Low-Rank Tensor Completion - Fundamental Limits and Efficient Algorithms

Morteza Ashraphijuo

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2020
ABSTRACT

Low-Rank Tensor Completion - Fundamental Limits and Efficient Algorithms

Morteza Ashraphijuo

This dissertation is motivated by the increasing applications of high-dimensional large-scale data sets in various fields and lack of theoretical understanding of the existing algorithms as well as lack of efficient algorithms in many cases. Hence, identifying the geometrical properties of data sets is essential for many data processing tasks, such as data retrieval and denoising.

In Part I, we derive the fundamental limits on the sampling rate required to study three important problems (i) low-rank data completion, (ii) rank estimation, and (iii) data clustering. In Chapter 2 we characterize the geometrical conditions on the sampling pattern, i.e., locations of the sampled entries, for finite and unique completability of a low-rank tensor, assuming that its rank vector is given or estimated. To this end, we propose a manifold analysis and study the independence of a set of polynomials defined based on the sampling pattern. Then, using the polynomial analysis, we derive a lower bound on the sampling rate such that it guarantees that the proposed conditions on the sampling patterns for finite and unique completability hold true with high probability. Then, in Chapter 3, we study the problem of rank estimation, where a data structure is partially sampled and we propose a geometrical analysis on the sampling pattern to estimate the true value of rank for various data structures by providing extremely tight lower and upper bounds on the rank value. And in Chapters 4 and 5, we make use of the developed tools to obtain a lower bound on the sampling rate to be able to correctly cluster a union of sampled matrices or tensors by identifying their corresponding unknown subspaces.

In Part II, first in Chapter 6, motivated by the algebraic tools developed in Part I, we develop a data completion algorithm based on solving a set of polynomial equations using Newton’s method, that is effective especially when the sampling rate is low. Then, in Chapter 7, we consider a data structure consisting of a union of nested low-rank matrix or tensor subspaces, and develop
a structured alternating minimization-based approach for completing such data, that is capable of taking advantage of multiple rank constraints simultaneously to achieve faster convergence and higher recovery accuracy.
Table of Contents

1 Introduction 1

I Fundamental Limits 7

2 Fundamental Limits For Low-Tucker-Rank Tensor Completion Problem 8
  2.1 Introduction ................................................. 9
  2.2 Background .................................................. 13
    2.2.1 Preliminaries and Notations ......................... 13
    2.2.2 Problem Statement and A Motivating Example .......... 13
  2.3 Deterministic Conditions for Finite Completablility .......... 16
    2.3.1 Condition for Finite Completablility Given the Core Tensor .... 16
    2.3.2 Geometry of Tucker Manifold ......................... 20
    2.3.3 Constraint Tensor ...................................... 23
    2.3.4 Algebraic Independence ................................. 25
    2.3.5 Finite Completablility Using Analysis on Tucker Manifold ....... 28
  2.4 Probabilistic Conditions for Finite Completablility ............ 30
    2.4.1 Lower Bound on Sampling Probability based on Analysis on Grassmannian
          Manifold .................................................. 31
    2.4.2 Lower Bound on Sampling Probability based on Analysis on Tucker Manifold 32
  2.5 Deterministic and Probabilistic Guarantees for Unique Completablility .......... 40
  2.6 Numerical Comparisons ..................................... 44
  2.7 Summary ..................................................... 46
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.8 Proof of Lemma 4</td>
<td>47</td>
</tr>
<tr>
<td>2.9 Proof of Theorem 6</td>
<td>48</td>
</tr>
<tr>
<td>2.10 Proof of Claim in Example 5</td>
<td>52</td>
</tr>
<tr>
<td>3 Rank Estimation For Low-Rank Sampled Data</td>
<td>55</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>56</td>
</tr>
<tr>
<td>3.2 Data Models and Problem Statement</td>
<td>58</td>
</tr>
<tr>
<td>3.2.1 Matrix Models</td>
<td>58</td>
</tr>
<tr>
<td>3.2.2 Tensor Models</td>
<td>59</td>
</tr>
<tr>
<td>3.2.3 Problem Statement</td>
<td>61</td>
</tr>
<tr>
<td>3.3 Scalar-Rank Cases</td>
<td>61</td>
</tr>
<tr>
<td>3.3.1 Single-View Matrix</td>
<td>61</td>
</tr>
<tr>
<td>3.3.2 CP-Rank Tensor</td>
<td>70</td>
</tr>
<tr>
<td>3.4 Vector-Rank Cases</td>
<td>76</td>
</tr>
<tr>
<td>3.4.1 Multi-View Matrix</td>
<td>76</td>
</tr>
<tr>
<td>3.4.2 Tucker-Rank Tensor</td>
<td>79</td>
</tr>
<tr>
<td>3.4.3 TT-Rank Tensor</td>
<td>83</td>
</tr>
<tr>
<td>3.5 Summary</td>
<td>87</td>
</tr>
<tr>
<td>4 Fundamental Limits For Clustering A Union of Low-Rank Matrices Of Different Dimensions With Missing Data</td>
<td>88</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>88</td>
</tr>
<tr>
<td>4.2 Background</td>
<td>90</td>
</tr>
<tr>
<td>4.3 Main Results</td>
<td>91</td>
</tr>
<tr>
<td>4.4 Numerical Experiments</td>
<td>97</td>
</tr>
<tr>
<td>4.5 Summary</td>
<td>98</td>
</tr>
<tr>
<td>5 Fundamental Limits For Clustering A Union of Low-Rank Tensor Spaces With Missing Data</td>
<td>99</td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>100</td>
</tr>
<tr>
<td>5.2 Preliminaries</td>
<td>101</td>
</tr>
</tbody>
</table>
5.3 Problem Statements And Matrix Approaches ........................................ 103
  5.3.1 Problem Statements ................................................................. 103
  5.3.2 Matrix Analysis Approach To The Clustering Problem ...................... 104
  5.3.3 Matrix Analysis Approach To The Completion Problem .................... 106
5.4 Tensor Space Clustering: Main Results .............................................. 106
  5.4.1 Useful Results on Tensor Completion .......................................... 106
  5.4.2 Conditions on Tensor Space Clustering ....................................... 110
5.5 Summary ............................................................................................. 115

II Efficient Algorithms .............................................................................. 116

6 Algorithm For Low-Rank Data Completion With Very Low Sampling Rate 117
  6.1 Introduction ......................................................................................... 118
  6.2 Matrix Completion ............................................................................ 120
    6.2.1 Background .................................................................................. 120
    6.2.2 Matrix Completion via Solving Polynomial Equations .................... 121
    6.2.3 The Algorithm .............................................................................. 124
  6.3 Convergence Analysis ........................................................................ 126
  6.4 Numerical Results ............................................................................. 129
    6.4.1 Noiseless Matrix .......................................................................... 129
    6.4.2 Noisy Matrix ................................................................................ 132
  6.5 Extensions .......................................................................................... 135
    6.5.1 Matrix Sensing ............................................................................. 135
    6.5.2 Tensor Completion ........................................................................ 140
  6.6 Summary ............................................................................................. 142

7 Structured Alternating Minimization For Union of Nested Low-Rank Subspaces

  Data Completion ......................................................................................... 143
    7.1 Introduction ......................................................................................... 144
    7.2 Completion of Union of Nested Low-Rank Matrices ............................ 146
      7.2.1 Problem Statement ............................................................... 146
7.2.2 Naive Approaches ................................................. 147
7.2.3 Structured Decomposition ........................................ 149
7.2.4 Proposed Structured Alternating Minimization Algorithm .... 151
7.3 Completion of Union of Nested Low-Rank Tensors .................. 153
  7.3.1 Background ...................................................... 154
  7.3.2 Problem Statement .............................................. 155
  7.3.3 Alternating Minimization For Tensor Completion .............. 156
  7.3.4 Naive Approaches .............................................. 157
  7.3.5 Structured Decomposition ..................................... 158
  7.3.6 Proposed Structured Alternating Minimization for Union of Nested Tensor Subspaces ............................................ 159
7.4 Simulation Results .................................................. 161
  7.4.1 Matrix Case ..................................................... 161
  7.4.2 Tensor Case .................................................... 167
7.5 Summary ............................................................ 172

8 Conclusions .......................................................... 173

Bibliography ............................................................ 174
Acknowledgments

First and foremost, I would like to express my deepest gratitudes to my advisor, Professor Xiaodong Wang, for his generosity and endless support throughout this journey. I would like to thank him for directing me towards a variety of interesting research areas and for his care about my development as an independent researcher and for the continuous support of my Ph.D study and his patience and immense knowledge. I have grown and learned so much under his watchful eye and been introduced to a broad range of disciplines that shaped my academic life. It has truly been an honor and privilege for me to have Professor Xiaodong Wang as my Ph.D. advisor.

I would also like to express my thanks to professors John Paisley, James Anderson, Asaf Cidon and Charles Elkan for serving as the members of my dissertation committee, and providing me with valuable feedback and comments.

Moreover, I am grateful to my friends and colleagues at Columbia University, who accompanied me in the last few years.

Finally, I would like to extend my deepest gratitude, love and affection to my beloved family, for loving me, believing in me and wishing the best for me. I owe all my achievements to their pure hearts; such pure hearts that any thing they dream for me comes true.
To my family
Chapter 1

Introduction

The new trends in technology and computational platforms created data deluge in various areas in different forms such as images, videos, product recommendations and rankings, etc. Most of the data that we collect or create in various applications and forms are better represented with multiple dimensions to capture the correlations across different attributes, and this emphasizes the importance of having a good understanding of tensors, which are the generalization of matrices to higher dimensions. Although many machine learning algorithms have been developed and caused tremendous empirical success, our understanding on the theoretical aspect of the developed algorithms is very limited. In most applications, we can further move boundaries of the existing theoretical results as well as the practical algorithms by making use of a blessing of dimensionality observed in high-dimensional data sets, which is the key idea behind many of the algebraic and statistical approaches to large-scale data sets.

A popular data science problem is to retrieve the missing entries of a data set which is partially collected while we are given the dimension of the subspace that the data set is chosen or created from. This problem is called low-rank tensor completion problem, has many applications in various fields and is one of the main focuses of this dissertation. Another related problem in this area that we studied is the rank estimation problem, in which we need to estimate the rank of a partially sampled data. This problem can be considered as a preprocessing step in analyzing the geometry of the sampled data before retrieving the non-sampled part of the data. Another important problem studied in this dissertation involves data structures that are a union of low-rank matrices or tensors, meaning that each part of the data is chosen from an unknown subspace. Identifying the source
CHAPTER 1. INTRODUCTION

of each part of such data structures, i.e., clustering a union of data sets, can be counted as a preprocessing step of retrieving the missing entries, and is studied in this dissertation.

In order to characterize the fundamental limits on the sampling rate to tackle tensor completion, rank estimation and data clustering problems, geometrical analyses of the sampling pattern are studied in Part I of this dissertation. Particularly, geometry of the sampling pattern and data structure is studied to (i) retrieve the sampled data, (ii) develop a method to approximate the rank, (iii) correctly cluster a union of tensor spaces, given the minimum number of sampled entries. Additionally, in Part II of this dissertation, first, an algorithm is developed to approach the fundamental limits obtained in Part I for low-rank data retrieval problem, and then, an efficient alternating minimization-based algorithm is developed to retrieve a union of low-rank tensors.

The following is a summary of the contributions of each chapter of this dissertation.

Part I: Fundamental Limits

The focus of this part of the dissertation is to find the fundamental limits on the sampling rate for the three problems of data completion, rank estimation and data clustering. To this end, geometrical analyses are developed to characterize the necessary and sufficient conditions on the sampling pattern for solving any of the three mentioned problems with a certain accuracy. This part contains the followings:

Chapter 2: Fundamental Limits For Low-Tucker-Rank Tensor Completion Problem

In this chapter, we characterize the fundamental conditions on the sampling pattern, i.e., locations of the sampled entries, for finite completability of a low-rank tensor given some components of its Tucker rank. In order to find the deterministic necessary and sufficient conditions, we propose an algebraic geometric analysis on the Tucker manifold, which allows us to incorporate multiple rank components in the proposed analysis in contrast with the conventional geometric approaches on the Grassmannian manifold. This analysis characterizes the algebraic independence of a set of polynomials defined based on the sampling pattern, which is closely related to finite completability of the sampled tensor, where finite completability simply means that the number of possible
completions of the sampled tensor is finite. Probabilistic conditions are then studied and a lower bound on the sampling probability is given, which guarantees that the proposed deterministic conditions on the sampling patterns for finite completability hold with high probability. Furthermore, using the proposed geometric approach for finite completability, we propose a sufficient condition on the sampling pattern that ensures there exists exactly one completion of the sampled tensor [Ashraphijuo et al., 2019a].

Chapter 3: Rank Estimation For Low-Rank Sampled Data

In this chapter, we make use of the recent fundamental conditions on the sampling patterns that have been obtained for finite completability of low-rank matrices or tensors given the corresponding ranks in another practical problem. We consider the scenario where the rank is not given and we aim to approximate the unknown rank based on the location of sampled entries and some given completion. We consider a number of data models, including single-view matrix, multi-view matrix, CP tensor, tensor-train tensor and Tucker tensor. For each of these data models, we provide an upper bound on the rank when an arbitrary low-rank completion is given. We characterize these bounds both deterministically, i.e., with probability one given that the sampling pattern satisfies certain combinatorial properties, and probabilistically, i.e., with high probability given that the sampling probability is above some threshold. Moreover, for both single-view matrix and CP tensor, we are able to show that the obtained upper bound is exactly equal to the unknown rank if the lowest-rank completion is given. Furthermore, we provide numerical experiments for the case of single-view matrix, where we use nuclear norm minimization to find a low-rank completion of the sampled data and we observe that in most of the cases the proposed upper bound on the rank is equal to the true rank [Ashraphijuo et al., 2017b].

Chapter 4: Fundamental Limits For Clustering A Union of Low-Rank Matrices Of Different Dimensions With Missing Data

In this chapter, we derive fundamental conditions for clustering a union of low-rank subspaces with missing data. In particular, given an incomplete matrix, assuming its columns are drawn from $K$ different subspaces with different dimensions, the subspace clustering problem is to cluster the
columns that belong to the same subspace. We derive a lower bound on the number of columns from each subspace such that the columns can be clustered correctly with high probability. The analysis focuses on the subspace with the lowest dimension and is a generalization of the corresponding results in [Pimentel-Alarcón and Nowak, 2016b] that assumes the subspaces are independent and with the same dimension [Ashraphijuo and Wang, 2019a].

Chapter 5: Fundamental Limits For Clustering A Union of Low-Rank Tensor Spaces With Missing Data

In this chapter, 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 [Ashraphijuo and Wang, 2020b].

Part II: Efficient Algorithms

The focus of this part of dissertation is to provide efficient algorithms that are faster and require a lower number of samples for data completion in comparison with the existing methods in the literature. We first provide an algorithm for low-rank matrix or tensor completion problem that performs better than the existing methods in the literature in terms of the required number of sampled entries as well as the output accuracy. Then, we consider a nested union of low-rank data structures and provide a structured alternating minimization method that converges to a solution that satisfies several rank constraints simultaneously. This part contains the followings:
Chapter 6: Algorithm For Low-Rank Data Completion With Very Low Sampling Rate

Newton’s method is a widely applicable and empirically efficient method for finding the solution to a set of equations. The recently developed algebraic geometric analyses provide information-theoretic bounds on the sampling rate to ensure the existence of a unique completion with a high probability. We study a remained open question from these works, which is whether we can retrieve the sampled data when the sampling rate is very close to the mentioned information-theoretic bounds. This work is concerned with proposing algorithms to retrieve the sampled data when the sampling rate is too small and close to the mentioned information-theoretic bounds. Hence, we propose a new approach for recovering a partially sampled low-rank matrix or tensor when the number of samples is only slightly more than the dimension of the corresponding manifold, by solving a set of polynomial equations using Newton’s method. In particular, we consider low-rank matrix completion, matrix sensing and tensor completion. Each observed entry contributes one polynomial equation in terms of the factors in the rank factorization of the data. By exploiting the specific structures of the resulting set of polynomial equations, we analytically characterize the convergence regions of the Newton’s method for matrix completion and matrix sensing. Through extensive numerical results, we show that the proposed approach outperforms the well known methods such as nuclear norm minimization and alternating minimization in terms of the success rate of data recovery (noiseless case) and peak signal-to-noise ratio (noisy case), especially when the sampling rate is very low. In terms of the running time, the proposed Newton’s method is significantly faster than the nuclear norm minimization method, but it is slower than the alternating minimization method [Ashraphijuo et al., 2019b].

Chapter 7: Structured Alternating Minimization For Union of Nested Low-Rank Subspaces Data Completion

In this chapter, we consider a particular data structure consisting of a union of several nested low-rank subspaces with missing data entries. Given the rank of each subspace, we treat the data completion problem, i.e., to estimate the missing entries. Starting from the case of two-dimensional
data, i.e., matrices, we show that the union of nested subspaces data structure leads to a structured
decomposition $\mathbf{U} = \mathbf{X} \mathbf{Y}$ where the factor $\mathbf{Y}$ has blocks of zeros that are determined by the rank
values. Moreover, for high-dimensional data, i.e., tensors, we show that a similar structured CP
decomposition also exists, $\mathcal{U} = \sum_{l=1}^{r} a_1^l \otimes a_2^l \otimes \ldots \otimes a_d^l$, where $\mathbf{A}_d = [a_1^d \ldots a_d^d]$ contains blocks
of zeros determined by the rank values. Based on such structured decompositions, we develop
efficient alternating minimization algorithms for both matrix and tensor completions, by enforcing
the above structures in each iteration including the initialization. Compared with naive approaches
where either the additional rank constraints are ignored, or data completion is performed part by
part, the proposed structured alternating minimization approaches exhibit faster convergence and
higher recovery accuracy.
Part I

Fundamental Limits
Chapter 2

Fundamental Limits For
Low-Tucker-Rank Tensor Completion Problem

We investigate the fundamental conditions on the sampling pattern, i.e., locations of the sampled entries, for finite completability of a low-rank tensor given some components of its Tucker rank. In order to find the deterministic necessary and sufficient conditions, we propose an algebraic geometric analysis on the Tucker manifold, which allows us to incorporate multiple rank components in the proposed analysis in contrast with the conventional geometric approaches on the Grassmannian manifold. This analysis characterizes the algebraic independence of a set of polynomials defined based on the sampling pattern, which is closely related to finite completability of the sampled tensor, where finite completability simply means that the number of possible completions of the sampled tensor is finite. Probabilistic conditions are then studied and a lower bound on the sampling probability is given, which guarantees that the proposed deterministic conditions on the sampling patterns for finite completability hold with high probability. Furthermore, using the proposed geometric approach for finite completability, we propose a sufficient condition on the sampling pattern that ensures there exists exactly one completion of the sampled tensor.
CHAPTER 2. FUNDAMENTAL LIMITS FOR LOW-TUCKER-RANK TENSOR COMPLETION PROBLEM

2.1 Introduction

Tensors are generalizations of vectors and matrices: a vector is a first-order tensor and a matrix is a second-order tensor. Most data around us are better represented with multiple dimensions to capture the correlations across different attributes. For example, a colored image can be considered as a third-order tensor, two of the dimensions (rows and columns) being spatial, and the third being spectral (color); while a colored video sequence can be considered as a fourth-order tensor, with time being the fourth dimension besides spatial and spectral dimensions. Similarly, a colored 3-D MRI image across time can be considered as a fifth-order tensor. In many applications, part of the data may be missing. This chapter investigates the fundamental conditions on the locations of the non-missing entries such that the multi-dimensional data can be recovered in finite and/or unique choices. In particular, we investigate deterministic and probabilistic conditions on the sampling pattern for finite or unique solution to a low-rank tensor completion problem given the sampled tensor and some of its Tucker rank components, i.e., ranks of some of its matricizations.

There are numerous applications of low-rank data completion in various areas including image or signal processing [Candès et al., 2013; Ji et al., 2010], data mining [Eldén, 2007], network coding [Harvey et al., 2005], compressed sensing [Lim and Comon, 2010; Sidiropoulos and Kyrillidis, 2012; Gandy et al., 2011], reconstructing the visual data [Liu et al., 2013b], seismic data processing [Kreimer et al., 2013; Ely et al., 2013; Wang et al., 2016], RF fingerprinting [Liu et al., 2016], and reconstruction of cellular data [Aggarwal et al., 2016].

The majority of the literature on matrix and tensor completion are concerned with developing various optimization-based algorithms under some assumptions such as incoherence [Jain et al., 2013], etc., to construct a completion. In particular, low-rank matrix completion has been widely studied and many algorithms based on convex relaxation of rank [Candès and Recht, 2009; Candès and Tao, 2010; Cai et al., 2010], non-convex optimization [Ashraphiju et al., 2019b] and alternating minimization [Jain et al., 2013], etc., have been proposed. Also, a generalization of the low-rank matrix completion, which is completion from several low-rank sources has attracted attention recently. For the tensor completion problem various solutions have been proposed that are based on convex relaxation of rank constraints [Kreimer et al., 2013; Gandy et al., 2011; Tomioka et al., 2010; Signoretto et al., 2014; Romera-Paredes and Pontil, 2013], alternating minimization [Wang et al., 2016] and other heuristics [Liu et al., 2016; Kressner et al., 2014; Krishnamurthy and Singh, 2013;
In the existing literature on optimization-based matrix or tensor completions, in addition to meeting the lower bound on the sampling probability, conditions such as incoherence [Jain et al., 2013; Goldfarb and Qin, 2014; Recht, 2011], which constrains the values of the matrix or tensor entries, are required to obtain a completion with high probability. On the other hand, fundamental completability conditions that are independent of the specific completion algorithms have also been investigated. In [Pimentel-Alarcón et al., 2016b; Ashraphijuo et al., 2018; Király et al., 2015] deterministic conditions on the locations of the sampled entries (sampling pattern) have been studied through algebraic geometry approaches on Grassmannian manifold that lead to finite and unique solutions to the matrix completion problem, where finite completability simply means that the number of possible completions of the sampled tensor is finite. Specifically, in [Pimentel-Alarcón et al., 2016b] a deterministic sampling pattern is proposed that is necessary and sufficient for finite completability of the sampled matrix of the given rank. Such an algorithm-independent condition can lead to a much lower sampling rate than that is required by the optimization-based completion algorithms. For example, the required number of samples per column in [Jain et al., 2013] is on the order of $O(\log(n)r^{2.5}\log(\|X\|_F))$, where $X$ is the unknown matrix with $n$ rows and of rank $r$, while the required number of samples per column in [Pimentel-Alarcón et al., 2016b] is on the order of $O(\max\{\log(n),r\})$. The analysis on Grassmannian manifold in [Pimentel-Alarcón et al., 2016b] is not capable of incorporating more than one rank constraint, and therefore this method is not efficient for solving the same problem for a tensor given multiple rank components. In this chapter, we propose a geometric analysis on Tucker manifold to obtain deterministic and probabilistic conditions that lead to finite or unique completability for low-rank tensors when multiple rank components are given. Moreover, other related problems have been studied using algebraic geometry analysis, including high-rank matrix completion [Balzano et al., 2012], rank estimation [Ashraphijuo et al., 2017b], and subspace clustering with missing data [Ashraphijuo and Wang, 2019a; Pimentel-Alarcón and Nowak, 2016b; Yang et al., 2015; Ashraphijuo et al., 2017a].

This work is inspired by [Pimentel-Alarcón et al., 2016b], where the analysis on Grassmannian manifold is proposed for a single-view matrix. Specifically, in [Pimentel-Alarcón et al., 2016b] a novel approach is proposed to consider the rank factorization of a matrix and to treat each observed entry as a polynomial in terms of the entries of the components of the rank factorization. Then,
under the genericity assumption, the algebraic independence among the mentioned polynomials is studied. In this chapter, we consider the low-Tucker-rank tensor and follow the general approach that is similar to that in [Pimentel-Alarcón et al., 2016b]. We mention some of the main differences: (i) geometry of the manifold, (ii) the equivalence class for the core tensor and consequently (iii) the canonical core tensor, (iv) structure of the polynomials, etc. are fundamentally different from those in [Pimentel-Alarcón et al., 2016b]. Moreover, (v) the idea of using more than one rank constraint simultaneously in the algebraic geometry approach is also new. Hence, the manifold structure for the low-Tucker-rank tensor is fundamentally different from the Grassmannian manifold and we need to develop almost every step anew.

Tucker decomposition is a well-known method to represent a tensor [Kolda, 2001; Grasedyck, 2010]. In this chapter, we use this decomposition to represent the sparsity of a tensor and use Tucker rank to model the low-rank structure of the tensor. There are several other well-known decompositions of a tensor as well, including polyadic decomposition [ten Berge and Sidiropoulos, 2002; Stegeman and Sidiropoulos, 2007], tensor-train decomposition [Oseledets, 2011; Holtz et al., 2012], hierarchical Tucker representation [Ballani et al., 2013; Hackbusch and Kühn, 2009], tubal rank decomposition [Kilmer et al., 2013] and others.

This chapter focuses on the low-rank tensor completion problem, given a portion of the rank vector of the tensor. Specifically, we investigate the following three problems:

- **Problem (i):** Characterizing the necessary and sufficient conditions on the sampling pattern to have finitely many tensor completions for the given rank.

To solve this fundamental problem, we propose a geometric analysis framework on Tucker manifold. Specifically, we obtain a set of polynomials based on the location of the sampled entries in tensor and identify the condition on the sampling pattern for ensuring sufficient number of algebraically independent polynomials in the mentioned set. Given any nonempty proper subset of the Tucker rank vector, this analysis leads to the necessary and sufficient condition on the sampling patterns for finite completability of the tensor. Given the entire Tucker rank vector this condition is sufficient for finite completability.

- **Problem (ii):** Characterizing conditions on the sampling pattern to ensure that there is exactly one completion for the given rank.
We use our proposed geometric analysis for finite completability of low-rank tensors to obtain a sufficient conditions on the sampling patterns to ensure unique completability, which is milder than the sufficient condition for unique completability obtained through matricization analysis and applying the matrix method in [Pimentel-Alarcón et al., 2016b].

- **Problem (iii):** If the elements in the tensor are sampled independently with probability $p$, what are the conditions on $p$ such that the conditions in Problems (i) and (ii) are satisfied with high probability?

We bound the number of needed samples to ensure the proposed sampling patterns for finite and unique tensor completability hold with high probability. Even though we follow a similar approach to [Pimentel-Alarcón et al., 2016b] for the matrix case, we develop a generalization of Hall’s theorem for bipartite graphs which is needed to prove the correctness of the bounds for both the tensor and the matrix cases. Moreover, it is seen that our proposed analysis on Tucker manifold leads to a much lower sampling rate than the corresponding analysis on Grassmannian manifold for both finite and unique tensor completions.

The remainder of this chapter is organized as follows. In Section 2.2, some preliminaries and notations are presented, and also an example is given that illustrates the advantage of tensor analysis over analyzing matricizations of a tensor. In Section 2.3, Problem (i) is studied and the sampling patterns that ensure finite completions are found using tensor algebra. In Section 2.4, we study Problem (iii) for the case of finite completion and the key to solving this problem is the proof of the generalized Hall’s theorem, which is an independent result in graph theory. Section 2.5 considers Problem (ii) to give a sufficient condition on the sampling pattern for unique completability. Further, Problem (iii) for unique completion is also studied. Some numerical results are provided in Section 2.6 to compare the sampling rates for finite and unique completions based on our proposed tensor analysis versus the matricization method. Finally, Section 2.7 summarizes the chapter.
CHAPTER 2. FUNDAMENTAL LIMITS FOR LOW-TUCKER-RANK TENSOR COMPLETION PROBLEM

2.2 Background

2.2.1 Preliminaries and Notations

In this chapter, it is assumed that a $d$th-order tensor $U \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is chosen generically from the manifold of $n_1 \times \cdots \times n_d$ tensors of the given Tucker rank (will be explained rigorously later). For the sake of simplicity in notation, define $N \triangleq \prod_{j=1}^{d} n_j$ and $N_{-i} \triangleq \frac{N}{n_i}$. Also, for any real number $x$, define $x^+ \triangleq \max\{0, x\}$. Let $U_{(i)} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be the $i$-th matricization of the tensor $U$ such that $U_{(i)}(x_1, \ldots, x_i, \ldots, x_d) = U(x_1, \ldots, x_i-1, x_i+1, \ldots, x_d)$, where $M_{(i)}$ is an arbitrary bijective mapping $M_{(i)} : (x_1, \ldots, x_i-1, x_i+1, \ldots, x_d) \to \{1, 2, \ldots, N_{-i}\}$ and $U_{(i)}(x)$ represents an entry of the tensor $U$ with coordinate $x = (x_1, \ldots, x_d)$.

Given $U \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and $X \in \mathbb{R}^{n_i \times n'_{-i}}$, $U' \triangleq U \times_i X \in \mathbb{R}^{n_1 \times \cdots \times n_i-1 \times n'_{i} \times n_{i+1} \times \cdots \times n_d}$ is defined as

$$U'(x_1, \ldots, x_{i-1}, k_i, x_{i+1}, \ldots, x_d) \triangleq \sum_{x_i=1}^{n_i} U(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_d)X(x_i, k_i). \quad (2.1)$$

Throughout this chapter, we use Tucker rank as the rank of a tensor, which is defined as $\text{rank}(U) = (r_1, \ldots, r_d)$ where $r_i = \text{rank}(U_{(i)})$. The Tucker decomposition of a tensor $U$ is given by

$$U = C \times_1 T_1, \quad (2.2)$$

where $C \in \mathbb{R}^{r_1 \times \cdots \times r_d}$ is the core tensor and $T_i \in \mathbb{R}^{r_i \times n_i}$ are $d$ orthogonal matrices. Then, (2.2) can be written as

$$U(\vec{x}) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} C(k_1, \ldots, k_d)T_1(k_1, x_1) \cdots T_d(k_d, x_d). \quad (2.3)$$

The dimension of the space of fixed Tucker-rank and size tensors is $\sum_{i=1}^{d} (n_i \times r_i - r_i^2) + \prod_{i=1}^{d} r_i$. Denote $\Omega$ as the binary sampling pattern tensor that is of the same size as $U$ and $\Omega(\vec{x}) = 1$ if $U(\vec{x})$ is observed and $\Omega(\vec{x}) = 0$ otherwise. For each subtensor $U'$ of the tensor $U$, define $N_\Omega(U')$ as the number of observed entries in $U'$ according to the sampling pattern $\Omega$.

2.2.2 Problem Statement and A Motivating Example

We are interested in finding deterministic conditions on the sampling pattern tensor $\Omega$ under which there are infinite, finite, or unique completions of the sampled tensor $U$ that satisfy $\text{rank}(U) = (r_1, r_2, \ldots, r_d)$. Moreover, we are interested in finding probabilistic sampling strategies that ensure
the obtained conditions for finite and unique completability hold, respectively, with high probability. The matrix version of this problem has been treated in [Pimentel-Alarcón et al., 2016b]. In this chapter, we investigate this problem for general order tensors.

In this subsection, we intend to compare the following two approaches in an example to emphasize the necessity of our analysis for general order tensors: (i) analyzing each matricization individually with the rank constraint of the corresponding matricization, (ii) analyzing via Tucker decomposition. In particular, we will show via an example that analyzing each of the matricizations separately is not enough to guarantee finite completability when multiple rank components are given. On the other hand, we show that for the same example Tucker decomposition ensures finite completability. Hence, this example illustrates that matricization analysis does not take advantage of the full information of given Tucker rank and thus fails to provide a necessary and sufficient condition for finite completability when more than one component of the rank vector is given.

Consider a 3rd-order tensor $U \in \mathbb{R}^{2 \times 2 \times 2}$ with Tucker rank $(1, 1, 1)$. First, we show that having any 4 entries of $U$, there are infinitely many completions of any matricization with the corresponding rank constraint. Hence, along each dimension there exist a set of infinite completions given the corresponding rank constraint. Note that the analysis on Grassmannian manifold in [Pimentel-Alarcón et al., 2016b] is not capable of incorporating more than one rank constraint. However, as we show it is possible that the intersection of the mentioned three infinite sets is a finite set and that is why we need an analysis that is able to incorporate more than one rank constraint. Without loss of generality, it suffices to show the claim only for its first matricization. Therefore, the claim reduces to the following statement:

**Statement:** Having any 4 entries of a rank-1 matrix $U \in \mathbb{R}^{2 \times 4}$, there are infinitely many completions for it.

In order to prove the above statement, we need to consider the following four possible scenarios:

1. The 4 observed entries are in a row. In this case, clearly, there are infinitely many completions for the other row as it can be any scalar multiplied by the first row.

2. The 4 observed entries are such that there is a column in which there is no observed entries. In this case, there are infinitely many completions for this column as it can be any scalar multiplied by the other columns.
3. The 4 observed entries are such that there is one observed entry in each column, and also each row has exactly two observed entries. Assume that the two observed entries in the second row are the pair \((a, b)\). In this case, for every pair \((ka, kb)\) as the value of the two non-observed entries of the first row (where \(k\) is an arbitrary scalar) there is a unique completion for the rest of the entries. As a result, there are infinitely many completions for this matrix.

4. The 4 observed entries are such that there is one observed entry in each column, and also the first and second rows have 3 and 1 observed entries, respectively. In this case, for each value of the only non-observed entry of the first row there is a unique completion. Therefore, there are infinitely many completions for this matrix.

Assume that the entries \(U(1, 1, 1), U(2, 1, 1), U(1, 2, 1),\) and \(U(1, 1, 2)\) are observed. Now, we take advantage of all elements of Tucker rank simultaneously, in order to show there are only finitely many tensor completions. Using Tucker decomposition (2.2), and given the rank is \((1, 1, 1)\), without loss of generality, assume that the scalar \(C = 1\) and \(T_1 = (x, x'), T_2 = (y, y')\) and \(T_3 = (z, z')\), and then the following equalities hold

\[
\begin{align*}
U(1, 1, 1) &= xyz, & U(2, 2, 1) &= x'y'z, \\
U(2, 1, 1) &= x'yz, & U(2, 1, 2) &= x'y'z', \\
U(1, 2, 1) &= xy'z, & U(1, 2, 2) &= xy'z', \\
U(1, 1, 2) &= xyz', & U(2, 2, 2) &= x'y'z'.
\end{align*}
\]

The unknown entries can be determined uniquely in terms of the 4 observed entries as

\[
\begin{align*}
U(2, 2, 1) &= x'y'z = \frac{U(2, 1, 1)U(1, 2, 1)}{U(1, 1, 1)}, \\
U(2, 1, 2) &= x'yz' = \frac{U(2, 1, 1)U(1, 1, 2)}{U(1, 1, 1)}, \\
U(1, 2, 2) &= xy'z' = \frac{U(1, 2, 1)U(1, 1, 2)}{U(1, 1, 1)}, \\
U(2, 2, 2) &= x'y'z' = \frac{U(2, 1, 1)U(1, 2, 1)U(1, 1, 2)}{U(1, 1, 1)U(1, 1, 1)}.
\end{align*}
\]

Therefore, considering the Tucker decomposition, there is only one (finite) completion(s) having this particular 4 observed entries as above. Note that only given \(r_2 = r_3 = 1\), it can be verified using Tucker decomposition similarly that the completion is still unique.
2.3 Deterministic Conditions for Finite Completableness

This section characterizes the connection between the sampling pattern and the number of solutions of a low-rank tensor completion. In Section 2.3.1, we define a polynomial based on each observed entry. Then, for a given subset of the rank components we transform the problem of finite completability of $U$ to the problem of finite completability of the core tensor in the Tucker decomposition of $U$. In Section 2.3.2, we propose a geometric analysis on Tucker manifold, by defining a structure for the core tensor of the Tucker decomposition such that we can determine if two core tensors span the same space. In Section 2.3.3, we construct a constraint tensor based on the sampling pattern $\Omega$. This tensor is useful for analyzing the algebraic independency of a subset of polynomials among all defined polynomials. In Section 2.3.4, we show the relationship between the number of algebraically independent polynomials in the mentioned set of polynomials and finite completability of the sampled tensor. Finally, Section 2.3.5 characterizes finite completability in terms of the sampling pattern instead of the algebraic variety for the defined set of polynomials.

