$, and $y_4 = 4 - r_2 = 2$. Hence, $\breve{\mathcal{Y}}_1 \in \mathbb{R}^{4\times4\times2}$, $\breve{\mathcal{Y}}_2 \in \mathbb{R}^{4\times4\times1}$, $\breve{\mathcal{Y}}_3 = \emptyset$, and $\breve{\mathcal{Y}}_4 \in \mathbb{R}^{4\times4\times2}$, and therefore the constraint tensor $\breve{\Omega} \in \mathbb{R}^{4\times4\times5}$.

Also, assume that the entries that we use to obtain $\mathbf{A}_3^k$, in terms of the entries of $\mathbf{A}_1^k$ and $\mathbf{A}_2^k$, for $k = 1$ and 2, are $(1,1,1)$, $(1,1,2)$, $(1,3,3),(3,2,3)$, $(2,2,4)$ and $(2,4,4)$. Note that $\breve{\mathcal{Y}}_1(1,1,1) = \breve{\mathcal{Y}}_1(1,1,2) = 1$ (correspond to entries of $\mathcal{Y}_1$ that have been used to obtain $\mathbf{A}_3^k$'s), and also for the two other observed entries we have $\breve{\mathcal{Y}}_1(2,3,1) = 1$ (correspond to $\mathcal{U}(2,3,1)$) and $\breve{\mathcal{Y}}_1(4,3,2) = 1$ (correspond to $\mathcal{U}(4,3,1)$) and the rest of the entries of $\breve{\mathcal{Y}}_1$ are equal to zero. Similarly, $\breve{\mathcal{Y}}_2(1,1,1) = \breve{\mathcal{Y}}_2(2,4,1) = 1$ and the rest of the entries of $\breve{\mathcal{Y}}_2$ are equal to zero. $\breve{\mathcal{Y}}_3 = \emptyset$. $\breve{\mathcal{Y}}_4(2,2,1) = \breve{\mathcal{Y}}_4(2,2,2) = \breve{\mathcal{Y}}_4(2,4,1) = \breve{\mathcal{Y}}_4(2,4,2) = 1$ (correspond to entries of $\mathcal{Y}_4$ that have been used to obtain $\mathbf{A}_3^k$'s), and also for the two other observed entries we have $\breve{\mathcal{Y}}_4(3,4,1) = 1$ (correspond to $\mathcal{U}(3,4,4)$) and $\breve{\mathcal{Y}}_4(4,1,2) = 1$ (correspond to $\mathcal{U}(4,1,4)$) and the rest of the entries of $\breve{\mathcal{Y}}_4$ are equal to zero.

Then, $\breve{\Omega}(x, y, z) = 1$ if $(x, y, z) \in \breve{\mathcal{S}}$ and $\breve{\Omega}(x, y, z) = 0$ otherwise, where

$$\begin{aligned}\breve{\mathcal{S}} = \{&(1,1,1),(1,1,2),(2,3,1),(4,3,2),(1,1,3),(2,4,3),\\ &(2,2,4),(2,2,5),(2,4,4),(2,4,5),(3,4,4),(4,1,5)\}.\end{aligned}$$

Note that if we just construct the constraint tensor based on only one rank constraint $\text{rank}(\mathcal{U}) = r_K$, as given in (Ashraphijuo and Wang, 2017), $\breve{\Omega}$ would have been totally different. For example, in this case it is required to use two observed entries of $\mathcal{Y}_1, \ldots, \mathcal{Y}_4$ and therefore in this case $\breve{\Omega} \in \mathbb{R}^{4\times4\times3}$, as $\breve{\mathcal{Y}}_1 \in \mathbb{R}^{4\times4\times1}$, $\breve{\mathcal{Y}}_2 = \emptyset$, $\breve{\mathcal{Y}}_3 = \emptyset$, and $\breve{\mathcal{Y}}_4 \in \mathbb{R}^{4\times4\times2}$. To mention another fundamental difference between these two cases, according to the above definitions even if we assume that $\mathcal{U}(2,4,2)$ is not observed anymore, still Assumption 1 holds and we can construct the constraint tensor and it would belong to $\mathbb{R}^{4\times4\times4}$. But in the case that we only use the rank constraint $\text{rank}(\mathcal{U}) = r_K$ and the definition of constraint tensor in (Ashraphijuo and Wang, 2017), the constraint tensor cannot be even built because Assumption 1 in (Ashraphijuo and Wang, 2017) is not satisfied.

### 5.4. Finite Completability

Denote $D = (n_1 + \cdots + n_{d-1})r_K - \sum_{k=1}^{K} r_k(r_k - r_{k-1}) - r_K(d-2)$. Then, $D$ is the number of unknown entries of $\bar{\mathbf{a}}_i^l$'s for $1 \leq i \leq d-1$ and $1 \leq l \leq r_K$, i.e., the number of entries excluding the 0's and 1's in the structure of Property II.

**Lemma 31** *$\mathcal{U}_\Omega$ is finitely completable with probability one if and only if the maximum number of algebraically independent polynomials in $\mathcal{P}(\Omega)$ is equal to $D$.*

**Proof** Assume that $\mathcal{L}$ denotes the set of all possible $\mathbf{a}_i^l$'s for $i = 1, \dots, d-1$ and $1 \leq l \leq r_K$ given $\mathbf{a}_d^l$'s for $1 \leq l \leq r_K$ without any observed entry, i.e., without any polynomial constraint of the form of (23) and the dimension of $\mathcal{L}$ is equal to the number of unknown entries $D$. The rest of the proof follows easily as each algebraically independent polynomial reduces the dimension of the set of solution by one and also finite completability is equivalent to the dimension of the set of solution being zero, similar to the proof of Lemma 4 in (Ashraphijuo and Wang, 2017). ∎

The set of polynomials corresponding to $\breve{\Omega}'$, i.e., $\mathcal{P}(\breve{\Omega}')$ is called minimally algebraically dependent if the polynomials in $\mathcal{P}(\breve{\Omega}')$ are algebraically dependent but polynomials in every of its proper subsets are algebraically independent. The following lemma which is Lemma 7 in (Ashraphijuo and Wang, 2017), provides a useful property about a set of minimally algebraically dependent $\mathcal{P}(\breve{\Omega}')$, which will be used later to derive the maximum number of algebraically independent polynomials in $\mathcal{P}(\breve{\Omega}')$.

**Lemma 32** *Given Assumption 1, consider a subtensor $\breve{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times t}$ of the constraint tensor $\breve{\Omega}$. Assume that polynomials in $\mathcal{P}(\breve{\Omega}')$ are minimally algebraically dependent. Then, the number of variables (unknown entries) of $\mathbf{a}_i^l$'s for $1 \leq i \leq d-1$ and $1 \leq l \leq r_K$ that are involved in $\mathcal{P}(\breve{\Omega}')$ is equal to $t-1$.*

In order to obtain the maximum number of algebraically independent polynomials in a set of polynomials, we first need to derive the number of involved entries of the CP-decomposition in the polynomials (Lemma 33) and then, obtain the number of involved variables (unknowns) of the CP-decomposition in the polynomials (Lemma 34).

**Lemma 33** *Given Assumption 1, consider a subtensor $\breve{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times t}$ of the constraint tensor $\breve{\Omega}$. The number of entries of the first $d-1$ components of CP-decomposition that are involved in at least one of the polynomials in $\mathcal{P}(\breve{\Omega}')$ is*

$$\sum_{k=1}^{K} (r_k - r_{k-1}) (\sum_{i=1}^{d-1} m_i(\breve{\Omega}'_{k:K})). \tag{24}$$

**Proof** It is easily verified from (1) and Property I that the entries of the $j$-th row of $\mathbf{A}_i^{k_0}$ are involved in a certain polynomial if and only if, first the $i$-th coordinate of the corresponding sampled entry is equal to $j$ and second, the sampled entry belongs to $\mathcal{U}_k$ for some $k \geq k_0$. Note that the first condition is a result of (1) and the second condition is a result of Property I. Moreover, note that if a sampled entry is such that its $i$-th coordinate is equal to $j$ and also the sampled entry belongs to $\mathcal{U}_k$

for some $k \geq k_0$, then we know that the $j$-th row of the $i$-th unfolding of $\breve{\Omega}'_{k:K}$ has a non-zero entry (corresponding the mentioned sampled entry).

Recall that according to Remark 29, the number of rows of $\mathbf{A}_i^k$ that are involved in the polynomials in $\mathcal{P}(\breve{\Omega}')$ is equal to $m_i(\breve{\Omega}'_{k:K})$ and therefore, the number of entries of $\mathbf{A}_i^k$ that are involved in the polynomials in $\mathcal{P}(\breve{\Omega}')$ is equal to $(r_k - r_{k-1})m_i(\breve{\Omega}'_{k:K})$ since $\mathbf{A}_i^k$ has $r_k - r_{k-1}$ columns. ■

**Lemma 34** *Given Assumption* 1*, consider a subtensor* $\breve{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times t}$ *of the constraint tensor* $\breve{\Omega}$*. The maximum number of algebraically independent polynomials in* $\mathcal{P}(\breve{\Omega}')$ *is at most*

$$\begin{aligned} f(\breve{\Omega}') \equiv & \sum_{k=1}^{K} (r_k - r_{k-1}) (\sum_{i=1}^{d-1} m_i(\breve{\Omega}'_{k:K})) \\ & - \max_{1 \leq i \leq d-1} \left\{ \sum_{k=1}^{K} \min \left\{ (r_k - r_{k-1}) m_i(\breve{\Omega}'_{k:K}), r_k(r_k - r_{k-1}) \right\} \right\} \\ & -(d-2) \sum_{k=1}^{K} \min \left\{ r_k - r_{k-1}, (r_k - r_{k-1}) m_1(\breve{\Omega}'_{k:K}) \right\}. \end{aligned} \tag{25}$$

**Proof** The maximum number of algebraically independent polynomials in a subset of polynomials of $\mathcal{P}(\breve{\Omega}')$ is at most equal to the total number of variables that are involved in the corresponding polynomials as each polynomial reduces the dimension of the set of solutions by one. As mentioned in Remark 26, due to the structure of each polynomial obtained from a sampled entry in $\mathcal{U}_k$ (given in (22) or (23)), exactly one entry (corresponding coordinate) of each $\mathbf{a}_i^l$ for $1 \leq i \leq d-1$ and $1 \leq l \leq r_k$ is involved in the polynomial. According to Lemma 33, the number of entries of $\mathbf{A}_i^k$ that are involved in the polynomials $\mathcal{P}(\breve{\Omega}')$ is equal to $(r_k - r_{k-1})m_i(\breve{\Omega}'_{k:K})$ and hence, the number of entries of $\mathbf{a}_i^l$'s for $1 \leq i \leq d-1$ and $1 \leq l \leq r_K$ that are involved in the polynomials $\mathcal{P}(\breve{\Omega}')$ is equal to $\sum_{k=1}^{K}(r_k - r_{k-1})(\sum_{i=1}^{d-1} m_i(\breve{\Omega}'_{k:K}))$. However, some of the entries of $\mathbf{a}_i^l$'s are known as in Definition 22, and we should subtract them from the total number of involved entries.

Note that any permutation of rows of $[\mathbf{A}_i^1|\dots|\mathbf{A}_i^K]$ in the canonical structure preserves the same property as in Lemma 24 and therefore, assuming that $i_0$ is a fixed number in Definition 22, we can simply observe that the maximum number of known entries of $[\mathbf{A}_{i_0}^1|\dots|\mathbf{A}_{i_0}^K]$ that are involved in the polynomials $\mathcal{P}(\breve{\Omega}')$ is

$$\sum_{k=1}^{K} \min \left\{ (r_k - r_{k-1}) m_{i_0}(\breve{\Omega}'_{k:K}), r_k(r_k - r_{k-1}) \right\}. \tag{26}$$

This is because as long as $m_{i_0}(\breve{\Omega}'_{k:K}) \leq r_k$, all $m_{i_0}(\breve{\Omega}'_{k:K})$ nonzero rows can be the 0's and 1's in the canonical structure and when $m_{i_0}(\breve{\Omega}'_{k:K}) > r_k$, all the 0's and 1's in the canonical structure ($r_k(r_k - r_{k-1}$ entries) are involved, since for a permutation of rows of $[\mathbf{A}_i^1|\dots|\mathbf{A}_i^K]$ in the canonical structure can maximize the number involved known entries. In other words, for a permutation of rows of $[\mathbf{A}_i^1|\dots|\mathbf{A}_i^K]$ in the canonical structure, the number of involved known entries reaches its possible maximum, which is given in (26). Therefore, by changing $i_0$ from 1 to $d-1$ we can obtain the maximum number of known entries of $\mathbf{A}_{i_0}^1$ as

$$\max_{1\leq i\leq d-1}\left\{\sum_{k=1}^{K}\min\left\{(r_k-r_{k-1})m_i(\breve{\Omega}'_{k:K}),r_k(r_k-r_{k-1})\right\}\right\}. \tag{27}$$

Similarly, we can obtain the maximum number of involved known entries in the polynomials for the other $(d-2)$ dimensions as the following. On the other hand, there is one single known entry in all other $\mathbf{a}_i^l$'s for $i\in\{1,\dots,d-1\}\backslash i_0$ in the canonical structure. Again by permuting the corresponding row, it is easily verified that the corresponding known entry of $\mathbf{A}_i^k$ can be involved in at least one of the polynomials in $\mathcal{P}(\breve{\Omega}')$ if $m_1(\breve{\Omega}'_{k:K})$. Note that $m_1(\breve{\Omega}'_{k:K})\geq 1$ is equivalent to $m_i(\breve{\Omega}'_{k:K})\geq 1$ for any $i$ as existence of one polynomial corresponding to a sampled entry in $\mathcal{U}_k$ results that the number of nonzero rows in any of the matricizations of $\mathcal{U}_k$ is at least one. Therefore, the maximum number of involved known entries for the these $(d-2)$ dimensions combined is

$$(d-2)\sum_{k=1}^{K}\min\left\{r_k-r_{k-1},(r_k-r_{k-1})m_1(\breve{\Omega}'_{k:K})\right\}. \tag{28}$$

Hence, for a canonical pattern, the number of variables, i.e., the unknown entries of $\mathbf{a}_i^l$'s for $1\leq i\leq d-1$ and $1\leq l\leq r_K$ that are involved in the polynomials $\mathcal{P}(\breve{\Omega}')$ is equal to $f(\breve{\Omega}')$ given in (25). ■

Given a subtensor $\breve{\Omega}'\in\mathbb{R}^{n_1\times n_2\times\cdots\times n_{d-1}\times t}$ of the constraint tensor $\breve{\Omega}$, we are interested in obtaining the maximum number of algebraically independent polynomials in $\mathcal{P}(\breve{\Omega}')$ based on the structure of nonzero entries of $\breve{\Omega}'$. The next lemma can be used to characterize this number in terms of a simple geometric structure of nonzero entries of $\breve{\Omega}'$.