2.3.1 Condition for Finite Completableness Given the Core Tensor

Assume that the sampled tensor is $U \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ and rank components $\{r_{j+1}, \ldots, r_d\}$ are given, where $j \in \{1, 2, \ldots, d-1\}$ is an arbitrary fixed number. Without loss of generality assume that $r_{j+1} \geq \ldots \geq r_d$ throughout the chapter. Define $\mathbb{P}_0$ as the Lebesgue measure on $\mathbb{R}^{r_{j+1} \times r_{j+2} \times \cdots \times r_d}$ and $\mathbb{P}_i$ as the Lebesgue measure on $\mathbb{R}^{r_i \times n_i}$, $i = 1, \ldots, d$. We assume that $U$ is chosen generically from the manifold corresponding to rank vector $(r_1, \ldots, r_d)$, or in other words, the entries of $U$ are drawn independently with respect to Lebesgue measure on the corresponding manifold. Hence, any statement that holds for $U$, it basically holds for almost every (with probability one) tensor of the same size and Tucker-rank with respect to the product measure $\mathbb{P}_0 \times \mathbb{P}_1 \times \cdots \times \mathbb{P}_d$.

Let the $d$th-order tensor $C \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times r_{j+1} \times r_{j+2} \times \cdots \times r_d}$ be a core tensor of the sampled tensor $U \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$. Then, there exist full-rank matrices $T_i$’s with $T_i \in \mathbb{R}^{r_i \times n_i}$ such that

$$U = C \times_{i=j+1}^d T_i,$$

or equivalently

$$U(\vec{x}) = \sum_{k_{j+1}=1}^{r_{j+1}} \cdots \sum_{k_d=1}^{r_d} C(x_1, \ldots, x_j, k_{j+1}, \ldots, k_d) T_{j+1}(k_{j+1}, x_{j+1}) \cdots T_d(k_d, x_d).$$
For notational simplicity, define $T = (T_{j+1}, \ldots, T_d)$. Figure 2.1 represents a Tucker decomposition for a 3rd-order tensor given the second and third components of its rank vector.

Figure 2.1: Tucker decomposition with $j = 1$ and $d = 3$.

Here, we briefly mention some key points to highlight the fundamentals of our proposed analysis.

- **Note 1:** As it can be seen from (2.7), any observed entry $U(\vec{x})$ results in an equation that involves $\Pi_{i=j+1}^d r_i$ entries of $C$ and also $r_i$ entries of $T_i$, $i = j + 1, \ldots, d$. Considering the entries of core tensor $C$ and tuple $T$ as variables (right-hand side of (2.7)), each observed entry results in a polynomial in terms of these variables.

- **Note 2:** For any observed entry $U(\vec{x})$, the tuple $(x_1, \ldots, x_j)$ specifies the coordinates of the $\Pi_{i=j+1}^d r_i$ entries of $C$ that are involved in the corresponding polynomial.

- **Note 3:** For any observed entry $U(\vec{x})$, the value of $x_i$ specifies the column of the $r_i$ entries of $T_i$ that are involved in the corresponding polynomial, $i = j + 1, \ldots, d$.

- **Note 4:** Given all observed entries $\{U(\vec{x}) : \Omega(\vec{x}) = 1\}$, we are interested in finding the number of possible solutions in terms of entries of $(C, T)$ (infinite, finite or unique) via investigating the algebraic independence among these polynomials.

Given $C$, we are interested to find a subset of the mentioned polynomials that guarantees tuple $T$ can be determined finitely. The following definition will be used to determine the number of involved variables in a set of polynomials.
**Definition 1**: For any \( i \in \{j + 1, \ldots, d\} \) and nonempty \( S_i \subseteq \{1, \ldots, n_i\} \), define \( U(S_i) \) as a set containing the locations of the entries of \(|S_i|\) rows (corresponding to the elements of \( S_i \)) of \( U(i) \). Moreover, define \( U(S_{j+1}, \ldots, S_d) = U(S_{j+1}) \cup \ldots \cup U(S_d) \). Let \( \tau \) be a subset of the locations of the entries of \( U \). Then, \( U(S_{j+1}, \ldots, S_d) \) is called the minimal hull of \( \tau \) if any \( U(S_i) \) includes exactly only those rows of \( U(i) \) that include at least one of the locations of the entries in \( \tau \).

**Example 1.** Consider a tensor \( U \in \mathbb{R}^{4 \times 4 \times 4} \) with \( j = 1, r_2 = 2 \) and \( r_3 = 3 \). Then we have

\[
U(\tilde{x}) = \sum_{k_2=1}^{2} \sum_{k_3=1}^{3} C(x_1, k_2, k_3) T_2(k_2, x_2) T_3(k_3, x_3),
\]

where \( C \in \mathbb{R}^{4 \times 2 \times 3}, T_2 \in \mathbb{R}^{2 \times 4} \) and \( T_3 \in \mathbb{R}^{3 \times 4} \). Define \( \tau = \{(1, 2, 2), (2, 2, 3), (3, 4, 2)\} \). Hence, among the four rows of the second matricization of \( U \), i.e. \( U(2) \), row numbers 2 and 4 include at least one entry in \( \tau \) (and row numbers 2 and 3 for \( U(3) \)). Then, for \( S_2 = \{2, 4\} \) \( S_3 = \{2, 3\} \), \( U(S_2, S_3) \) is the minimal hull of \( \tau \). Note that due to the definition, the minimal hull is unique. 

**Remark 1.** Consider any set of polynomials \( \{p_1, \ldots, p_k\} \) in form of (2.7). Given the core tensor \( C \) these polynomials are in terms of entries of \( T \) and let \( \tau \) be the set of corresponding entries to these polynomials in \( U \) and \( U(S_{j+1}, \ldots, S_d) \) be the minimal hull of \( \tau \). Let \( S \) denote the set of all variables (entries of \( T \)) that are involved in at least one of the polynomials \( \{p_1, \ldots, p_k\} \). Recall that according to Note 3, if an entry of \( T_i \) is involved in a polynomial, all entries of the column that includes that entry are also involved in that polynomial. Therefore, \( |S| = \sum_{i=j+1}^{d} |S_i|r_i \).

Note that each observed entry results in a scalar equation of the form of (2.7). Given the core tensor \( C \), we need at least \( \sum_{i=j+1}^{d} (n_i r_i) \) polynomials to ensure the number of possible tuples \( T \) is not infinite since the number of variables (entries of \( T \)) is \( \sum_{i=j+1}^{d} (n_i r_i) \) in total. On the other hand, the \( \sum_{i=j+1}^{d} (n_i r_i) \) mentioned polynomials should be algebraically independent to ensure the finiteness of tuples \( T \) since any algebraically independent polynomial reduces the dimension of the set of solutions by one. To ensure this independency, any subset of \( t \) polynomials of the set of polynomials corresponding to the \( \sum_{i=j+1}^{d} (n_i r_i) \) observed entries, should involve at least \( t \) variables. The following assumption will be used frequently as we show it satisfies the mentioned property.

**Assumption A1:** Anywhere that this assumption is stated, there exist \( \sum_{i=j+1}^{d} (n_i r_i) \) observed entries such that for any nonempty \( S_i \subseteq \{1, \ldots, n_i\} \) for \( i \in \{j + 1, \ldots, d\} \), \( U(S_{j+1}, \ldots, S_d) \) includes at most \( \sum_{i=j+1}^{d} |S_i|r_i \) of the mentioned \( \sum_{i=j+1}^{d} (n_i r_i) \) observed entries.
CHAPTER 2. FUNDAMENTAL LIMITS FOR LOW-TUCKER-RANK TENSOR
COMPLETION PROBLEM

Example 2. Consider a tensor $\mathcal{U} \in \mathbb{R}^{3 \times 3 \times 3}$ with $j = 1$, $r_2 = 1$ and $r_3 = 2$. Define the following two sets of observed entries each including $\sum_{i=j+1}^{d} (n_i r_i) = 9$ entries

$$
\mathcal{I}_1 = \{(1,1,1), (1,2,1), (1,1,3), (2,1,1), (2,2,1), (2,2,3), (2,3,1), (2,3,2), (3,1,1)\},
$$
$$
\mathcal{I}_2 = \{(1,1,1), (1,1,2), (1,2,1), (1,1,3), (2,2,2), (2,2,3), (2,3,1), (2,3,2), (3,3,3)\}.
$$

First, we can show that $\mathcal{I}_1$ does not satisfy Assumption $A_1$. For $\mathcal{S}_2 = \{1,2\}$ and $\mathcal{S}_3 = \{1\}$ we have

$$
\{(1,1,1), (1,2,1), (2,1,1), (2,2,1), (3,1,1)\} \subset \mathcal{U}^{(\mathcal{S}_2, \mathcal{S}_3)}.
$$

Hence, $\mathcal{U}^{(\mathcal{S}_2, \mathcal{S}_3)}$ includes 5 entries belonging to $\mathcal{I}_1$ and $\sum_{i=j+1}^{d} |\mathcal{S}_i| r_i = 4 < 5$. Therefore, $\mathcal{I}_1$ does not satisfy Assumption $A_1$ since there exists $(\mathcal{S}_2, \mathcal{S}_3)$ that violates the mentioned condition.

Second, it is easy to verify that $\mathcal{I}_2$ satisfies Assumption $A_1$ by checking all possible pairs $(\mathcal{S}_2, \mathcal{S}_3)$.

Remark 2. Assume that each column of $\mathcal{U}_{(1)}$ includes at least one observed entry, and also $\sum_{i=j+1}^{d} (n_i r_i) < n_j + \ldots + n_d$, for some $j \in \{1, \ldots, d-1\}$. Then, for any tuple $(x_2, x_3, \ldots, x_d)$ that $x_i \in \{1, \ldots, n_i\}$, there exists at least one observed entry among the set

$$
\{(1, x_2, x_3, \ldots, x_d), (2, x_2, x_3, \ldots, x_d), \ldots, (n_1, x_2, x_3, \ldots, x_d)\}.
$$

(2.9)

Hence, there exist $\sum_{i=j+1}^{d} (n_i r_i)$ observed entries that satisfy Assumption $A_j$. This is because all possible tuples $(x_{j+1}, \ldots, x_d)$ are available to be selected. For example, assuming that each column of $\mathcal{U}_{(1)}$ includes at least one observed entry and $\sum_{i=2}^{d} (n_i r_i) < n_2 + \ldots + n_d$, we can choose $\sum_{i=2}^{d} (n_i r_i) < n_2 + \ldots + n_d = N-1$ observed entries in different columns of $\mathcal{U}_{(1)}$ to satisfy Assumption $A_1$, by choosing either zero or one observed entry in each column.

Given the core tensor, the following lemma characterizes the necessary and sufficient condition on observed entries that leads to finite completability.

Lemma 1. Assume that in (2.6) the core tensor $\mathcal{C} \in \mathbb{R}^{n_1 \times \cdots \times n_j \times r_{j+1} \times \cdots \times r_d}$ is given and $\mathbf{T}_i \in \mathbb{R}^{r_i \times n_i}$ are variables. Then, for almost every $\mathcal{U}$ with probability one, there are at most finitely many possible tuples $\mathcal{I}$ that satisfy (2.6) if and only if Assumption $A_j$ holds.

Proof. The $\sum_{i=j+1}^{d} (n_i r_i)$ observed entries results in $\sum_{i=j+1}^{d} (n_i r_i)$ scalar polynomials in terms of entries of $\mathcal{T}$ as in (2.6)-(2.7). We claim that any subset of these $\sum_{i=j+1}^{d} (n_i r_i)$ polynomials with $t$ members involves at least $t$ variables in total and the sufficiency holds.
CHAPTER 2. FUNDAMENTAL LIMITS FOR LOW-TUCKER-RANK TENSOR COMPLETION PROBLEM

In order to prove the necessity, by contradiction, assume that there exists a subset of polynomials \( \{p_1, \ldots, p_t\} \) that involves at most \( t - 1 \) variables in total. Let \( \tau \) be the subset of entries of \( \mathcal{U} \) that result in polynomials \( \{p_1, \ldots, p_t\} \) and denote the minimal hull of \( \tau \) by \( \mathcal{U}(S_{j+1}, \ldots, S_d) \). Observe that according to Remark 1, \( \sum_{i=j+1}^{d} |S_i| r_i \leq t - 1 \). On the other hand, Assumption \( A_j \) results that the number of polynomials in \( \{p_1, \ldots, p_t\} \) is at most \( \sum_{i=j+1}^{d} |S_i| r_i \), i.e., \( t \leq \sum_{i=j+1}^{d} |S_i| r_i \). Hence, we have a contradiction, which completes the proof of the lemma.

**Remark 3.** Assumption \( A_j \) results that given a core tensor there are finitely many tuples \( T \) such that (2.6) holds. Consequently, in what follows without loss of generality, we analyze the finite completability of core tensor \( C \) for one particular tuple \( T \) among all finitely many tuples.

Since \( \sum_{i=j+1}^{d} (n_i r_i) \) entries of the sampled tensor \( \mathcal{U} \) are used to determine \( T \), in what follows we will use the polynomials corresponding to the set of the rest \( N_\Omega(\mathcal{U}) - \sum_{i=j+1}^{d} (n_i r_i) \) observed entries, denoted by \( \mathcal{P}(\Omega) \), to obtain \( C \). Note that since \( T \) is already solved in terms of \( C \), each polynomial in \( \mathcal{P}(\Omega) \) is in terms of elements of \( C \).

### 2.3.2 Geometry of Tucker Manifold

We need to define the following equivalence class in order to characterize a condition on the sampling pattern to study the algebraic independency of the polynomials in \( \mathcal{P}(\Omega) \). This equivalence class leads to a geometric structure for core tensors which specifies exactly one core tensor among all core tensors that span the same space.

**Definition 1.** Define an equivalence class for all core tensors \( C \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times r_{j+1} \times \cdots \times r_d} \) of the sampled tensor \( \mathcal{U} \) such that two core tensors \( C_1 \) and \( C_2 \) belong to the same class if and only if there exist full rank matrices \( D_i \in \mathbb{R}^{r_i \times r_i} \), \( i = j + 1, \ldots, d \), such that

\[
C_2 = C_1 \times_{i=j+1}^{d} D_i.
\]

A subtensor \( \mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times 1 \times \cdots \times 1} \) of \( \mathcal{U} \) can be represented by a core tensor \( C \) if there exist vectors \( \theta_i \in \mathbb{R}^{r_i \times 1} \), \( i = j + 1, \ldots, d \), such that

\[
\mathcal{Y} = C \times_{i=j+1}^{d} \theta_i.
\]

According to Definition 1, it is easy to verify that two core tensors are in the same class if and only if one of them can represent each subtensor in \( \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times 1 \times \cdots \times 1} \) of the other one.
Definition 2. Let $N_i = n_1 n_2 \ldots n_i$, $\bar{N}_i = n_{i+1} n_{i+2} \ldots n_d$ and define the matrix $\bar{U}_{(i)} \in \mathbb{R}^{N_i \times \bar{N}_i}$ as the $i$-th unfolding of the tensor $\mathcal{U}$, such that $\mathcal{U}(\bar{x}) = \bar{U}_{(i)} (\bar{\mathcal{M}}_{(i)}(x_1, \ldots, x_i), \bar{\mathcal{M}}_{(i)}(x_{i+1}, \ldots, x_d))$, where $\bar{\mathcal{M}}_{(i)}$ and $\bar{\mathcal{M}}_{(i)}$ are two bijective mappings $\bar{\mathcal{M}}_{(i)} : (x_1, \ldots, x_i) \to \{1, 2, \ldots, N_i\}$ and $\bar{\mathcal{M}}_{(i)} : (x_{i+1}, \ldots, x_d) \to \{1, 2, \ldots, \bar{N}_i\}$.

We make the following assumption which will be referred to, when it is needed.

Assumption $B_j$: $n_1 n_2 \ldots n_j \geq \sum_{i=j+1}^{d} r_i$.

Consider an arbitrary entry $C(\bar{x})$ of core tensor $C$. Note that the tuple $(x_{j+1}, \ldots, x_d)$ specifies the column number of this entry in $\bar{C}(\bar{y})$. Furthermore, $x_i$ specifies the row number of this entry in $C(i)$. Consequently, each column of $\bar{C}(\bar{y})$ indexed by $(x_{j+1}, \ldots, x_d)$ belongs to the $x_i$-th row of $C(i)$ for $i \geq j + 1$.

We are interested in providing a structure on the core tensor $C$ such that exactly one core tensor in any class satisfies it. We present this structure on $\bar{C}(\bar{y})$ ($j$-th unfolding of $C$).

Definition 3. Consider any $(d - j)$ disjoint submatrices\(^1\) $(P_{j+1}, \ldots, P_d)$ of $\bar{C}(\bar{y})$ such that (i) $P_i \in \mathbb{R}^{r_i \times r_i}$, $i = j + 1, \ldots, d$, (ii) the $\sum_{i=j+1}^{d} r_i$ rows of $\bar{C}(\bar{y})$ corresponding to rows of these submatrices are disjoint, (iii) the $r_i$ columns of $\bar{C}(\bar{y})$ corresponding to columns of $P_i$ belong to $r_i$ disjoint rows of $C(i)$, $i = j + 1, \ldots, d$. Then, $C$ is said to have a proper structure if $P_i = I_{r_i}$, $i = j + 1, \ldots, d$.

Assumption $B_j$ ensures the existence of a proper structure since the number of rows in $\bar{C}(\bar{y})$ should be at least $\sum_{i=j+1}^{d} r_i$. Note that given a proper structure there exists one core tensor in each class that satisfies it. Moreover, we can permute the rows of $P_i$'s and obtain another proper structure. Consider the following specific structure. Define

$$
P_i^{\text{can}}(x_1, x_i) = C(x_1, 1, \ldots, 1, x_i, 1, \ldots, 1) \in \mathbb{R}^{r_i \times r_i}, \quad i = j + 1, \ldots, d, \tag{2.12}
$$

where $x_1' = x_1 - \sum_{s=j+1}^{i-1} r_s$, $1 + \sum_{s=j+1}^{i-1} r_s \leq x_1 \leq \sum_{s=j+1}^{i} r_s$ and $1 \leq x_i \leq r_i$. It is easily verified all three properties in Definition 3 are satisfied by $P_i^{\text{can}}$, $i = j + 1, \ldots, d$.

Definition 4. (Canonical core tensor) We call $C$ a canonical core tensor if for $i \geq j + 1$ we have $P_i^{\text{can}} = I_{r_i}$, where $I_{r_i}$ is the $r_i \times r_i$ identity matrix.

\(^1\)Specified by a subset of rows and a subset of columns (not necessarily consecutive).
Example 3. Assume $\mathcal{U} \in \mathbb{R}^{7 \times 3 \times 4}$, $j = 1$, $r_2 = 2$ and $r_3 = 3$. For simplicity in representing the canonical core tensor, we consider bijective mappings $\mathcal{M}(i)$ and $\tilde{\mathcal{M}}(i)$ (in Definition 2) that result in the structure of $\tilde{\mathcal{C}}(1)$ shown below. Observe that any permutation of rows of $\tilde{\mathcal{C}}(1)$ results in a proper structure. However, only those permutations of columns of $\tilde{\mathcal{C}}(1)$ that satisfy property (iii) in Definition 3 result in a proper structure.

$$
\begin{array}{c|c|c}
1 & 0 & 0 \\
0 & 1 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{array}
\quad
\begin{array}{c|c|c|c|c|c}
1 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 \\
\end{array}
$$

Note that (2.10) leads to the fact that the dimension of all core tensors $\mathcal{C}$ that span different spaces (without any polynomial restrictions in $\mathcal{P}(\Omega)$) is equal to $\left(\prod_{i=1}^{j} n_i\right) \left(\prod_{i=j+1}^{d} r_i\right) \left(\sum_{i=j+1}^{d} r_i^2\right)$, as the total number of entries of $D_i$'s is equal to $\left(\sum_{i=j+1}^{d} r_i^2\right)$. Moreover, observe that a core tensor with a proper structure has $\left(\sum_{i=j+1}^{d} r_i^2\right)$ known entries, and therefore the number of unknown entries is equal to $\left(\prod_{i=1}^{j} n_i\right) \left(\prod_{i=j+1}^{d} r_i\right) \left(\sum_{i=j+1}^{d} r_i^2\right)$.

Remark 4. In order to prove there are finitely many completions for tensor $\mathcal{U}$, it suffices to prove that there are finitely many canonical core tensors that fit in $\mathcal{U}$.

Suppose $\mathcal{C}$ has a proper structure. Let $g_{j+1}(x)$ denote the maximum number of known entries among any $x$ rows of $\tilde{\mathcal{C}}(j)$. As will be seen in Section 2.3.3, $g_{j+1}(x)$ plays an important role in expressing the maximum number of algebraically independent polynomials in a subset of $\mathcal{P}(\Omega)$. Note that in exactly $r_i$ rows of $\tilde{\mathcal{C}}(j)$ there are exactly $r_i$ known entries, i.e., entries of $P_i$, $i = j + 1, \ldots, d$. Also, there are $\sum_{i=j+1}^{d} r_i$ rows that include known entries in $\tilde{\mathcal{C}}(j)$, i.e., rows of $P_i$, $i = j + 1, \ldots, d$.

Recall the assumption $r_{j+1} \geq \ldots \geq r_d$. Therefore, as long as $x \leq r_{j+1}$, the maximum number of known entries is $g_{j+1}(x) = r_{j+1} x$ by selecting the $x$ rows of $\tilde{\mathcal{C}}(j)$ to cover any $x$ rows of $P_{j+1}$.
On the other hand, if \( r_{j+1} \leq x \leq r_{j+1} + r_{j+2} \), the maximum number of known entries is \( g_{j+1}(x) = r_{j+1}^2 + r_{j+2}(x - r_{j+1}) \) by selecting the \( x \) rows of \( \mathbf{C}(j) \) to cover all rows of \( \mathbf{P}_{j+1} \) and any \( (x - r_{j+1}) \) rows of \( \mathbf{P}_{j+2} \). Then, in general we have

\[
g_{j+1}(x) = \sum_{i=j+1}^{d} \min \left\{ r_i, \left( x - \sum_{i'=j+1}^{i-1} r_{i'} \right)^+ \right\} r_i. \tag{2.13}
\]

### 2.3.3 Constraint Tensor

In the following, we propose a procedure to construct a \((j+1)^{th}\) order binary tensor \( \mathbf{\Omega} \) based on \( \Omega \) such that \( \mathcal{P}(\mathbf{\Omega}) = \mathcal{P}(\Omega) \). Using \( \mathbf{\Omega} \), we are able to recognize the observed entries that have been used to obtain the tuple \( \mathbb{T} \), and we can easily verify if two polynomials in \( \mathcal{P}(\Omega) \) are in terms of the same set of variables. Then, in Section 2.3.4, we characterize the relationship between the maximum number of algebraically independent polynomials in \( \mathcal{P}(\mathbf{\Omega}) \) and \( \mathbf{\Omega} \).

For any subtensor \( \mathbf{Y} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times 1 \times \cdots \times 1} \) of the tensor \( \mathbf{U} \), there exist row vectors \( \theta_i \in \mathbb{R}^{r_i \times 1} \), \( i = j+1, \ldots, d \), such that (2.11) holds or equivalently

\[
\mathbf{Y}(x_1, \ldots, x_j, \mathbf{I}_{d-j}) = \sum_{k_{j+1}=1}^{r_{j+1}} \cdots \sum_{k_d=1}^{r_d} \mathbf{C}(x_1, \ldots, x_j, k_{j+1}, \ldots, k_d) \theta_{j+1}(k_{j+1}, 1) \cdots \theta_d(k_d, 1), \tag{2.14}
\]

where \( \mathbf{I}_{d-j} \) is an all-1 \((d-j)\)-dimensional row vector.

For each subtensor \( \mathbf{Y} \) of the sampled tensor \( \mathbf{U} \), let \( N_\Omega(\mathbf{Y}^\mathbb{T}) \) denote the number of sampled entries in \( \mathbf{Y} \) that have been used to obtain the tuple \( \mathbb{T} \). Then, \( \mathbf{Y} \) contributes \( N_\Omega(\mathbf{Y}) - N_\Omega(\mathbf{Y}^\mathbb{T}) \) polynomial equations in terms of the entries of the core tensor \( \mathbf{C} \) among all \( N_\Omega(\mathbf{U}) - \sum_{i=j+1}^{d} (n_ir_i) \) polynomials in \( \mathcal{P}(\Omega) \).

The sampled tensor \( \mathbf{U} \) includes \( n_{j+1}n_{j+2} \cdots n_d \) subtensors that belong to \( \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times 1 \times \cdots \times 1} \) and we label these subtensors by \( \mathbf{Y}_{(t_{j+1}, \ldots, t_d)} \) where \((t_{j+1}, \ldots, t_d)\) represents the coordinate of the subtensor. Define a binary valued tensor \( \hat{\mathbf{Y}}_{(t_{j+1}, \ldots, t_d)} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times 1 \times \cdots \times 1 \times k} \), where \( k = N_\Omega(\mathbf{Y}_{(t_{j+1}, \ldots, t_d)}) - N_\Omega(\mathbf{Y}^\mathbb{T}_{(t_{j+1}, \ldots, t_d)}) \) and its entries are described as the following. We can look at \( \hat{\mathbf{Y}}_{(t_{j+1}, \ldots, t_d)} \) as \( k \) tensors each belongs to \( \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times 1 \times \cdots \times 1} \). For each of the mentioned \( k \) tensors in \( \hat{\mathbf{Y}}_{(t_{j+1}, \ldots, t_d)} \) we set the entries corresponding to the \( N_\Omega(\mathbf{Y}^\mathbb{T}_{(t_{j+1}, \ldots, t_d)}) \) observed entries that are used to obtain \( \mathbb{T} \) in (2.6) equal to 1. For each of the other \( k \) observed entries, we pick one of the \( k \) tensors of \( \hat{\mathbf{Y}}_{(t_{j+1}, \ldots, t_d)} \) 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.
CHAPTER 2. FUNDAMENTAL LIMITS FOR LOW-TUCKER-RANK TENSOR COMPLETION PROBLEM

For the sake of simplicity in notation, we treat tensors \( \tilde{Y}_{(t_{j+1}, \ldots, t_d)} \) as a member of \( \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_j \times k} \), instead of \( \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_j \times 1 \times 1 \times \ldots \times 1 \times k} \). Now, by putting together all \( n_{j+1}n_{j+2} \cdots n_d \) tensors in dimension \((j+1)\), we construct a binary valued tensor \( \tilde{\Omega} \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_j \times K_j} \), where \( K_j = N_{\Omega}(\mathcal{U}) - \sum_{l=j+1}^{d} (n_l r_l) \) and call it the constraint tensor. In order to shed some light on the above procedure we give an illustrative example in the following.

**Example 4.** Consider an example in which \( d = 3, j = 2, r_3 = 2 \) and \( \mathcal{U} \in \mathbb{R}^{3 \times 2 \times 2} \). Assume that \( \Omega(x, y, z) = 1 \) if \((x, y, z) \in \mathcal{S} \) and \( \Omega(x, y, z) = 0 \) otherwise, where

\[
\mathcal{S} = \{(1, 1, 1), (1, 2, 1), (2, 2, 1), (3, 1, 1), (1, 1, 2), (2, 1, 2), (3, 2, 2)\},
\]

represents the set of observed entries. Also, assume that \( N_{\Omega}(\tilde{Y}_{(1)}^{T}) = N_{\Omega}(\tilde{Y}_{(2)}^{T}) = 2 \) and the entries that we use to obtain \( \mathcal{T} \) are \((1, 1, 1), (1, 2, 1), (1, 1, 2) \) and \((3, 2, 2)\). Hence, \( \tilde{Y}_{(1)} \in \mathbb{R}^{3 \times 2 \times 1 \times 2}, \tilde{Y}_{(2)} \in \mathbb{R}^{3 \times 2 \times 1 \times 1} \), and therefore the constraint tensor \( \tilde{\Omega} \) belongs to \( \mathbb{R}^{3 \times 2 \times 3} \).

Note that \( \tilde{Y}_{(1)}(1, 1, 1, 1) = \tilde{Y}_{(1)}(1, 1, 1, 2) = \tilde{Y}_{(1)}(1, 2, 1, 1) = \tilde{Y}_{(1)}(1, 2, 1, 2) = 1 \), and also for the two other observed entries we have \( \tilde{Y}_{(1)}(2, 2, 1, 1) = 1 \) and \( \tilde{Y}_{(1)}(3, 1, 1, 2) = 1 \) and the rest of the entries of \( \tilde{Y}_{(1)} \) are equal to zero. Moreover, \( \tilde{Y}_{(2)}(1, 1, 1, 1) = \tilde{Y}_{(2)}(2, 1, 1, 1) = \tilde{Y}_{(2)}(3, 2, 1, 1) = 1 \) and the rest of the entries of \( \tilde{Y}_{(2)} \) are equal to zero.

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

\[
\mathcal{S}' = \{(1, 1, 1), (1, 2, 1), (2, 2, 1), (1, 1, 2), (1, 2, 2), (3, 1, 2), (1, 1, 3), (3, 2, 3), (2, 1, 3)\}.
\]

Note that each subtensor of \( \tilde{\Omega} \) that belongs to \( \mathbb{R}^{n_1 \times \ldots \times n_j \times 1} \) represents one of the polynomials in \( \mathcal{P}(\Omega) \) besides showing the polynomials that have been used to obtain \( \mathcal{T} \). More specifically, consider a subtensor of \( \tilde{\Omega} \) that belongs to \( \mathbb{R}^{n_1 \times \ldots \times n_j \times 1} \) with \( l + 1 \) nonzero entries. Observe that exactly \( l \) of them correspond to the observed entries that have been used to obtain \( \mathcal{T} \). Hence, this subtensor represents a polynomial after replacing entries of \( \mathcal{T} \) by the expressions in terms of entries of \( \mathcal{C} \), i.e., a polynomial in \( \mathcal{P}(\Omega) \).

Recall that the tuple \((x_{j+1}, \ldots, x_d)\) specifies the column number of entry \( C(x) \) in \( \tilde{C}_{(j)} \). Hence, if the \( i \)-th column of \( \tilde{\Omega}_{(j+1)} \) is nonzero, then there exists a polynomial in \( \mathcal{P}(\Omega) \) that involves all entries of core tensor corresponding to the entries of the \( i \)-th row of \( \tilde{C}_{(j)} \).
2.3.4 Algebraic Independence

In this subsection, we derive the required number of algebraically independent polynomials in \( P(\Omega) \) for finite completability. Then, a sampling pattern on the constraint tensor is proposed to obtain the maximum number of algebraically independent polynomials in \( P(\bar{\Omega}) \).

The following lemma determines the required number of algebraically independent polynomials in \( P(\Omega) \) that is needed to ensure finite completability of the core tensor.

**Lemma 2.** Assume that Assumptions \( A_j \) and \( B_j \) hold. For almost every \( U \), there exist only finitely many completions of \( U \) if and only if there exist

\[
\left( \prod_{i=1}^{j} n_i \right) \left( \prod_{i=j+1}^{d} r_i \right) - \left( \sum_{i=j+1}^{d} r_i^2 \right)
\]

algebraically independent polynomials in \( P(\Omega) \).

**Proof.** Assume that \( C \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_j \times r_{j+1} \times \ldots \times r_d} \) is a core tensor for the sampled tensor \( U \). Since assumption \( A_j \) holds, Lemma 1 results that there exist finitely many tuples \( T \) such that (2.6) holds. However, according to Remark 3, it suffices to assume \( T \) is fixed and then prove the statement. Let \( P(\Omega) = \{p_1, \ldots, p_m\} \) and define \( S_i \) as the set of all core tensors that satisfy polynomial restrictions \( \{p_1, \ldots, p_i\}, i = 0, \ldots, m \) (\( S_0 \) is the set of all core tensors without any polynomial restriction).

Observe that each algebraically independent polynomial reduces the dimension (degree of freedom) of the set of solutions by one. In other words, \( \dim(S_i) = \dim(S_{i-1}) \) if the maximum number of algebraically independent polynomials in sets \( \{p_1, \ldots, p_i\} \) and \( \{p_1, \ldots, p_{i-1}\} \) are the same and \( \dim(S_i) = \dim(S_{i-1}) - 1 \) otherwise. Moreover, with probability one, \( |S_m| \) is finite if and only if there are \( \dim(C) = \dim(S_0) \) algebraically independent polynomial restrictions on the entries of the core tensor \( C \), i.e., \( |S_m| \) is finite if and only if \( \dim(S_m) = 0 \) [Pimentel-Alarcón et al., 2016b]. Hence, there are finitely many completions of the sampled tensor \( U \) if and only if there exist

\[
\left( \prod_{i=1}^{j} n_i \right) \left( \prod_{i=j+1}^{d} r_i \right) - \left( \sum_{i=j+1}^{d} r_i^2 \right)
\]

algebraically independent polynomials in \( P(\Omega) \). \( \Box \)

**Remark 5.** The proof of [Pimentel-Alarcón et al., 2016b, Lemma 2] is based on the claim that the independency of a set of polynomials is equivalent to those polynomials being a regular sequence. Very recently, as this dissertation was being completed, in [Tsakiris, 2020], the validity of this statement for every set of polynomial was questioned.

As a result of Lemma 2, we can certify finite completability based on the maximum number of algebraically independent polynomials in \( P(\Omega) = P(\bar{\Omega}) \).
Definition 5. Let $\hat{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times t}$ be a subtensor of the constraint tensor $\hat{\Omega}$. Let $m_i(\hat{\Omega}')$ denote the number of nonzero columns of $\hat{\Omega}'_{(i)}$. Also, let $\mathcal{P}(\hat{\Omega}')$ denote the set of polynomials that correspond to nonzero entries of $\hat{\Omega}'$.

Recall Note 2 regarding the number of involved entries of core tensor in a set of polynomials. However, as mentioned earlier, some of the entries of $C$ are known, i.e., $(P_{j+1}, \ldots, P_d)$. Therefore, in order to find the number of variables (unknown entries of $C$) in a set of polynomials, we should subtract the number of known entries in the corresponding pattern from the total number of involved entries.

For any subtensor $\hat{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times t}$ of the constraint tensor, the next theorem states an upper bound on the number of algebraically independent polynomials in the set $\mathcal{P}(\hat{\Omega}')$.

Theorem 1. Assume that Assumption $B_j$ holds. For any subtensor $\hat{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times t}$ of the constraint tensor, the maximum number of algebraically independent polynomials in $\mathcal{P}(\hat{\Omega}')$ is no more than

$$\left(\prod_{i=j+1}^d r_i\right) m_{j+1}(\hat{\Omega}') - g_{j+1}(m_{j+1}(\hat{\Omega}')), \tag{2.15}$$

where $g_{j+1}(\cdot)$ is given in (2.13).

Proof. Observe that the number of algebraically independent polynomials in a subset of polynomials of $\mathcal{P}(\hat{\Omega}')$ is at most equal to the total number of variables that are involved in the corresponding polynomials. According to (2.7), for each observed entry $\mathcal{U}(\vec{x})$, we have a polynomial in terms of $\prod_{i=j+1}^d r_i$ entries of the core tensor $C$ corresponding to the first $j$ coordinates of the location of the observed entry. Therefore, the number of such variables in $\mathcal{P}(\hat{\Omega}')$ will be $\prod_{i=j+1}^d r_i$ times the number of different tuples $(x_1, \ldots, x_j)$ among the corresponding observed entries. The number of nonzero columns of $\hat{\Omega}'_{(j+1)}$ is exactly equal to the number such tuples. Therefore, the number of involved entries of $C$ (known and unknown) in polynomials in $\mathcal{P}(\hat{\Omega}')$ is equal to $\left(\prod_{i=j+1}^d r_i\right) m_{j+1}(\hat{\Omega}')$.