**Lemma 35** *Given Assumption* 1*, consider a subtensor $\breve{\Omega}'\in\mathbb{R}^{n_1\times n_2\times\cdots\times n_{d-1}\times t}$ of the constraint tensor. The polynomials in $\mathcal{P}(\breve{\Omega}')$ are algebraically independent if and only if for any $t'\in\{1,\dots,t\}$ and any subtensor $\breve{\Omega}''\in\mathbb{R}^{n_1\times n_2\times\cdots\times n_{d-1}\times t'}$ of $\breve{\Omega}''$ we have $f(\breve{\Omega}'')\geq t'$.*

**Proof** By contradiction, assume that the polynomials in $\mathcal{P}(\breve{\Omega}')$ are algebraically dependent. Hence, there exists a subset of the polynomials that are minimally algebraically dependent and let us denote it by $\mathcal{P}(\breve{\Omega}'')\in\mathbb{R}^{n_1\times n_2\times\cdots\times n_{d-1}\times t'}$, which is a subtensor of $\breve{\Omega}'$. As showed in Lemma 34, the number involved variables (unknown entries) is at least $f(\breve{\Omega}'')$. According to Lemma 32, the number of involved variables in polynomials in $\mathcal{P}(\breve{\Omega}'')$ is equal to $t'-1$. Consequently, $f(\breve{\Omega}'')\leq t'-1$.

Now, assume that there exists a subtensor $\breve{\Omega}''\in\mathbb{R}^{n_1\times\cdots\times n_{d-1}\times t'}$ of $\breve{\Omega}'$ that $f(\breve{\Omega}'')<t'$. Note that according to the definitions, $\mathcal{P}(\breve{\Omega}'')$ includes $t'$ polynomials. Moreover, according to Lemma 34 the maximum number of algebraically independent polynomials in $\mathcal{P}(\breve{\Omega}'')$ is at most $f(\breve{\Omega}'')$. Hence, the polynomials in $\mathcal{P}(\breve{\Omega}'')$ (and therefore in $\mathcal{P}(\breve{\Omega}')$) are not algebraically independent. ■

Finally, the following theorem characterizes the necessary and sufficient condition on $\breve{\Omega}$ for finite completability of the sampled tensor $\mathcal{U}$.

**Theorem 36** *Given Assumption* 1*, with probability one, $\mathcal{U}_\Omega$ is finitely many completable if and only if $\breve{\Omega}$ contains a subtensor $\breve{\Omega}'\in\mathbb{R}^{n_1\times\cdots\times n_{d-1}\times D}$ such that for any $D'\in\{1,\dots,D\}$ and any subtensor $\breve{\Omega}''\in\mathbb{R}^{n_1\times\cdots\times n_{d-1}\times D'}$ of $\breve{\Omega}'$, $f(\breve{\Omega}'')\geq D'$.*

**Proof** This theorem is an easy conclusion of Lemmas 31 and 35. Specifically, according to Lemma 31, $\mathcal{U}$ is finitely many completable with probability one if and only if $\mathcal{P}(\breve{\Omega})$ includes $D$ algebraically independent polynomials. On the other hand, Lemma 35 results that polynomials corresponding to a subtensor $\breve{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times D}$ of the constraint tensor are algebraically independent if and only if for any $D' \in \{1, \dots, D\}$ and any subtensor $\breve{\Omega}'' \in \mathbb{R}^{n_1 \times \cdots \times n_{d-1} \times D'}$ of $\breve{\Omega}'$, $f(\breve{\Omega}'') \geq D'$. ■

**Remark 37** *If we set $K = 1$, then using* (25)*, the condition in Theorem 36 simply reduces to*

$$r_K \left( \left( \sum_{i=1}^{d-1} m_i(\breve{\Omega}'') \right) - \min \left\{ \max \left\{ m_1(\breve{\Omega}''), \dots, m_{d-1}(\breve{\Omega}'') \right\}, r_K \right\} - (d-2) \right) \geq D', \quad (29)$$

*and therefore, Theorem 36 reduces to Theorem 1 in (Ashraphijuo and Wang, 2017) for finite completability of a regular tensor.*

### 5.5. Unique Completability

In the previous subsection we characterized the deterministic conditions on the sampling pattern for finite completability in Theorem 36. However, knowing whether $\mathcal{U}$ is uniquely completable can be very useful. For example, having the unique completability property, any valid completion provided by an optimization algorithm is the original sampled union of tensor spaces; or as we observed in Section 4, the unique completability can be useful in clustering problems. According to Theorem 36 finite completability is equivalent to having $D$ algebraically independent polynomials. As a result, adding any single polynomial to these $D$ algebraically independent polynomials results in a set of algebraically dependent polynomials. Then, according to Lemma 38 below, a certain subset of the entries of $\mathbf{a}_i^l$'s can be determined uniquely and these additional polynomials are captured in the structure of new condition on the sampling pattern given in Theorem 40 below such that all entries of the canonical decomposition can be determined uniquely.

The following lemma is a re-statement of Lemma 25 in (Ashraphijuo and Wang, 2017), which shows that variables involved in a minimally algebraically dependent polynomials can be determined uniquely.

**Lemma 38** *Given Assumption* 1*, consider a subtensor $\breve{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_{d-1} \times t}$ of the constraint tensor $\breve{\Omega}$. Assume that polynomials in $\mathcal{P}(\breve{\Omega}')$ are minimally algebraically dependent. Then, all variables that are involved in $\mathcal{P}(\breve{\Omega}')$ can be determined uniquely.*

The following lemma is useful in proving Theorem 40.

**Lemma 39** *Consider a subtensor $\breve{\Omega}'_K \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times t}$ of $\breve{\Omega}_K$ such that $m_i(\breve{\Omega}'_K) = n_i$. Then, all the rows of $[\mathbf{A}_i^1 | \dots | \mathbf{A}_i^K]$ are involved in at least one of the polynomials in $\mathcal{P}(\breve{\Omega}'_K)$.*

**Proof** Recall from the proof of Lemma 33 that $m_i(\breve{\Omega}'_K)$ denotes the number of rows of $[\mathbf{A}_i^1 | \dots | \mathbf{A}_i^K]$ that are involved in at least one of the polynomials in $\mathcal{P}(\breve{\Omega}'_K)$ and hence, the proof is completed as $[\mathbf{A}_i^1 | \dots | \mathbf{A}_i^K]$ has $n_i$ rows. ■

**Theorem 40** *Suppose that Assumption* 1 *holds and also there exists a subtensor* $\breve{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times D}$ *of* $\breve{\Omega}$ *such that the condition given in Theorem 36 holds (we call it Condition (i)). Moreover, assume that there exists a subtensor* $\breve{\Omega}'_K \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times t}$ *of* $\breve{\Omega}_K$ *which is disjoint from* $\breve{\Omega}'$ *and the following Condition (ii) holds:* $m_i(\breve{\Omega}'_K) = n_i$ *for* $1 \leq i \leq d-1$*. Then, with probability one, there exists exactly one union of tensors that fits* $\mathcal{U}_\Omega$ *and satisfies the rank constraints for each tensor space (which is the original sampled union of tensor spaces* $\mathcal{U}$*).*

**Proof** Note that to complete the proof, it suffices to show that all unknown entries of $\mathbf{a}_i^l$'s for $1 \leq i \leq d-1$ and $1 \leq l \leq r_K$ can be determined uniquely. Because then, Lemma 27 results in the uniqueness of $\mathbf{a}_d^l$'s. According to Lemma 39, Condition (ii) ensures that for each entry of $\mathbf{a}_i^l$'s for $1 \leq i \leq d-1$ and $1 \leq l \leq r_K$, there exists at least one polynomial $p_0 \in \mathcal{P}(\breve{\Omega}'_K)$ that involves that particular entry. In the following, we show that we can obtain all the variables involved in any polynomial $p_0 \in \mathcal{P}(\breve{\Omega}'_K)$ uniquely, and consequently according to Lemma 39, all unknown entries of $\mathbf{a}_i^l$'s for $1 \leq i \leq d-1$ and $1 \leq l \leq r_K$ can be determined uniquely.

Therefore, we only need to show that all variables involved in any polynomial $p_0 \in \mathcal{P}(\breve{\Omega}'_K)$ can be determined uniquely. According to Theorem 36, condition (i) means that $\mathcal{P}(\breve{\Omega}')$ includes $D$ algebraically independent polynomials and denote these polynomials by $\{p_1, \dots, p_D\}$. By adding any polynomial $p_0$ to $\{p_1, \dots, p_D\}$ we will have a set of algebraically dependent polynomials as they are in terms of $D$ variables. Hence, there exists a subset of polynomials $P' \subset \{p_1, \dots, p_D\}$ such that polynomials in $P' \cup p_0$ are minimally algebraically dependent polynomials. By Lemma 38 we can obtain all variables involved in $P' \cup p_0$ uniquely, and consequently, all variables involved in $p_0$ can be determined uniquely. ∎

## 6. Union of Tensor Spaces Completion: Probabilistic Analysis

In this section, we obtain the probabilistic versions of Theorem 36 (finite completability) and Theorem 40 (unique completability) in Theorems 42 and 43, respectively. Recall that $\breve{\Omega}'_k$ is a subtensor of $\breve{\Omega}'_{k:K}$. First note that we relax the statements in Theorems 36 and 40 in terms of the number of nonzero rows of $\breve{\Omega}'_k$ instead of $\breve{\Omega}'_{k:K}$.

The following lemma is a re-statement of Lemma 19 in (Ashraphijuo and Wang, 2017) and provides a bound on the number of samples that ensures the sampling pattern satisfies a certain condition as in (31). In particular, it connects the relationship between the number of sampled entries and a geometrical condition on the non-zero entries of the constraint tensor. Note that in the following lemma, $r$ and $r'_i$'s are not "rank" values but are just numbers that satisfy some given properties (similar to Lemma 19 in (Ashraphijuo and Wang, 2017)). Moreover, note that when we consider a submatrix $\widetilde{\mathbf{\Omega}}'_{(d-1)}$ of the $(d-1)$-th unfolding of the sampling pattern, we denote its corresponding submatrix of the $(d-1)$-th unfolding of the constraint tensor by $\breve{\widetilde{\mathbf{\Omega}}}'_{(d-1)}$.

**Lemma 41** *Assume that* $(\min_{1 \leq i \leq d-1} n_i) > 200$ *and also* $\widetilde{\mathbf{\Omega}}_{(d-1)}$ *includes at least* $r(n_i - r'_i)$ *columns, where* $r'_i \leq r \leq \frac{n_i}{6}$*. Assume that each column of* $\widetilde{\mathbf{U}}_{(d-1)}$ *includes at least* $l$ *nonzero entries, where*

$$l > \max \left\{ 27 \ \log \left( \frac{\max_{1 \leq i \leq d-1} n_i}{\epsilon} \right) + 9 \ \log \left( \frac{2rs}{\epsilon} \right) + 18, 6r'_i \right\}. \tag{30}$$

*Then, there exists an $N_{d-1} \times r(n_i - r'_i)$ matrix $\breve{\widetilde{\mathbf{\Omega}}}'_{(d-1)}$ such that: each column has exactly $r'_i + 1$ entries equal to one, and if $\breve{\widetilde{\mathbf{\Omega}}}'_{(d-1)}(x, y) = 1$ then we have $\widetilde{\mathbf{\Omega}}'_{(d-1)}(x, y) = 1$ and also it satisfies the following property: with probability at least $1 - \frac{\epsilon}{s}$, every subset $\breve{\widetilde{\mathbf{\Omega}}}''_{(d-1)}$ of columns of $\breve{\widetilde{\mathbf{\Omega}}}'_{(d-1)}$ satisfies the following inequality*

$$r\left(m_i(\breve{\Omega}'') - r'_i\right) \geq t, \tag{31}$$

*where $t$ is the number of columns of $\breve{\widetilde{\mathbf{\Omega}}}''_{(d-1)}$ and $\breve{\Omega}''$ is the tensor corresponding to unfolding $\breve{\widetilde{\mathbf{\Omega}}}''_{(d-1)}$.*

In the following theorem, we are interested to find the required number of samples to ensure that $\breve{\Omega}$ contains a subtensor $\breve{\Omega}' \in \mathbb{R}^{n_1 \times \cdots \times n_{d-1} \times D}$ such that for any $D' \in \{1, \ldots, D\}$ and any subtensor $\breve{\Omega}'' \in \mathbb{R}^{n_1 \times \cdots \times n_{d-1} \times D'}$ of $\breve{\Omega}'$, $D' \leq f(\breve{\Omega}'')$ as shown in (43). To this end we use Lemma 41 along different dimensions several times (Eq. (33)) and then combine them (Eq. (39)) by using the following fact; assuming that $a \geq a_0$ and $b \geq b_0$ with probabilities at least $1 - \epsilon_1$ and $1 - \epsilon_2$, respectively, then $a + b \geq a_0 + b_0$ with probability at least $1 - \epsilon_1 - \epsilon_2$.

Note that the conditions $r_k \leq \frac{n_k}{6}$ and $\min_{1 \leq i \leq d-1} n_i > 200$ in the theorem below are required for the probabilistic analysis but they are not required in the deterministic analysis.

**Theorem 42** *Assume that $r_k \leq \frac{n_k}{6}$ and $(\min_{1 \leq i \leq d-1} n_i) > 200$ and $c_k \geq (r_k - r_{k-1})(\sum_{i=1}^{d-1} n_i) - r_k(r_k - r_{k-1}) - (d-2)(r_k - r_{k-1})$ for $1 \leq k \leq K$. Assume that each column of $\widetilde{\mathbf{\Omega}}_{(d-1)}$ includes at least $l$ nonzero entries, where*

$$l > \max\left\{27 \log\left(\frac{\max_{1 \leq i \leq d-1} n_i}{\epsilon}\right) + 9 \log\left(\frac{2K(d-1)\left(\max_{1 \leq k \leq K}\{r_k - r_{k-1}\}\right)}{\epsilon}\right) + 18, 6r_K\right\}. \tag{32}$$

*Then, with probability at least $1 - \epsilon$, $\mathcal{U}_\Omega$ is finitely many completable.*

**Proof** Since the proof is long, we first provide an overview. First, we consider disjoint sets of columns $\widetilde{\mathbf{\Omega}}'^{(i,k)}_{(d-1)}$ of $\widetilde{\mathbf{\Omega}}^k_{(d-1)}$, where $\widetilde{\mathbf{\Omega}}'^{(i,k)}_{(d-1)} \in \mathbb{R}^{N_{d-1} \times (r_k - r_{k-1})(n_i - 1)}$ for $i \in \{1, \ldots, d-1\} \backslash \{i_0\}$ and $\widetilde{\mathbf{\Omega}}'^{(i_0,k)}_{(d-1)} \in \mathbb{R}^{N_{d-1} \times (r_k - r_{k-1})(n_{i_0} - r_k)}$. Then, we use Lemma 42 to conclude (33) holds. Then, we put these columns together and show that (39) holds. Finally, we derive (43) from (39), which completes the proof according to Theorem 36.