On the other hand, among the $\sum_{i=j+1}^d r_i^2$ known entries corresponding to $(P_{j+1}, \ldots, P_d)$ in $\hat{C}_{(j)}$, $g_{j+1}(m_{j+1}(\hat{\Omega}'))$ of them are involved in polynomials of $\mathcal{P}(\hat{\Omega}')$. This is because entries of the $i$-th row of $\hat{C}_{(j)}$ are involved in a polynomials if and only if the $i$-th column of $\hat{\Omega}'_{(j+1)}$ includes at least one nonzero entry. Hence, the number of variables that are involved in the set of polynomials
\( \mathcal{P}(\tilde{\Omega}') \) is given by (2.15) for a particular proper structure and proof is complete as (2.15) is an upper bound for the number of algebraically independent polynomials.

We are also interested in finding a condition on \( \tilde{\Omega}' \) which results that \( \mathcal{P}(\tilde{\Omega}') \) is minimally algebraically dependent, i.e., the polynomials in \( \mathcal{P}(\tilde{\Omega}') \) are algebraically dependent but polynomials in every of its proper subset are algebraically independent. This can help obtain the maximum number of algebraically independent polynomials in \( \mathcal{P}(\tilde{\Omega}') \) as Theorem 1 only provides an upper bound. The next lemma will be used in Theorem 2 in order to find a condition on \( \tilde{\Omega}' \) which results that the set of polynomials in \( \mathcal{P}(\tilde{\Omega}') \) is minimally algebraically dependent. The following lemma is a re-statement of Lemma 7 in [Ashraphijuo and Wang, 2017a].

**Lemma 3.** Assume that Assumption \( B_j \) holds. Suppose that \( \tilde{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_j \times t} \) is a subtensor of the constraint tensor such that \( \mathcal{P}(\tilde{\Omega}') \) is minimally algebraically dependent. Then, for almost every \( \mathcal{U} \), the number of variables that are involved in the set of polynomials \( \mathcal{P}(\tilde{\Omega}') \) is \( t - 1 \).

Finally, the next theorem provides a relationship between the exact number of algebraically independent polynomials in \( \mathcal{P}(\tilde{\Omega}) \) and a geometric property on \( \tilde{\Omega} \).

**Theorem 2.** Assume that Assumption \( B_j \) holds. The polynomials in the set \( \mathcal{P}(\tilde{\Omega}) \) are algebraically dependent if and only if \( \left( \prod_{i=j+1}^{d} m_{j+1}(\tilde{\Omega}') - g_{j+1}(m_{j+1}(\tilde{\Omega}')) \right) < t \) for some subtensor \( \tilde{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_j \times t} \) of the constraint tensor \( \tilde{\Omega} \).

**Proof.** If the polynomials in set \( \mathcal{P}(\tilde{\Omega}) \) are algebraically dependent, then there exists a subset of the polynomials that are minimally algebraically dependent. According to Lemma 3, if \( \tilde{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_j \times t} \) is the corresponding subtensor to this minimally algebraically dependent set of polynomials, the number of variables that are involved in \( \mathcal{P}(\tilde{\Omega}') = \{ p_1, p_2, \ldots, p_t \} \) is equal to \( t - 1 \).

On the other hand, \( \left( \prod_{i=j+1}^{d} m_{j+1}(\tilde{\Omega}') - g_{j+1}(m_{j+1}(\tilde{\Omega}')) \right) \) is the minimum possible number of involved variables in \( \mathcal{P}(\tilde{\Omega}') \) since \( g_{j+1}(m_{j+1}(\tilde{\Omega}')) \) is the maximum number of known entries of core tensor that are involved in \( \mathcal{P}(\tilde{\Omega}') \). Therefore, \( \left( \prod_{i=j+1}^{d} m_{j+1}(\tilde{\Omega}') - g_{j+1}(m_{j+1}(\tilde{\Omega}')) \right) \leq t - 1 \).

In order to prove the other side of the statement, assume that \( \left( \prod_{i=j+1}^{d} m_{j+1}(\tilde{\Omega}') - g_{j+1}(m_{j+1}(\tilde{\Omega}')) \right) < t \) for some subtensor \( \tilde{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_j \times t} \) of the constraint tensor \( \tilde{\Omega} \). Recall that \( t \) is the number of polynomials in \( \mathcal{P}(\tilde{\Omega}') \). On the other hand, according to Theorem 1, \( \left( \prod_{i=j+1}^{d} m_{j+1}(\tilde{\Omega}') - g_{j+1}(m_{j+1}(\tilde{\Omega}')) \right) \) is the maximum number of algebraically independent polynomials, and therefore the polynomials in \( \mathcal{P}(\tilde{\Omega}') \) are not algebraically independent and it completes the proof.
CHAPTER 2. FUNDAMENTAL LIMITS FOR LOW-TUCKER-RANK TENSOR COMPLETION PROBLEM

2.3.5 Finite Completability Using Analysis on Tucker Manifold

Theorem 2 together with Lemma 2 can lead to a necessary and sufficient condition on the constraint tensor $\tilde{\Omega}$ in order to ensure that there are finitely many completions for the sampled tensor $U$, as stated by the next theorem.

**Theorem 3.** Assume that Assumptions $A_j$ and $B_j$ hold. Then, for almost every $U$, there are only finitely many tensors that fit in the sampled tensor $U$, and have tensor rank components $r_i$ for $i = j + 1, \ldots, d$ if and only if the following two conditions hold:

1. There exists a subtensor $\tilde{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times n}$ of the constraint tensor such that
   $$n = \left(\prod_{i=1}^{j} n_i\right) \left(\prod_{i=j+1}^{d} r_i\right) - \left(\sum_{i=j+1}^{d} r_i^2\right),$$

2. For any $t \in \{1, \ldots, n\}$ and any subtensor $\tilde{\Omega}'' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times t}$ of the tensor $\tilde{\Omega}'$ (in condition (i)), the following inequality holds
   $$\left(\prod_{i=j+1}^{d} r_i\right) m_{j+1}(\tilde{\Omega}'') - g_{j+1}(m_{j+1}(\tilde{\Omega}'')) \geq t.$$  \hspace{1cm} (2.16)

**Proof.** According to Lemma 2, for almost every $U$, there are finitely many completions of $U$ if and only if there exist algebraically independent polynomials in $P(\tilde{\Omega})$. On the other hand, according to Theorem 2 we conclude that a set of polynomials are algebraically independent if and only if condition (ii) in the statement of the theorem holds. Hence, for almost every $U$, there are finitely many completions of $U$ if and only if conditions (i) and (ii) hold.

**Remark 6.** As a sanity check, we next show that when $d = 2$ (matrix case), Theorem 3 reduces to Theorem 1 in [Pimentel-Alarcón et al., 2016b]. To see this, note that for $d = 2$, we only have one rank component and denote it by $r$. Then, Condition (i) states that there exists a submatrix $\tilde{\Omega}' \in \mathbb{R}^{n_1 \times (n_1 r - r^2)}$ (basically $n = n_1 r - r^2$ in Condition (i)) of the constraint matrix $\tilde{\Omega}$. And Condition (ii) on the property of submatrix $\tilde{\Omega}'$ becomes: for any $t \in \{1, \ldots, n_1 r - r^2\}$ and any submatrix $\tilde{\Omega}'' \in \mathbb{R}^{n_1 \times t}$ of the matrix $\tilde{\Omega}'$, the following inequality holds
   $$r m_1(\tilde{\Omega}'') - g_1(m_1(\tilde{\Omega}'')) \geq t.$$  \hspace{1cm} (2.17)

Note that due to the way that we constructed the constraint matrix (tensor) each column of $\tilde{\Omega}''$ has exactly $r + 1$ non-zero entries. Therefore, $\tilde{\Omega}''$ has at least $r + 1$ non-zero rows, i.e., $m_1(\tilde{\Omega}'') \geq$
Then, according to the definition of $g_1$ in (2.13), $g_1(m_1(\mathcal{\tilde{Y}}'')) = r \min\{r, m_1(\mathcal{\tilde{Y}}'') - 0\} = r^2$. Therefore, (2.17) becomes

$$rm_1(\mathcal{\tilde{Y}}'') - r^2 \geq t,$$  \hspace{1cm} (2.18)

or equivalently,

$$m_1(\mathcal{\tilde{Y}}'') \geq \frac{t}{r} + r.$$  \hspace{1cm} (2.19)

Hence, Theorem 3 for $d = 2$ is exactly the same as Theorem 1 in [Pimentel-Alarcón et al., 2016b].

There are a few observations based on Theorem 3:

- In theorem 3, we characterize all of the sampling patterns that ensure there are only finitely many completions such that if only one single sample from that pattern is missed, then there are infinitely many completions.

- If we use the Grassmannian analysis on each dimension individually, we will only obtain a sufficient condition on the sampling patterns for finite completability, given multiple rank constraints. However, our proposed analysis on Tucker manifold results in a necessary and sufficient condition on the sampling patterns for finite completability when multiple rank components are given. Hence, our proposed analysis on Tucker manifold in general requires much less number of samples to ensure finite completability in comparison with the analysis on Grassmannian manifold when more than one rank component is given.

- It is also important to observe the advantage of the proposed method as we decrease the value of $j$ since we are incorporating more rank components. Intuitively, for the case of $j = 1$, the polynomials obtained in (2.6) involve much more variables in comparison with the low-order analysis on Grassmannian manifold, i.e., $j = d - 1$. Therefore, the case of $j = 1$ requires much less number of samples in order to have sufficient number of algebraically independent polynomials for finite completability.

- Note that our analysis is not valid when $j = 0$ (given all rank components) since the defined proper structure does not have meaning any more as zero unfolding does not exist, and therefore we are not able to characterize the necessary and sufficient condition on sampling.
pattern for finite completability. However, if all rank components are given we can simply ignore one of them and characterize the necessary and sufficient condition for \( j = 1 \) which results in a sufficient condition for \( j = 0 \).

If the number of completions given \((r_{j+1},\ldots,r_d)\) is finite, it can be concluded that the number of completions given \((r_j,\ldots,r_d)\) is finite as well. Therefore, this property should be verifiable through the geometric property (2.16) proposed in Theorem 3. In the following lemma, using the pigeonhole principle, we show this result without analyzing the algebraic variety.

**Lemma 4.** Assume that Assumptions \( A_j, B_j, A_{j-1} \) and \( B_{j-1} \) hold. Suppose that the sampling pattern is such that the properties (i) and (ii) in the statement of Theorem 3 hold for \( j \). Then, the properties (i) and (ii) in the statement of Theorem 3 hold for \( j - 1 \).

**Proof.** The proof is given in Appendix 2.8.

### 2.4 Probabilistic Conditions for Finite Completability

In this section, two different lower bounds on the sampling probability are proposed and analyzed to ensure finite completability. The first bound is obtained by applying [Pimentel-Alarcón et al., 2016b, Theorem 3] and the second bound is obtained through the proposed geometric approach in Theorem 3. We will observe later that our proposed analysis on Tucker manifold leads to a better lower bound through numerical analysis.

Here we briefly outline the key steps of the second approach. Lemmas 6 and 7 each provides a lower bound on the sampling probability that results in a geometric property for \( \Omega \), i.e., inequalities (2.23) and (2.30), respectively. Then, Theorem 5 takes advantage of the above lemmas to propose a bound on the sampling probability to guarantee property (2.16) for \( \Omega \). Finally, Lemma 8 shows that (2.16) also holds for the constraint tensor \( \bar{\Omega} \). In order to show Lemma 8 we develop a generalization of Hall’s Theorem on bipartite graphs in Theorem 6.

We use the approach similar to [Pimentel-Alarcón et al., 2016b, Lemma 9] in order to apply Theorem 3 and obtain a lower bound on the sampling probability to ensure that there are only finitely many completions. According to our earlier discussion, Theorem 3 for the case of \( j = 1 \) results in the mildest condition on the sampling patterns for ensuring finite completability. In other words, setting \( j = 1 \) in Theorem 3 results in a tighter lower bound.
CHAPTER 2. FUNDAMENTAL LIMITS FOR LOW-TUCKER-RANK TENSOR COMPLETION PROBLEM

2.4.1 Lower Bound on Sampling Probability based on Analysis on Grassmannian Manifold

We first state a lemma, whose corollary (Corollary 1) is used extensively in this section in order to find a lower bound on the sampling probability to guarantee a lower bound on the number of sampled entries with high probability.

Lemma 5. Consider a vector with $n_i$ entries and assume each entry is observed with probability $p$ and independently from the other entries. Then, with probability at least $1 - \exp\left(-\frac{n_i^2}{2c^2}\right)$ at least \((p - \frac{1}{c})n_i\) entries are observed.

Proof. Azuma’s inequality states that for a martingale \(\{X_k : k = 0, 1, 2, \ldots\}\) that \(|X_k - X_{k-1}| \leq 1\) holds, we have \(P(X_n - E[X_n] > t) \leq \exp\left(-\frac{t^2}{2n}\right)\) and \(P(E[X_n] - X_n > t) \leq \exp\left(-\frac{t^2}{2n}\right)\) [Janson, 2002]. Therefore, using Azuma’s inequality and the fact that sampling has Bernoulli distribution with parameter $p$, it can be seen that with probability at most $\exp\left(-\frac{t^2}{2n}\right)$ the number of observed entries is less than $n_i p - t$. Now, by setting $t$ equal to $\frac{n_i}{c}$ the proof is complete. \(\square\)

Corollary 1. Consider a vector with $n_i$ entries where each entry is observed with probability $p$ independently from the other entries. If $p > p' = \frac{2r_i}{n_i} + \frac{1}{\sqrt{n_i}}$, with probability at least \(1 - \exp\left(-\frac{n_i}{2}\right)\), more than $2r_i$ entries are observed. Similarly, if $p > p'' = \frac{12\log\left(e n_i\epsilon\right)}{n_i} + \frac{1}{\sqrt{n_i}}$, with probability at least \(1 - \exp\left(-\frac{n_i}{2}\right)\), more than $12\log\left(e n_i\epsilon\right)$ entries are observed.

Proof. Both parts follow by setting $c = \sqrt{n_i}$ in Lemma 5. \(\square\)

The following theorem uses the matrix result for single rank constraint and provides a sufficient condition on the sampling pattern for finite completability of tensor. We will assume that

\[r_i \leq \frac{n_i}{6}, \quad N_i \geq r_i(n_i - r_i) \quad \text{for} \quad i = 1, 2, \ldots, d. \tag{2.20}\]

Theorem 4. Consider a randomly sampled tensor $U$ with Tucker rank \((r_1, \ldots, r_d)\). Suppose that the inequalities in (2.20) hold. Moreover, assume that the sampling probability satisfies

\[p > \min_{1 \leq i \leq d} \left(\max\left(\frac{2r_i}{n_i}, \frac{12\log\left(e n_i\epsilon\right)}{n_i}\right) + \frac{1}{\sqrt{n_i}}\right). \tag{2.21}\]

Then, there are only finitely many completions of the sampled tensor $U$ with the given rank vector with probability at least $(1 - 2\epsilon)$. 
Proof. Denote $\ell = \arg \min_{1 \leq i \leq d} \left( \frac{2r_i}{n_i}, \frac{12 \log(\frac{en_i}{\epsilon})}{n_i} + \frac{1}{\sqrt{n_i}} \right)$. By (2.21) and using Corollary 1, with probability almost one the number of observed entries at each column of $U_{(\ell)}$ is at least $\max\{2r_\ell, 12 \log(\frac{en_\ell}{\epsilon})\}$. Then, with (2.20) and using [Pimentel-Alarcón et al., 2016b, Theorem 3], it follows that there are finitely many matrix completions for the observed entries of $U_{(\ell)}$ with rank $r_\ell$. As a result, there are finitely many tensors with the $\ell$-th component of rank being $r_\ell$ that agree with the observed entries. The proof is complete since the set of tensors that are a completion with the rank vector $(r_1, \ldots, r_d)$ is a subset of the set of tensors whose finiteness is shown above. \hfill \Box

2.4.2 Lower Bound on Sampling Probability based on Analysis on Tucker Manifold

We are interested in taking advantage of Theorem 3 to obtain another lower bound on the sampling probability for finite completability. We assume $j = 1$, and therefore the constraint tensor $\tilde{\Omega}$ is a second-order tensor, i.e., an $n_1 \times K_1$ matrix.

**Lemma 6.** Consider an arbitrary set $\Omega'_{(1)}$ of $n_1 - 1$ columns of $\Omega_{(1)}$ (first matricization of $\Omega$). Assume that each column of $\Omega_{(1)}$ includes at least $l$ nonzero entries, where

$$l > 6 \log(n_1) + 2 \log\left(\frac{k}{\epsilon}\right) + 4. \quad (2.22)$$

Then, with probability at least $1 - \frac{\epsilon}{k}$, every subset $\Omega''_{(1)}$ of columns of $\Omega'_{(1)}$ satisfies

$$m_2(\Omega''_{(1)}) - 1 \geq t, \quad (2.23)$$

where $t$ is the number of columns of $\Omega''_{(1)}$ and $m_2(\Omega''_{(1)})$ is the number of nonzero rows of $\Omega''_{(1)}$ (observe that second matricization of a matrix is its transpose).

**Proof.** The proof is similar to the proof of [Pimentel-Alarcón et al., 2016b, Lemma 9] with some delicate modifications to improve the result for this case. Note that (2.22) can be rewritten as

$$l > 2 \log\left(\frac{n_1k}{\epsilon}\right) + 4 \log(n_1) + 4 \quad (2.24)$$

$$> \log\left(\frac{n_1e^2k}{\epsilon}\right) + 2. \quad (2.25)$$

Define $\mathcal{E}$ as the event that for some submatrix $\Omega''_{(1)} \in \mathbb{R}^{n_1 \times t}$ of the matrix $\Omega'_{(1)}$ (2.23) does not hold. We are interested in finding an upper bound on the probability of $\mathcal{E}$. Then, from the proof
of [Pimentel-Alarcón et al., 2016b, Lemma 9] and by setting $r = 1$ in inequalities (12) and (13) in [Pimentel-Alarcón et al., 2016b], we have

$$P(\mathcal{E}) < \sum_{n=1}^{\frac{n_1}{2}} \binom{n_1}{n} \left( \frac{n}{n_1} \right)^{ln_1} + \sum_{n=1}^{\frac{n_1}{2}} \left( \frac{n_1 - n}{n_1} \right)^{l(n_1-n)}. \quad (2.26)$$

Note that

$$\binom{s}{q} = s(s-1)\ldots(s-q+1) \frac{q!}{s^q} \leq \left( \frac{se}{q} \right)^q, \quad (2.27)$$

where the last inequality holds since $e^q = \sum_{i=0}^{\infty} \frac{q^i}{i!} \leq \frac{q^q}{q!}$. Applying (2.27) to each term of the first summation of (2.26) we obtain

$$\binom{n_1}{n} \left( \frac{n}{n_1} \right)^{ln_1} < e^{2n} \left( \frac{n}{n_1} \right)^{(l-2)n} \leq \left( e^{2l+2} \right)^n < \frac{e}{n_1 k}, \quad (2.28)$$

where the last inequality follows from (2.25). This directly results that the first summation in (2.26) is less than $\frac{e}{n_1 k}$.

Moreover, again by applying (2.27) to each term of the second summation in (2.26) we obtain

$$\left( \frac{n_1}{n_1 - n} \right)^{l(n_1-n)} \left( \frac{n_1 - n}{n_1} \right)^{ln_1} < (n_1 e)^{2n} \left( \frac{n_1 - n}{n_1} \right)^{ln_1} \leq (n_1 e)^{2n} e^{-\frac{ln_1}{2}} = \left( e^{2\log n_1 + 2 - \frac{l}{2}} \right)^n < \frac{e}{n_1 k}, \quad (2.29)$$

where the last inequality follows from (2.24). This directly results that the second summation in (2.26) is less than $\frac{e}{n_1 k}$ and that completes the proof.

Lemma 7. Consider an arbitrary set $\Omega'(1)$ of $n_1$ columns of $\Omega(1)$ (first matricization of $\Omega$). Assume that each column of $\Omega(1)$ includes at least $l$ nonzero entries, where $l > 6 \log (n_1) + 2 \log \left( \frac{2k}{\epsilon} \right) + 4$. Then, with probability at least $1 - \frac{e}{n_1 k}$, every subset $\Omega''(1)$ of columns of $\Omega'(1)$ satisfies

$$m_2(\Omega''(1)) \geq t, \quad (2.30)$$

where $t$ is the number of columns of $\Omega''(1)$ and $m_2(\Omega''(1))$ is the number of nonzero rows of $\Omega''(1)$.

Proof. By setting $r = 0$ in inequalities (12) and (13) in [Pimentel-Alarcón et al., 2016b], we have

$$P(\mathcal{E}) < \sum_{n=1}^{\frac{n_1}{2}} \binom{n_1}{n} \left( \frac{n}{n_1} \right)^{ln_1+1} + \sum_{n=1}^{\frac{n_1}{2}} \left( \frac{n_1 - n}{n_1} \right)^{l(n_1-n+1)}. \quad (2.31)$$

Since $n < n_1$ it then follows that $P(\mathcal{E}) < \text{RHS of (2.31)} < \text{RHS of (2.26)} < \frac{e}{n_1 k}$, where the last inequality follows from Lemma 6.
The next theorem gives a lower bound on the number of needed samples to ensure finite completable for tensor $\mathbf{U}$. More specifically, we show under some mild assumptions if (2.32) holds, all conditions and assumptions in the statement of Theorem 3 hold with high probability, and therefore the tensor is finitely completable.

**Theorem 5.** Assume that $\sum_{i=2}^{d} (n_i r_i) < n_2 \ldots n_d$, $\sum_{i=2}^{d} r_i^2 \leq \Pi_{i=2}^{d} r_i$, $\Pi_{i=2}^{d} n_i \geq n_1 \Pi_{i=2}^{d} r_i - \sum_{i=2}^{d} r_i^2$, and Assumption $B_1$ hold. Furthermore, assume that each column of $\mathbf{U}^{(1)}$ includes at least $l$ observed entries, where

$$l > 6 \log(n_1) + 2 \max \left\{ \log \left( \frac{2 \sum_{i=2}^{d} r_i^2}{\epsilon} \right), \log \left( \frac{2 \Pi_{i=2}^{d} r_i - 2 \sum_{i=2}^{d} r_i^2}{\epsilon} \right) \right\} + 4. \quad (2.32)$$

Then, with probability at least $1 - \epsilon$, there exist only finitely many completions, given the rank components $r_2, \ldots, r_d$.

**Proof.** Since each column of $\mathbf{U}^{(1)}$ includes at least one nonzero entry, there exist $\sum_{i=2}^{d} n_i r_i$ entries in different columns such that they satisfy Assumption $A_1$ (as mentioned in Remark 2). Also, $\mathbf{U}^{(1)}$ has $\Pi_{i=2}^{d} n_i$ columns and by assumption $\Pi_{i=2}^{d} n_i \geq n_1 \Pi_{i=2}^{d} r_i - \sum_{i=2}^{d} r_i^2$, therefore there exist $\Pi_{i=2}^{d} r_i$ disjoint sets $\mathcal{S}_1, \ldots, \mathcal{S}_{\Pi_{i=2}^{d} r_i}$ such that each $\mathcal{S}_\ell$ consists of $n_1 - 1$ columns of $\mathbf{U}^{(1)}$ for $1 \leq \ell \leq \sum_{i=2}^{d} r_i^2$ and each $\mathcal{S}_\ell$ consists of $n_1$ columns of $\mathbf{U}^{(1)}$ for $\sum_{i=2}^{d} r_i^2 + 1 \leq \ell \leq \Pi_{i=2}^{d} r_i$.

Note that by (2.32), we have $l > 6 \log(n_1) + 2 \log \left( \frac{2 \sum_{i=2}^{d} r_i^2}{\epsilon} \right) + 4$. Using Lemma 6, with $k = 2 \sum_{i=2}^{d} r_i^2$, for each $1 \leq \ell \leq \sum_{i=2}^{d} r_i^2$, with probability at least $1 - \frac{\epsilon}{2 \sum_{i=2}^{d} r_i^2}$, (2.23) holds for every subset of columns of $\mathcal{S}_\ell$. Hence, with probability at least $1 - \frac{\epsilon}{k}$, (2.23) holds for every subset of columns of $\mathcal{S}_\ell$ for $1 \leq \ell \leq \sum_{i=2}^{d} r_i^2$, simultaneously. According to Remark 9 and Lemma 8 below and by setting $r = 1$, as a subset of columns $\mathcal{S}_\ell$ of $\mathbf{U}^{(1)}$ satisfies (2.23), there exist a subset of columns $\mathcal{S}_\ell$ of the constraint matrix $\hat{\mathbf{U}}$ (corresponding columns to the columns of $\mathcal{S}_\ell$) that satisfies (2.23) as well, for $1 \leq \ell \leq \sum_{i=2}^{d} r_i^2$. Denote $\mathcal{S}(0) = \cup_{\ell=1}^{\sum_{i=2}^{d} r_i^2} \mathcal{S}_\ell$.

Consider any subset of columns of $\mathcal{S}(0)$ and denote it by $\mathcal{S}'$. Let $\hat{\mathcal{S}}_\ell'$ denote the columns of $\hat{\mathcal{S}}'$ that belong in $\mathcal{S}_\ell$ and, without loss of generality, assume that $|\hat{\mathcal{S}}_1'| \geq \cdots \geq |\hat{\mathcal{S}}_{\sum_{i=2}^{d} r_i^2}'|$, where $|\hat{\mathcal{S}}_\ell'|$ denotes the number of columns of $\hat{\mathcal{S}}_\ell'$. As a result, we obtain

$$|\hat{\mathcal{S}}'| = \sum_{\ell=1}^{\sum_{i=2}^{d} r_i^2} |\hat{\mathcal{S}}_\ell'| \leq \left( \sum_{i=2}^{d} r_i^2 \right) |\hat{\mathcal{S}}_1'| \leq \left( \sum_{i=2}^{d} r_i^2 \right)^2 (m_2(\hat{\mathcal{S}}_1') - 1) \leq \left( \sum_{i=2}^{d} r_i^2 \right) (m_2(\hat{\mathcal{S}}') - 1), \quad (2.33)$$
and consequently

\[ |\tilde{S}'| \leq \left( \sum_{i=2}^{d} r_i^2 \right) m_2(\tilde{S}') - \sum_{i=2}^{d} r_i^2, \quad \forall \tilde{S}' \subseteq \tilde{S}^{(0)}. \tag{2.34} \]

Moreover, by (2.32), we have \( l > 6 \log (n_1) + 2 \log \left( \frac{2n_1^2d^2r_1^2 - 2d^2r_i^2}{\epsilon} \right) + 4 \). Using Lemma 7, with \( k = 2n_1^2d^2r_i^2 - 2d^2r_i^2 \), for each \( \sum_{i=2}^{d} r_i^2 + 1 \leq \ell \leq n_1^2d^2r_i^2 \), with probability at least \( 1 - \frac{\epsilon}{2} \), (2.30) holds for every subset of columns of \( S_\ell \). As a result, with probability at least 1, (2.30) holds for every subset of columns of \( S_\ell \) for \( \sum_{i=2}^{d} r_i^2 + 1 \leq \ell \leq n_1^2d^2r_i^2 \), simultaneously. According to Remark 9 and Lemma 8 below and by setting \( r = 0 \), as a subset of columns \( S_\ell \) of \( \Omega_{(1)} \) satisfies (2.30), the corresponding subset of columns \( \tilde{S}_\ell \) of the constraint matrix \( \bar{\Omega} \) satisfies (2.30) as well, \( \sum_{i=2}^{d} r_i^2 + 1 \leq \ell \leq n_1^2d^2r_i^2 \). Denote \( \tilde{S}^{(1)} = \cup_{\ell=\sum_{i=2}^{d} r_i^2+1}^{n_1^2d^2r_i^2} \tilde{S}_\ell \). Similarly, we have

\[ |\tilde{S}'| \leq \left( \prod_{i=2}^{d} r_i^2 - \sum_{i=2}^{d} r_i^2 \right) m_2(\tilde{S}'), \quad \forall \tilde{S}' \subseteq \tilde{S}^{(1)}. \tag{2.35} \]

For any subset of columns \( \bar{\Omega}' \) of \( \bar{\Omega}' = S^{(0)} \cup S^{(1)} \) we define \( S^{(0)'} \) and \( S^{(1)'} \) as the set of columns of \( \bar{\Omega}' \) that belong to \( \tilde{S}^{(0)} \) and \( \tilde{S}^{(1)} \), respectively. Observe that as \( \bar{\Omega}' \) has exactly \( \sum_{i=1}^{n_1^2d^2r_i^2} |S_i| = n_1^2d^2r_i^2 - \sum_{i=2}^{d} r_i^2 \) columns, condition (i) of Theorem 3 for \( j = 1 \) is satisfied. Then, by (2.34), (2.35) and the assumption that \( \sum_{i=2}^{d} r_i^2 \leq n_1^2d^2r_i^2 \), we have (recall \( |\bar{\Omega}'| \) denotes the number of columns of \( \bar{\Omega}' \) here)

\[ |\bar{\Omega}'| = |\tilde{S}^{(0)'}| + |\tilde{S}^{(1)'}| \leq \left( \sum_{i=2}^{d} r_i^2 \right) m_2(\tilde{S}^{(0)'} - \sum_{i=2}^{d} r_i^2 + \left( \prod_{i=2}^{d} r_i - \sum_{i=2}^{d} r_i^2 \right) m_2(\tilde{S}^{(1)'}) \tag{2.36} \]

\[ \leq \left( \prod_{i=2}^{d} r_i^2 \right) m_2(\tilde{\Omega}') - \sum_{i=2}^{d} r_i^2 \leq \left( \prod_{i=2}^{d} r_i \right) m_2(\tilde{\Omega}') - g_2(m_2(\tilde{\Omega}')). \]

Therefore, any subset of columns of \( \bar{\Omega}' \) satisfies (2.16) (condition (ii) of Theorem 3 for \( j = 1 \)), with probability at least \( 1 - \epsilon \). Then, according to Theorem 3, with probability at least \( 1 - \epsilon \), there are only finitely many completions that fit in the sampled tensor \( \mathcal{U} \), given the rank components \( r_2, \ldots, r_d \).

\[ \square \]

**Remark 7.** In the case that only a subset of rank components are given, e.g., \( r_{j+1}, \ldots, r_d \), we can treat the first \( j \) dimensions of the tensor as one single dimension, and therefore the above result is
still applicable. In particular, assume that each column of $\tilde{U}_{(j)}$ ($j$-th unfolding) includes at least $l$ observed entries, where (recall $N_j = n_1 \ldots n_j$)

$$l > 6 \log(N_j) + 2 \max \left\{ \log \left( \frac{2 \sum_{i=j+1}^d r_i^2}{\epsilon} \right), \log \left( \frac{2 \Pi_{i=j+1}^d r_i}{\epsilon} - 2 \sum_{i=j+1}^d r_i^2 \right) \right\} + 4.$$  

(2.37)

Given that $\sum_{i=j+1}^d r_i^2 \leq \Pi_{i=j+1}^d r_i$, $\Pi_{i=j+1}^d n_i \geq N_j \Pi_{i=j+1}^d r_i - \sum_{i=j+1}^d r_i^2$, and Assumption $B_j$ hold, with probability at least $1 - \epsilon$, there exist only finitely many completions, given the rank components $r_{j+1}, \ldots, r_d$.

Remark 8. Assume that there exist only finitely many completions, given the rank components $r_{j+1}, \ldots, r_d$, where $j \geq 1$. Then, there exist only finitely many completions of rank $(r_1, \ldots, r_d)$.

As in the proof of Theorem 5 we referred to Lemma 8, we need to show that if a subset of columns of $\Omega_{(1)}$ satisfies (2.23) or (2.30), the same property holds for a corresponding subset of columns of the constraint matrix $\tilde{\Omega}$.

Remark 9. Recall that for each column of $\Omega_{(1)}$ that has $k + r$ nonzero entries and $r$ of them have been used to obtain $T$, there exist $k$ columns in the constraint matrix $\tilde{\Omega}$, each with exactly $r + 1$ nonzero entries.

Recall that as we showed after stating Assumption $A_j$, we can choose $\sum_{i=2}^d n_i r_i$ nonzero entries of $\Omega_{(1)}$ to obtain $T$, i.e., to satisfy Assumption $A_1$ such that in each column of $\Omega_{(1)}$ either we choose one or zero nonzero entry. In that case, each column of $\tilde{\Omega}$ includes $0 + 1$ or $1 + 1$ nonzero entries.

As a result, in the proof of Theorem 5, we only need Lemma 8 for $r = 0$ and $r = 1$. However the general statement is needed to complete a proof in [Pimentel-Alarcón et al., 2016b], as we explain in Remark 11.

Consequently, the matrix $\tilde{\Omega}'$ defined in Lemma 8 for $r = 0$ and $r = 1$ corresponds to the columns of the constraint matrix $\tilde{\Omega}$.

Lemma 8. Let $r$ be a given nonnegative integer. Assume that there exists an $n_1 \times (n_1 - r)$ matrix $\Omega'$ composed of $n_1 - r$ columns of $\Omega_{(1)}$ such that each column of $\Omega'$ includes at least $r + 1$ nonzero entries and satisfies the following property.\textsuperscript{2}

\textsuperscript{2}This property is exactly property (ii) in [Pimentel-Alarcón et al., 2016b].
• Denote an $n_1 \times t$ matrix (for any $1 \leq t \leq n_1 - r$) composed of any $t$ columns of $\Omega'$ by $\Omega''$. Then

$$m_2(\Omega'') - r \geq t,$$

where $m_2(\Omega'')$ is the number of nonzero rows of $\Omega''$.

Then, there exists an $n_1 \times (n_1 - r)$ matrix $\hat{\Omega}'$ such that: each column has exactly $r + 1$ entries equal to one, and if $\hat{\Omega}'(x,y) = 1$ then we have $\Omega'(x,y) = 1$. Moreover, $\hat{\Omega}'$ satisfies the above-mentioned property.

Proof. Let $\Omega'$ denote a $n_1 \times (n_1 - r)$ matrix that consists of $n_1 - r$ columns of $\Omega_{(1)}$ and satisfies the mentioned property in the statement of the lemma. Note that the given property in the statement of the lemma is equivalent to the following statement

• The matrix obtained by choosing any subset of the columns of $\Omega'$ includes at least $r + t$ nonzero rows, where $t$ is the number of the selected columns of $\Omega'$.

Consider a bipartite graph $G$ with the two sets of nodes $T_1$ with $n_1 - r$ nodes corresponding to the columns and $T_2$ with $n_1$ nodes corresponding to the rows. We add an edge between the $i$-th node of $T_1$ and the $i'$-th node of $T_2$ if and only if the $(i,i')$-th entry of $\Omega'$ is equal to one. For a set of nodes $S$, let $N(S)$ denote the set of their neighbors. Note that according to the assumption, any subset of the columns of $\Omega'$ includes at least $r + t$ nonzero rows ($r + t$ edges in the defined graph), where $t$ is the number of the selected columns of $\Omega_r$. Therefore, every subset of nodes of $T_1$ has the following property

$$N(S) \geq |S| + r.$$

The statement of the lemma is equivalent to proving that there exists a spanning subgraph $G'$ of $G$ such that every node of $T_1$ is of degree $r + 1$, and also for each subset of nodes of $T_1$, (2.39) holds in $G'$.