In order to prove this theorem, we need to show that $\breve{\Omega}$ contains a subtensor that satisfies the property described in the statement of Theorem 36. The assumption $c_k \geq (r_k - r_{k-1})(\sum_{i=1}^{d-1} n_i) - r_k(r_k - r_{k-1}) - (d-2)(r_k - r_{k-1})$ results that there exist disjoint sets of columns $\widetilde{\mathbf{\Omega}}'^{(i,k)}_{(d-1)}$ of $\widetilde{\mathbf{\Omega}}^k_{(d-1)}$, where $\widetilde{\mathbf{\Omega}}'^{(i,k)}_{(d-1)} \in \mathbb{R}^{N_{d-1} \times (r_k - r_{k-1})(n_i - 1)}$ for $i \in \{1, \ldots, d-1\} \backslash \{i_0\}$ and $\widetilde{\mathbf{\Omega}}'^{(i_0,k)}_{(d-1)} \in \mathbb{R}^{N_{d-1} \times (r_k - r_{k-1})(n_{i_0} - r_k)}$.

Define $r'_{(i,k)} = 1$ if $i \in \{1, \ldots, d-1\} \backslash \{i_0\}$ and $r'_{(i,k)} = r_k$ if $i = i_0$. By Lemma 41 we conclude that there exists $\breve{\widetilde{\mathbf{\Omega}}}'^{(i,k,1)}_{(d-1)} \in \mathbb{R}^{N_{d-1} \times (r_k - r_{k-1})(n_i - r'_{(i,k)})}$ such that each column of it includes $r'_{(i,k)} + 1$ nonzero entries and if $\breve{\widetilde{\mathbf{\Omega}}}'^{(i,k,1)}_{(d-1)}(x, y) = 1$ then we have $\widetilde{\mathbf{\Omega}}'^{(i,k)}_{(d-1)}(x, y) = 1$ and moreover, with probability at least $1-\frac{\epsilon}{(d-1)K}$, every subset $\tilde{\breve{\mathbf{\Omega}}}''^{(i,k,1)}_{(d-1)} \in \mathbb{R}^{N_{d-1}\times t}$ of columns of $\tilde{\breve{\mathbf{\Omega}}}'^{(i,k,1)}_{(d-1)}$ satisfies the following inequality

$$(r_k - r_{k-1})\left(m_i(\breve{\Omega}''_{(i,k,1)}) - r'_{(i,k)}\right) \geq t. \tag{33}$$

Next, note that according to the definition of the constraint tensor, each column of $\tilde{\breve{\mathbf{\Omega}}}^{k}_{(d-1)}$ has exactly $r_k+1$ nonzero entries. On the other hand, $r'_{(i,k)}+1 = 2$ or $r'_{(i,k)}+1 = r_k+1$ and hence, the columns of $\tilde{\breve{\mathbf{\Omega}}}^{k}_{(d-1)}$ corresponding to $\tilde{\breve{\mathbf{\Omega}}}'^{(i,k,1)}_{(d-1)}$ also satisfy (33) with probability at least $1-\frac{\epsilon}{(d-1)K}$, as those columns of $\tilde{\breve{\mathbf{\Omega}}}^{k}_{(d-1)}$ have more or equal number of nonzero entries in addition to the nonzero entries of $\tilde{\breve{\mathbf{\Omega}}}'^{(i,k,1)}_{(d-1)}$. Let us denote these columns by $\tilde{\breve{\mathbf{\Omega}}}'^{(i,k)}_{(d-1)}$. Therefore,

$$\tilde{\breve{\mathbf{\Omega}}}'^{k}_{(d-1)} = [\tilde{\breve{\mathbf{\Omega}}}'^{(1,k)}_{(d-1)}|\dots|\tilde{\breve{\mathbf{\Omega}}}'^{(d-1,k)}_{(d-1)}] \in \mathbb{R}^{N_{d-1}\times\left((r_k-r_{k-1})(\sum_{i=1}^{d-1} n_i) - r_k(r_k-r_{k-1}) - (d-2)r_k\right)}, \tag{34}$$

is a submatrix of $\tilde{\breve{\mathbf{\Omega}}}^{k}_{(d-1)}$ and with probability at least $1-\frac{\epsilon}{(d-1)K}$, every subset $\tilde{\breve{\mathbf{\Omega}}}''^{(i,k)}_{(d-1)} \in \mathbb{R}^{N_{d-1}\times t}$ of columns of $\tilde{\breve{\mathbf{\Omega}}}'^{(i,k)}_{(d-1)}$ satisfies

$$(r_k - r_{k-1})\left(m_i(\breve{\Omega}''_{(i,k)}) - r'_{(i,k)}\right) \geq t. \tag{35}$$

Note that $D = (n_1+\dots+n_{d-1})r_K - \sum_{k=1}^{K} r_k(r_k-r_{k-1}) - r_K(d-2) = \sum_{k=1}^{K}(r_k - r_{k-1})(n_{i_0}-r_k) + \sum_{i=1,i\neq i_0}^{d-1}\sum_{k=1}^{K}(r_k-r_{k-1})(n_i-1)$. Let $\tilde{\breve{\mathbf{\Omega}}}'_{(d-1)}$ denote the union of $\tilde{\breve{\mathbf{\Omega}}}'^{k}_{(d-1)}$'s for $1\leq k\leq K$, i.e.,

$$\tilde{\breve{\mathbf{\Omega}}}'_{(d-1)} = [\tilde{\breve{\mathbf{\Omega}}}'^{1}_{(d-1)}|\dots|\tilde{\breve{\mathbf{\Omega}}}'^{K}_{(d-1)}] \in \mathbb{R}^{N_{d-1}\times D}. \tag{36}$$

Note that (35) holds for each $\tilde{\breve{\mathbf{\Omega}}}'^{(i,k)}_{(d-1)}$ with probability at least $1-\frac{\epsilon}{(d-1)K}$ and hence, it holds for all $1\leq i\leq d-1$ and $1\leq k\leq K$ simultaneously with probability at least $1-\epsilon$. In the rest of the proof we show that the tensor corresponding to unfolding $\tilde{\breve{\mathbf{\Omega}}}'_{(d-1)}$, or in other words, $\breve{\Omega}' \in \mathbb{R}^{n_1\times\dots\times n_{d-1}\times D}$ satisfies the property described in the statement of Theorem 36.

Let $\tilde{\breve{\mathbf{\Omega}}}''_{(d-1)} \in \mathbb{R}^{N_{d-1}\times D'}$ denote a subset of columns of $\tilde{\breve{\mathbf{\Omega}}}'_{(d-1)}$. Moreover, define $\tilde{\breve{\mathbf{\Omega}}}''^{(i,k)}_{(d-1)} \in \mathbb{R}^{N_{d-1}\times D'_{(i,k)}}$ and $\tilde{\breve{\mathbf{\Omega}}}''^{k}_{(d-1)} \in \mathbb{R}^{N_{d-1}\times D'_k}$ as the matrices containing those columns of $\tilde{\breve{\mathbf{\Omega}}}''_{(d-1)}$ that also belong to $\tilde{\breve{\mathbf{\Omega}}}'^{(i,k)}_{(d-1)}$ and $\tilde{\breve{\mathbf{\Omega}}}'^{k}_{(d-1)}$, respectively. Hence, $D' = \sum_{k=1}^{K} D'_k$ and $D'_k = \sum_{i=1}^{d-1} D'_{(i,k)}$. Note that we only need to show that $D' \leq f(\breve{\Omega}'')$.

Recall that

$$(r_k - r_{k-1})\left(m_i(\breve{\Omega}''_{(i,k)}) - r'_{(i,k)}\right) \geq D'_{(i,k)}, \qquad \text{if } D'_{(i,k)} \neq 0, \tag{37}$$

since $D'_{(i,k)}$ may be zero as we consider a subset of the union of columns. Since each column of $\breve{\widetilde{\mathbf{\Omega}}}'^{(i,k)}_{(d-1)}$ has exactly $r_k + 1$ nonzero entries and $r_k + 1 \geq r'_{(i,k)} + 1$, $\left(m_i(\breve{\Omega}''_{(i,k)}) - r'_{(i,k)}\right)^+ = 0$ if $m_i(\breve{\Omega}''_{(i,k)}) = 0$ or equivalently $D'_{(i,k)} = 0$ and $\left(m_i(\breve{\Omega}''_{(i,k)}) - r'_{(i,k)}\right)^+ > 0$ otherwise. Hence,

$$(r_k - r_{k-1}) \left(m_i(\breve{\Omega}''_{(i,k)}) - r'_{(i,k)}\right)^+ \geq D'_{(i,k)}. \tag{38}$$

Therefore, we have

$$\begin{aligned} D'_k = \sum_{i=1}^{d-1} D'_{(i,k)} \quad &\leq \quad \sum_{i=1}^{d-1} (r_k - r_{k-1}) \left(m_i(\breve{\Omega}''_{(i,k)}) - r'_{(i,k)}\right)^+ \overset{(a)}{\leq} \sum_{i=1}^{d-1} (r_k - r_{k-1}) \left(m_i(\breve{\Omega}''_{k}) - r'_{(i,k)}\right)^+ \\ &\overset{(b)}{=} \quad (r_k - r_{k-1}) \left( (\sum_{i=1}^{d-1} m_i(\breve{\Omega}''_k)) - \min\{r_k, m_{i_0}(\breve{\Omega}''_k)\} - \sum_{i=1, i\neq i_0}^{d-1} \min\{1, m_i(\breve{\Omega}''_k)\} \right), \end{aligned} \tag{39}$$

where (a) follows from the fact that $\breve{\Omega}''_{(i,k)}$ is a subset of $\breve{\Omega}''_k$ and (b) follows from $(x-y)^+ = x - \min\{x,y\}$. Then

$$\begin{aligned} D' = \sum_{k=1}^{K} D'_k \leq \sum_{k=1}^{K} (r_k - r_{k-1}) (\sum_{i=1}^{d-1} m_i(\breve{\Omega}''_k)) \\ - \sum_{k=1}^{K} \min\left\{(r_k - r_{k-1}) r_k, (r_k - r_{k-1}) m_{i_0}(\breve{\Omega}''_k)\right\} \\ - \sum_{k=1}^{K} \sum_{i=1, i\neq i_0}^{d-1} \min\left\{r_k - r_{k-1}, (r_k - r_{k-1}) m_i(\breve{\Omega}''_k)\right\}. \end{aligned} \tag{40}$$

Note that $i_0$ can be any number in $\{1, \dots, d-1\}$. Moreover, $m_1(\breve{\Omega}'_k) \geq 1$ is equivalent with $m_i(\breve{\Omega}'_k) \geq 1$ for any $i$ as existence of one polynomial corresponding to a sampled entry in $\mathcal{U}_k$ results that the number of nonzero rows in any of the matricizations of $\mathcal{U}_k$ is at least one. Hence, independent from the choice of $i_0$,

$$\sum_{k=1}^{K} \sum_{i=1, i\neq i_0}^{d-1} \min\left\{r_k - r_{k-1}, (r_k - r_{k-1}) m_i(\breve{\Omega}''_k)\right\} = (d-2) \sum_{k=1}^{K} \min\left\{r_k - r_{k-1}, (r_k - r_{k-1}) m_1(\breve{\Omega}''_k)\right\}. \tag{41}$$

As a result, in order to obtain the tightest bound on the RHS of (40), we chose $i_0$ such that

$$\begin{aligned} \sum_{k=1}^{K} \min\left\{(r_k - r_{k-1}) r_k, (r_k - r_{k-1}) m_{i_0}(\breve{\Omega}''_k)\right\} = \\ \max_{1\leq i\leq d-1} \left\{ \sum_{k=1}^{K} \min\left\{(r_k - r_{k-1}) r_k, (r_k - r_{k-1}) m_i(\breve{\Omega}''_k)\right\} \right\}, \end{aligned} \tag{42}$$

and consequently

$$
\begin{aligned}
D' \leq & \sum_{k=1}^{K}(r_k - r_{k-1})(\sum_{i=1}^{d-1} m_i(\breve{\Omega}''_k)) \\
& - \max_{1\leq i\leq d-1} \left\{ \sum_{k=1}^{K} \min\left\{ (r_k - r_{k-1})r_k, (r_k - r_{k-1})m_i(\breve{\Omega}''_k) \right\} \right\} \\
& -(d-2)\sum_{k=1}^{K} \min\left\{ r_k - r_{k-1}, (r_k - r_{k-1})m_1(\breve{\Omega}''_k) \right\} \leq f(\breve{\Omega}'').
\end{aligned} \tag{43}
$$

■

**Theorem 43** *Assume that $r_k \leq \frac{n_k}{6}$ and $(\min_{1\leq i\leq d-1} n_i) > 200$, $c_k \geq (r_k - r_{k-1})(\sum_{i=1}^{d-1} n_i) - r_k(r_k - r_{k-1}) - (d-2)(r_k - r_{k-1})$ for $1 \leq k \leq K-1$ and $c_K \geq (\max_{1\leq i\leq d-1} n_i) + (r_K - r_{K-1})(\sum_{i=1}^{d-1} n_i) - r_K(r_K - r_{K-1}) - (d-2)(r_K - r_{K-1})$. Assume that each column of $\widetilde{\mathbf{\Omega}}_{(d-1)}$ includes at least $l$ nonzero entries, where*

$$
l > \max\left\{ 27\ \log\left(\frac{\max_{1\leq i\leq d-1} n_i}{\epsilon}\right) + 9\ \log\left(\frac{4K(d-1)\left(\max_{1\leq k\leq K}\{r_k - r_{k-1}\}\right)}{\epsilon}\right) + 18, 6r_K \right\}. \tag{44}
$$

*Then, with probability at least $1-\epsilon$, $\mathcal{U}_\Omega$ is uniquely completable.*