First consider the scenario that $r = 0$. Then, according to Hall’s theorem [Kierstead, 1983], there is a perfect matching (since inequality (2.39) holds). Observe that a perfect matching in the described graph is exactly equivalent to what we are looking for when $r = 0$. For other values of $r > 0$, the proof is a direct result of Theorem 6 which is a generalization of the Hall’s theorem.
CHAPTER 2. FUNDAMENTAL LIMITS FOR LOW-TUCKER-RANK TENSOR COMPLETION PROBLEM

An example of such spanning subgraph for a graph with $n_1 = 6$ and $r = 1$ is shown in Figure 2.2. In Figure 2.2(a) a bipartite graph is given such that (2.39) holds for $r = 1$. Figure 2.2(b) gives an spanning subgraph of the given graph in Figure 2.2(a) such that degree of each node in $T_1$ is equal to 2 and still (2.39) holds for $r = 1$.

![Figure 2.2: An example of the corresponding bipartite graph with $r = 1$ and $n_1 = 6$.](image)

**Theorem 6.** (Generalized Hall’s Theorem) Consider a bipartite graph $G$ with two sets of nodes, $T_1$ with $x$ nodes and $T_2$ with $x + r$ nodes, where $x, r \in \mathbb{N}$. Suppose that for each subset $S$ of $T_1$ the following inequality holds

$$|N(S)| \geq |S| + r. \tag{2.40}$$

Then, there exists a spanning subgraph $G'$ of $G$ such that every node of $T_1$ is of degree $r + 1$, and also for each subset of nodes of $T_1$, the inequality (2.40) holds in $G'$ as well.

**Proof.** The proof is given in Appendix 2.9. \hfill \Box

**Corollary 2.** Assume that $\sum_{i=j+1}^d r_i^2 \leq \prod_{i=j+1}^d r_i$, $\prod_{i=j+1}^d n_i \geq N_j \prod_{i=j+1}^d r_i - \sum_{i=j+1}^d r_i^2$, and Assumption $B_j$ hold. Furthermore, assume that we observe each entry of $U$ with probability $p$, where

$$p > \frac{1}{N_j} \left( 6 \log (N_j) + 2 \log \left( \max \left\{ \frac{2 \sum_{i=j+1}^d r_i^2}{\epsilon}, \frac{2 \prod_{i=j+1}^d r_i - 2 \sum_{i=j+1}^d r_i^2}{\epsilon} \right\} \right) + 4 \right) + \frac{1}{\sqrt{N_j}} \tag{2.41}$$
Then, with probability at least $1 - 2\epsilon$, there are only finitely many completions for $\mathcal{U}$, given the rank components $r_{j+1}, \ldots, r_d$.

**Proof.** The proof is straightforward given Theorem 5 and Corollary 1.

**Remark 10.** In Theorem 4 which is obtained by applying the analysis in [Pimentel-Alarcón et al., 2016b], the number of needed samples per column of the $i$-th matricization to guarantee finite completability with high probability is on the order of $\mathcal{O}(\max\{\log(n_i), r_i\})$. Therefore, the number of needed samples in total is on the order of $\mathcal{O}(n_i \max\{\log(n_i), r_i\})$. On the other hand, in Theorem 5 for the case of $d > 2$ (order of the tensor is at least 3), the number of needed samples per column for the $i$-th unfolding becomes $\mathcal{O}(\max\{\log(n_i), \log(r_{i+1} \ldots r_d)\})$, and the number of needed samples in total is on the order of $\mathcal{O}(N_i \max\{\log(n_i), \log(r_{i+1} \ldots r_d)\})$, if $r_{i+1} \ldots r_d < N_i^{d-j}$. As an example, consider the case that $n_i = n, i = 1, \ldots, d$. Also, assume that $r_i = r, i = j + 1, \ldots, d$, where $r \leq \sqrt{n}$. The matricization analysis requires $\mathcal{O}(n^{d-1} \max\{\log(n), r\})$ samples for any $j$. However, according to Theorem 5, if $j = \frac{d}{3}$ it requires $\mathcal{O}(n^{\frac{2d}{3}} \log(n^{\frac{d}{3}}))$ samples for finite completability. Note that under the assumption $r_{j+1} \ldots r_d < \frac{N_i^n}{N_i^{d-j}}, i.e., r^{d-j} < n^{d-2j}, j = \frac{d}{3}$ results in the best possible bound as $r = \sqrt{n}$ since increasing $j$ violates the assumption.

**Remark 11.** We note that the proof in [Pimentel-Alarcón et al., 2016b, Theorem 3] is incomplete and Lemma 8 is needed to complete that proof.

First, observe that sampling pattern matrix $\Omega$ represents the observed entries and it has at least $l$ nonzero elements per column. On the other hand, $\tilde{\Omega}$ represents the constraint matrix defined in [Pimentel-Alarcón et al., 2016b] and it has exactly $(r + 1)$ nonzero elements per column, where $r$ is the given rank.

Secondly, observe that in [Pimentel-Alarcón et al., 2016b, Lemma 1] each $\hat{\Omega}_\tau$ is a subset of the constraint matrix $\hat{\Omega}$ (not a subset of $\Omega$).

Finally, in the proof and statement of [Pimentel-Alarcón et al., 2016b, Lemma 9] the authors consider a subset of $\Omega$. However, in the proof of [Pimentel-Alarcón et al., 2016b, Theorem 3] they refer to [Pimentel-Alarcón et al., 2016b, Lemma 1], which considers the columns as the subsets of $\hat{\Omega}$. Now, considering $\hat{\Omega}$ instead of $\Omega$ in [Pimentel-Alarcón et al., 2016b, Lemma 9], the equation below the equation (11) in [Pimentel-Alarcón et al., 2016b] does not hold. This is because the assumption right below equation (11) in [Pimentel-Alarcón et al., 2016b] that $\tilde{\Omega}$ has at least $l$
nonzero elements per column is incorrect (this is a property of \( \Omega \) but not necessarily \( \hat{\Omega} \)).

As a result, a part of the proof in [Pimentel-Alarcón et al., 2016b, Theorem 3] is missing which is Lemma 8 in this chapter.

\[ \square \]

2.5 Deterministic and Probabilistic Guarantees for Unique Completability

In this section, we first provide an example with exactly two completions to emphasize that finite completability does not necessarily result in unique completability. Then, we provide the conditions on the sampling pattern to guarantee unique completability.

**Example 5.** Assume that the sampled matrix \( U \in \mathbb{R}^{5 \times 4} \) is given as the incomplete matrix on the left below:

\[
\begin{array}{ccc}
1 & 1 & -\frac{1}{2} \\
-4 & 2 & -1 \\
0 & 1 & 2 \\
1 & 4 & \\
4 & -2 & \frac{3}{2}
\end{array}
\Rightarrow
\begin{array}{ccc}
1 & -\frac{21}{32} & 1 & -\frac{1}{2} \\
-4 & 2 & -1 & \frac{3}{4} \\
0 & 1 & \frac{24}{5} & 2 \\
1 & -\frac{41}{32} & 4 & -\frac{7}{4} \\
-8 & 4 & -2 & \frac{3}{2}
\end{array}
\text{and}
\begin{array}{ccc}
1 & -2 & 1 & -\frac{1}{2} \\
-4 & 2 & -1 & -10 \\
0 & 1 & -\frac{1}{2} & 2 \\
1 & -8 & 4 & -\frac{25}{2} \\
-\frac{39}{27} & 4 & -2 & \frac{3}{2}
\end{array}
\]

Moreover, assume that \( \text{rank}(U) = 2 \). In Appendix 2.10, it is shown that there exist exactly two completions of \( U \) as given by the two complete matrices on the right.

The following assumptions is a stronger version of Assumption \( A_j \) to ensure that there exists only one tuple \( T \) given the core tensor.

**Assumption \( A_j^+ \):** Anywhere that this assumption is stated, there exist \( \sum_{i=j+1}^d (n_i (r_i + 1)) \) observed entries such that for any \( S_i \subseteq \{1, \ldots, n_i\} \) for \( i \in \{j+1, \ldots, d\} \), \( U^{(S_{j+1}, \ldots, S_d)} \) includes at most \( \sum_{i=j+1}^d |S_i|(r_i + 1) \) of the mentioned \( \sum_{i=j+1}^d (n_i (r_i + 1)) \) observed entries.

**Remark 12.** In Lemma 1 we showed that Assumption \( A_j \) results that tuple \( T \) can be determined finitely. Note that Assumption \( A_j^+ \) implies Assumption \( A_j \) and therefore using only \( \sum_{i=j+1}^d (n_i r_i) \) of the observed entries tuple \( T \) can be determined finitely. Moreover, observe that the rest of the sampled entries result in a set of polynomials that involve any of variables at least once and, tuple \( T \) can be determined uniquely, with probability one.

\[ \square \]
The following lemma is a re-statement of Lemma 25 in [Ashraphijuo and Wang, 2017a] (also an adaptation of Lemma 7 in [Pimentel-Alarcón et al., 2016b]) and gives the conditions under which a subset of entries of the core tensor $C$ can be determined uniquely.

**Lemma 9.** Assume that Assumption $B_j$ holds. Suppose that $\tilde{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times t}$ is a subtensor of the constraint tensor such that $\mathcal{P}(\tilde{\Omega}')$ is minimally algebraically dependent. Then, for almost every $\mathcal{U}$, all variables that are involved in the set of polynomials $\mathcal{P}(\tilde{\Omega}')$ can be uniquely determined.

**Proof.** According to Lemma 3, the number of involved variables in polynomials in $\mathcal{P}(\tilde{\Omega}') = \{p_1, p_2, \ldots, p_t\}$ is $t - 1$ and are denoted by $\{x_1, \ldots, x_{t-1}\}$. Moreover, as mentioned in the proof of Lemma 3, $\mathcal{P}(\tilde{\Omega}') \setminus p_i$ is a set of algebraically independent polynomials and the number of involved variables is $t - 1$, $i = 1, \ldots, t$. Consider an arbitrary variable $x_1$ that is involved in polynomials in $\mathcal{P}(\tilde{\Omega}')$ and without loss of generality, assume that $x_1$ is involved in $p_1$.

On the other hand, as mentioned all variables $\{x_1, \ldots, x_{t-1}\}$ are involved in algebraically independent polynomials in $\mathcal{P}(\tilde{\Omega}') \setminus p_1$. Hence, tuple $(x_1, \ldots, x_{t-1})$ can be determined finitely. Given that one of these finitely many tuples $(x_1, \ldots, x_{t-1})$ satisfy polynomial equation $p_1$, with probability one there does not exist another tuple $(x_1, \ldots, x_{t-1})$ among them that satisfies polynomial equation $p_1$. This is because with probability zero a tuple satisfy a polynomial equation in which the coefficients are chosen generically, and also the fact that the number of such tuples is finite.

The next theorem provides two conditions on the sampling pattern to ensure unique completability. The first one is as the same as the condition proposed in Theorem 3 for finite completability. The second condition takes advantage of finite completability and leads to the unique core tensor.

**Theorem 7.** Suppose that assumptions $A_j^+$ and $B_j$ hold. Also, assume that there exist two disjoint subtensors $\tilde{\Omega}' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times n}$ and $\tilde{\Omega}'_0 \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times n_0}$ of the constraint tensor such that $n = \left(\prod_{i=1}^{d} n_i\right) \left(\prod_{i=j+1}^{d} m_i\right) - \left(\sum_{i=j+1}^{d} r_i^2\right)$ and $n_0 = \left(\prod_{i=1}^{d} n_i\right) - \left[\sum_{i=j+1}^{d} r_i^2\prod_{i=j+1}^{d} m_i\right]$ with the following conditions:

(i) For any $t \in \{1, \ldots, n\}$ and any subtensor $\tilde{\Omega}'' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times t}$ of the tensor $\tilde{\Omega}'$, the following inequality holds

\[ \left(\prod_{i=j+1}^{d} m_i\right) m_{j+1}(\tilde{\Omega}'') - g_{j+1}(m_{j+1}(\tilde{\Omega}'')) \geq t. \]  

\[ (2.42) \]
(ii) For any \( t' \in \{1, \ldots, n_0\} \) and any subtensor \( \mathbf{\tilde{\Omega}}_0'' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times t'} \) of the tensor \( \mathbf{\tilde{\Omega}}_0' \), the following inequality holds

\[
\left( \prod_{i=j+1}^{d} r_i \right) m_{j+1}(\mathbf{\tilde{\Omega}}_0'') - g_{j+1}(m_{j+1}(\mathbf{\tilde{\Omega}}_0'')) \geq \left( \prod_{i=j+1}^{d} r_i \right) t' - \left( \sum_{i=j+1}^{d} r_i^2 \right) (t' - n_0 + 1)^+ \quad (2.43)
\]

Then, for almost every \( \mathcal{U} \) with probability one, there exists exactly one tensor that fits in the sampled tensor, and also \( \text{rank}(\mathcal{U}_{(i)}) = r_i, \ i = j + 1, \ldots, d \). Therefore there is a unique completion of the sampled tensor with rank of \( (r_1, r_2, \ldots, r_d) \).

**Proof.** As mentioned Assumption \( A_j^+ \) results that \( \mathcal{T} \) can be determined uniquely. In order to complete the proof it suffices to show the core tensor \( \mathcal{C} \) can be determined uniquely as well. As defined earlier, \( \mathcal{P}(\mathbf{\tilde{\Omega}}') \) and \( \mathcal{P}(\mathbf{\tilde{\Omega}}_0') \) denote the set of polynomials obtained from the sampled entries corresponding to \( \mathbf{\tilde{\Omega}}' \) and \( \mathbf{\tilde{\Omega}}_0' \), respectively. According to Theorem 3, \( \mathcal{P}(\mathbf{\tilde{\Omega}}') \) results in finite completability of the sampled tensor \( \mathcal{U} \), and also \( n \) algebraically independent polynomials. Hence, adding any of the polynomials in \( \mathcal{P}(\mathbf{\tilde{\Omega}}_0') \) to \( \mathcal{P}(\mathbf{\tilde{\Omega}}') \) results in an algebraically dependent set of polynomials. Using \( n \) algebraically independent polynomials in \( \mathcal{P}(\mathbf{\tilde{\Omega}}') \) and polynomials in \( \mathcal{P}(\mathbf{\tilde{\Omega}}_0') \), we show all entries of core tensor can be determined uniquely.

Let \( p_0 \) denote an arbitrary polynomial in \( \mathcal{P}(\mathbf{\tilde{\Omega}}_0') \). Consider \( \mathcal{P}'(p_0) \subset \mathcal{P}(\mathbf{\tilde{\Omega}}') \) such that \( \mathcal{P}'(p_0) \cup p_0 \) is a minimally algebraically dependent set of polynomials. Let tuple \((x_1, \ldots, x_j)\) denote the first \( j \) coordinates of the corresponding observed entry to polynomial \( p_0 \). Then, according to Note 2, all \( \prod_{i=j+1}^{d} r_i \) entries of core tensor \( \mathcal{C} \) with the first \( j \) coordinates as \((x_1, \ldots, x_j)\) are involved in polynomial \( p_0 \), and therefore they can be determined uniquely, according to Lemma 9. However, a subset of these \( \prod_{i=j+1}^{d} r_i \) entries of core tensor may be the elements of the proper structure which are known entries.

Similarly, repeating this procedure for any subtensor \( \mathbf{\tilde{\Omega}}_0''' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times t'} \) of the tensor \( \mathbf{\tilde{\Omega}}_0' \) results in \( \left( \prod_{i=j+1}^{d} r_i \right) t' \) polynomials in terms of the \( \left( \prod_{i=j+1}^{d} r_i \right) m_{j+1}(\mathbf{\tilde{\Omega}}_0''') - g_{j+1}(m_{j+1}(\mathbf{\tilde{\Omega}}_0''')) \) unknown entries of the core tensor. These polynomials are algebraically independent if

\[
\left( \prod_{i=j+1}^{d} r_i \right) m_{j+1}(\mathbf{\tilde{\Omega}}_0''') - g_{j+1}(m_{j+1}(\mathbf{\tilde{\Omega}}_0''')) \geq \left( \prod_{i=j+1}^{d} r_i \right) t',
\]

for any subtensor \( \mathbf{\tilde{\Omega}}_0''' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times t'} \) of the tensor \( \mathbf{\tilde{\Omega}}_0' \). In order to include \( \text{dim}(\mathcal{C}) = \left( \prod_{i=1}^{d} r_i \right) - \left( \sum_{i=j+1}^{d} r_i^2 \right) \) algebraically independent polynomials, \( \mathbf{\tilde{\Omega}}_0' \) should include at least \( n_0 \) columns otherwise for any subtensor \( \mathbf{\tilde{\Omega}}_0'' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times t'} \) of the tensor \( \mathbf{\tilde{\Omega}}_0' \) we have \( \text{dim}(\mathcal{C}) > \left( \prod_{i=j+1}^{d} r_i \right) t' \).
CHAPTER 2. FUNDAMENTAL LIMITS FOR LOW-TUCKER-RANK TENSOR COMPLETION PROBLEM

since $t' < n_0$. Therefore, inequality (2.43) results that all $n$ variables (unknown entries) of the core tensor can be determined uniquely.

We are also interested to obtain a lower bound on the sampling probability, which ensures the unique completability. The following lemma leads to Lemma 11 that characterizes the number of required sampled entries to ensure the unique completability with high probability.

**Lemma 10.** Consider an arbitrary set $\Omega'_{(1)}$ of $n_1$ columns of $\Omega_{(1)}$ (first matricization of $\Omega$). Assume that each column of $\Omega_{(1)}$ includes at least $l$ nonzero entries, where $l > 6 \log(n_1) + 2 \log\left(\frac{n_1^k}{\epsilon}\right) + 4$. Then, with probability at least $1 - \frac{\epsilon}{k}$ every proper subset $\Omega''_{(1)}$ of columns of $\Omega'_{(1)}$ satisfies

$$m_2(\Omega''_{(1)}) - 1 \geq t,$$

(2.45)

where $t$ is the number of columns of $\Omega''_{(1)}$ and $m_2(\Omega''_{(1)})$ is the number of nonzero rows of $\Omega''_{(1)}$.

**Proof.** According to Lemma 6, every subset $\Omega''_{(1)}$ of columns of $\Omega'_{(1)}$ with $n_1 - 1$ columns satisfies property (2.45) with probability at least $1 - \frac{\epsilon}{n_1k}$. There are exactly $n_1$ subsets with $n_1 - 1$ columns, and therefore property (2.45) holds for all of them with probability at least $1 - \frac{\epsilon}{k}$.

**Lemma 11.** Assume that $\sum_{i=2}^d (n_i r_i) < n_2 \ldots n_d$, $\sum_{i=2}^d r_i^2 \leq \Pi_{i=2}^d r_i$, $\Pi_{i=2}^d n_i \geq n_1 (\Pi_{i=2}^d r_i + 1) - \sum_{i=2}^d r_i^2$, and Assumption $B_1$ hold. Furthermore, assume that each column of $U_{(1)}$ includes at least $l$ observed entries, where

$$l > 6 \log(n_1) + 2 \max \left\{ \log\left(\frac{\sum_{i=2}^d r_i^2}{\epsilon}\right), \log\left(\frac{\Pi_{i=2}^d r_i - \sum_{i=2}^d r_i^2}{\epsilon}\right), \log\left(\frac{n_1}{\epsilon}\right) \right\} + 8.$$  

(2.46)

Then, with probability at least $1 - \epsilon$, there exists only one completion of rank $(r_1, \ldots, r_d)$.

**Proof.** Observe that since $\Pi_{i=2}^d n_i \geq n_1 (\Pi_{i=2}^d r_i + 1) - \sum_{i=2}^d r_i^2$ there exist $n_1 (\Pi_{i=2}^d r_i) - \left(\sum_{i=2}^d r_i^2\right)$ and $n_1$ disjoint columns in $\Omega_{(1)}$ denoted by $\check{\Omega}'$ and $\check{\Omega}'_0$, respectively. In order to show unique completability, it suffices to show $\check{\Omega}'$ and $\check{\Omega}'_0$ satisfy properties (i) and (ii) in the statement of Theorem 7, respectively, with probability at least $1 - \frac{\epsilon}{k}$.

Note that having (2.46), it is easy to see that (2.32) holds with $\epsilon$ replaced by $\frac{\epsilon}{k}$. Therefore, according to Theorem 5, $\check{\Omega}'$ satisfies property (i) in the statement of Theorem 7 with probability at least $1 - \frac{\epsilon}{k}$.
On the other hand, \( \sum_{i=2}^{d} r_i^2 \leq \Pi_{i=2}^{d} r_i \), and according to Lemma 10 (with \( k = 2 \)), with probability at least \( 1 - \frac{\epsilon}{2} \), any subset of columns of \( \tilde{\Omega}_0' \) satisfies (2.45). For any proper subset of column \( \tilde{\Omega}_0'' \) of \( \tilde{\Omega}_0' \), we have \( m_2(\tilde{\Omega}_0') - 1 \geq t' \) or equivalently \( (t' = \text{the number of columns of } \tilde{\Omega}_0') \)

\[
\left( \Pi_{i=2}^{d} r_i \right) m_{j+1}(\tilde{\Omega}_0') - \Pi_{i=2}^{d} r_i \geq \left( \Pi_{i=2}^{d} r_i \right) t',
\]

which results in (2.43) as \( g_{j+1}(m_{j+1}(\tilde{\Omega}_0')) \leq \sum_{i=2}^{d} r_i^2 \leq \Pi_{i=2}^{d} r_i \) and \( (t' - n_0 + 1)^+ = 0 \). Hence, in order to show property (ii) in the statement of Theorem 7 holds with probability at least \( 1 - \frac{\epsilon}{2} \) (which completes the proof) we only need to show that (2.43) holds when \( \tilde{\Omega}_0'' = \tilde{\Omega}_0' \).

Note that \( m_{j+1}(\tilde{\Omega}_0') \leq m_{j+1}(\tilde{\Omega}_0') \). Therefore, as \( m_2(\tilde{\Omega}_0') - 1 \geq t' \) holds for any proper subset of column \( \tilde{\Omega}_0'' \) of \( \tilde{\Omega}_0' \), we have \( m_{j+1}(\tilde{\Omega}_0') \geq n_1 \) (by setting \( t' = n_1 - 1 \)). Then, (2.43) holds as \( g_{j+1}(m_{j+1}(\tilde{\Omega}_0')) \leq \sum_{i=2}^{d} r_i^2 \) and \( (t' - n_0 + 1)^+ = 1. \)

Finally, in the following, we use Corollary 1 and the above lemma to propose a lower bound on the sampling probability to guarantee the unique completability with high probability.

**Corollary 3.** Assume that \( \sum_{i=2}^{d} r_i^2 \leq \Pi_{i=2}^{d} r_i \), \( \Pi_{i=2}^{d} n_i \geq N_j(\Pi_{i=2}^{d} r_i + 1) - \sum_{i=2}^{d} r_i^2 \), and Assumption B_j hold. Furthermore, assume that we observe each entry of the tensor \( U \) with probability \( p \), where

\[
p > \frac{1}{N_j} \left( 6 \log (N_j) + 2 \log \left( \max \left\{ \frac{\sum_{i=2}^{d} r_i^2}{\epsilon}, \frac{\Pi_{i=2}^{d} r_i - \sum_{i=2}^{d} r_i^2}{\epsilon}, \frac{N_j}{\epsilon} \right\} \right) + 8 \right) + \frac{1}{\sqrt{N_j}} \tag{2.48}
\]

Then, with probability at least \( 1 - 2\epsilon \), there exists only one completion for \( U \), given the rank components \( r_{j+1}, \ldots, r_d \).

**Proof.** The proof is straightforward given Lemma 11 and Corollary 1. \( \square \)

### 2.6 Numerical Comparisons

In order to show the advantage of our proposed method over matrix analysis, we compare the lower bound on the sampling probability obtained by matricization analysis with the bound obtained by tensor analysis.

In the first example, numerical comparisons are performed on a 4th-order tensor \( U \in \mathbb{R}^{900 \times 900 \times 900 \times 900} \). Figure 2.3 plots the bounds given in (2.21) (Grassmannian manifold) and (2.41) (Tucker manifold).
Figure 2.3: Lower bounds on sampling probability for a 4th-order tensor.

for finite completability, where $r_1 = \cdots = r_4 = r$, and $\epsilon = 0.0001$. As the second example, we consider a 6th-order tensor $\mathbf{U} \in \mathbb{R}^{900 \times 900 \times 900 \times 900 \times 900 \times 900}$. Figure 2.4 plots the bounds given in (2.21) (Grassmannian manifold) and (2.41) (Tucker manifold) for finite completability, where $r_3 = r_4 = r_5 = r_4 = r$, and $\epsilon = 0.0001$. In this example, a significant reduction in the sampling probability is seen for the tensor model.

We have the following observations:

- In Figure 2.3, the bound obtained through the analysis on Tucker manifold is less than the bound obtained through the low-order analysis on Grassmannian manifold. This improvement significantly increases as the value of the rank increases.

- In Figure 2.3, the restriction $r \leq \frac{n}{6}$ in the analysis on Grassmannian manifold makes the bound valid only for $r \leq 150$ (and that is why the corresponding curve is dotted for $r > 150$ in Figure 2.3). On the other hand, the restriction $r(d-1) \leq n$ in our proposed approach ensures the validity of the bound for $r \leq 300$. Similarly, in Figure 2.4, the restriction $\Pi_{i=j+1}^d n_i \geq N_j \Pi_{i=j+1}^d r_i - \sum_{i=j+1}^d r_i^2$ in the analysis on Tucker manifold makes the bound valid only for $r \leq 30$. 
CHAPTER 2. FUNDAMENTAL LIMITS FOR LOW-TUCKER-RANK TENSOR COMPLETION PROBLEM

Figure 2.4: Lower bounds on sampling probability for a 6th-order tensor.

- Since the bounds for finite completability (2.48) and unique completability (2.41) result in almost the same curves in the above examples, we only plot bounds for finite completability in the figures. In general, the main difference between (2.41) and (2.48) is an additional term \( \log \left( \frac{N_j}{\epsilon} \right) \) in the second term inside of maximum operator.

2.7 Summary

This chapter aims to find fundamental conditions on the sampling pattern for finite completability of a low-rank partially sampled tensor. To solve this problem, a novel geometric approach on Tucker manifold is proposed. Specifically, a set of polynomials based on the location of the sampled entries are first defined, and then a relationship between a geometric pattern on the sampled entries and the number of algebraically independent polynomials among all of the defined polynomials is characterized. Moreover, an extension to Hall’s theorem in graph theory is provided which is key to obtaining probabilistic conditions for finite and unique completabilities. Using these developed tools, we have addressed three problems in this chapter: (i) Characterizing the necessary and sufficient conditions on the sampling pattern to have finitely many tensor completions for the
CHAPTER 2. FUNDAMENTAL LIMITS FOR LOW-TUCKER-RANK TENSOR COMPLETION PROBLEM

47
given rank, (ii) Characterizing conditions on the sampling pattern to ensure that there is exactly
one completion for the given rank, (iii) Lower bounding the sampling probability such that the
conditions in Problems (i) and (ii) are satisfied with high probability.

Finally, through numerical examples it is seen that our proposed analysis on Tucker manifold
for finite tensor completableility leads to a much lower sampling rate than the matricization approach
that is based on analysis on Grassmannian manifold.

2.8 Proof of Lemma 4

As mentioned before, finiteness of the number of completions given \((r_{j+1}, \ldots, r_d)\) results finiteness
of the number of completions given \((r_j, \ldots, r_d)\). We are interested to show this statement holds in
Theorem 3, i.e, satisfying the properties (i) and (ii) for \(j\) in the statement of Theorem 3 results in
satisfying the properties (i) and (ii) for \(j - 1\). Before we present this result, the following notation
is introduced for the simplicity.

Definition 6. Let set \(S\) be a set of \(d\)-tuples such that each element of \(S\) consists of \(d\) natural
numbers \((x_1, \ldots, x_d)\). Define \(f_j(S)\) as the number of different tuples of the first \(j\) components of
the elements of \(S\).

For example, let \(S = \{(2, 2, 1), (2, 3, 3), (2, 2, 2), (1, 2, 1), (2, 3, 1)\}\). Then, according to the defi-
nition \(f_2(S) = |\{(2, 2), (2, 3), (1, 2)\}| = 3\).

Now, we are ready to provide the proof of lemma. It is easy to see that Lemma 4 is exactly
equivalent to the following statement:

Assume that we sample a tensor \(U \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}\). Suppose that there exist
\(n = \left(\prod_{i=1}^{d} n_i\right) \left(\prod_{i=j+1}^{d} r_i\right) - \left(\sum_{i=j+1}^{d} r_i^2\right)\) of the sampled entries such that any subset of them
(called \(S\)) with \(t\) observed entires (for any \(1 \leq t \leq n\)) satisfies the following inequality

\[
\left(\prod_{i=j+1}^{d} r_i\right) f_j(S) - g_{j+1}(f_j(S)) \geq t. \tag{2.49}
\]

Then, there exist \(n' = \left(\prod_{i=1}^{j-1} n_i\right) \left(\prod_{i=j}^{d} r_i\right) - \left(\sum_{i=j}^{d} r_i^2\right)\) of the sampled entries such that any
subset of them (called \(S'\)) with \(t\) observed entires (for any \(1 \leq t \leq n'\)) satisfies the following
inequality

\[
\left(\prod_{i=j}^{d} r_i\right) f_{j-1}(S') - g_j(f_{j-1}(S')) \geq t. \tag{2.50}
\]
Therefore, it suffices to show the above statement. For \( r_j = 1 \), the statement can be easily verified, and therefore we consider the case of \( r_j > 1 \) in the rest of proof. Also recall the assumption \( n_j > \sum_{i=1}^{d} r_i \). We partition all of the sampled entries into \( n_j \) groups such that group \( i \) includes all of the sampled entries that \( j \)-th component of their coordinate is equal to \( i \). Let \( S_i \) denote the \( i \)-th group. Observe that every subset of the sampled entries of each group satisfies inequality (2.49). Moreover, according to pigeonhole principle we know that between the defined \( n_j \) groups there exist \( r_j \) groups (without loss of generality assume \( S_1, \ldots, S_{r_j} \)) such that \( \sum_{i=1}^{r_j} |S_i| \geq \frac{r_j}{n_j} n \).

Since \( \frac{r_j}{n_j} n = \left( \prod_{i=1}^{j-1} n_i \right) \left( \prod_{i=j}^{d} r_i \right) - \frac{r_j}{n_j} \left( \sum_{i=j+1}^{d} r_i^2 \right) \geq n' \), we only need to show that any subset of the elements of the groups \( S_1, \ldots, S_{r_j} \) satisfies inequality (2.50). Consider an arbitrary subset of the elements of groups \( S_1, \ldots, S_{r_j} \) and denote it by \( S' \). Assume that \( S'_i \) denotes the elements of \( S' \) that belong in \( S_i \) for \( 1 \leq i \leq r_j \). Recall that every subset of the sampled entries of each group satisfies inequality (2.49). Moreover, due to the fact that in each group the \( j \)-th component of the coordinates are the same, for each subset of the entries of a group \( S_i \) we have \( f_j(S'_i) = f_{j-1}(S'_i) \).

As a result, we have

\[
\left( \prod_{i=j+1}^{d} r_i \right) f_{j-1}(S'_i) - g_{j+1}(f_{j-1}(S'_i)) \geq |S'_i|.
\] (2.51)

Without loss of generality, assume \( f_j(S'_1) \geq f_j(S'_2) \geq \ldots \geq f_j(S'_{r_j}) \). Also, observe that for each \( 1 \leq i \leq r_j \) we have \( f_j(S'_i) \leq f_j(S') \). Therefore, by adding up the above inequalities for \( 1 \leq i \leq r_j \) we have

\[
\left( \prod_{i=j}^{d} r_i \right) f_{j-1}(S'_1) - g_{j+1}(f_{j-1}(S'_1))r_j \geq |S'|.
\] (2.52)

For the case that \( 1 < r_j < \sum_{i=j+1}^{d} r_i^2 \) holds, we can see that \( r_j \left( \sum_{i=j+1}^{d} r_i^2 \right) \geq \sum_{i=j}^{d} r_i^2 \) also holds, and therefore using (2.52) we can obtain

\[
\left( \prod_{i=j}^{d} r_i \right) f_{j-1}(S') - g_j(f_{j-1}(S')) \geq |S'|,
\] (2.53)

which completes the proof. If \( r_j \geq \sum_{i=j+1}^{d} r_i^2 \) similarly using the above inequality and by ignoring \( r_j^2 \) of entries proof can be completed.

### 2.9 Proof of Theorem 6

The proof of this theorem is based on strong induction on \( x \). For \( x = 1 \) the theorem is easy to verify. Assume that the statement of the theorem holds for \( 1 \leq x \leq x_0 \). Then, it suffices to show
that the statement also holds for the case that $x = x_0 + 1$. There are two different scenarios that we need to show the statement holds for separately.

**Scenario 1.** There exists a proper and nonempty subset of nodes $S_1$ of $T_1$ such that $|N(S_1)| = |S_1| + r$:

Consider the induced subgraph of $G$ with the set of nodes $S_1 \cup N(S_1)$ and denote it by $G_1$. Induction hypothesis results in a spanning subgraph $G'_1$ of $G_1$ that satisfies the desired properties in the statement of the theorem for the corresponding subgraph $G_1$.

Now, consider the induced subgraph of $G$ with the set of nodes $\{T_1 \cup T_2\} \setminus \{S_1\}$ and denote it by $G_2$. Since $|N(S_1)| = |S_1| + r$, we conclude that for each subset of nodes of $\{T_1\} \setminus \{S_1\}$ we have $|N(S) \cap \{T_2 \setminus N(S_1)\}| \geq |S|$. The reason is that if we choose a set of nodes including the members of $S \subset \{T_1\} \setminus \{S_1\}$ plus all nodes in $S_1$ and use the assumption in the statement of the theorem, it results that $|N(S) \cap \{T_2 \setminus N(S_1)\}| \geq |S|$. Moreover, induction hypothesis results that there exists a spanning subgraph that satisfies the desired properties in the statement of the theorem for the corresponding subgraph $G_2$. Now, Lemma 12 results that there exists a spanning subgraph $G'_2$ of the graph $G_2$ so that every node of $T_1 \setminus S_1$ is of degree $r + 1$, and also for each subset of nodes of $T_1 \setminus S_1$, the inequality (2.40) holds and in addition, $G'_2$ includes a perfect matching between the nodes in $T_1 \setminus S_1$ and $T_2 \setminus N(S_1)$.

Now, consider a spanning subgraph $G'$ of the graph $G$ that only includes all edges of $G'_1$ and $G'_2$. We can easily observe that every node of $T_1$ is of degree $r + 1$. Moreover, for each subset of nodes of $S_1$ the inequality (2.40) holds, and also for each subset of nodes of $T_1 \setminus S_1$, the inequality (2.40) holds. Now, consider a subset of nodes $S$ of $T_1$. It suffices to show the inequality (2.40) holds for $S$.

Define $S' = S \cap S_1$ and $S'' = S \cap \{T_1 \setminus S_1\}$. Since for each subset of nodes of $S_1$ the inequality (2.40) holds, we have $|N(S) \cap N(S_1)| \geq |S'| + r$. On the other hand, $G'$ includes a perfect matching between the nodes in $T_1 \setminus S_1$ and $T_2 \setminus N(S_1)$, and therefore $|N(S) \cap \{T_2 \setminus N(S_1)\}| \geq |S''|$. As a result, $|N(S)| \geq |S'| + r + |S''| = |S| + r$ and the proof is complete for this scenario.