**Proof** According to the proof of Theorem 42, existence of $(r_k - r_{k-1})(\sum_{i=1}^{d-1} n_i) - r_k(r_k - r_{k-1}) - (d-2)(r_k - r_{k-1})$ columns in $\widetilde{\mathbf{\Omega}}^k_{(d-1)}$ and the assumption (44) result that (35) holds with probability at least $1 - \frac{\epsilon}{(d-1)2K}$ (instead of $1 - \frac{\epsilon}{(d-1)K}$ in Theorem 42) and therefore, it results in finite completability of $\mathcal{U}_\Omega$, i.e., condition (i) in Theorem 40, with probability at least $1 - \frac{\epsilon}{2}$. Hence, to complete the proof, it suffices to show that having $(\max_{1\leq i\leq d-1} n_i)$ more columns in $\widetilde{\mathbf{\Omega}}^K_{(d-1)}$, condition (ii) in Theorem 40 holds with probability at least $1-\frac{\epsilon}{2}$. By Lemma 41 (for $r=1$ and $r'_i = 0$), $m_i(\breve{\Omega}'_K) = n_i$ with probability at least $1 - \frac{\epsilon}{2K(d-1)\left(\max_{1\leq k\leq K}\{r_k - r_{k-1}\}\right)}$ for any $1 \leq i \leq d-1$. Therefore, $m_i(\breve{\Omega}'_K) = n_i$ with probability at least $1 - \frac{\epsilon}{2K\left(\max_{1\leq k\leq K}\{r_k - r_{k-1}\}\right)} \geq 1 - \frac{\epsilon}{2}$ for all $1 \leq i \leq d-1$. ■

Note that if $\mathcal{U}$ is finitely (uniquely) completable given $r_K$ (ignoring the union of tensors structure), then $\mathcal{U}$ is finitely (uniquely) completable given $r_1, \ldots, r_K$. Hence, for a looser bound, we can simply apply tensor analysis by invoking Lemma 11.

**Remark 44** *Assuming that $\mathcal{O}(r_k - r_{k-1}) = 1$, the required number of samples per column of the $(d-1)$-th unfolding for unique completability using the matrix analysis* (7) *, the tensor analysis* (8) *and our proposed union of tensors analysis* (44) *are of orders $\mathcal{O}(\log(Kn_1 \ldots n_{d-1}))$, $\mathcal{O}(\log(dr_K \max n_i))$, and $\mathcal{O}(\log(K \max n_i))$, respectively. The orders for the tensor analysis and our proposed union of tensors analysis are similar (this is expected as intuitively a few more rank constraints should not change the order of the fundamental limits) but still when we compare the exact numbers in the following example, we see the advantage of an efficient analysis that takes advantage of all the rank constraints.*

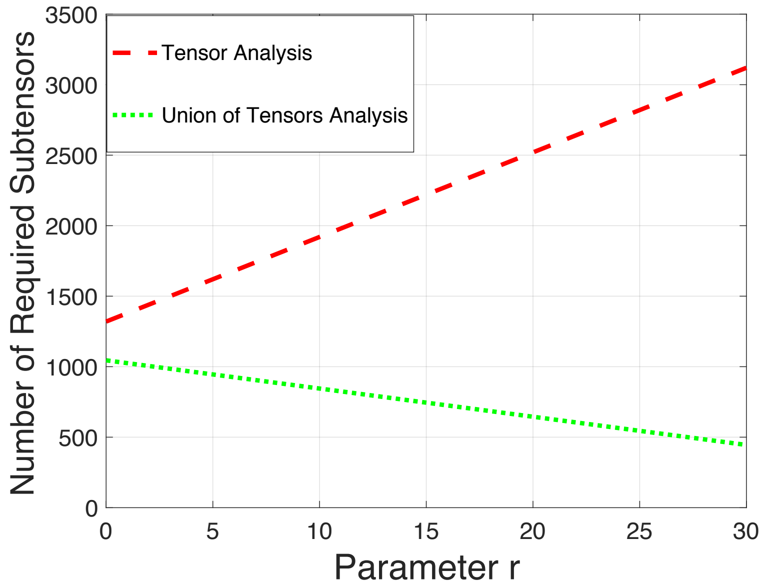


Figure 3: The total number of subtensors $c_1 + \cdots + c_K$ required for unique completability.

**Example 5** *Consider an example in which* $d = 7$, $K = 10$, $n_1 = \cdots = n_6 = 300$, $\epsilon = 0.1$. *Also, let* $r_k = 2k + r$ *for* $k = 1, \ldots, 10$, *where* $r$ *varies from* $0$ *to* $30$, *i.e.,* $r_{\max}$ *varies from* $20$ *to* $50$.

*First, we compare the total number of subtensors, i.e., structured columns, needed for the three approaches. For the matrix analysis, Lemma 10 requires the condition* $c_k \geq (r_k - r_{k-1} + 1)(N_{-d} - r_k) = 3 \times 10^6 - 6k - 3r \approx 3 \times 10^6$, *and hence the total number of subtensors is approximately* $3 \times 10^7$ *since* $K = 10$. *For the union of tensor spaces analysis, Theorem 43 requires* $c_k \geq (r_k - r_{k-1})(\sum_{i=1}^{d-1} n_i) - r_k(r_k - r_{k-1}) - (d-2)(r_k - r_{k-1}) = 110 - 4k - 2r$ *and hence,* $c_1 + \cdots + c_K \geq 110K - 4\sum_{k=1}^{K} k - 2Kr = 1045 - 20r$. *The tensor analysis in Lemma 11 requires only one condition (instead of* $K$ *conditions for all* $c_k$*'s) and it is* $c_1 + \cdots + c_K \geq (r_K + 2)(\sum_{i=1}^{d-1} n_i) = 1320 + 60r$. *In Figure 3, we plot the tensor and union of tensors cases and it is seen that in terms of the total number of subtensors, the requirement of the union of tensor spaces analysis is much less than the tensor analysis, and both are orders of magnitude less than the matrix analysis.*

*Next, we compare the required number of samples per subtensor, i.e., column of the* $(d-1)$*-th unfolding, using the matrix analysis* (7), *the tensor analysis* (8) *and our proposed union of tensors analysis* (44) *in Figure 4. It is seen that the union of tensors analysis requires the least number of samples per subtensor followed by the tensor analysis, and the matrix analysis requires the most.*

## 7. Conclusions

We have investigated the generalization of the problems of union of two-dimensional subspace clustering/retrieval to higher dimensions. In order to develop a clustering analysis for a union of tensor spaces, we made use of the condition on unique completability of a sampled tensor and developed an approach for identifying which tensor space correctly fits a certain tensor component of the union of tensor spaces, given that the sampling rate is higher than our obtained fundamental limit. Moreover, we investigated the completion problem for the case that the tensor spaces satisfy certain geometrical properties. Combinatorial conditions on the sampling pattern are characterized

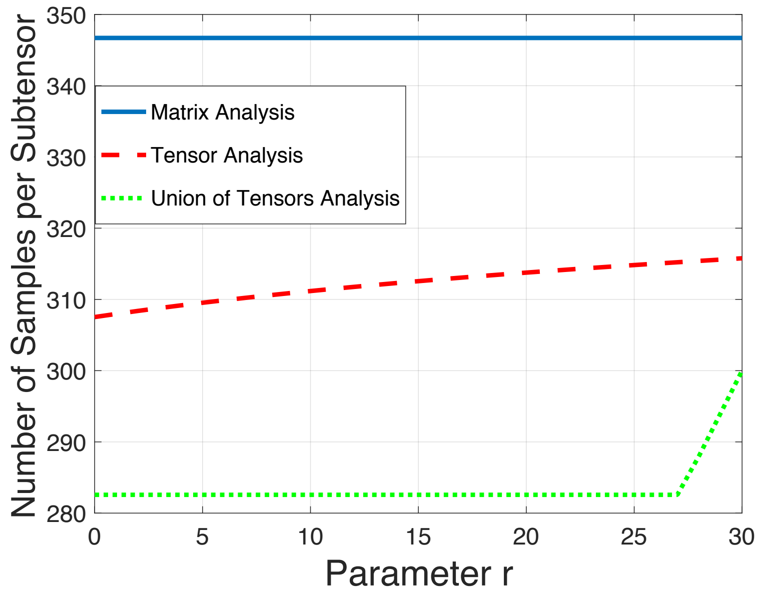


Figure 4: The required number of samples per subtensor for unique completability.

to ensure finite/unique completability of the data with probability one. And finally, the sampling rates that ensure finite/unique completability of the data with high probability are derived.

To the best of our knowledge, this work is the first to provide a fundamental theoretical analysis for the two important problems of low-rank tensor clustering and completion. There are a number of avenues for future investigations. First, this work is based on the assumption that the tensors are chosen generically from certain tensor spaces. One future direction is to develop similar/weaker results without this assumption. Secondly, the deterministic analysis in this paper characterizes the necessary and sufficient conditions on the sampling pattern for finite completability and therefore, it cannot be improved. However, the deterministic analysis for unique completability only provides sufficient conditions and it could be improved. Moreover, the probabilistic analysis provides sufficient conditions on the sampling rate and they could also be potentially improved. Further, conditions such as $r_k \leq \frac{n_k}{6}$ and $\min_{1\leq i\leq d-1} n_i > 200$ are the limitations in the probabilistic analysis due to the combinatorial analysis and another future direction is to achieve similar results with weaker restrictions.

## Acknowledgments

This work was supported in part by the U.S. National Science Foundation under Grant CCF-1814803 and in part by the U.S. Office of Naval Research under Grant N000141712827.

## References

Morteza Ashraphijuo and Xiaodong Wang. Fundamental conditions for low-CP-rank tensor completion. *Journal of Machine Learning Research*, 18 (63):1–29, 2017.

Morteza Ashraphijuo and Xiaodong Wang. A characterization of sampling patterns for union of low-rank subspaces retrieval problem. *International Symposium on Artificial Intelligence and*

*Mathematic*, pages 1–8, 2018.

Morteza Ashraphijuo and Xiaodong Wang. Clustering a union of low-rank subspaces of different dimensions with missing data. *Pattern Recognition Letters*, 120:31–35, 2019a.

Morteza Ashraphijuo and Xiaodong Wang. Fundamental conditions on the sampling pattern for union of low-rank subspaces retrieval. *Annals of Mathematics and Artificial Intelligence*, 87(4): 373–393, 2019b.

Morteza Ashraphijuo, Xiaodong Wang, and Vaneet Aggarwal. Rank determination for low-rank data completion. *The Journal of Machine Learning Research*, 18(1):3422–3450, 2017.

Morteza Ashraphijuo, Xiaodong Wang, and Junwei Zhang. Low-rank data completion with very low sampling rate using Newton's method. *IEEE Transactions on Signal Processing*, 67(7): 1849–1859, 2019.

Laura Balzano, Brian Eriksson, and Robert Nowak. High rank matrix completion and subspace clustering with missing data. In *Conference on Artificial Intelligence and Statistics*, 2012.

Robert Gardner Bartle and Donald R Sherbert. *Introduction to Real Analysis*. Hoboken, NJ: Wiley, 2011.

Jian Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

Emmanuel J Candès and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

Emmanuel J Candès, Yonina C Eldar, Thomas Strohmer, and Vladislav Voroninski. Phase retrieval via matrix completion. *SIAM Journal on Imaging Sciences*, 6(1):199–225, 2013.

Lars Eldén. *Matrix Methods in Data Mining and Pattern Recognition*, volume 4. SIAM, 2007.

Ehsan Elhamifar and René Vidal. Sparse subspace clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2790–2797. IEEE, 2009.

Ehsan Elhamifar and Rene Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765–2781, 2013.

Guojun Gan, Chaoqun Ma, and Jianhong Wu. *Data clustering: theory, algorithms, and applications*, volume 20. Siam, 2007.

Silvia Gandy, Benjamin Recht, and Isao Yamada. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):1–19, 2011.

Nicholas JA Harvey, David R Karger, and Kazuo Murota. Deterministic network coding by matrix completion. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 489–498, 2005.

Wei Hong, John Wright, Kun Huang, and Yi Ma. Multiscale hybrid linear models for lossy image representation. *IEEE Transactions on Image Processing*, 15(12):3655–3671, 2006.

Hui Ji, Chaoqiang Liu, Zuowei Shen, and Yuhong Xu. Robust video denoising using low rank matrix completion. In *CVPR*, pages 1791–1798, 2010.

Tamara G Kolda, Brett W Bader, and Joseph P Kenny. Higher-order web link analysis using multilinear algebra. In *Fifth IEEE International Conference on Data Mining (ICDM'05)*, pages 8–pp. IEEE, 2005.

Daniel Kressner, Michael Steinlechner, and Bart Vandereycken. Low-rank tensor completion by riemannian optimization. *BIT Numerical Mathematics*, 54(2):447–468, 2014.

Nan Li and Baoxin Li. Tensor completion for on-board compression of hyperspectral images. In *2010 IEEE International Conference on Image Processing*, pages 517–520. IEEE, 2010.

Lek-Heng Lim and Pierre Comon. Multiarray signal processing: Tensor decomposition meets compressed sensing. *Comptes Rendus Mecanique*, 338(6):311–320, 2010.

Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2013.

Kedar A Patwardhan, Guillermo Sapiro, and Marcelo Bertalmío. Video inpainting under constrained camera motion. *IEEE Transactions on Image Processing*, 16(2):545–553, 2007.

Pouya Pezeshkpour, Liyan Chen, and Sameer Singh. Embedding multimodal relational data for knowledge base completion. *arXiv preprint arXiv:1809.01341*, 2018.

Daniel Pimentel-Alarcón and Robert Nowak. The information-theoretic requirements of subspace clustering with missing data. In *International Conference on Machine Learning*, pages 802–810, 2016.

Daniel Pimentel-Alarcón, R Nowak, and Laura Balzano. On the sample complexity of subspace clustering with missing data. In *IEEE Workshop on Statistical Signal Processing*, pages 280–283. IEEE, 2014.

Daniel Pimentel-Alarcón, Laura Balzano, and Robert Nowak. Necessary and sufficient conditions for sketched subspace clustering. In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1335–1343. IEEE, 2016a.

Daniel Pimentel-Alarcón, Nigel Boston, and Robert Nowak. A characterization of deterministic sampling patterns for low-rank matrix completion. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):623–636, 2016b.

Daniel Pimentel-Alarcón, Laura Balzano, Roummel Marcia, Robert Nowak, and Rebecca Willett. Mixture regression as subspace clustering. In *2017 International Conference on Sampling Theory and Applications (SampTA)*, pages 456–459. IEEE, 2017.

Jasson DM Rennie and Nathan Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *International Conference on Machine Learning*, pages 713–719. ACM, 2005.

Anne C Sauve, Alfred O Hero, W Leslie Rogers, Scott J Wilderman, and Neal H Clinthorne. 3d image reconstruction for a compton spect camera model. *IEEE Transactions on Nuclear Science*, 46(6):2075–2084, 1999.

Nicholas D Sidiropoulos and Anastasios Kyrillidis. Multi-way compressed sensing for sparse low-rank tensors. *IEEE Signal Processing Letters*, 19(11):757–760, 2012.

Congyuan Yang, Daniel Robinson, and Rene Vidal. Sparse subspace clustering with missing entries. In *International Conference on Machine Learning*, pages 2463–2472, 2015.

Changqing Zhang, Huazhu Fu, Si Liu, Guangcan Liu, and Xiaochun Cao. Low-rank tensor constrained multiview subspace clustering. In *Proceedings of the IEEE international conference on computer vision*, pages 1582–1590, 2015.