**Scenario 2:** For any proper and nonempty subset of nodes of $T_1$ we have $|N(S)| \geq |S| + r + 1$:

Consider an arbitrary node $v_1$ and observe that $|N(v_1)| \geq r + 1$. Define $S_1 = T_1 \setminus \{v_1\}$ and let $G_1$ denote the induced subgraph of $G$ with set of nodes $S_1 \cup T_2$ (which is all nodes of graph $G$ except for the node $v_1$). Induction hypothesis results that there exists a spanning subgraph $G'_1$ of
the graph $G_1$ such that every node of $S_1$ is of degree $r + 1$, and also for each subset of nodes of $S_1$, the inequality (2.40) holds. Now, consider a spanning subgraph $G'$ of the graph $G$ that includes only all edges of $G'_1$, and also $r + 1$ random edges among all edges that are connected to $v_1$. It is clear that $G'$ satisfies all of the desired properties mentioned in the statement of the theorem.

In order to complete the proof of Theorem 6, we need the following lemma as it was mentioned before.

**Lemma 12.** Consider a bipartite graph $G$ with the two sets of nodes, $T_1$ with $x$ nodes and $T_2$ with $x + r + y$ nodes, where $x, y, r \in \mathbb{N}$. Suppose that for each subset of the nodes of $T_1$ we have

$$|N(S)| \geq |S| + r.$$  (2.54)

Moreover, assume that there exists a subset of nodes $S_0$ of $T_2$ such that $|S_0| = x$, and also for every subset of nodes $S$ of $T_1$ we have $|N(S) \cap S_0| \geq |S|$. Assume that there exists a spanning subgraph $G_0$ of $G$ such that every node of $T_1$ is of degree $r + 1$, and also for each subset of nodes of $T_1$, the inequality (2.54) holds. Then, there exists a spanning subgraph $G'$ of $G$ such that every node of $T_1$ is of degree $r + 1$, and also for each subset of nodes of $T_1$, the inequality (2.54) holds and in addition, $G'$ includes a perfect matching between the nodes in $T_1$ and $S_0$.

**Proof.** We prove the lemma using strong induction on $x$. For $x = 1$ the lemma is easy to verify. Assume that the statement of lemma holds for $1 \leq x \leq x_0$. Then, we only need to show that lemma holds for the case that $x = x_0 + 1$. There are two scenarios and we prove the statement for these two scenarios separately as follows.

**Scenario 1:** There exists a proper and nonempty subset of nodes $S_1$ of $T_1$ such that $|N(S_1) \cap S_0| = |S_1|$:

For simplicity in notation, define $S'_1 \triangleq T_1 \setminus S_1$. In this scenario, we use the induction hypothesis for $S_1$ and $S'_1$ separately (since $|S_1| \leq x_0$ and $|S'_1| \leq x_0$). Define the sets $S'_0 = N(S_1) \cap S_0$ and $S''_0 = S_0 \setminus S'_0$. Consider the induced subgraph $G_1$ of the graph $G$ where the set of vertices of $G_1$ is $T_2 \cup S_1$. Assumption of the lemma results that there exists a spanning subgraph of $G_1$ such that every node of $S_1$ is of degree $r + 1$, and also for each subset of nodes of $S_1$, the inequality (2.54) holds (by considering all the edges of the induced subgraph with vertices $T_2 \cup S_1$ that also exist in $G_0$). Induction hypothesis results that there exists a spanning subgraph $G'_0$ of $G_1$ such that every
node of $S_1$ is of degree $r + 1$, and also for each subset of nodes of $S_1$, the inequality (2.54) holds and in addition, $G'_0$ includes a perfect matching between nodes of $S_1$ and nodes of $S'_0$.

Consider the induced subgraph $G'_1$ of the graph $G$ where the set of vertices of $G_2$ is $T_2 \cup S'_1$. Again, assumption of the lemma results that there exists a spanning subgraph of $G'_1$ such that every node of $S'_1$ is of degree $r + 1$, and also for each subset of nodes of $S'_1$, the inequality (2.54) holds (by considering all the edges of the induced subgraph with vertices $T_2 \cup S'_1$ that also exist in $G_0$). Also, observe that since $|N(S_1) \cap S_0| = |S_1|$, for every subset of nodes $S$ of $S'_1$ we have $|N(S) \cap S'_0| \geq |S|$. As a result, induction hypothesis results that there exists a spanning subgraph $G''_0$ of $G'_1$ such that every node of $S'_1$ is of degree $r + 1$, and also for each subset of nodes of $S'_1$, the inequality (2.54) holds. In addition, $G''_0$ includes a perfect matching between nodes of $S'_1$ and nodes of $S'_0$.

Now, consider a spanning subgraph $G'$ of the graph $G$ that only includes all edges of $G'_0$ and $G''_0$. It can be verified that $G'$ satisfies all of the desired properties in the statement of the lemma and therefore the proof is complete for this case.

**Scenario 2:** For any proper and nonempty subset of nodes of $T_1$ we have $|N(S) \cap S_0| \geq |S| + 1$:

In this case, consider an arbitrary vertex of $T_1$ and denote it by $v_0$, and also define $S_1 = T_1 \setminus \{v_0\}$. Hence, we have $|N(S_1) \cap S_0| \geq |S_1| + 1$ and since for any $S$ we have $|N(S) \cap S_0| \leq |S_0|$ we conclude that $|N(S_1) \cap S_0| = |S_1| + 1$. Also, according to the assumptions of the lemma we know $N(v_0) \geq r + 1$ and $|N(v_0) \cap S_0| \geq 1$. We choose an arbitrary node among the nodes in $|N(v_0) \cap S_0|$ and denote it by $u_0$. Construct a graph $G''_0$ with nodes $T_2 \cup \{v_0\}$ and $r + 1$ edges among edges that are connected to $v_0$ including the edge between $v_0$ and $u_0$, i.e., $(v_0, u_0)$.

Now, consider the induced subgraph $G_1$ of the graph $G$ where the set of vertices of $G_1$ is $T_2 \cup S_1$, i.e., all nodes in $G$ except for the node $v_0$. Assumption of the lemma results that there exists a spanning subgraph of $G_1$ such that every node of $S_1$ is of degree $r + 1$, and also for each subset of nodes of $S_1$, the inequality (2.54) holds (by considering all edges of the induced subgraph with vertices $T_2 \cup S_1$ that also exist in $G_0$). Induction hypothesis results that there exists a spanning subgraph $G'_0$ of $G_1$ such that every node of $S_1$ is of degree $r + 1$, and also for each subset of nodes of $S_1$, the inequality (2.54) holds and in addition, $G'_0$ includes a perfect matching between nodes of $S_1$ and nodes of $S_0 \setminus \{u_0\}$.

Finally, consider a spanning subgraph $G'$ of the graph $G$ that only includes all edges of $G'_0$ and $G''_0$. The constructed graph $G'$ satisfies all of the mentioned properties in the lemma. $\square$
2.10 Proof of Claim in Example 5

In the following we show that given the rank-2 sampled matrix in Example 5, there are exactly two completions. Note that the following decomposition always holds for some values of \( r_1, r_2, \ldots, r_6 \) and \( x_1, x_2, \ldots, x_8 \):

\[
\begin{bmatrix}
1 & 1 & -\frac{1}{2} \\
-4 & 2 & -1 \\
0 & 1 & 2 \\
1 & 4 & \\
4 & -2 & \frac{3}{2}
\end{bmatrix} = 
\begin{bmatrix}
1 & 0 \\
0 & 1 \\
r_1 & r_2 \\
r_3 & r_4 \\
r_5 & r_6
\end{bmatrix} \times
\begin{bmatrix}
x_1 & x_2 & x_3 & x_4 \\
x_5 & x_6 & x_7 & x_8
\end{bmatrix}
\]

Note that the 2\times2 identity matrix in the above decomposition represents canonical basis defined in Definition 2.12. The first two rows of \( U \) in the above decomposition result the following:

\[
\begin{align*}
1 &= x_1, \\
1 &= x_3, \\
-\frac{1}{2} &= x_4, \\
-4 &= x_5, \\
2 &= x_6, \\
-1 &= x_7.
\end{align*}
\]

Then, the decomposition can be simplified as

\[
\begin{bmatrix}
0 & 1 & 2 \\
1 & 4 & \\
4 & -2 & \frac{3}{2}
\end{bmatrix} = 
\begin{bmatrix}
r_1 & r_2 \\
r_3 & r_4 \\
r_5 & r_6
\end{bmatrix} \times
\begin{bmatrix}
1 & x_2 & 1 & -\frac{1}{2} \\
-4 & 2 & -1 & x_8
\end{bmatrix}
\]
Therefore, we have the following system of equations

\[ 0 = r_1 - 4r_2, \quad (2.56a) \]
\[ 1 = x_2r_1 + 2r_2, \quad (2.56b) \]
\[ 2 = -\frac{1}{2}r_1 + x_8r_2, \quad (2.56c) \]
\[ 1 = r_3 - 4r_4, \quad (2.56d) \]
\[ 4 = r_3 - r_4, \quad (2.56e) \]
\[ 4 = x_2r_5 + 2r_6, \quad (2.56f) \]
\[ -2 = r_5 - r_6, \quad (2.56g) \]
\[ \frac{3}{2} = -\frac{1}{2}r_5 + x_8r_6. \quad (2.56h) \]

Observe that \( r_3 = 5 \) and \( r_4 = 1 \) can be determined uniquely from (2.56d) and (2.56e). Note that \( r_1 = 4r_2 \) and \( r_5 = (r_6 - 2) \) can be concluded from (2.56a) and (2.56g), respectively. Then, by substituting \( r_1 \leftarrow 4r_2 \) in (2.56b) and (2.56c) and substituting \( r_5 \leftarrow (r_6 - 2) \) in (2.56f) and (2.56h) we obtain:

\[ 1 = 4x_2r_2 + 2r_2, \quad (2.57a) \]
\[ 2 = -2r_2 + x_8r_2, \quad (2.57b) \]
\[ 4 = x_2(r_6 - 2) + 2r_6, \quad (2.57c) \]
\[ \frac{3}{2} = -\frac{1}{2}(r_6 - 2) + x_8r_6. \quad (2.57d) \]

Similarly, \( r_2 = \frac{2}{x_8 - 2} \) and \( r_6 = \frac{1}{x_8 - 1} \) can be concluded from (2.57b) and (2.57d), respectively. Hence, substituting \( r_2 \leftarrow \frac{2}{x_8 - 2} \), \( r_6 \leftarrow \frac{1}{x_8 - 1} \) in (2.57a) and (2.57c) results in the following system of equations

\[ x_8 = 8x_2 + 6, \quad (2.58a) \]
\[ 8x_8 - 4 = x_2 - 2x_2(2x_8 - 1) + 2. \quad (2.58b) \]

Finally, (2.58a) results \( x_8 = 8x_2 + 6 \), and therefore substituting \( x_8 \leftarrow 8x_2 + 6 \) in (2.58b) results

\[ 8(8x_2 + 6) - 4 = x_2 - 2x_2(2(8x_2 + 6) - 1) + 2, \quad (2.59) \]

which can be simplified as

\[ 32x_2^2 + 85x_2 + 42 = 0. \quad (2.60) \]
Therefore, $x_2 \in \{-2, -\frac{21}{32}\}$. Given $x_2$, all other variables can be determined uniquely recursively as above, and the following completions of $U$ can be obtained as the only possible rank-2 completions of $U$.

\[
\begin{pmatrix}
1 & -\frac{21}{32} & 1 & -\frac{1}{2} \\
-4 & 2 & -1 & \frac{3}{4} \\
0 & 1 & -\frac{24}{5} & 2 \\
1 & -\frac{41}{32} & 4 & -\frac{7}{4} \\
-8 & 4 & -2 & \frac{3}{2}
\end{pmatrix}
= \begin{pmatrix}
1 & 0 \\
0 & 1 \\
-\frac{32}{5} & -\frac{8}{5} \\
5 & 1 \\
0 & 2
\end{pmatrix}
\times
\begin{pmatrix}
1 & -\frac{21}{32} & 1 & -\frac{1}{2} \\
-4 & 2 & -1 & \frac{3}{4}
\end{pmatrix}
\]

and

\[
\begin{pmatrix}
1 & -2 & 1 & -\frac{1}{2} \\
-4 & 2 & -1 & -10 \\
0 & 1 & \frac{1}{2} & 2 \\
1 & -8 & 4 & -\frac{25}{2} \\
\frac{39}{21} & 4 & -2 & \frac{3}{2}
\end{pmatrix}
= \begin{pmatrix}
1 & 0 \\
0 & 1 \\
-\frac{2}{3} & -\frac{1}{6} \\
5 & 1 \\
-\frac{43}{21} & -\frac{1}{21}
\end{pmatrix}
\times
\begin{pmatrix}
1 & -2 & 1 & -\frac{1}{2} \\
-4 & 2 & -1 & -10
\end{pmatrix}
\]
Chapter 3

Rank Estimation For Low-Rank Sampled Data

Recently, fundamental conditions on the sampling patterns have been obtained for finite completability of low-rank matrices or tensors given the corresponding ranks. In this chapter, we consider the scenario where the rank is not given and we aim to approximate the unknown rank based on the location of sampled entries and some given completion. We consider a number of data models, including single-view matrix, multi-view matrix, CP tensor, tensor-train tensor and Tucker tensor. For each of these data models, we provide an upper bound on the rank when an arbitrary low-rank completion is given. We characterize these bounds both deterministically, i.e., with probability one given that the sampling pattern satisfies certain combinatorial properties, and probabilistically, i.e., with high probability given that the sampling probability is above some threshold. Moreover, for both single-view matrix and CP tensor, we are able to show that the obtained upper bound is exactly equal to the unknown rank if the lowest-rank completion is given. Furthermore, we provide numerical experiments for the case of single-view matrix, where we use nuclear norm minimization to find a low-rank completion of the sampled data and we observe that in most of the cases the proposed upper bound on the rank is equal to the true rank.
3.1 Introduction

Developing methods and algorithms to study large high-dimensional data is becoming more indispensable as hyperspectral images and videos, product ranking datasets and other applications of big datasets are attracting more attention recently. Moreover, in order to guarantee the same level of efficiency in images or videos, a minor increment in dimensionality in the datasets entails a significant increment in the amount of the data, and this fact causes modeling and also computational challenges to analyze big high-dimensional datasets. Consequently, providing a statistically rigorous result requires a massive amount of data that grows exponentially with the dimension. The low-rank data completion problem is concerned with completing a matrix or tensor given a subset of its entries and some rank constraints. Various applications can be found in many fields including image and signal processing [Candès et al., 2013; Ji et al., 2010], data mining [Eldén, 2007], network coding [Harvey et al., 2005], compressed sensing [Lim and Comon, 2010; Sidiropoulos and Kyrillidis, 2012; Gandy et al., 2011; Ashraphijuo and Wang, 2017b], reconstructing the visual data [Liu et al., 2013b], etc. There is an extensive literature on developing various optimization methods to treat this problem including minimizing a convex relaxation of rank [Candès and Recht, 2009; Candès and Tao, 2010; Cai et al., 2010; Gandy et al., 2011; Ashraphijuo and Wang, 2017b], non-convex approaches [Recht and Ré, 2013], and alternating minimization [Jain et al., 2013; Ge et al., 2016], etc. More recently, deterministic conditions on the sampling patterns have been studied for subspace clustering in [Pimentel-Alarcón et al., 2015; Pimentel-Alarcón and Nowak, 2016b]. Moreover, the fundamental conditions on the sampling pattern that lead to different numbers of completion (unique, finite, or infinite) for different data structures given the corresponding rank constraints have been investigated in [Pimentel-Alarcón et al., 2016b; Ashraphijuo and Wang, 2017a; Ashraphijuo et al., 2019a].

However, in many practical low-rank data completion problems, the rank may not be known a priori. In this chapter, we investigate this problem and we aim to approximate the rank based on the given entries, where it is assumed that the original data is generically chosen from the manifold corresponding to the unknown rank. The only existing work that treats this problem for a single-view matrix data based on the sampling pattern is [Pimentel-Alarcón and Nowak, 2016a], which requires some strong assumptions including the existence of a completion whose rank $r$ is a lower bound on the unknown true rank $r^*$, i.e., $r^* \geq r$. We start by investigating the single-view
matrix to provide a new analysis that does not require such assumption and also we can extend our approach to treat the CP rank tensor model. Moreover, we further generalize our approach to treat vector rank data models including the multi-view matrix, the Tucker rank tensor and the tensor-train (TT) rank tensor. For each of these data models, we obtain the upper bound on the scalar rank or component-wise upper bound on the unknown vector rank, deterministically based on the sampling pattern and the rank of a given completion. We also obtain such bound that holds with high probability based on the sampling probability. Moreover, for the single-view matrix, we provide some numerical results to show how tight our probabilistic bounds on the rank are (in terms of the sampling probability). In particular, we used nuclear norm minimization to find a completion and demonstrate our proposed method in obtaining a tight bound on the unknown rank.

In general, providing a completion requires much less samples than recovering the original sampled data. The goal of this chapter is to solve the fundamental problem of rank determination for the original sampled data given an arbitrary low-rank data completion. One possible application scenario is to improve upon the low-rank completion obtained by the convex relaxation methods. Specifically, using convex optimization (minimization of nuclear and atomic norms or summation of nuclear norms of matricizations and unfoldings) or any other methods in the literature, we may be able to find a fairly low-rank “completion” of the original data, which is not necessarily equal (or even close) to the original sampled data. Then, under some circumstances, the rank of the obtained completion using any rank independent method can be an upper bound on the rank of the original sampled data (and sometimes the obtained rank can be exactly equal to the rank of the original sampled data).

We take advantage of the geometric analysis on the manifold of the corresponding data which leads to the fundamental conditions on the sampling pattern (independent of the value of entries) [Pimentel-Alarcón et al., 2016b; Ashraphijuo et al., 2020; Ashraphijuo and Wang, 2017a; Ashraphijuo et al., 2019a; Ashraphijuo and Wang, 2020a] such that given an arbitrary low-rank completion we can provide a tight upper bound on the rank. To illustrate how such approximation is even possible consider the following example. Assume that an \(n_1 \times n_2\) rank-2 matrix is chosen generically from the corresponding manifold. Hence, any \(2 \times 2\) submatrix of this matrix is full-rank with probability one (due to the genericity assumption). Moreover, note that any \(3 \times 3\) subma-
matrix of this matrix is not full-rank. As a result, by observing the sampled entries we can find some bounds on the rank. Using the analysis in [Pimentel-Alarcón et al., 2016b; Ashraphijuo et al., 2020; Ashraphijuo and Wang, 2017a; Ashraphijuo et al., 2019a; Ashraphijuo and Wang, 2020a] on finite completability of the sampled data (finite number of completions) for different data models, we characterize both deterministic and probabilistic bounds on the unknown rank.

The remained of the chapter is organized as follows. In Section 3.2, we introduce the data models and problem statement. In Sections 3.3 and 3.4 we characterize our deterministic and probabilistic bounds for scalar-rank cases (single-view matrix and CP tensor) and vector-rank cases (multi-view matrix, Tucker tensor and TT tensor), respectively. Finally, Section 3.5 summarizes the chapter.

3.2 Data Models and Problem Statement

3.2.1 Matrix Models

3.2.1.1 Single-View Matrix

Assume that the sampled matrix $U$ is chosen generically from the manifold of the $n_1 \times n_2$ matrices of rank $r^*$, where $r^*$ is unknown. The matrix $V \in \mathbb{R}^{n_1 \times r^*}$ is called a basis for $U$ if each column of $U$ can be written as a linear combination of the columns of $V$. Denote $\Omega$ as the binary sampling pattern matrix that is of the same size as $U$ and $\Omega(\vec{x}) = 1$ if $U(\vec{x})$ is observed and $\Omega(\vec{x}) = 0$ otherwise, where $\vec{x} = (x_1, x_2)$ represents the entry corresponding to row number $x_1$ and column number $x_2$. Moreover, define $U_\Omega$ as the matrix obtained from sampling $U$ according to $\Omega$, i.e.,

$$U_\Omega(\vec{x}) = \begin{cases} U(\vec{x}) & \text{if } \Omega(\vec{x}) = 1, \\ 0 & \text{if } \Omega(\vec{x}) = 0. \end{cases} \quad (3.1)$$

3.2.1.2 Multi-View Matrix

The matrix $U \in \mathbb{R}^{n \times (n_1+n_2)}$ is sampled. Denote a partition of $U$ as $U = [U_1|U_2]$ where $U_1 \in \mathbb{R}^{n \times n_1}$ and $U_2 \in \mathbb{R}^{n \times n_2}$ represent the first and second views of data, respectively. The sampling pattern is defined as $\Omega = [\Omega_1|\Omega_2]$, where $\Omega_1$ and $\Omega_2$ represent the sampling patterns corresponding to the first and second views of data, respectively. Assume that $\text{rank}(U_1) = r_1^*$, $\text{rank}(U_2) = r_2^*$ and $\text{rank}(U) = r^*$, and also $U$ is chosen generically from the manifold structure with above parameters. Denote $r^* = (r_1^*, r_2^*, r^*)$ which is assumed unknown.
CHAPTER 3. RANK ESTIMATION FOR LOW-RANK SAMPLED DATA

3.2.2 Tensor Models

Assume that a d-way tensor \( U \in \mathbb{R}^{n_1 \times \cdots \times n_d} \) is sampled. For the sake of simplicity in notation, define \( N_i \triangleq (\Pi_{j=1}^{i-1} n_j) \), \( \bar{N}_i \triangleq (\Pi_{j=i+1}^d n_j) \) and \( N_{-i} \triangleq \frac{N_d}{n_i} \). Denote \( \Omega \) as the binary sampling pattern tensor that is of the same size as \( U \) and \( \Omega(\bar{x}) = 1 \) if \( U(\bar{x}) \) is observed and \( \Omega(\bar{x}) = 0 \) otherwise, where \( U(\bar{x}) \) represents an entry of tensor \( U \) with coordinate \( \bar{x} = (x_1, \ldots, x_d) \). Moreover, define \( U_\Omega \) as the tensor obtained from sampling \( U \) according to \( \Omega \), i.e.,

\[
U_\Omega(\bar{x}) = \begin{cases} 
U(\bar{x}) & \text{if } \Omega(\bar{x}) = 1, \\
0 & \text{if } \Omega(\bar{x}) = 0.
\end{cases}
\] (3.2)

For each subtensor \( U' \) of the tensor \( U \), define \( N_\Omega(U') \) as the number of observed entries in \( U' \) according to the sampling pattern \( \Omega \).

Define the matrix \( \tilde{U}_{(i)} \in \mathbb{R}^{N_i \times \bar{N}_i} \) as the \( i \)-th unfolding of the tensor \( U \), such that \( U(\bar{x}) = \tilde{U}_{(i)}(\tilde{M}_i(x_1, \ldots, x_i), \tilde{M}_{-i}(x_{i+1}, \ldots, x_d)) \), where \( \tilde{M}_i : (x_1, \ldots, x_i) \rightarrow \{1, 2, \ldots, N_i\} \) and \( \tilde{M}_{-i} : (x_{i+1}, \ldots, x_d) \rightarrow \{1, 2, \ldots, \bar{N}_i\} \) are two bijective mappings.

Let \( U_{(i)} \in \mathbb{R}^{n_1 \times N_{-i}} \) be the \( i \)-th matricization of the tensor \( U \), such that \( U(\bar{x}) = U_{(i)}(M_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)) \), 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. Observe that for any arbitrary tensor \( A \), the first matricization and the first unfolding are the same, i.e., \( A_{(1)} = \tilde{A}_{(1)} \).

In what follows, we introduce three different tensor ranks, i.e., the CP rank, Tucker rank and TT rank.

3.2.2.1 CP Decomposition

The CP rank of a tensor \( U \), \( \text{rank}_{CP}(U) = r \), is defined as the minimum number \( r \) such that there exist \( a_1^l \in \mathbb{R}^{n_1} \) for \( 1 \leq i \leq d \) and \( 1 \leq l \leq r \), such that

\[
U = \sum_{l=1}^r a_1^l \otimes a_2^l \otimes \ldots \otimes a_d^l,
\] (3.3)
or equivalently,

\[
U(x_1, x_2, \ldots, x_d) = \sum_{l=1}^r a_1^l(x_1)a_2^l(x_2)\ldots a_d^l(x_d),
\] (3.4)

where \( \otimes \) denotes the tensor product (outer product) and \( a_i^l(x_i) \) denotes the \( x_i \)-th entry of vector \( a_i^l \). Note that \( a_1^l \otimes a_2^l \otimes \ldots \otimes a_d^l \in \mathbb{R}^{n_1 \times \cdots \times n_d} \) is a rank-1 tensor, \( l = 1, 2, \ldots, r \).
3.2.2.2 Tucker Decomposition

Given $U \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and $X \in \mathbb{R}^{n_i \times n'_i}$, the product $U' \triangleq U \times_i X \in \mathbb{R}^{n_1 \times \cdots \times n_{i-1} \times n'_i \times n_{i+1} \times \cdots \times n_d}$ is defined as

$$U'(x_1, \ldots, x_{i-1}, k_i, x_{i+1}, \ldots, x_d) \triangleq \sum_{x_i=1}^{n_i} U(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_d)X(x_i, k_i).$$

(3.5)

The Tucker rank of a tensor $U$ is defined as $\text{rank}_{\text{Tucker}}(U) = r_\text{T} = (m_1, \ldots, m_d)$ where $m_i = \text{rank}(U(i))$, i.e., the rank of the $i$-th matricization, $i = 1, \ldots, d$. The Tucker decomposition of $U$ is given by

$$U(\vec{x}) = \sum_{k_1=1}^{m_1} \cdots \sum_{k_d=1}^{m_d} C(k_1, \ldots, k_d)T_1(k_1, x_1) \cdots T_d(k_d, x_d),$$

(3.6)

or in short

$$U = C \times^d_{i=1} T_i,$$

(3.7)

where $C \in \mathbb{R}^{m_1 \times \cdots \times m_d}$ is the core tensor and $T_i \in \mathbb{R}^{m_i \times n_i}$ are $d$ orthogonal matrices.

3.2.2.3 TT Decomposition

The separation or TT rank of a tensor is defined as $\text{rank}_{\text{TT}}(U) = r_\text{TT} = (u_1, \ldots, u_{d-1})$ where $u_i = \text{rank}(U(i))$, i.e., the rank of the $i$-th unfolding, $i = 1, \ldots, d - 1$. Note that $u_i \leq \max\{N_i, \bar{N}_i\}$ in general and also $u_1$ is simply the conventional matrix rank when $d = 2$. The TT decomposition of a tensor $U$ is given by

$$U(\vec{x}) = \sum_{k_1=1}^{u_1} \cdots \sum_{k_{d-1}=1}^{u_{d-1}} U^{(1)}(x_1, k_1) \left( \prod_{i=2}^{d-1} U^{(i)}(k_{i-1}, x_i, k_i) \right) U^{(d)}(k_{d-1}, x_d),$$

(3.8)

or in short

$$U = U^{(1)} \cdots U^{(d)},$$

(3.9)

where the 3-way tensors $U^{(i)} \in \mathbb{R}^{u_{i-1} \times n_i \times u_i}$ for $i = 2, \ldots, d - 1$ and matrices $U^{(1)} \in \mathbb{R}^{n_1 \times u_1}$ and $U^{(d)} \in \mathbb{R}^{u_{d-1} \times n_d}$ are the components of this decomposition.

For each matrix or tensor model, we assume that the true rank of $U$ or $U'$ is $r^*$ or $r^*$ which is unknown, and also $U$ or $U'$ is chosen generically from the corresponding manifold. The table below represents the mentioned symbols in brief.
### 3.2.3 Problem Statement

In this chapter, we assume that there exists a full rank completion of the sampled data (i.e., the data is not over-sampled). For each one of the above data models, we are interested in obtaining the upper bound on the unknown scalar-rank $r^*$ or component-wise upper bound on the unknown vector-rank $\ell^*$, deterministically based on the sampling pattern $\Omega$ or $\Omega$ and the rank of a given completion. Also, we aim to provide such bound that holds with high probability based only on the sampling probability of the entries and the rank of a given completion. Moreover, for the single-view matrix model and CP-rank tensor model, where the rank is a scalar, we provide both deterministic and probabilistic conditions such that the unknown rank can be exactly determined.

### 3.3 Scalar-Rank Cases

#### 3.3.1 Single-View Matrix

Previously, this problem has been treated in [Pimentel-Alarcón and Nowak, 2016a], where strong assumptions including the existence of a completion with rank $r \leq r^*$ have been used. In this section, we provide an analysis that does not require such assumption and moreover our analysis can be extended to multi-view data and tensors in the following sections. Furthermore, we show the tightness of our theoretical bounds via numerical examples.
Assume that \( U \in \mathbb{R}^{n_1 \times n_2} \) is the sampled matrix. Let \( P_1 \) and \( P_2 \) denote the Lebesgue measures on \( \mathbb{R}^{n_1 \times r^*} \) and \( \mathbb{R}^{r^* \times n_2} \), respectively. In this chapter, we assume that the matrix \( U \) is chosen generically from the manifold of \( n_1 \times n_2 \) matrices of rank \( r^* \), i.e., the entries of \( U \) are drawn independently with respect to Lebesgue measure on the corresponding manifold. Hence, the probability measures of all statements in this subsection are \( P_1 \times P_2 \).

### 3.3.1.1 Deterministic Rank Analysis

The following condition will be used frequently in this subsection.

**Condition \( A_r \):** Each column of the sampled matrix includes at least \( r \) sampled entries.

Consider an arbitrary column of the sampled matrix \( U(:,i) \), where \( i \in \{1, \ldots, n_2\} \). Let \( l_i = N_{\Omega}(U(:,i)) \) denote the number of observed entries in the \( i \)-th column of \( U \). Condition \( A_r \) results that \( l_i \geq r \).

We construct a binary valued matrix called **constraint matrix** \( \bar{\Omega}_r \) based on \( \Omega \) and a given number \( r \). Specifically, we construct \( l_i - r \) columns with binary entries based on the locations of the observed entries in \( U(:,i) \) such that each column has exactly \( r + 1 \) entries equal to one. Assume that \( x_1, \ldots, x_{l_i} \) are the row indices of all observed entries in this column. Let \( \Omega^i_r \) be the corresponding \( n_1 \times (l_i - r) \) matrix to this column which is defined as the following: for any \( j \in \{1, \ldots, l_i - r\} \), the \( j \)-th column has the value 1 in rows \( \{x_1, \ldots, x_r, x_{r+j}\} \) and zeros elsewhere. Define the binary constraint matrix as

\[
\bar{\Omega}_r = [\Omega^1_r | \Omega^2_r | \ldots | \Omega^{n_2}_r] \in \mathbb{R}^{n_1 \times K_r} \quad \text{[Pimentel-Alarcón et al., 2016b],}
\]

where \( K_r = N_{\Omega}(U) - n_2r \).

**Condition \( B_r \):** There exists a submatrix\(^1\) \( \bar{\Omega}_r' \in \mathbb{R}^{n_1 \times K} \) of \( \bar{\Omega}_r \) such that \( K = n_1r - r^2 \) and for any \( K' \in \{1, 2, \ldots, K\} \) and any submatrix \( \bar{\Omega}_r'' \in \mathbb{R}^{n_1 \times K'} \) of \( \bar{\Omega}_r' \) we have

\[
rf(\bar{\Omega}_r'') - r^2 \geq K',
\]

where \( f(\bar{\Omega}_r'') \) denotes the number of nonzero rows of \( \bar{\Omega}_r'' \).

Note that exhaustive enumeration is needed in order to check whether or not Condition \( B_r \) holds. Hence, the deterministic analysis cannot be used in practice for large-scale data. However, it serves as the basis of the subsequent probabilistic analysis that will lead to a simple lower bound.

---

\(^1\)Specified by a subset of rows and a subset of columns (not necessarily consecutive).
on the sampling probability such that Condition $B_r$ holds with high probability, which is of practical value.

In the following, we restate Theorem 1 in [Pimentel-Alarcón et al., 2016b] which will be used later.

**Lemma 13.** With probability one, there are finitely many completions of the sampled matrix if and only if Conditions $A_{r^*}$ and $B_{r^*}$ hold.

Recall that the true rank $r^*$ is assumed unknown.

**Definition 7.** Let $S_\Omega$ denote the set of all natural numbers $r$ such that both Conditions $A_r$ and $B_r$ hold.

**Lemma 14.** There exists a number $r_\Omega$ such that $S_\Omega = \{1, 2, \ldots, r_\Omega\}$.

**Proof.** Assume that $1 < r \leq \min\{n_1, n_2\}$ and $r \in S_\Omega$. It suffices to show $r - 1 \in S_\Omega$. By contradiction, assume that $r - 1 \notin S_\Omega$. Therefore, according to Lemma 13, there exist infinitely many completions of $U$ of rank $r - 1$ and there exist at most finitely many completions of $U$ of rank $r$.

Consider a rank $r - 1$ completion $U_{r-1}$. Note that changing one single entry (a non-observed entry) of $U_{r-1}$, for example $U_{r-1}(1, 1) = x$, to a random number in $y \in \mathbb{R}$ changes the rank of this matrix by at most 1 and basically since we are changing to a random number, it can be easily seen that the rank does not decrease with probability one. Hence, the rank of the modified matrix $U'_{r-1}$ would be either $r - 1$ or $r$. Assume that the rank has been increased to $r$. Then, we show there exist infinitely many completions of rank $r$, which contradicts the assumption. In fact, for any value of $U_{r-1}(1, 1)$ except $x$, this matrix would be a rank $r$ completion. To observe this more clearly, consider the $r \times r$ submatrix of $U'_{r-1}$ whose determinant is not zero due to changing the value of $U_{r-1}(1, 1)$. It is easily observed that this submatrix includes $U'_{r-1}(1 : r, 1 : r)$, and therefore the determinant of $U'_{r-1}(2 : r, 2 : r)$ is a nonzero number (otherwise the rank would not increase by changing the value of $U_{r-1}(1, 1)$). Hence, it is easy to see that for any value of $U_{r-1}(1, 1)$ except $x$, $U'_{r-1}$ would be a rank $r$ completion, and therefore there exist infinitely many completions of rank $r$ and proof is complete in this scenario.

Now, assume that changing any of the non-observed entries does not increase the rank of $U_{r-1}$. 
CHAPTER 3. RANK ESTIMATION FOR LOW-RANK SAMPLED DATA

Then, this contradicts the assumption that there exists a full rank completion of the sampled data since there does not exist any completion of rank higher than \( r - 1 \).

The following theorem provides a relationship between the unknown rank \( r^* \) and \( r_\Omega \).

**Theorem 8.** With probability one, exactly one of the following statements holds

(i) \( r^* \in S_\Omega = \{1, 2, \ldots, r_\Omega \} \);

(ii) For any arbitrary completion of the sampled matrix \( U \) of rank \( r \), we have \( r \notin S_\Omega \).

**Proof.** Suppose that there does not exist a completion of the sampled matrix \( U \) of rank \( r \) such that \( r \in S_\Omega \). Therefore, it is easily verified that statement (ii) holds and statement (i) does not hold. On the other hand, assume that there exists a completion of the sampled matrix \( U \) of rank \( r \), where \( r \in S_\Omega \). Hence, statement (ii) does not hold and to complete the proof it suffices to show that with probability one, statement (i) holds.

Observe that \( r_\Omega \in S_\Omega \), and therefore Condition \( A_{r_\Omega} \) holds. Hence, each column of \( U \) includes at least \( r_\Omega + 1 \) observed entries. On the other hand, the existence of a completion of the sampled matrix \( U \) of rank \( r \in S_\Omega \) results in the existence of a basis \( X \in \mathbb{R}^{n_1 \times r} \) such that each column of \( U \) is a linear combination of the columns of \( X \), and thus there exists \( Y \in \mathbb{R}^{r \times n_2} \) such that \( U_\Omega = (XY)_\Omega \). Hence, given \( X \), each observed entry \( U(i, j) \) results in a degree-1 polynomial in terms of the entries of \( Y \) as the following

\[
U(i, j) = \sum_{l=1}^{r} X(i, l)Y(l, j). \tag{3.11}
\]

Consider the first column of \( U \) and recall that it includes at least \( r_\Omega + 1 \geq r + 1 \) observed entries.

The genericity of the coefficients of the above-mentioned polynomials results that using \( r \) of the observed entries the first column of \( Y \) can be determined uniquely. This is because there exists a unique solution for a system of \( r \) linear equations in \( r \) variables that are linearly independent.

Then, there exists at least one more observed entry besides these \( r \) observed entries in the first column of \( U \) and it can be written as a linear combination of the \( r \) observed entries that have been used to obtain the first column of \( Y \). Let \( U(i_1, 1), \ldots, U(i_r, 1) \) denote the \( r \) observed entries that have been used to obtain the first column of \( Y \) and \( U(i_{r+1}, 1) \) denote the other observed entry. Hence, the existence of a completion of the sampled matrix \( U \) of rank \( r \in S_\Omega \) results in an equation
as the following

\[ U(i_{r+1}, 1) = \sum_{l=1}^{r} t_l U(i_l, 1), \quad (3.12) \]

where \( t_l \)'s are constant scalars, \( l = 1, \ldots, r \). Assume that \( r^* \notin S_\Omega \), i.e., statement (i) does not hold. Then, note that \( r^* \geq r + 1 \) and \( U \) is chosen generically from the manifold of \( n_1 \times n_2 \) rank-\( r^* \) matrices, and therefore an equation of the form of (3.12) holds with probability zero. Moreover, according to Lemma 13 there exist at most finitely many completions of the sampled matrix of rank \( r \). Therefore, there exist a completion of \( U \) of rank \( r \) with probability zero, which contradicts the initial assumption that there exists a completion of the sampled matrix \( U \) of rank \( r \), where \( r \in S_\Omega \).

Note that the existing work that treats the similar problem for a single-view matrix data based on the sampling pattern is [Pimentel-Alarcón and Nowak, 2016a], which requires some strong assumptions including the existence of a completion whose rank \( r \) is a lower bound on the unknown true rank \( r^* \), i.e., \( r^* \geq r \). We provide a new analysis that does not require such assumption and also based on our new analysis, we can extend our approach to treat other data structures.

**Corollary 4.** Consider an arbitrary number \( r' \in S_\Omega \). Similar to Theorem 8, it follows that with probability one, exactly one of the followings holds

(i) \( r^* \in \{1, 2, \ldots, r'\} \);

(ii) For any arbitrary completion of the sampled matrix \( U \) of rank \( r \), we have \( r \notin \{1, 2, \ldots, r'\} \).

As a result of Corollary 4, we have the following.

**Corollary 5.** Assuming that there exists a rank-\( r \) completion of the sampled matrix \( U \) such that \( r \in S_\Omega \), then with probability one \( r^* \leq r \).

**Corollary 6.** Let \( U^* \) denote an optimal solution to the following NP-hard optimization problem

\[
\begin{align*}
\text{minimize}_{U' \in \mathbb{R}^{n_1 \times n_2}} & \quad \text{rank}(U') \\
\text{subject to} & \quad U'_\Omega = U_\Omega.
\end{align*}
\quad (3.13)
\]

Also, let \( \hat{U} \) denote a suboptimal solution to the above optimization problem. Then, Corollary 4 results the following statements:
(i) If \( \text{rank}(U^*) \in S_\Omega \), then \( r^* = \text{rank}(U^*) \) with probability one.

(ii) If \( \text{rank}(\hat{U}) \in S_\Omega \), then \( r^* \leq \text{rank}(\hat{U}) \) with probability one.

**Remark 13.** One challenge of applying Corollary 6 or any of the other obtained deterministic results is the computation of \( S_\Omega \), which involves exhaustive enumeration to check Condition \( B_r \). Next, for each number \( r \), we provide a lower bound on the sampling probability in terms of \( r \) that ensures \( r \in S_\Omega \) with high probability. Consequently, we do not need to compute \( S_\Omega \) but instead we can certify the above results with high probability.

### 3.3.1.2 Probabilistic Rank Analysis

The following lemma is a re-statement of Theorem 3 in [Pimentel-Alarcón et al., 2016b], which is the probabilistic version of Lemma 13.

**Lemma 15.** Suppose \( r \leq \frac{n_1}{6} \) and that each column of the sampled matrix is observed in at least \( l \) entries, uniformly at random and independently across entries, where

\[
l > \max \left\{ 12 \log \left( \frac{n_1}{\epsilon} \right) + 12, 2r \right\}. \tag{3.14}\]

Also, assume that \( r(n_1 - r) \leq n_2 \). Then, with probability at least \( 1 - \epsilon \), \( r \in S_\Omega \).

The following lemma is taken from [Ashraphijuo et al., 2019a] and will be used to derive a lower bound on the sampling probability that leads to the similar statement as Theorem 8 with high probability.

**Lemma 16.** Consider a vector with \( n \) entries where each entry is observed with probability \( p \) independently from the other entries. If \( p > p' = \frac{k}{n} + \frac{1}{\sqrt{n}} \), then with probability at least \( 1 - \exp(-\frac{\sqrt{n}}{2}) \), more than \( k \) entries are observed.

The following proposition characterizes the probabilistic version of Theorem 8.

**Proposition 1.** Suppose \( r \leq \frac{n_1}{6} \), \( r(n_1 - r) \leq n_2 \) and that each entry of the sampled matrix is observed uniformly at random and independently across entries with probability \( p \), where

\[
p > \frac{1}{n_1} \max \{ 12 \log \left( \frac{n_1}{\epsilon} \right) + 12, 2r \} + \frac{1}{\sqrt{n_1}}. \tag{3.15}\]

Then, with probability at least \( (1 - \epsilon) \left( 1 - \exp(-\frac{\sqrt{n_1}}{2}) \right)^{n_2} \), we have \( r \in S_\Omega \).
Proof. Consider an arbitrary column of $U$ and note that resulting from Lemma 16 the number of observed entries at this column of $U$ is greater than $\max\left\{12 \log \left(\frac{n_1}{\epsilon}\right) + 12, 2r\right\}$ with probability at least $\left(1 - \exp\left(-\frac{\sqrt{n_1}}{2}\right)\right)$. Therefore, the number of sampled entries at each column satisfies

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

(3.16)

with probability at least $\left(1 - \exp\left(-\frac{\sqrt{n_1}}{2}\right)\right)^{n_2}$. Thus, resulting from Lemma 15 with probability at least $(1 - \epsilon)\left(1 - \exp\left(-\frac{\sqrt{n_1}}{2}\right)\right)^{n_2}$, we have $r \in S_\Omega$. 

Finally, we have the following probabilistic version of Corollary 6.

Corollary 7. Assume that $\text{rank}(U^*) \leq \frac{n_1}{6}$ and $\text{rank}(U^*)(n_1 - \text{rank}(U^*)) \leq n_2$ and (3.15) holds for $r = \text{rank}(U^*)$, where $U^*$ denotes an optimal solution to the optimization problem (3.13). Then, according to Proposition 1 and Corollary 6, with probability at least $(1 - \epsilon)\left(1 - \exp\left(-\frac{\sqrt{n_1}}{2}\right)\right)^{n_2}$, $r^* = \text{rank}(U^*)$. Similarly, assume that $\text{rank}(\hat{U}) \leq \frac{n_1}{6}$ and $\text{rank}(\hat{U})(n_1 - \text{rank}(\hat{U})) \leq n_2$ and (3.15) holds for $r = \text{rank}(\hat{U})$, where $\hat{U}$ denotes a suboptimal solution to the optimization problem (3.13). Then, with probability at least $(1 - \epsilon)\left(1 - \exp\left(-\frac{\sqrt{n_1}}{2}\right)\right)^{n_2}$, $r^* \leq \text{rank}(\hat{U})$.

3.3.1.3 Numerical Results

In Fig. 3.1 and Fig. 3.2, the x-axis represents the sampling probability, and the y-axis denotes the value of $r$. The color scale represents the lower bound on the probability of event $r \in S_\Omega$. For example, as we can observe in Fig. 3.1, for any $r \in \{1, \ldots, 44\}$ we have $r \in S_\Omega$ with probability at least 0.6 (approximately based on the color scale since the corresponding points are orange) given that $p = 0.54$.

We consider the sampled matrix $U \in \mathbb{R}^{300 \times 15000}$ and $U \in \mathbb{R}^{1200 \times 240000}$ in Fig. 3.1 and Fig. 3.2, respectively. In particular, for fixed values of sampling probability $p$ and $r$, we first find a "small" $\epsilon$ that (3.15) holds by trial-and-error. Then, according to Proposition 1, we conclude that with probability at least $(1 - \epsilon)\left(1 - \exp\left(-\frac{\sqrt{n_1}}{2}\right)\right)^{n_2}$, $r \in S_\Omega$.

The purpose of Figs. 3.3–3.6 is to show how tight our proposed upper bounds on rank can be. Here, we first generate an $n_1 \times n_2$ random matrix of a given rank $r$ by multiplying a random (entries
are drawn according to a uniform distribution on real numbers within an interval) \( n_1 \times r \) matrix and \( r \times n_2 \) matrix. Then, each entry of the randomly generated matrix is sampled uniformly at random and independently across entries with some sampling probability \( p \). Afterwards, we apply the nuclear norm minimization method proposed in [Candès and Recht, 2009] for matrix completion, where the non-convex objective function in (3.13) is relaxed by using nuclear norm, which is the convex hull of the rank function, as follows

\[
\begin{align*}
\text{minimize} & \quad \|U'|_* \\
\text{subject to} & \quad U'_\Omega = U_\Omega,
\end{align*}
\]  

(3.17)

where \( \|U'|_* \) denotes the nuclear norm of \( U' \). Let \( \hat{U}^* \) denote an optimal solution to (3.17) and recall that \( U^* \) denotes an optimal solution to (3.13). Since (3.17) is a convex relaxation to (3.13), we conclude that \( \hat{U}^* \) is a suboptimal solution to (3.13), and therefore \( \text{rank}(U^*) \leq \text{rank}(\hat{U}^*) \). We used the Matlab program found online [Shabat, 2015] to solve (3.17).

As an example, we generate a random matrix \( U \in \mathbb{R}^{300 \times 15000} \) (the same size as the matrix in Fig. 3.1) of rank \( r \) as described above for \( r \in \{1, \ldots, 50\} \) and some values of the sampling probability \( p \). Then, we obtain the rank of the completion given by (3.17) and denote it by \( r' \). Due to the randomness of the sampled matrix, we repeat this procedure 5 times. We calculate the “gap” \( r' - r \) in each of these 5 runs and denote the maximum and minimum among these 5 numbers by \( d_{\text{max}} \) and \( d_{\text{min}} \), respectively. Hence, \( d_{\text{max}} \) and \( d_{\text{min}} \) represent the loosest (worst) and tightest
Figure 3.2: Probability of $r \in \mathcal{S}_\Omega$ as a function of sampling probability for $U \in \mathbb{R}^{1200 \times 240000}$.

(best) gaps between the rank obtained by (3.17) and rank of the original sampled matrix over 5 runs, respectively. In Figs. 3.3–3.6, the maximum and minimum gaps are plotted as a function of rank of the matrix, for different sampling probabilities.

We have the following observations.

- According to Fig. 3.1, for $p = 0.54$ and $p = 0.58$ we can ensure that the rank of any completion is an upper bound on the rank of the sampled matrix or $r^*$ with probability at least 0.6 and 0.8, respectively.

- As we can observe in Figs. 3.3–3.6, the defined gap is always a nonnegative number, which is consistent with previous observation that for $p = 0.54$ and $p = 0.58$ we can certify that with high probability ($\geq 0.6$) the rank of any completion is an upper bound on the rank of the sampled matrix or $r^*$.

- For $p = 0.54$ and $p = 0.58$ that we have theoretical results (as mentioned in the first observation) the gap obtained by (3.17) is very close to zero. This phenomenon (that we do not have a rigorous justification for) shows that as soon as we can certify our proposed theoretical results (i.e., as soon as the rank of a completion provides an upper bound on the rank of the sampled matrix or $r^*$) by increasing the sampling probability, the upper bound found through (3.17) becomes very tight; in some cases this bound is exactly equal to $r^*$ (red curves) and
in some cases this bound is almost equal to \( r^* \) (blue curves). However, these gaps are not small (specially blue curves) for \( p = 0.46 \) and \( p = 0.50 \) and note that according to Fig. 3.1, for these values of \( p \) we cannot guarantee the bounds on the value of rank hold with high probability.

### 3.3.2 CP-Rank Tensor

Let \( \mathbb{P}_i \) denote the Lebesgue measures on \( \mathbb{R}^{n_i \times r^*} \), \( i = 1, \ldots, d \). In this subsection, we assume that the sampled tensor \( \mathcal{U} \in \mathbb{R}^{n_1 \times \cdots \times n_d} \) is chosen generically from the manifold of tensors of rank \( r^* = \text{rank}_{\text{CP}}(\mathcal{U}) \) (where \( r^* \) is unknown), or in other words, the entries of \( \mathcal{U} \) are drawn independently with respect to Lebesgue measure on the corresponding manifold. Hence, the probability measures of all statements in this subsection are \( \mathbb{P}_1 \times \mathbb{P}_2 \times \cdots \times \mathbb{P}_d \).

**Condition \( A_r \):** Each row of the \( d \)-th matricization of the sampled tensor, i.e., \( \mathcal{U}_{(d)} \) includes at least \( r \) observed entries.

We construct a binary valued tensor called constraint tensor \( \hat{\Omega}_r \) based on \( \Omega \) and a given number \( r \). Consider any subtensor \( \mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times 1} \) of the tensor \( \mathcal{U} \). The sampled tensor \( \mathcal{U} \) includes \( n_d \) subtensors that belong to \( \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times 1} \) and let \( \mathcal{Y}_i \) for \( 1 \leq i \leq n_d \) denote these \( n_d \) subtensors. Define a binary valued tensor \( \hat{\mathcal{Y}}_i \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times k_i} \), where \( k_i = N_\Omega(\mathcal{Y}_i) - r \)
Figure 3.4: The gaps between the rank of the obtained matrix via (3.17) and that of the original sampled matrix for $p = 0.50$.

and its entries are described as the following. We can look at $\tilde{Y}_i$ as $k_i$ tensors each belongs to $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times 1}$. For each of the mentioned $k_i$ tensors in $\tilde{Y}_i$ we set the entries corresponding to $r$ of the observed entries equal to $1$. For each of the other $k_i$ observed entries, we pick one of the $k_i$ tensors of $\tilde{Y}_i$ 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 $k_i = 0$ we simply ignore $\tilde{Y}_i$, i.e., $\tilde{Y}_i = \emptyset$.

By putting together all $n_d$ tensors in dimension $d$, we construct a binary valued tensor $\tilde{\Omega}_r \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times K}$, where $K = \sum_{i=1}^{n_d} k_i = N_{\Omega}(\mathcal{U}) - r n_d$ and call it the constraint tensor. Observe that each subtensor of $\tilde{\Omega}_r$ which belongs to $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times 1}$ includes exactly $r + 1$ nonzero entries. In [Ashraphijuo and Wang, 2017a], an example is given on the construction of $\tilde{\Omega}_r$.

**Condition $B_r$:** $\tilde{\Omega}_r$ consists a subtensor $\tilde{\Omega}_r' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times K'}$ such that $K = r(\sum_{i=1}^{d-1} n_i) - r^2 - r(d - 2)$ and for any $K' \in \{1, 2, \ldots, K\}$ and any subtensor $\tilde{\Omega}_r'' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times K''}$ of $\tilde{\Omega}_r'$ we have

$$r \left( \sum_{i=1}^{d-1} f_i(\tilde{\Omega}_r'') \right) - \min \left\{ \max \left\{ f_1(\tilde{\Omega}_r''), \ldots, f_{d-1}(\tilde{\Omega}_r'') \right\}, r \right\} - (d - 2) \geq K', \tag{3.18}$$

where $f_i(\tilde{\Omega}_r'')$ denotes the number of nonzero rows of the $i$-th matricization of $\tilde{\Omega}_r''$.

The following lemma is a re-statement of Theorem 1 in [Ashraphijuo and Wang, 2017a].
Figure 3.5: The gaps between the rank of the obtained matrix via (3.17) and that of the original sampled matrix for $p = 0.54$.

**Lemma 17.** With probability one, there are only finitely many rank-$r^*$ completions of the sampled tensor if and only if Conditions $A_{r^*}$ and $B_{r^*}$ hold.

**Definition 8.** Let $S_{\Omega}$ denote the set of all natural numbers $r$ such that both Conditions $A_r$ and $B_r$ hold.

**Lemma 18.** There exists a number $r_{\Omega}$ such that $S_{\Omega} = \{1, 2, \ldots, r_{\Omega}\}$.

**Proof.** The proof is similar to the proof of Lemma 14 since the dimension of the manifold of CP rank-$r$ tensors is $r(\sum_{i=1}^{d} n_i) - r^2 - r(d - 1)$, which is an increasing function in $r$.

The following theorem gives an upper bound on the unknown rank $r^*$.

**Theorem 9.** With probability one, exactly one of the following statements holds

(i) $r^* \in S_{\Omega} = \{1, 2, \ldots, r_{\Omega}\}$;

(ii) For any arbitrary completion of the sampled tensor $U$ of rank $r$, we have $r \notin S_{\Omega}$.

**Proof.** Similar to the proof of Theorem 8, it suffices to show that the assumption $r^* \notin S_{\Omega}$ results that there exists a completion of $U$ of CP rank $r$, where $r \in S_{\Omega}$, with probability zero. Define $\mathcal{V} = (V_1, \ldots, V_r)$ as the basis of the rank-$r$ CP decomposition of $U$ as in (3.3), where $V_l = a_l^1 \otimes a_l^2 \otimes a_l^3 \otimes \cdots \otimes a_l^n$.
CHAPTER 3. RANK ESTIMATION FOR LOW-RANK SAMPLED DATA

Figure 3.6: The gaps between the rank of the obtained matrix via (3.17) and that of the original sampled matrix for \( p = 0.58 \).

\[ \cdots \otimes a_{d-1}^l \in \mathbb{R}^{n_1 \times \cdots \times n_d-1} \] is a rank-1 tensor and \( a_i^l \) is defined in (3.3) for \( 1 \leq l \leq r \) and \( 1 \leq i \leq d \).

Define \( \mathcal{Y} = (a_1^0, \ldots, a_d^0) \) and \( \mathcal{V} \otimes_d \mathcal{Y} = \sum_{l=1}^r \mathcal{V}_l \otimes a_i^l \). Observe that \( \mathcal{U} = \sum_{l=1}^r \mathcal{V}_l \otimes a_i^l = \mathcal{V} \otimes_d \mathcal{Y} \).

Observe that each row of \( \mathcal{U}(d) \) includes at least \( r_\Omega + 1 \) observed entries since Condition \( A_{r_\Omega} \) holds. Moreover, the existence of a completion of the sampled tensor \( \mathcal{U} \) of rank \( r \in \mathcal{S}_\Omega \) results in the existence of a basis \( \mathcal{V} = (\mathcal{V}_1, \ldots, \mathcal{V}_r) \) such that there exists \( \mathcal{Y} = (a_1^0, \ldots, a_d^0) \) and \( \mathcal{U}_\Omega = (\mathcal{V} \otimes_d \mathcal{Y})_\Omega \).

As a result, given \( \mathcal{V} \), each observed entry of \( \mathcal{U} \) results in a degree-1 polynomial in terms of the entries of \( \mathcal{Y} \) as

\[
\mathcal{U}(\bar{x}) = \sum_{i=1}^r \mathcal{V}_l(x_1, \ldots, x_{d-1})a_i^l(x_d). \tag{3.19}
\]

Note that \( r_\Omega \geq r \) and each row of \( \mathcal{U}(d) \) includes at least \( r_\Omega + 1 \geq r + 1 \) observed entries. Consider \( r + 1 \) of the observed entries of the first row of \( \mathcal{U}(d) \) and we denote them by \( \mathcal{U}(\bar{x}_1), \ldots, \mathcal{U}(\bar{x}_{r+1}) \), where the last component of the vector \( \bar{x}_i \) is equal to one, \( 1 \leq i \leq r + 1 \). Similar to the proof of Theorem 8, genericity of \( \mathcal{U} \) results in

\[
\mathcal{U}(\bar{x}_{r+1}) = \sum_{l=1}^r t_l \mathcal{U}(\bar{x}_l), \tag{3.20}
\]

where \( t_l \)'s are constant scalars, \( l = 1, \ldots, r \). On the other hand, according to Lemma 17 there exist at most finitely many completions of the sampled tensor of rank \( r \). Therefore, there exist a
CHAPTER 3. RANK ESTIMATION FOR LOW-RANK SAMPLED DATA

completion of $U$ of rank $r$ with probability zero. Moreover, an equation of the form of (3.20) holds with probability zero as $r^* \geq r + 1$ and $U$ is chosen generically from the manifold of tensors of rank-$r^*$. Therefore, there exists a completion of rank $r$ with probability zero.

**Corollary 8.** Consider an arbitrary number $r' \in S_\Omega$. Similar to Theorem 9, it follows that with probability one, exactly one of the followings holds

(i) $r^* \in \{1, 2, \ldots, r'\}$;

(ii) For any arbitrary completion of the sampled tensor $U$ of rank $r$, we have $r \notin \{1, 2, \ldots, r'\}$.

**Corollary 9.** Assuming that there exists a CP rank-$r$ completion of the sampled tensor $U$ such that $r \in S_\Omega$, we conclude that with probability one $r^* \leq r$.

**Corollary 10.** Let $U^*$ denote an optimal solution to the following NP-hard optimization problem

$$
\begin{align*}
\text{minimize}_{U' \in \mathbb{R}^{n_1 \times \cdots \times n_d}} & \quad \text{rank}_{CP}(U') \\
\text{subject to} & \quad U'_\Omega = U_\Omega.
\end{align*}
$$

(3.21)

Assume that $\text{rank}_{CP}(U^*) \in S_\Omega$. Then, Corollary 9 results that $r^* = \text{rank}_{CP}(U^*)$ with probability one.

The following lemma is Lemma 15 in [Ashraphijuo and Wang, 2017a], which is the probabilistic version of Lemma 17 in terms of the sampling probability.

**Lemma 19.** Assume that $n_1 = n_2 = \cdots = n_d = n$, $d > 2$, $n > \max\{200, r(d-2)\}$ and $r \leq \frac{n}{6}$. Moreover, assume that the sampling probability satisfies

$$
p > \frac{1}{n^{d-2}} \max \left\{ 27 \log \left( \frac{n}{\epsilon} \right) + 9 \log \left( \frac{2r(d-2)}{\epsilon} \right) + 18, 6r \right\} + \frac{1}{\sqrt{n^{d-2}}}.
$$

(3.22)

Then, with probability at least $(1 - \epsilon) \left( 1 - \exp \left( -\frac{\sqrt{n^{d-2}}}{2} \right) \right)^{n^2}$, we have $r \in S_\Omega$.

The following corollary is the probabilistic version of Corollaries 9 and 10.

**Corollary 11.** Assuming that there exists a CP rank-$r$ completion of the sampled tensor $U$ such that the conditions given in Lemma 19 hold, with the sampling probability satisfying (3.22), we conclude that with probability at least $(1 - \epsilon) \left( 1 - \exp \left( -\frac{\sqrt{n^{d-2}}}{2} \right) \right)^{n^2}$ we have $r^* \leq r$. Therefore, given that (3.22) holds for $r = \text{rank}(U^*)$ and $U^*$ denotes an optimal solution to the optimization problem (3.21), with probability at least $(1 - \epsilon) \left( 1 - \exp \left( -\frac{\sqrt{n^{d-2}}}{2} \right) \right)^{n^2}$ we have $r^* = \text{rank}(U^*)$. 

3.3.2.1 Numerical Results

We generate a random tensor \( U \in \mathbb{R}^{8 \times 8 \times 8 \times 8 \times 8} \) of CP-rank 2 by adding two random rank-1 tensors. The color scale represents the lower bound on the probability that we can guarantee the rank of a given completion is an upper bound on the true value of rank. Then, we solve the following convex optimization problem for different values of the sampling probability.

\[
\begin{align*}
\text{minimize} & \quad \| \tilde{U}'_{(3)} \|_* \\
\text{subject to} & \quad U'_{\Omega} = U_{\Omega}.
\end{align*}
\]

(3.23)

Note that rank of any of the unfoldings of a tensor is a lower bound on the CP-rank of that tensor. Hence, we minimize the nuclear norm of the unfolding with the possible maximum rank among all unfoldings as \( \tilde{U}_{(3)} \in \mathbb{R}^{512 \times 512} \). Then, we use the Matlab toolbox found online “Tensorlab” to calculate the CP-rank of the obtained tensor via solving convex program (3.23) (there are other methods to calculate CP decomposition, e.g., [Pimentel-Alarcón, 2016]). In Figure 3.7, gap represents the CP-rank of the solution of (3.23) minus the CP-rank of the original sampled tensor.

![Figure 3.7: The rank gap as a function of sampling probability for \( U \in \mathbb{R}^{8 \times 8 \times 8 \times 8 \times 8} \) of CP-rank 2.](image)
3.4 Vector-Rank Cases

3.4.1 Multi-View Matrix

Let \( P_1 \) and \( P_2 \) denote the Lebesgue measures on \( \mathbb{R}^{n \times r_1^*} \) and \( \mathbb{R}^{r_1^* \times n_1} \), respectively. Moreover, let \( P_3 \) and \( P_4 \) denote the Lebesgue measures on \( \mathbb{R}^{n \times (r^* - r_1^*)} \) and \( \mathbb{R}^{r_2^* \times n_2} \), respectively. In this chapter, we assume that \( U \) is chosen generically from the manifold corresponding to rank vector \((r_1^*, r_2^*, r^*)\), i.e., the entries of \( U \) are drawn independently with respect to Lebesgue measure on the corresponding manifold. Hence, the probability measures of all statements in this subsection are \( P_1 \times P_2 \times P_3 \times P_4 \).

The following Conditions will be used frequently in this subsection.

**Condition** \( A_{r_1, r_2} \): Each column of \( U_1 \) and \( U_2 \) include at least \( r_1 \) and \( r_2 \) sampled entries, respectively.

We construct a binary valued matrix called **constraint matrix** for multi-view matrix \( U \) as \( \Omega_{r_1, r_2} = [\Omega_{r_1}, \Omega_{r_2}] \), where \( \Omega_{r_1} \) and \( \Omega_{r_2} \) represent the constraint matrix for single-view matrices \( U_1 \) and \( U_2 \) (defined in Section 3.3.1), respectively.

**Condition** \( B_{r_1, r_2, r} \): \( \Omega_{r_1, r_2} \) consists a submatrix \( \Omega_{r_1, r_2}' \in \mathbb{R}^{n \times K} \) such that \( K = n r - r^2 - r_1^2 - r_2^2 + r(r_1 + r_2) \) and for any \( K' \in \{1, 2, \ldots, K\} \) and any submatrix \( \Omega''_{r_1, r_2} \in \mathbb{R}^{n \times K'} \) of \( \Omega_{r_1, r_2}' \) we have

\[
(r - r_2) \left( f(\Omega''_{r_1}) - r_1 \right)^+ + (r - r_1) \left( f(\Omega''_{r_2}) - r_2 \right)^+ + (r_1 + r_2 - r) \left( f(\Omega''_{r_1, r_2}) - (r_1 + r_2 - r) \right)^+ \geq K',
\]

where \( f(X) \) denotes the number of nonzero rows of \( X \) for any matrix \( X \) and \( \Omega''_{r_1, r_2} = [\Omega''_{r_1}, \Omega''_{r_2}] \), and also \( \Omega''_{r_1} \) and \( \Omega''_{r_2} \) denote the columns of \( \Omega_{r_1, r_2}' \) corresponding to \( \Omega_{r_1} \) and \( \Omega_{r_2} \), respectively.

The following lemma is a re-statement of Theorem 2 in [Ashraphijuo et al., 2020].

**Lemma 20.** With probability one, there are only finitely many completions of the sampled multi-view data if and only if Conditions \( A_{r_1, r_2} \) and \( B_{r_1, r_2, r} \) hold.

**Definition 9.** Denote the rank vector \( r = (r_1, r_2, r) \). Define the generalized inequality \( r' \preceq r \) as the component-wise set of inequalities, e.g., \( r'_1 \leq r_1, r'_2 \leq r_2 \) and \( r' \leq r \).

**Definition 10.** Let \( S_\Omega \) denote the set of all \( r \) such that both Conditions \( A_{r_1, r_2} \) and \( B_{r_1, r_2, r} \) hold.

**Lemma 21.** Assume \( r \in S_\Omega \). Then, for any \( r' \preceq r \), we have \( r' \in S_\Omega \).
Proof. We consider the rank factorization of $U$ as in [Ashraphijuo et al., 2020] and similar to the single-view scenario in Lemma 14 each observed entry results in a polynomial in terms of the entries of the components of the decomposition. Note that the dimension of the manifold corresponding to rank vector $r$ is equal to $r_1n_1 + r_2n_2 - r_1^2 - r_2^2 + r(r_1 + r_2)$ [Ashraphijuo et al., 2020], and also observe that the fact that $\max\{r_1, r_2\} \leq r \leq r_1 + r_2 \leq \min\{2n, n_1 + n_2\}$ implies that reducing any of the values $r_1, r_2,$ and $r$ reduces the value of $r_1n_1 + r_2n_2 - r_1^2 - r_2^2 + r(r_1 + r_2)$. Hence, the dimension of the manifold corresponding to rank vector $r$ is larger than that for rank vector $r'$, given $r' \preceq r$, and thus similar to the proof of Lemma 14, finite completability of data with $r$ results finite completability of data with $r'$ with probability one. Then, using Lemma 20, the proof is complete.

The following theorem provides a relationship between the unknown rank vector $r^*$ and $S_\Omega$.

**Theorem 10.** With probability one, exactly one of the following statements holds

(i) $r^* \in S_\Omega$;

(ii) For any arbitrary completion of the sampled matrix $U$ of rank vector $r$, we have $r \notin S_\Omega$.

Proof. Similar to the proof of Theorem 8, suppose that there does not exist a completion of $U$ of rank vector $r$ such that $r \in S_\Omega$. Therefore, it is easily verified that statement (ii) holds and statement (i) does not hold. On the other hand, assume that there exists a completion of $U$ of rank vector $r$, where $r \in S_\Omega$. Hence, statement (ii) does not hold and to complete the proof it suffices to show that with probability one, statement (i) holds. Similar to Theorem 8, we show that assuming $r^* \notin S_\Omega$, there exists a completion of $U$ of rank vector $r$, where $r \in S_\Omega$, with probability zero.

Since $r^* \notin S_\Omega$, according to Lemma 21, for any $r \in S_\Omega$ at least one the following inequalities holds; $r_1 < r_1^*$, $r_2 < r_2^*$ and $r < r^*$. Note that assuming that there exists a completion of $U_1$ of rank $r_1$ with probability zero results that there exists a completion of $U$ of rank vector $r$ with probability zero and similar statement holds for $r_2$ and $r$. Hence, in any possible scenario ($r_1 < r_1^*$ or $r_2 < r_2^*$ or $r < r^*$) the similar proof as in Theorem 8 (for single-view matrix) results that there exists a completion of $U$ of rank vector $r$, where $r \in S_\Omega$, with probability zero.

**Corollary 12.** Consider a subset $S'_\Omega$ of $S_\Omega$ such that for any two members of $S_\Omega$ that $r' \preceq r''$ and $r'' \in S'_\Omega$ we have $r' \in S'_\Omega$. Then, with probability one, exactly one of the followings holds

(i) $r^* \in S'_\Omega$;
(ii) For any arbitrary completion of $U$ of rank vector $\underline{r}$, we have $\underline{r} \notin S_\Omega'$.

**Proof.** Note that the property in the statement of Lemma 21 holds for $S_\Omega'$ as well as $S_\Omega$. Moreover, as $S_\Omega' \subseteq S_\Omega$, for any $\underline{r} \in S_\Omega'$ there exists at most finitely many completions of $U$ of rank vector $\underline{r}$, and therefore the rest of the proof is the same as the proof of Theorem 10. \qed

**Corollary 13.** Assuming that there exists a completion of $U$ with rank vector $\underline{r}$ such that $\underline{r} \in S_\Omega$, then with probability one $\underline{r}^* \preceq \underline{r}$.

The following lemma which is a re-statement of Theorem 3 in [Ashraphijuo et al., 2020] gives the number of samples per column that is needed to ensure that Conditions $A_{r_1,r_2}$ and $B_{r_1,r_2,r}$ hold with high probability.

**Lemma 22.** Suppose that the following inequalities hold

\[
\frac{n}{6} \geq \max\{r_1, r_2, (r_1 + r_2 - r)\}, \quad \tag{3.25}
\]

\[
n_1 \geq (r - r_2)(n - r_1), \quad \tag{3.26}
\]

\[
n_2 \geq (r - r_1)(n - r_2), \quad \tag{3.27}
\]

\[
n_1 + n_2 \geq (r - r_2)(n - r_1) + (r - r_1)(n - r_2) + (r_1 + r_2 - r)(n - (r_1 + r_2 - r)). \tag{3.28}
\]

Moreover assume that each column of $U$ is observed in at least $l$ entries, uniformly at random and independently across entries, where

\[
l > \max\left\{9 \log\left(\frac{n}{\epsilon}\right) + 3 \log\left(\frac{3 \max\{r - r_2, r - r_1, r_1 + r_2 - r\}}{\epsilon}\right) + 6, 2r_1, 2r_2\right\}. \tag{3.29}
\]

Then, with probability at least $1 - \epsilon$, $\underline{r} \in S_\Omega$.

The following proposition is the probabilistic version of Theorem 10 in terms of the sampling probability instead of verifying Conditions $A_{r_1,r_2}$ and $B_{r_1,r_2,r}$.

**Proposition 2.** Suppose that (3.25)-(3.28) hold for $\underline{r}$ and that each entry of the sampled matrix is observed uniformly at random and independently across entries with probability $p$, where

\[
p > \frac{1}{n} \max\left\{9 \log\left(\frac{n}{\epsilon}\right) + 3 \log\left(\frac{3 \max\{r - r_2, r - r_1, r_1 + r_2 - r\}}{\epsilon}\right) + 6, 2r_1, 2r_2\right\} + \frac{1}{\sqrt{n}}. \tag{3.29}
\]

Then, with probability at least $(1 - \epsilon)\left(1 - \exp\left(-\frac{\sqrt{n}}{\epsilon}\right)\right)^{n_1+n_2}$, we have $\underline{r} \in S_\Omega$. 

CHAPTER 3. RANK ESTIMATION FOR LOW-RANK SAMPLED DATA

Proof. The proposition is easy to verify using Lemma 22 and Lemma 15 (similar to the proof for Proposition 1). □

Corollary 14. Assuming that there exists a completion of \( U \) of rank vector \( r \) such that (3.25)-(3.28) hold and the sampling probability satisfies (3.30), then with probability at least \((1 - \epsilon) \left( 1 - \exp\left(-\frac{\sqrt{n}}{2}\right) \right)^{n_1+n_2}\) we have \( r^* \leq r \).

3.4.2 Tucker-Rank Tensor

Let \( P_i \) denote the Lebesgue measures on \( \mathbb{R}^{n_i \times m_i^i} \), \( i = j+1, \ldots, d \), and \( P_0 \) denotes the Lebesgue measure on \( \mathbb{R}^{m_j^1 \times m_j^2 \times \cdots \times m_j^d} \). In this subsection, we assume that the sampled tensor \( U \in \mathbb{R}^{n_1 \times \cdots \times n_d} \) is chosen generically from the manifold of tensors of rank \( r^* = \text{rank}_\text{Tucker}(U) = (m_j^1, \ldots, m_d^j) \) (where \( r^* \) is unknown), or in other words, the entries of \( U \) are drawn independently with respect to Lebesgue measure on the corresponding manifold. Hence, the probability measures of all statements in this subsection are \( P_0 \times P_{j+1} \times P_{j+2} \times \cdots \times P_d \).

Without loss of generality assume that \( m_j^1 \geq \cdots \geq m_d^j \) throughout this subsection. Also, given \( r = (m_j^1, \ldots, m_d^j) \), define the following function

\[
g_r(x) = \sum_{i=j+1}^{d} \min_{r_i} \left\{ r_i, \left( x - \sum_{i'=j+1}^{i-1} r_{i'} \right)^+ \right\} \tag{3.30}
\]

**Definition 11.** For any \( i \in \{j+1, \ldots, d\} \) and \( S_i \subseteq \{1, \ldots, n_i\} \), define \( U(S_i) \) as a set containing the entries of \( |S_i| \) rows (corresponding to the elements of \( S_i \)) of \( U_{(i)} \). Moreover, define \( U(S_{j+1}, \ldots, S_d) = U(S_{j+1}) \cup \ldots \cup U(S_d) \).

**Condition** \( A^\text{Tucker}_E \): There exist \( \sum_{i=j+1}^{d} (n_i m_i) \) observed entries such that for any \( S_i \subseteq \{1, \ldots, n_i\} \) for \( i \in \{j+1, \ldots, d\} \), \( U(S_{j+1}, \ldots, S_d) \) includes at most \( \sum_{i=j+1}^{d} |S_i| m_i \) of the mentioned \( \sum_{i=j+1}^{d} n_i m_i \) observed entries.

Let \( P \) be a set of \( \sum_{i=j+1}^{d} (n_i m_i) \) observed entries such that they satisfy Condition \( A^\text{Tucker}_E \). Now, we construct a \((j+1)^{\text{th}}\)-order binary constraint tensor \( \tilde{\Omega}_E \) in some sense similar to that in Section 3.3.2. For any subtensor \( Y \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times 1 \times \cdots \times 1} \) of the tensor \( U \), let \( N_{\Omega}(YP) \) denote the number of sampled entries in \( Y \) that belong to \( P \).

The sampled tensor \( U \) includes \( n_{j+1} n_{j+2} \cdots n_d \) sub-tensors that belong to \( \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times 1 \times \cdots \times 1} \) and we label these sub-tensors by \( Y_{(t_{j+1}, \ldots, t_d)} \) where \( (t_{j+1}, \ldots, t_d) \) represents the coordinate of the
subtensor. Define a binary valued tensor \( \tilde{\mathcal{Y}}_{(t_j+1, \ldots, t_d)} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times 1 \times \cdots \times 1 \times 1 \times k} \), where \( k = N_\Omega(\mathcal{Y}_{(t_j+1, \ldots, t_d)}) - N_\Omega(\mathcal{Y}_{(t_j+1, \ldots, t_d)}^P) \) and its entries are described as the following. We can look at \( \tilde{\mathcal{Y}}_{(t_j+1, \ldots, t_d)} \) as \( k \) tensors each belongs to \( \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times 1 \times \cdots \times 1} \). For each of the mentioned \( k \) tensors in \( \tilde{\mathcal{Y}}_{(t_j+1, \ldots, t_d)} \) we set the entries corresponding to the \( N_\Omega(\mathcal{Y}_{(t_j+1, \ldots, t_d)}^P) \) observed entries that belong to \( \mathcal{P} \) equal to 1. For each of the other \( k \) observed entries, we pick one of the \( k \) tensors of \( \tilde{\mathcal{Y}}_{(t_j+1, \ldots, t_d)} \) 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.

For the sake of simplicity in notation, we treat tensors \( \tilde{\mathcal{Y}}_{(t_j+1, \ldots, t_d)} \) as a member of \( \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times k} \) instead of \( \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times 1 \times \cdots \times 1 \times k} \). Now, by putting together all \( n_j + n_{j+1} + \cdots + n_d \) tensors in dimension \((j+1)\), we construct a binary valued tensor \( \tilde{\Omega}_\mathcal{L} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times K_j} \), where \( K_j = N_\Omega(\mathcal{U}) - \sum_{i=j+1}^{d} (n_i m_i) \) and call it the constraint tensor [Ashraphijuo et al., 2019a]. In [Ashraphijuo et al., 2019a], an example is given on the construction of \( \tilde{\Omega}_\mathcal{L} \).

**Condition** \( \mathcal{B}_\mathcal{L}^{\text{Tucker}} \). The constraint tensor \( \tilde{\Omega}_\mathcal{L} \) consists a subtensor \( \tilde{\Omega}_\mathcal{L}' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_j \times K} \) such that \( K = \left( \prod_{i=1}^{j} n_i \right) \left( \prod_{i=j+1}^{d} m_i \right) - \sum_{i=j+1}^{d} m_i^2 \) and for any \( K' \in \{1, 2, \ldots, K\} \) and any subtensor \( \tilde{\Omega}_\mathcal{L}'' \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d-1 \times K'} \) of \( \tilde{\Omega}_\mathcal{L}' \) we have

\[
\left( \prod_{i=j+1}^{d} m_i \right) \left( f_{j+1}(\tilde{\Omega}_\mathcal{L}'') \right) - g_x \left( f_{j+1}(\tilde{\Omega}_\mathcal{L}'') \right) \geq K',
\]

where \( f_{j+1}(\tilde{\Omega}_\mathcal{L}'') \) denotes the number of nonzero columns of the \((j+1)\)-th matricization of \( \tilde{\Omega}_\mathcal{L}'' \).

The following lemma is a re-statement of Theorem 3 in [Ashraphijuo et al., 2019a].

**Lemma 23.** With probability one, there are only finitely many completions of rank \( r^* \) of the sampled tensor if and only if Conditions \( \mathcal{A}_\mathcal{L}^{\text{Tucker}} \) and \( \mathcal{B}_\mathcal{L}^{\text{Tucker}} \) hold.

**Definition 12.** Let \( \mathcal{S}_\Omega \) denote the set of all rank vectors \( \mathbf{r} \) such that both Conditions \( \mathcal{A}_\mathcal{L}^{\text{Tucker}} \) and \( \mathcal{B}_\mathcal{L}^{\text{Tucker}} \) hold.

**Lemma 24.** Assume \( \mathbf{r} \in \mathcal{S}_\Omega \). Then, for any rank vector \( \mathbf{r}' \preceq \mathbf{r} \), we have \( \mathbf{r}' \in \mathcal{S}_\Omega \).

**Proof.** Note that the dimension of the manifold corresponding to \( \mathbf{r} \) is \( \left( \prod_{i=1}^{j} n_i \right) \left( \prod_{i=j+1}^{d} m_i \right) + \sum_{i=j+1}^{d} n_i m_i - \sum_{i=j+1}^{d} m_i^2 \), and thus by reducing the value of \( m_{i_0} \) by one (for \( i_0 \in \{j + 1, \ldots, d\} \)), the value of the mentioned dimension reduces by at least \( \left( \prod_{i=1}^{j} n_i \right) + n_i - 2m_i + 1 \), which is greater than zero since \( m_i \leq n_i \). The rest of the proof is similar to the proof of Lemma 14. \( \square \)
**Definition 13.** Define $S_{\Omega}(r)$ as a subset of $S_{\Omega}$, which includes all $r' \in S_{\Omega}$ that $r' \preceq r$.

The following theorem gives a relationship between $r^*$ and $S_{\Omega}$.

**Theorem 11.** With probability one, exactly one of the following statements holds

(i) $r^* \in S_{\Omega}$;

(ii) For any arbitrary completion of the sampled tensor $U$ of rank $r$, we have $r \not\in S_{\Omega}(r^*)$.

**Proof.** Similar to the proof of Theorem 8, to complete the proof it suffices to show that the assumption $r^* \not\in S_{\Omega}$ results that there exists a completion of $U$ of rank $r$, where $r \in S_{\Omega}(r^*)$, with probability zero. Note that $r \preceq r^*$ and since $r^* \not\in S_{\Omega}$ we conclude that there exists $i_0 \in \{j + 1, \ldots, d\}$ such that $m_{i_0} < m^*_i$. As a result, $\sum_{i=j+1}^{d} n_i m_i < \sum_{i=j+1}^{d} n_i m^*_i$.

Condition $B^T_{\text{Tucker}}$ ensures there exists at least one more observed entry (otherwise the constraint tensor does not exist) besides the $\sum_{i=j+1}^{d} n_i m_i$ mentioned observed entries. Given the basis $C \in \mathbb{R}^{n_1 \times \cdots \times n_j \times m_{j+1} \times \cdots \times m_d}$ as in (3.7), there exist $\sum_{i=j+1}^{d} n_i m_i$ variables in the corresponding Tucker decomposition. However, we have $\sum_{i=j+1}^{d} n_i m_i + 1$ polynomials in terms these $\sum_{i=j+1}^{d} n_i m_i$ variables and therefore the last polynomials can be written as algebraic combination of the other $\sum_{i=j+1}^{d} n_i m_i$ polynomials. This leads to a linear equation in terms of the $\sum_{i=j+1}^{d} n_i m_i + 1$ corresponding observed entries. On the other hand, the $\sum_{i=j+1}^{d} n_i m_i$ observed entries satisfy the property stated as Condition $A^T_{\text{Tucker}}$ and it is easily verified that there exist $\sum_{i=j+1}^{d} n_i m^*_i$ entries (observed and non-observed) satisfying Condition $A^T_{\text{Tucker}}$ such that the union of the mentioned $\sum_{i=j+1}^{d} n_i m_i$ entries with any arbitrary other observed entry be a subset of those $\sum_{i=j+1}^{d} n_i m^*_i$ entries. However, $U$ is generically chosen from the manifold corresponding to $r^*$ and therefore a particular linear equation in terms of the mentioned $\sum_{i=j+1}^{d} n_i m^*_i$ entries holds with probability zero. The rest of the proof is similar to the proof of Theorem 8.

**Corollary 15.** Assuming that there exists a completion of $U$ with rank vector $r$ such that $r \in S_{\Omega}$, we conclude that with probability one $r^* \preceq r$.

The following lemma is Corollary 2 in [Ashraphijuo et al., 2019a], which ensures that Conditions $A^T_{\text{Tucker}}$ and $B^T_{\text{Tucker}}$ hold with high probability.
Lemma 25. Assume that \( \sum_{i=j+1}^d m_i^2 \leq \prod_{i=j+1}^d m_i \), \( \prod_{i=j+1}^d m_i \geq N_j \prod_{i=j+1}^d m_i - \sum_{i=j+1}^d m_i^2 \), \( \prod_{i=j+1}^d m_i \leq N_j \), where \( N_j = \prod_{i=1}^j n_i \). Furthermore, assume that we observe each entry of \( \mathcal{U} \) with probability \( p \), where

\[
p > \frac{1}{N_j} \left( 6 \log (N_j) + 2 \log \left( \max \left\{ \frac{2 \sum_{i=j+1}^d r_i^2}{\epsilon}, \frac{2 \prod_{i=j+1}^d r_i - 2 \sum_{i=j+1}^d r_i^2}{\epsilon} \right\} + 4 \right) + \frac{1}{\sqrt{N_j}}. \right)
\]

Then, with probability at least \( (1 - \epsilon) \left( 1 - \exp \left( -\frac{\sqrt{\prod_{i=1}^j n_i}}{2} \right) \right)^{\prod_{i=j+1}^d n_i} \), \( \mathcal{U} \in \mathcal{S}_\Omega \).

The following corollary is the probabilistic version of Theorem 11.

Corollary 16. Assuming that there exists a completion of the sampled tensor \( \mathcal{U} \) of Tucker rank \( r \) such that the assumptions in Lemma 25 hold and the sampling probability satisfies (3.32), then with probability at least \( (1 - \epsilon) \left( 1 - \exp \left( -\frac{\sqrt{\prod_{i=1}^j n_i}}{2} \right) \right)^{\prod_{i=j+1}^d n_i} \) we have \( r^* \leq r \).

3.4.2.1 Numerical Results

We generate a random tensor \( \mathcal{U} \in \mathbb{R}^{8 \times 8 \times 8 \times 8 \times 8} \) of Tucker-rank \((1, 3, 3, 2, 2)\). The color scale represents the lower bound on the probability that we can guarantee the rank of a given completion is a component-wise upper bound on the true rank. Then, we solve the following convex optimization problem for different values of the sampling probability.

\[
\begin{align*}
\text{minimize} \quad & \sum_{i=1}^d \| \mathcal{U}'(i) \|_* \\
\text{subject to} \quad & \mathcal{U}'_\Omega = \mathcal{U}_\Omega.
\end{align*}
\]

Then, we calculate rank of each matricization of the tensor obtained via solving (3.32) to find its Tucker-rank. In this scenario, for each component of the Tucker-rank, we find the percentage of error via \( \frac{m_i - m_i^*}{m_i} \times 100\% \), where \( n = 8 \), \( m_i \) and \( m_i^* \) are the \( i \)-th rank component of the obtained tensor and original tensor, respectively. Hence, 100\% error simply means that the corresponding matricization is full rank. In Figure 3.8, gap represents the average of the defined error over all components of Tucker-rank, i.e., over all matricizations.
Figure 3.8: The rank gap as a function of sampling probability for $U \in \mathbb{R}^{8 \times 8 \times 8 \times 8 \times 8}$ of Tucker-rank $(1, 3, 3, 2, 2)$.

### 3.4.3 TT-Rank Tensor

Let $\mathbb{P}_i$ denote the Lebesgue measures on $\mathbb{R}^{u^*_i \times n_i \times u^*_i}$, $i = 1, \ldots, d$, where $u^*_0 = u^*_d = 1$. In this subsection, we assume that the sampled tensor $U \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is chosen generically from the manifold of tensors of rank $r^* = \text{rank}_{\text{TT}}(U) = (u^*_1, \ldots, u^*_{d-1})$ (where $r^*$ is unknown), or in other words, the entries of $U$ are drawn independently with respect to Lebesgue measure on the corresponding manifold. Hence, the probability measures of all statements in this subsection are $\mathbb{P}_1 \times \ldots \times \mathbb{P}_d$.

**Condition $A^\text{TT}_d$:** Each row of the $d$-th matricization of the sampled tensor, i.e., $U_{(d)}$ includes at least $u_{d-1}$ observed entries.

We construct the $d$-way binary valued constraint tensor $\tilde{\Omega}_{u_{d-1}}$ similar to that in Section 3.3.2 as the following. Consider any subtensor $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times 1}$ of the tensor $U$. The sampled tensor $U$ includes $n_d$ sub-tensors that belong to $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times 1}$ and let $\mathcal{Y}_i$ for $1 \leq i \leq n_d$ denote these $n_d$ sub-tensors. Define a binary valued tensor $\tilde{\mathcal{Y}}_i \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times k_i}$, where $k_i = N_{\Omega}(\mathcal{Y}_i) - u_{d-1}$ and its entries are described as the following. We can look at $\tilde{\mathcal{Y}}_i$ as $k_i$ tensors each belongs to $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times 1}$. For each of the mentioned $k_i$ tensors in $\tilde{\mathcal{Y}}_i$ we set the entries corresponding to $u_{d-1}$ of the observed entries equal to 1. For each of the other $k_i$ observed entries, we pick one of the $k_i$ tensors of $\tilde{\mathcal{Y}}_i$ 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 \( k_i = 0 \) we simply ignore \( \tilde{Y}_i \), i.e., \( \tilde{Y}_i = \emptyset \).

By putting together all \( n_d \) tensors in dimension \( d \), we construct a binary valued tensor \( \tilde{\Omega}_{ud-1} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times K} \), where \( K = \sum_{i=1}^{n_d} k_i = N_{\Omega}(U) - u_{d-1}n_d \) and call it the constraint tensor. Observe that each subtensor of \( \tilde{\Omega}_{ud-1} \) which belongs to \( \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times K} \) includes exactly \( u_{d-1} + 1 \) nonzero entries. In [Ashraphijuo and Wang, 2020a], an example is given on the construction of \( \tilde{\Omega}_{ud-1} \).

**Condition** \( B_{\Sigma}^{TT} \): \( \tilde{\Omega}_{ud-1} \) consists of a subtensor \( \tilde{\Omega}'_{ud-1} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times K} \) such that \( K = \sum_{i=1}^{d-1} u_{i-1}n_iu_i - \sum_{i=1}^{d-1} u_i^2 \) and for any \( K' \in \{1, 2, \ldots, K\} \) and any subtensor \( \tilde{\Omega}''_{ud-1} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1} \times K'} \) of \( \tilde{\Omega}'_{ud-1} \) we have

\[
\sum_{i=1}^{d-1} \left( u_{i-1}f_i(\tilde{\Omega}''_{ud-1})u_i - u_i^2 \right) \geq K',
\]

where \( f_i(\tilde{\Omega}''_{ud-1}) \) denotes the number of nonzero rows of the \( i \)-th matricization of \( \tilde{\Omega}''_{ud-1} \).

The following lemma is a re-statement of Theorem 1 in [Ashraphijuo and Wang, 2020a].

**Lemma 26.** With probability one, there are only finitely many completions of rank \( \underline{r}^* \) of the sampled tensor if and only if Conditions \( A_{\Sigma}^{TT} \) and \( B_{\Sigma}^{TT} \) hold.

**Definition 14.** Let \( S_{\Omega} \) denote the set of all rank vectors \( \underline{r} \) such that both Conditions \( A_{\Sigma}^{TT} \) and \( B_{\Sigma}^{TT} \) hold.

The following lemma will be used in Lemma 28.

**Lemma 27.** \( u_i \leq \min\{u_{i-1}n_i, u_{i+1}n_{i+1}\} \) for \( 1 \leq i \leq d - 1 \).

**Proof.** We first show that \( u_i \leq u_{i-1}n_i \), which is easily verified for \( i = 1 \) as \( \tilde{U}_1 \) includes \( n_1 \) rows and \( u_0 = 1 \), and therefore assume that \( i > 1 \). Define the \( (d-1) \)-way tensor \( \tilde{U}^{l_i} \in \mathbb{R}^{n_1 \times \cdots \times n_{i-1} \times n_{i+1} \times \cdots \times n_d} \) such that \( \tilde{U}^{l_i}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d) = U(x_1, \ldots, x_{i-1}, l_i, x_{i+1}, \ldots, x_d) \) for \( 1 \leq i \leq d \) and \( 1 \leq l_i \leq n_i \). Also, recall that \( \tilde{U}^{l_i}_{(i-1)} \) denotes the \( (i-1) \)-th unfolding of \( \tilde{U}^{l_i} \). Observe that \( \tilde{U}^{l_i}_{(i-1)} \) is a subset of columns of matrix \( \tilde{U}_{(i-1)} \) (those columns that correspond to the entries of \( U \) with the \( i \)-th component of the location equal to \( l_i \)). Therefore, \( \text{rank}(\tilde{U}^{l_i}_{(i-1)}) \leq \text{rank}(\tilde{U}_{(i-1)}) = u_{i-1} \).

On the other hand, observe that \( \tilde{U}^{l_i}_{(i-1)} \) is a subset of rows of \( \tilde{U}_{(i)} \) (those rows that correspond to the entries of \( U \) with the \( i \)-th component of the location equal to \( l_i \)). Hence, the union of rows of
Theorem 12. With probability one, exactly one of the following statements holds:

(i) $\mathbf{r}^* \in \hat{\mathcal{S}}_\Omega$;

(ii) For any arbitrary completion of the sampled tensor $\mathbf{U}$ of rank $\mathbf{r}$, we have $\mathbf{r} \notin \hat{\mathcal{S}}_\Omega$.

Proof. Similar to the proof of Theorem 8, to complete the proof it suffices to show that the assumption $\mathbf{r}^* \notin \hat{\mathcal{S}}_\Omega$ results that there exists a completion of $\mathbf{U}$ of rank $\mathbf{r}$, where $\mathbf{r} \in \hat{\mathcal{S}}_\Omega$, with probability zero. Define the multiplication $\mathbf{U}^{(1)} \ldots \mathbf{U}^{(d-1)}$ in (3.9) as the basis of the rank $\mathbf{r}$ TT decomposition of $\mathbf{U}$. Then, by considering the $(d-1)$-th unfolding of $\mathbf{U}^{(1)} \ldots \mathbf{U}^{(d-1)}$ in TT decomposition we obtain a matrix factorization of the $(d-1)$-th unfolding of $\mathbf{U}$. The rest of the proof is similar to the proof of Theorem 8. □

Corollary 17. Consider a subset $\hat{\mathcal{S}}_\Omega'$ of $\hat{\mathcal{S}}_\Omega$ such that for any two members of $\hat{\mathcal{S}}_\Omega$ that $\mathbf{r}'' \preceq \mathbf{r}'$ and $\mathbf{r}' \in \hat{\mathcal{S}}_\Omega'$ we have $\mathbf{r}'' \in \hat{\mathcal{S}}_\Omega'$. Then, with probability one, exactly one of the followings holds

(i) $\mathbf{r}^* \in \hat{\mathcal{S}}_\Omega'$;

(ii) For any arbitrary completion of $\mathbf{U}$ of rank vector $\mathbf{r}$, we have $\mathbf{r} \notin \hat{\mathcal{S}}_\Omega'$.

Corollary 18. Assuming that there exists a completion of $\mathbf{U}$ with rank vector $\mathbf{r}$ such that $\mathbf{r} \in \hat{\mathcal{S}}_\Omega$, we conclude that with probability one $\mathbf{r}^* \preceq \mathbf{r}$. 

CHAPTER 3. RANK ESTIMATION FOR LOW-RANK SAMPLED DATA

The following lemma is Lemma 14 in [Ashraphijuo and Wang, 2020a], which ensures that Conditions $A^\text{TT}_2$ and $B^\text{TT}_2$ hold with high probability.

**Lemma 29.** Define $m = \sum_{k=1}^{d-2} u_{k-1} u_k$, $M = n \sum_{k=1}^{d-2} u_{k-1} u_k - \sum_{k=1}^{d-2} u_k^2$ and $u' = \max \left\{ \frac{u_1}{u_0}, \ldots, \frac{u_{d-2}}{u_{d-3}} \right\}$. Assume that $n_1 = n_2 = \cdots = n_d = n$, $n > \max\{m, 200\}$ and $u' \leq \min\left\{ \frac{n}{6}, u_{d-2} \right\}$ hold. Moreover, assume that the sampling probability satisfies

$$p > \frac{1}{n^{d-2}} \max \left\{ 27 \log \left( \frac{n}{\epsilon} \right) + 9 \log \left( \frac{2M}{\epsilon} \right) + 18, 6u_{d-2} \right\} + \frac{1}{\sqrt{n^{d-2}}} \tag{3.34}$$

Then, with probability at least $(1 - \epsilon) \left( 1 - \exp \left( -\frac{\sqrt{n^{d-2}}}{2} \right) \right)^n$, we have $r \in S_\Omega$.

The following corollary is the probabilistic version of Corollary 18.

**Corollary 19.** Assuming that there exists a completion of the sampled tensor $U$ of TT rank $r$ such that the assumptions in Lemma 29 hold and the sampling probability satisfies (3.34), then with probability at least $(1 - \epsilon) \left( 1 - \exp \left( -\frac{\sqrt{n^{d-2}}}{2} \right) \right)^n$ we have $r^* \preceq r$.

### 3.4.3.1 Numerical Results

We generate a random tensor $U \in \mathbb{R}^{8 \times 8 \times 8 \times 8 \times 8}$ of TT-rank $(1, 2, 4, 1, 1)$. The color scale represents the lower bound on the probability that we can guarantee the rank of a given completion is a component-wise upper bound on the true rank. Then, we solve the following convex optimization problem for different values of the sampling probability.

$$\min_{U' \in \mathbb{R}^{n_1 \times \cdots \times n_d}} \left\| \sum_{i=1}^{d-1} \tilde{U}'_{(i)} \right\|_* \quad \text{(3.35)}$$

subject to $U'_{\Omega} = U_{\Omega}$.

Then, we calculate rank of each unfolding of the tensor obtained via solving (3.35) to find its TT-rank. In this scenario, for each component of the TT-rank, we find the percentage of error via

$$\frac{u_i - u_i^*}{\min\{n^i, n^{d-i}\} - u_i^*} \times 100\%$$

where $n = 8$, $d = 6$, $u_i$ and $u_i^*$ are the $i$-th rank component of the obtained tensor and original tensor, respectively. Hence, 100% error simply means that the corresponding unfolding is full rank. In Figure 3.9, gap represents the average of the defined error over all components of TT-rank, i.e., over all unfoldings.


Figure 3.9: The rank gap as a function of sampling probability for $U \in \mathbb{R}^{8 \times 8 \times 8 \times 8 \times 8}$ of TT-rank $(1, 2, 4, 1, 1)$.

### 3.5 Summary

We make use of the recently developed algebraic geometry analyses that study the fundamental conditions on the sampling patterns for finite completability under a number of low-rank matrix and tensor models to treat the problem of rank approximation for a partially sampled data. Particularly, the goal is to approximate the unknown scalar or vector rank based on the sampling pattern and the rank of a given completion. A number of data models have been treated, including single-view matrix, multi-view matrix, CP tensor, tensor-train tensor and Tucker tensor. First we have provided an upper bound on the unknown scalar rank (for single-view matrix and CP tensor) and an component-wise upper bound on the vector rank (for multi-view matrix, Tucker tensor and TT tensor) with probability one assuming that the sampling pattern satisfies the proposed combinatorial conditions. Moreover, we have also provided probabilistic versions of such bounds that hold with high probability assuming that the sampling probability is above a threshold. In addition, for single-view matrix and CP tensor, these upper bounds can be exactly equal to the unknown scalar rank given the lowest-rank completion. To illustrate how tight our proposed upper bounds are, we have provided some numerical results for the single-view matrix case in which we applied the nuclear norm minimization to find a low-rank completion of the sampled data and observe that the proposed upper bound is almost equal to the true unknown rank.
Chapter 4

Fundamental Limits For Clustering A Union of Low-Rank Matrices Of Different Dimensions With Missing Data

We derive fundamental conditions for clustering a union of low-rank subspaces with missing data. In particular, given an incomplete matrix, assuming its columns are drawn from $K$ different subspaces with different dimensions, the subspace clustering problem is to cluster the columns that belong to the same subspace. We derive a lower bound on the number of columns from each subspace such that the columns can be clustered correctly with high probability. The analysis focuses on the subspace with the lowest dimension and is a generalization of the corresponding results in [Pimentel-Alarcón and Nowak, 2016b] that assumes the subspaces are independent and with the same dimension.

4.1 Introduction

In many practical applications, we need to analyze a collection of datasets like images, text documents, etc. To model such data structures, we can consider a matrix $U \in \mathbb{R}^{n_1 \times n_2}$ whose
columns are chosen from one of $K$ unknown subspaces. The problem of subspace clustering aims to cluster the columns of this matrix to $K$ 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; Elhamifar and Vidal, 2013; Liu et al., 2013a]. The problem is much more challenging with missing data, i.e., when the matrix $U$ is incomplete, which is an important problem in subspace learning for real-world scenarios and is studied broadly [Balzano et al., 2012; Pimentel-Alarcón et al., 2016a; Pimentel-Alarcón and Nowak, 2016b; Pimentel-Alarcón et al., 2015; Yang et al., 2015]. Subspace clustering can be also used as a prepossessing step for data completion problem. Subspace clustering has many applications in various fields including image processing [Hong et al., 2006], recommender systems [Rennie and Srebro, 2005], etc.

In [Pimentel-Alarcón and Nowak, 2016b], it is assumed that all the $K$ unknown subspaces have the same dimension and are chosen independently from the Grassmannian manifold $\text{Gr}(n_1, r)$ (set of all $r$-dimensional subspaces of the $n_1$-dimensional space). It is shown that if the number of samples per column is above a threshold, and assuming that there exists an $r$-dimensional subspace that fits enough number of columns of the sampled matrix, then it is ensured that this subspace is one of the $K$ unknown subspaces and all the covered columns belong to that subspace. The key condition for this to hold is that the number of columns of $U$ drawn from each of the $K$ subspaces should be more than $(r + 1)(n_1 - r + 1) = \mathcal{O}(rn_1)$. This bound is interesting since before it was only known to be necessary when each column includes $r + 1$ sampled entries. The similar algebraic geometry approaches as in [Pimentel-Alarcón and Nowak, 2016b] have been studied in [Ashraphijuo et al., 2018; Ashraphijuo and Wang, 2017a; Ashraphijuo et al., 2017b] for data completion and sensing problem.

In this chapter, we consider the general scenario that the $K$ unknown subspaces are chosen (not necessarily independently) from $K$ different Grassmannian manifolds with different dimensions $\text{Gr}(n_1, r_1), \ldots, \text{Gr}(n_1, r_K)$. Our main result states that if at least $K(r_{\text{max}} + 1)(n_1 - r_{\text{max}} + 1)$ columns are drawn from each subspace, where $r_{\text{max}} = \max_{1 \leq k \leq K} r_k$, then the columns can be correctly clustered with high probability. The key approach in our analysis is to cluster the subspaces from the lowest dimension to the highest.
CHAPTER 4. FUNDAMENTAL LIMITS FOR CLUSTERING A UNION OF LOW-RANK MATRICES OF DIFFERENT DIMENSIONS WITH MISSING DATA

4.2 Background

Given positive integers $r_1, r_2, \ldots, r_K$, we consider $K$ different subspaces $S_1, \ldots, S_K$ chosen from the Grassmannian manifolds $\text{Gr}(n_1, r_k), k = 1, \ldots, K$. Let $I_k$ be a set of $c_k$ columns chosen generically from the mentioned $r_k$-dimensional subspace (drawn independently according to a continuous distribution with respect to the Lebesgue measure on the mentioned $r_k$-dimensional subspace), $k = 1, \ldots, K$. Assume that $U \in \mathbb{R}^{n_1 \times n_2}$ is a matrix such that its $n_2 = \sum_{k=1}^{K} c_k$ columns are the union of all columns in $\{I_k, k = 1, \ldots, K\}$. However, these $n_2$ columns are blended so that we do not know the source subspace of each column.

We assume that $U$ is randomly sampled, i.e., each entry of $U$ is independently sampled with probability $0 < p < 1$. Let $\Omega$ be an $n_1 \times n_2$ binary sampling matrix such that $\Omega(i, j) = 1$ if $U(i, j)$ is sampled and $\Omega(i, j) = 0$ otherwise. Let $U_\Omega$ denote the incomplete matrix consisting of only the sampled entries of $U$. We are interested in clustering the columns of the sampled matrix $U_\Omega$ into $K$ groups such that the columns in each group belong to one subspace with high probability. The number of columns in $I_k$, i.e., $c_k$ and the sampling probability $p$ or the number of sampled entries in $U_\Omega$ are key parameters in this problem.

In [Pimentel-Alarcón and Nowak, 2016b], the above subspace clustering problem with missing data is studied for the special case that $r_1 = r_2 = \cdots = r_K = r$ and the subspaces $S_1, \ldots, S_K$ are independent. Here, we restate the main result of [Pimentel-Alarcón and Nowak, 2016b] (i.e., Theorems 1 and 3 in [Pimentel-Alarcón and Nowak, 2016b]) using our notations in Theorem 14, which provides a lower bound on $c_k$’s and the number of samples per column (which can be translated in terms of $p$) such that the columns of $U_\Omega$ chosen from the same subspace, can be correctly clustered with high probability. First we restate Theorem 3 in [Pimentel-Alarcón et al., 2016b] that characterizes a condition on unique completability of $U_\Omega$, i.e., a condition that ensures there exists a unique way to complete $U_\Omega$ while satisfying the rank constraint, as the following theorem. This theorem is used to show the main result in [Pimentel-Alarcón and Nowak, 2016b], i.e., Theorem 14 below, as well as our new result, i.e., Theorem 15 in Section 4.3.

**Theorem 13.** [Pimentel-Alarcón et al., 2016b] Assume that a generic rank-$r$ matrix $U \in \mathbb{R}^{n_1 \times n_2}$ with $r \leq \frac{n_1}{6}$ and $n_2 \geq (r+1)(n_1-r+1)$ is randomly sampled such that each column of $U_\Omega$ includes
at least \( l \) sampled entries where

\[
l > \max\{12\log\left(\frac{n_1(r+1)}{\epsilon}\right) + 1, 2r\}
\]

for some \( 0 < \epsilon < 1 \). Then, the sampled matrix \( U_\Omega \) is uniquely completable with probability at least \( 1 - \epsilon \).

**Definition 16.** Consider a subspace \( S \in \text{Gr}(n_1, r) \) and a sampled column \( u_\Omega \in \mathbb{R}^{n_1 \times 1} \). We say that \( S \) fits \( u_\Omega \) or \( u_\Omega \) can be covered (generated) by \( S \) if there exists at least one completion of \( u_\Omega \) that belongs to \( S \).

**Theorem 14.** [Pimentel-Alarcón and Nowak, 2016b] Assume that the subspaces \( S_1, \ldots, S_K \) are independently chosen from \( \text{Gr}(n_1, r) \), \( r = r_1 = r_2 = \cdots = r_K \leq \frac{n_1}{6} \) and \( c_k \geq (r+1)(n_1 - r + 1) \), \( k = 1, \ldots, K \). Moreover suppose that each column of \( U_\Omega \) includes at least \( \epsilon \) sampled entries such that (4.1) holds. Let \( \bar{S} \) denote an \( r \)-dimensional subspace that fits exactly \( \bar{c} \) columns of \( U_\Omega \) (i.e., \( \bar{c} \) is the maximum number of columns of \( U_\Omega \) that can be covered by \( \bar{S} \)) 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 \( U_\Omega \) covered by \( \bar{S} \) belong to one source \( I_{k_0} \) for some \( 1 \leq k_0 \leq K \) and the rest of the columns of \( U_\Omega \) do not belong to \( I_{k_0} \) and moreover, \( \bar{c} = c_{k_0} \) and \( \bar{S} = S_{k_0} \).

### 4.3 Main Results

We are interested in generalizing Theorem 14 to the general scenario when \( r_1, \ldots, r_K \) are not necessarily equal and also, the subspaces \( S_1, \ldots, S_K \) are not chosen independently.

We start by stating some basic properties as a consequence of the genericity assumption. Consider a matrix \( X \in \mathbb{R}^{n_1 \times n_2} \) whose columns are drawn generically from a subspace that belongs to \( \text{Gr}(n_1, r) \), where \( n_1 \geq r \) and \( n_2 \geq r \). Then, with probability one, \( X \) is a rank-\( r \) matrix. More specifically, with probability one, any \( r \) columns (or any \( r \) rows) of \( X \) are linearly independent; and any \( r \times r \) submatrix of \( X \) is full-rank. Further, given two different subspaces \( S_1 \subset S_2 \) and a column \( u \) that is drawn generically from \( S_2 \), we have \( u \notin S_1 \) with probability one.

The following three lemmas are instrumental to the proof of our clustering result, i.e., Theorem 15.
Lemma 30. Let $X_0$ be a rank-$(r - 1)$ matrix and $X_0(i, j) = x$ be an entry of this matrix. Assume that changing the value of entry $X_0(i, j)$ from $x$ to $y$ results in $X'_0$, which is a rank-$r$ matrix. Then, there are infinitely many scalars $z$ such that changing the value of entry $X_0(i, j)$ from $x$ to $z$ results in a rank-$r$ matrix.

Proof. Since $X_0$ is a rank-$(r - 1)$ matrix, the determinant of any $r \times r$ submatrix of $X_0$ is zero. Moreover, $X'_0$ is a rank-$r$ matrix and hence, there exists an $r \times r$ submatrix $X_r$ of $X_0$ such that changing the value of the corresponding entry of $X_r$ from $x$ to $y$ results in a non-zero determinant. Since changing the order of the rows and columns does not affect the values of rank and determinant, we can assume that $X_r(1, 1) = X_0(i, j) = x$. Hence, $\det(X_r) = 0$ and changing the value of $X_r(1, 1)$ from $x$ to $y$ makes the determinant of $X_r$ non-zero. On the other hand, we have

$$0 = \det(X_r) = X_r(1, 1)\det(X_r(2 : r, 2 : r)) - \Sigma_{i=2}^{r}(-1)^i X_r(1, i)\det(X_r(2 : r, \{1, \ldots, r\} \setminus \{i\})),$$

or equivalently,

$$x\det(X_r(2 : r, 2 : r)) = \Sigma_{i=2}^{r}(-1)^i X_r(1, i)\det(X_r(2 : r, \{1, \ldots, r\} \setminus \{i\})).$$

(4.2)

Moreover, we have

$$y\det(X_r(2 : r, 2 : r)) \neq \Sigma_{i=2}^{r}(-1)^i X_r(1, i)\det(X_r(2 : r, \{1, \ldots, r\} \setminus \{i\})).$$

(4.3)

Note that if $\det(X_r(2 : r, 2 : r)) = 0$, then both sides of (4.3) are zero; and hence both sides of (4.4) are zero; which contradicts the inequality in (4.4). Hence, $\det(X_r(2 : r, 2 : r)) \neq 0$, then (4.3) and (4.4) can be written as

$$y \neq x = \frac{\Sigma_{i=2}^{r}(-1)^i X_r(1, i)\det(X_r(2 : r, \{1, \ldots, r\} \setminus \{i\}))}{\det(X_r(2 : r, 2 : r))}.$$

(4.5)

Therefore, changing the value of $X_r(1, 1)$ (i.e., $X_0(i, j)$) from $x$ to any $z \neq x$ leads $X_r$ to an $r \times r$ full-rank matrix. As a result, there are infinitely many scalars $z$ such that changing the value of
CHAPTER 4. FUNDAMENTAL LIMITS FOR CLUSTERING A UNION OF LOW-RANK MATRICES OF DIFFERENT DIMENSIONS WITH MISSING DATA

$X_0(i,j)$ from $x$ to $z$ results in the existence of a full-rank $r \times r$ submatrix, i.e., results in a matrix with rank at least $r$.

On the other hand, changing the value of only one entry of a matrix can affect the rank of the matrix by at most one, i.e., the rank can decrease or increase by one or stay the same. This is because changing one single entry of the matrix affects only one column of the matrix. Hence, the rank cannot decrease or increase by more than one. Therefore, there are infinitely many scalars $z$ such that changing the value of $X_0(i,j)$ from $x$ to $z$ results in a rank-$r$ matrix.

Lemma 31. Consider a sampled matrix $X_\Omega$ such that there exist at least one rank-$(r-1)$ and one rank-$r$ completion for some $r > 1$. Then, there exists infinitely many rank-$r$ completions of $X_\Omega$.

Proof. Note that changing the value of only one entry of a matrix results in changing the rank of the matrix by at most one. Let $X_1$ and $X_2$ denote the rank-$(r-1)$ and rank-$r$ completions, respectively. $X_1$ and $X_2$ are the same over the sampled entries, i.e., $(X_1)_\Omega = (X_2)_\Omega$, and their difference is only over some of the non-sampled entries. We start changing the value of non-sampled entries of $X_1$ one by one to the value of the corresponding non-sampled entries of $X_2$, which will eventually result in $X_2$ 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 matrix is $r-1$ 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 $X_3$ of the sampled matrix $X_\Omega$ such that changing the value of some entry $X_3(i,j)$ from $v$ to $v'$ increases the rank to $r$ for some scalars $v$ and $v'$. The rest of the proof is straight-forward due to Lemma 30.

Lemma 32. Consider a sampled matrix $X_\Omega$ such that there exist at least one rank-$(r-i)$ and one rank-$r$ completion for some $i < r$. Then, there exist infinitely many rank-$r$ completions of $X_\Omega$.

Proof. Using the same process described in the proof of Lemma 31, i.e., changing the values of the non-sampled entries of the rank-$(r-i)$ completion to reach to the rank-$r$ completion, it is trivial to see that there exists at least one rank-$(r-1)$ completion as well. Hence, according to Lemma 31, there exists infinitely many rank-$r$ completions of $X_\Omega$.

The following theorem extends Theorem 14 to the general case and provides the conditions to
Theorem 15. Without loss of generality, assume that \( r_1 \leq r_2 \leq \ldots \leq r_K \) and denote \( r_{\text{max}} = \max_{1 \leq k \leq K} r_k = r_K \). Assume further that \( r_{\text{max}} \leq \frac{n_1}{6}, \ c_k \geq K(r_{\text{max}} + 1)(n_1 - r_{\text{max}} + 1), \ k = 1, \ldots, K, \) and also, each column of \( U_{\Omega} \) includes at least \( l \) sampled entries such that \( l > \max\{12(\log(\frac{n_1(r_{\text{max}} + 1)}{\epsilon})) + 1, 2r_{\text{max}}\} \). Let \( \tilde{S} \) denote an \( r_1 \)-dimensional subspace that fits exactly \( \bar{c} \) columns of \( U_{\Omega} \) (i.e., \( \bar{c} \) is the maximum number of columns of \( U_{\Omega} \) that can be covered by \( \tilde{S} \)) and assume that \( \bar{c} \geq K(r_{\text{max}} + 1)(n_1 - r_{\text{max}} + 1) \). Then, with probability at least \( 1 - \epsilon \) the following statement holds: All the \( \bar{c} \) columns of \( U_{\Omega} \) covered by \( \tilde{S} \) belong to one source \( I_{k_0} \) for some \( 1 \leq k_0 \leq K \) 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 \( U_{\Omega} \) do not belong to \( I_{k_0} \) and moreover, \( \bar{c} = c_{k_0} \) and \( \tilde{S} = S_{k_0} \).

Proof. According to pigeonhole principle, at least \( \lceil \frac{n_1}{r_1} \rceil \geq (r_{\text{max}} + 1)(n_1 - r_{\text{max}} + 1) \) columns of the \( \bar{c} \) covered columns by \( \tilde{S} \) are chosen from one source \( I_{k_0} \). Note that due to the assumptions \( r_{\text{max}} \geq r_{k_0} \) and \( r_{\text{max}} \leq \frac{n_1}{6} \) we have \( (r_{\text{max}} + 1)(n_1 - r_{\text{max}} + 1) \geq (r_{k_0} + 1)(n_1 - r_{k_0} + 1) \) and hence, there are at least \( (r_{k_0} + 1)(n_1 - r_{k_0} + 1) \) columns covered by \( \tilde{S} \) that are chosen from one source \( I_{k_0} \). Then, according to Theorem 13, there exists a unique rank-\( r_{k_0} \) completion for the mentioned \( (r_{k_0} + 1)(n_1 - r_{k_0} + 1) \) columns with probability at least \( 1 - \epsilon \). In the rest of the proof, we assume the mentioned unique completability holds and show the mentioned statement holds with probability one.

First, we show that \( r_{k_0} = r_1 \). By contradiction, assume otherwise that \( r_1 < r_{k_0} \). Recall that \( \tilde{S} \) is an \( r_1 \)-dimensional subspace that fits the mentioned \( (r_{k_0} + 1)(n_1 - r_{k_0} + 1) \) columns and hence, there exists a rank-\( r_1 \) completion of these columns. Hence, according to Lemma 32, there exist infinitely many rank-\( r_{k_0} \) completions of these columns, which contradicts the earlier uniqueness assumption. As a result, we have \( r_{k_0} = r_1 \) with probability one.

Therefore, again according to the uniqueness of rank-\( r_{k_0} \) completion assumption, and due to the fact that both subspaces \( \tilde{S} \) and \( S_{k_0} \) are \( r_1 \)-dimensional (since \( r_{k_0} = r_1 \)) and they both fit the mentioned \( (r_{k_0} + 1)(n_1 - r_{k_0} + 1) \) columns, we simply conclude \( \tilde{S} = S_{k_0} \). Consequently, \( \tilde{S} \) covers all \( c_{k_0} \) columns of \( U_{\Omega} \) that belong to \( I_{k_0} \). In order to complete the proof, it suffices to show that \( \bar{c} = c_{k_0} \), i.e., \( \tilde{S} \) does not cover any other column of \( U_{\Omega} \) that belongs to other sources \( I_k \) for \( k \neq k_0 \), with probability one.
CHAPTER 4. FUNDAMENTAL LIMITS FOR CLUSTERING A UNION OF LOW-RANK MATRICES OF DIFFERENT DIMENSIONS WITH MISSING DATA

Note that any column chosen from sources other than $I_{k_0}$ does not belong to $S_{k_0}$ with probability one (this statement is not valid if $r_{k_0} \neq \min\{r_1, r_2, \ldots, r_K\}$ as will be discussed in Remark 14). This is because none of the other subspaces can be a subspace of $S_{k_0}$ as $r_{k_0} = r_1 = \min\{r_1, r_2, \ldots, r_K\}$.

By contradiction, assume that a column $u_\Omega$ of $U_\Omega$ is chosen from $I_{k_1}$ (for some $k_1 \neq k_0$) and it can be covered by $\bar{S}$. Recall that $t > \max\{12(\log(n_1(r_{\max}+1)) + 1), 2r_{\max}\}$ holds and therefore, $u_\Omega$ includes at least $2r_{\max} \geq 2r_1$ sampled entries. Now, consider $r_1$ random columns of $U_\Omega$ that belong to $I_{k_0}$ and denote it by $U_{0\Omega}$. Also, let the unique completion of $U_{0\Omega}$ be $U_0$. Then, define $U_{1\Omega} = [U_0|u_\Omega] \in \mathbb{R}^{n_1 \times (r_1+1)}$ (where $U_0$ denotes the corresponding $r_1$ columns of the unique completion that is not given to us and $u_\Omega$ is an incomplete column; so only the last column of $U_{1\Omega}$ is incomplete) and consider an $(r_1 + 1) \times (r_1 + 1)$ submatrix of $U_{1\Omega}$ that includes $r_1 + 1$ of the sampled entries of $u_\Omega$ and denote it by $U_1'$. Since $r_{k_1} \geq r_{k_0} = r_1$ and $S_{k_1} \neq S_{k_0}$ (because $k_1 \neq k_0$), we conclude that $\text{rank}(U_1) = r_1 + 1$ and hence, $\text{rank}(U_1') = r_1 + 1$ with probability one, where $U_1$ denotes the original (before sampling) matrix corresponding to $U_{1\Omega}$. Hence, for any completion of $U_{1\Omega}$, there exists a “fixed” and full-rank $(r_1 + 1) \times (r_1 + 1)$ submatrix. Therefore, $\bar{S}$ cannot fit $u_\Omega$ with probability one (since $\bar{S}$ is an $r_1$-dimensional subspace) and the proof is complete due to this contradiction.

\[\square\]

Remark 14. Note that the above proof is valid since $r_{k_0} = r_1 = \min\{r_1, r_2, \ldots, r_K\}$, as mentioned in the last part of the proof. Moreover, we can show that if $r_{k_0} \neq \min\{r_1, r_2, \ldots, r_K\}$, the statement of the theorem does not hold. For example, consider the scenario when $r_1 < r_2 < \cdots < r_K$ and $S_k$ is a subspace of $S_{k+1}$ (this can happen as the subspaces are not necessarily independent), $k = 1, \ldots, K - 1$. Now, assume that $\bar{S}$ in the statement of the above theorem is $r_2$-dimensional instead of $r_1$-dimensional. Then if $\bar{S} = S_2$, $\bar{S}$ also fits the columns drawn from $S_1$ (recall that $S_1$ is a subspace of $S_2$) and hence, we cannot distinguish the columns drawn from $S_1$ and $S_2$.

Remark 15. Theorem 15 requires $K$ times more columns from each unknown subspace in comparison with Theorem 14 to identify the columns of one subspace. However, Theorem 15 does not require all ranks to be the same or the independent subspace assumption. Moreover, the probability of clustering failure in Theorem 15 is $K$ times less than that in Theorem 14.

After identifying all columns chosen from an $r_1$-dimensional subspace correctly, we can exclude the identified columns from the sampled matrix. Then, the problem reduces to the similar problem
with $K - 1$ subspaces of ranks $r_2 \leq \ldots \leq r_K$ and a smaller number of columns for the sampled matrix. Hence, the same analysis is applicable again.

Specifically, let $\tilde{S}_1, \ldots, \tilde{S}_K'$ (for some $1 \leq K' < K$) denote different $r_1, \ldots, r_K'$-dimensional subspaces that fit exactly $\tilde{c}_1, \ldots, \tilde{c}_K'$ columns of $U_{\Omega}$ (i.e., $\tilde{c}_k$ is the maximum number of columns of $U_{\Omega}$ that can be covered by $\tilde{S}_k$), respectively, and assume that $\tilde{c}_k \geq K(r_{\max} + 1)(n_1 - r_{\max} + 1)$, $k = 1, \ldots, K'$. Moreover, assume that there exist $K(r_{\max} + 1)(n_1 - r_{\max} + 1)$ columns covered by $\tilde{S}_k$ that cannot be covered by any of $\tilde{S}_1, \ldots, \tilde{S}_{k-1}$, $k = 1, \ldots, K'$.

Then, according to Theorem 15, we have $\tilde{S}_1 = S_{k_1}$ and $\tilde{c}_1 = c_{k_1}$ with probability at least $1 - \epsilon$. Moreover, we can exclude all the $c_{k_1}$ columns from the sampled matrix and the identified subspace $S_{k_1}$. Then, the new sampled matrix is $n_1 \times (n_2 - c_{k_1})$ and the columns of this matrix are chosen from the $K - 1$ remaining subspaces. Then, similarly, we apply Theorem 15 for the $r_2$-dimensional subspace $\tilde{S}_2$ that has the lowest dimension now (because one $r_1$-dimensional subspace has been excluded).

Now, assuming that the clustering of the $c_{k_1}$ columns in the previous step was correct, we can cluster the columns of the next subspace correctly with probability at least $1 - \epsilon$. This can be done because due to the assumption, after excluding the columns of the first cluster, there exist $K(r_{\max} + 1)(n_1 - r_{\max} + 1)$ columns covered by $\tilde{S}_2$ that cannot be covered by $\tilde{S}_1$. Hence, we apply Theorem 15 again and therefore, with probability at least $(1 - \epsilon)^2$ the following statement holds: All the $\tilde{c}_k$ columns of $U_{\Omega}$ covered by $\tilde{S}_k$ belong to one source $\mathcal{I}_{k'}$ such that $r_{k'} = r_k$ and the rest of the columns of $U_{\Omega}$ do not belong to $\mathcal{I}_{k'}$ and moreover, $\tilde{c}_k = c_{k'}$ and $\tilde{S}_k = S_{k'}$, $k = 1, 2$. By simply repeating this procedure, we conclude the following corollary.

**Corollary 20.** Without loss of generality, assume that $r_1 \leq r_2 \leq \ldots \leq r_K$ and denote $r_{\max} = \max_{1 \leq k \leq K} r_k = r_K$. Assume further that $r_{\max} \leq \frac{n_1}{n}$, $c_k \geq K(r_{\max} + 1)(n_1 - r_{\max} + 1)$, $k = 1, \ldots, K$, and also, each column of $U_{\Omega}$ includes at least $l$ sampled entries such that $l > \max\{12(\log(\frac{21(r_{\max} + 1)}{\epsilon}) + 1), 2r_{\max}\}$. Let $\tilde{S}_1, \ldots, \tilde{S}_K'$ (for some $1 \leq K' < K$) denote different $r_1, \ldots, r_K'$-dimensional subspaces that fit exactly $\tilde{c}_1, \ldots, \tilde{c}_K'$ columns of $U_{\Omega}$ (i.e., $\tilde{c}_k$ is the maximum number of columns of $U_{\Omega}$ that can be covered by $\tilde{S}_k$), respectively, and assume that $\tilde{c}_k \geq K(r_{\max} + 1)(n_1 - r_{\max} + 1)$, $k = 1, \ldots, K'$. Moreover, assume that there exist $K(r_{\max} + 1)(n_1 - r_{\max} + 1)$ columns covered by $\tilde{S}_k$ that cannot be covered by any of $\tilde{S}_1, \ldots, \tilde{S}_{k-1}$, $k = 1, \ldots, K'$. Then, with probability at least $(1 - \epsilon)^K'$ the following statement holds: All the $\tilde{c}_k$ columns of $U_{\Omega}$ covered by $\tilde{S}_k$ belong to one
CHAPTER 4. FUNDAMENTAL LIMITS FOR CLUSTERING A UNION OF LOW-RANK MATRICES OF DIFFERENT DIMENSIONS WITH MISSING DATA

source $I_k'$ such that $r_{k'} = r_k$ and the rest of the columns of $U_{\Omega}$ do not belong to $I_k'$ and moreover, $c_k = c_{k'}$ and $S_k = S_{k'}$, $k = 1, \ldots, K'$.

We would like to emphasize the advantage of our results when the number of sampled entries are as low as $O(n_1 r_{\max})$ per column. Please refer to [Pimentel-Alarcón and Nowak, 2016b] to see the discussion on how tight our information-theoretic bounds on the number of samples are in comparison with the theoretical bounds in the existing works on subspace clustering with missing data. Moreover, our results in this chapter not only improved the bound on the number of sampled entries in [Pimentel-Alarcón and Nowak, 2016b], but also removed the strong restrictions such as independency of the subspaces or subspaces being of the same size.

4.4 Numerical Experiments

Assume that $n_1 = 1000$ and $c_1 = c_2 = c_3 = 600000$. We construct $K = 3$ matrices of rank $r_i$ by multiplying a random $n_1 \times r_i$ matrix by a random $r_i \times c_i$ matrix. We assume each entry is sampled uniformly and independently with some sampling probability $p$. Since in our probabilistic analysis, only the maximum rank $r_{\max}$ matters (our bounds and analyses are based on the maximum rank), the $x$-axis in Figure 4.1 represents the maximum rank. Also, the $y$-axis represents the required sampling probability. Then, using Corollary 20, the average number of required samples to guarantee the correct clustering with probability at least $1 - \epsilon$ is

$$\frac{(\max\{12(\log(n_1(r_{\max}+1)/\epsilon) + 1), 2r_{\max}\})}{n_1}.$$ 

Hence, in Figure 4.1, we have provided several curves to represent the value of sampling probability and certainty value using our analysis. Each curve represents the probability of sampling (for different rank value) such that according to Corollary 20, we can guarantee the correct clustering with probability at least $1 - \epsilon$, for different values of $\epsilon$.

Note that our analysis is more efficient for relatively low-rank scenarios. This is because as long as $2r_{\max} < 12(\log(n_1(r_{\max}+1)/\epsilon) + 1)$, we basically provide a very tight bound on the number of samples to for correctly clustering with probability $1 - \epsilon$. However, as we need $2r_{\max}$ samples as well in Corollary 20 (since we used Theorem 13), we can observe that by increasing the value of rank to a very large number (high-rank scenarios) the bound can be slightly weak and $\epsilon$ disappears in the curves as it means we can guarantee the correct clustering with probability almost 1.
4.5 Summary

We have developed a generalization to the low-rank subspace clustering conditions in [Pimentel-Alarcón and Nowak, 2016b]. In particular, given an incomplete matrix whose columns are drawn from $K$ independent subspaces with the same dimension, a lower bound on the number of columns from each subspace is given in [Pimentel-Alarcón and Nowak, 2016b], such that, with high probability, the columns are clustered correctly. In order to treat the general case that the subspaces are not independently chosen, and their dimensions can be different, we have provided a new analysis that leads to the lower bound on the number of columns from each subspace, for the general case, which is $K$ times that in [Pimentel-Alarcón and Nowak, 2016b]; however, the probability of clustering failure is reduced by a factor of $K$ compared with that in [Pimentel-Alarcón and Nowak, 2016b]. The key approach in our analysis is to focus on the subspace of the lowest dimension.
Chapter 5

Fundamental Limits For Clustering A Union of Low-Rank Tensor Spaces With Missing Data

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.
5.1 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 $U \in \mathbb{R}^{n_1 \times n_2}$ 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., 2013a; Elhamifar and Vidal, 2013]. The problem is much more challenging with missing data, i.e., when the matrix $U$ is incomplete, which is an important problem in subspace learning for real-world scenarios and has been treated broadly [Balzano et al., 2012; Yang et al., 2015]. In particular, the information-theoretic bounds on clustering the union of subspaces is investigated in [Pimentel-Alarcón et al., 2016a; Pimentel-Alarcón et al., 2017a; Pimentel-Alarcón and Nowak, 2016b], based on the geometrical analysis for matrix completion problems [Pimentel-Alarcón et al., 2016b].

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 chapter. In general, subspace clustering has many applications in various fields including image processing [Hong et al., 2006], recommender systems [Rennie and Srebro, 2005], etc. 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.
CHAPTER 5. FUNDAMENTAL LIMITS FOR CLUSTERING A UNION OF LOW-RANK TENSOR SPACES WITH MISSING DATA

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].

The remainder of this chapter is organized as follows. In Section 5.2, some preliminaries are presented. In Section 5.3, we state the problem treated in this chapter, i.e., tensor space clustering with missing data, and provide the matrix analysis approach to this problem. Tensor analysis for clustering tensor spaces is presented in Section 5.4. Finally, Section 5.5 summarizes the chapter.

5.2 Preliminaries

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

$$U = \sum_{l=1}^{r} a^l_1 \otimes a^l_2 \otimes \ldots \otimes a^l_d,$$

or equivalently,

$$U(x_1, x_2, \ldots, x_d) = \sum_{l=1}^{r} a^l_1(x_1)a^l_2(x_2)\ldots a^l_d(x_d),$$

where $\otimes$ denotes the tensor product (outer product) and $U(x_1, x_2, \ldots, x_d)$ denotes the entry of tensor $U$ with coordinate $\vec{x} = (x_1, x_2, \ldots, x_d)$ and $a^l_i(x_i)$ denotes the $x_i$-th entry of vector $a^l_i$. Note that $a^l_1 \otimes a^l_2 \otimes \ldots \otimes a^l_d \in \mathbb{R}^{n_1 \times \ldots \times n_d}$ is a rank-1 tensor, $l = 1, 2, \ldots, r$.

For notational convenience, define $N_{-i} \triangleq \prod_{j=1}^{i-1} n_j$, $N_i \triangleq \left(\prod_{j=1}^{i} n_j\right)$, $\bar{N}_i \triangleq \left(\prod_{j=i+1}^{d} n_j\right)$. Also, define $x^+ \triangleq \max\{0, x\}$.

Definition 17. Define the matrix $\bar{U}_{(i)} \in \mathbb{R}^{N_i \times \bar{N}_i}$ as the $i$-th unfolding of tensor $U$, such that $U(\vec{x}) = \bar{U}_{(i)}(\bar{M}_i(x_1, \ldots, x_i), \bar{M}_i(x_{i+1}, \ldots, x_d))$, where $\bar{M}_i : (x_1, \ldots, x_i) \rightarrow \{1, 2, \ldots, N_i\}$ 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 18.** Define the matrix $U_{(i)} \in \mathbb{R}^{n_i \times N_i}$ as the $i$-th matricization of tensor $U$, such that $U(\vec{x}) = U_{(i)}(x_i, M_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d))$, 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 19.** We call a column $u \in \mathbb{R}^{n_i}$ an “$i$-th unfolded column,” $i = 1, \ldots, d - 1$, if there exist $u_s \in \mathbb{R}^{n_s}$ for $s = 1, \ldots, i$, such that $u = \text{vec}(u_1 \otimes \ldots \otimes u_i)$, where $\text{vec}(\cdot)$ uses the same bijective mappings that is used for the $i$-th unfolding, i.e., $M_i : (x_1, \ldots, x_i) \rightarrow \{1, 2, \ldots, N_i\}$, and hence $u(M_i((x_1, \ldots, x_i)) = u_1(x_1) \ldots u_i(x_i)$. 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 33.** The matrix rank of any unfolding or matricization of $U$ 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 $\text{rank}(\tilde{U}_{(i)}) \leq r$, it suffices to show that there exist $b_1^l \in \mathbb{R}^{N_i}$ and $b_2^l \in \mathbb{R}^N$ for $1 \leq l \leq r$ such that

$$\tilde{U}_{(i)} = \sum_{l=1}^{r} b_1^l \otimes b_2^l. \quad (5.3)$$

Recall the CP decomposition in (5.1). Then, we define $A_1^l = a_1^l \otimes \ldots \otimes a_d^l$ and $A_2^l = a_{i+1}^l \otimes \ldots \otimes a_d^l$ for $1 \leq l \leq l$ and let $b_1^l$ and $b_2^l$ denote the vectorizations of $A_1^l$ and $A_2^l$ with the bijective mappings $\tilde{M}_i : (x_1, \ldots, x_i) \rightarrow \{1, 2, \ldots, N_i\}$ and $\tilde{M}_i : (x_{i+1}, \ldots, x_d) \rightarrow \{1, 2, \ldots, \tilde{N}_i\}$ of the unfolding $\tilde{U}_{(i)}$, respectively. Hence, there exist $b_1^l \in \mathbb{R}^{N_i}$ and $b_2^l \in \mathbb{R}^{\tilde{N}_i}$ for $1 \leq l \leq r$ such that (5.3) holds. Similarly, for the matricization $U_{(i)}$, we can write $U_{(i)} = \sum_{l=1}^{r} c_1^l \otimes c_2^l$, where $c_1^l = a_1^l \in \mathbb{R}^{n_i}$ and $c_2^l \in \mathbb{R}^{N_i}$ is the vectorization of $a_1^l \otimes \ldots \otimes a_{i-1}^l \otimes a_{i+1}^l \otimes \ldots \otimes a_d^l$ 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 $U_{(i)}$. \hfill \square

**Lemma 34.** The CP-rank of a tensor $U$ is equal to the minimum number of structured columns that span all columns of $\tilde{U}_{(d-1)}$.

**Proof.** Recall from Lemma 33 that there exist $b_1^l \in \mathbb{R}^{N_i}$ and $b_2^l \in \mathbb{R}^{N_i}$ for $1 \leq l \leq r$ such that (5.3) holds. Define $B_1 = [b_1^1 | b_1^2 | \ldots | b_1^r] \in \mathbb{R}^{N_i \times r}$ and $B_2 = [b_2^1 | b_2^2 | \ldots | b_2^r]^T \in \mathbb{R}^{r \times N_i}$. Then, (5.3)
can be rewritten as $\widetilde{U}(i) = B_1B_2$ and therefore, there exist $r$ columns such that each one is an $i$-th unfolded column (columns of $B_1$) and each column of the $i$-th unfolding of $U$ can be written as a linear combination of the mentioned $r$ columns. Moreover, the columns of $B_2^\top$ 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 $B_1$ are “structured columns” and the columns of $B_2^\top$ do no have any particular structure, i.e., $B_2$ is an arbitrary matrix. Therefore, $\text{rank}(U) = r$ means that $r$ is the minimum number of structured columns that span all columns of $\widetilde{U}(d-1)$.

**Remark 16.** $\text{rank}(U) = r$ concludes that there exists a set $S$ consisting of $r$ structured columns whose column span (denoted by $T$) includes any column of $\widetilde{U}(d-1)$. In other words, the column span of these $r$ structured columns, i.e., $T$, is an unfolded tensor space of rank $r$. In fact, the image of the bijective mapping $(\widetilde{M}_{d-1}, \widetilde{M}_{d-1})$ with the domain of space of all $d$-way tensors generated by this unfolded tensor space is $T$.

### 5.3 Problem Statements And Matrix Approaches

#### 5.3.1 Problem Statements

Assume that $U_k \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_{d-1} \times c_k}$ is a $d$-way tensor, $k = 1, 2, \ldots, K$. Define $\mathcal{U} \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_d}$ 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 $U_k$, $k = 1, 2, \ldots, K$. We assume that $\mathcal{U}$ is (or $U_k$’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 $\Omega(\bar{x}) = 1$ if $\mathcal{U}(\bar{x})$ is sampled and $\Omega(\bar{x}) = 0$ otherwise. Let $\mathcal{U}_\Omega$ denote the incomplete tensor consisting of only the sampled entries of $\mathcal{U}$. We assume that the rank values $r_1, \ldots, r_K$ and the sampled tensor $\mathcal{U}_\Omega$ are given. Furthermore, we assume that $U_k$ 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 $U_k$ is drawn independently according to a continuous uniform distribution with respect to the Lebesgue measure on $\mathbb{R}$, $k = 1, \ldots, K$.

For the sake of notational simplicity, instead of working with tensors, we look into the image of the bijective mapping $(\widetilde{M}_{d-1}, \widetilde{M}_{d-1})$ with the domain $\mathcal{U}$, i.e., $\widetilde{U}(d-1)$. Specifically, consider the set $S_k$ consisting of $r_k$ structured columns (i.e., $(d-1)$-th unfolded columns) that are chosen generically,
CHAPTER 5. FUNDAMENTAL LIMITS FOR CLUSTERING A UNION OF LOW-RANK TENSOR SPACES WITH MISSING DATA

$k = 1, \ldots, K$. 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 $T_k$ denote the column span of $S_k$. Then, consider a matrix $\tilde{U}_k^{(d-1)} \in \mathbb{R}^{n_{d-1} \times c_k}$ such that $c_k \geq r_k$ and the columns of $\tilde{U}_k^{(d-1)}$ are drawn generically from the column span of $S_k$. Hence, the folded $d$-way tensor corresponding to $\tilde{U}_k^{(d-1)}$ (denoted by $U_k$) is of CP-rank $r_k$ and the column span of $S_k$ provides an unfolded tensor space of rank $r_i$, $i = 1, \ldots, k$.

5.3.1.1 Tensor Space Clustering With Missing Data

In Section 5.4, we are interested in clustering the subtensors of $U$ of size $\mathbb{R}^{n_1 \times n_2 \times \ldots \times n_{d-1} \times 1}$ in $K$ groups such that members of each group belong to the same $d$-dimensional source of size $n_1 \times n_2 \times \ldots \times n_{d-1} \times c_k$ and rank $r_k$ with high probability. This is equivalent to clustering the columns of $\tilde{U}_k^{(d-1)}$ such that their corresponding tensors (folded versions) belong to the same tensor space.

Assuming that $I_k$ denotes a set of $c_k$ columns generically chosen from $T_k$, i.e., the column span of the unfolded tensor basis $S_k$, $k = 1, \ldots, K$, and we call $\{I_1, \ldots, I_K\}$ the $K$ unknown sources. In fact, the $(d - 1)$-th unfolding $\tilde{U}_k^{(d-1)} \in \mathbb{R}^{n_{d-1} \times c_k}$ of tensor $U_k$ consists of the $c_k$ columns in $I_k$. As mentioned earlier, in the clustering problem, we assume that $\sum_{k=1}^K c_k$ structured columns of $I_k$’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 $I_k$ that each structured column is chosen from with high probability.

5.3.2 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 $I_k$ and simply apply the subspace clustering results for matrices.

First, we restate the main result of [Pimentel-Alarcón and Nowak, 2016b] (i.e., Theorems 1 and 3 in [Pimentel-Alarcón and Nowak, 2016b]). Consider a similar problem as described in Section 5.3.1 for two-dimensional data, or matrix. Let $T_1, \ldots, T_K$ be subspaces independently chosen from $\text{Gr}(n_1, r)$ (set of all $r$-dimensional subspaces of the $n_1$-dimensional space), and source $I_k$ includes $c_k$ columns generically chosen from $T_k$, $k = 1, \ldots, K$. The matrix $U \in \mathbb{R}^{n_1 \times n_2}$ is such that $n_2 = \sum_{k=1}^K c_k$ and includes all columns of $I_k$ for $k = 1, \ldots, K$. $\Omega$ is an $n_1 \times n_2$ binary sampling
Lemma 35. [Pimentel-Alarcón and Nowak, 2016b] Assume that the subspaces \( \mathcal{T}_1, \ldots, \mathcal{T}_K \) are independently chosen from \( \text{Gr}(n_1, r) \), \( r \leq \frac{n_1}{6} \) and \( c_k \geq (r + 1)(n_1 - r + 1) \), \( k = 1, \ldots, K \). Moreover suppose that each column of \( U_{\Omega} \) includes at least \( l \) sampled entries such that

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

(5.4)

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

Remark 17. 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 5.3.1, assume that \( r_1 = \cdots = r_K = r \) and the tensor spaces are chosen independently. Then, we can simply apply Lemma 35 to \( \tilde{U}_{(d-1)} \) to cluster one of the tensor spaces if

\[
c_k \geq (r_{\text{max}} + 1)(N_{-d} - r_{\text{max}} + 1), \quad k = 1, \ldots, K,
\]

(5.5)

and

\[
l > \max\{12 \log \left( \frac{N_{-d}(r_{\text{max}} + 1)}{\epsilon} \right) + 12, 2r_{\text{max}} \}.
\]

(5.6)

Note that the assumptions such as \( r_1 = \cdots = r_K = r \), the tensor spaces are chosen independently, and (5.5) are very strong and we are interested in clustering the tensor spaces without requiring such assumptions.

Remark 18. The reason that the above naive method is not very efficient and requires a very strong assumption (5.5) is that we completely ignored the “unfolded structure” of the columns of \( \tilde{U}_{(d-1)} \) 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.
5.3.3 Matrix Analysis Approach To The Completion Problem

A naive approach is again to ignore the unfolding structure of the columns of $\tilde{U}_{(d-1)}$ and treat it as a union of low-rank matrices completion problem. For notational clarity, denote $V \triangleq \tilde{U}_{(d-1)} \in \mathbb{R}^{N_d \times (c_1 + \cdots + c_k)}$, $V_k \triangleq \tilde{U}_{k(d-1)} \in \mathbb{R}^{N_d \times (c_1 + \cdots + c_k)}$. 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 $\text{rank} (V_k) \leq r_k, k = 1, \ldots, K$. Moreover, the tensor space structure for the completion problem $(S_1 \subseteq S_2 \subseteq \ldots \subseteq S_K)$ simply results that span $(V_1) \subseteq \text{span} (V_2) \subseteq \ldots \subseteq \text{span} (V_K)$. Now, let us consider the special case when $\text{rank} (V_k) = r_k, k = 1, \ldots, K$ and we know that given span $(V_1) \subseteq \text{span} (V_2) \subseteq \ldots \subseteq \text{span} (V_K)$, the following lemma found in [Ashraphijuo and Wang, 2018] gives a lower bound on the number of samples such that $V$ is uniquely completable.

**Lemma 36.** Assume that $r_k \leq N_d$ and $c_k \geq (r_k - r_{k-1} + 1)(N_d - r_k)$ for $k = 1, \ldots, K$, and each column of $\Omega$ (i.e., $\tilde{\Omega}_{(d-1)}$ for $U$) includes at least $l$ nonzero entries where

$$l > \max \left\{ 9 \log \left( \frac{N_d}{\epsilon} \right) + 3 \log \left( \frac{\max_{1 \leq k \leq K} \{r_k - r_{k-1}\} 2K}{\epsilon} \right) + 6, 2r_K \right\}. \quad (5.7)$$

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

Recall that $N_d = n_1 n_2 \ldots n_{d-1}$ 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. (5.5)].

5.4 Tensor Space Clustering: Main Results

5.4.1 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, 2017a], 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 $U \in \mathbb{R}^{n_1 \times \ldots \times n_d}$ 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, 2017a].

**Lemma 37.** 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 $\tilde{U}_{(d-1)}$ includes at least $l$ nonzero entries, where

$$l > \max \left\{ 27 \log \left( \frac{2 \max_{1 \leq i \leq d-1} n_i}{\epsilon} \right) + 9 \log \left( \frac{8r(d-1)}{\epsilon} \right) + 18, 6r \right\}. \quad (5.8)$$

Then, with probability at least $1 - \epsilon$, there exist only one completion of the sampled tensor $U$ 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 41.

**Lemma 38.** Let $U$ be a rank-$(r-1)$ tensor and $U(\bar{x}) = y$ be an entry of this tensor. Assume that changing the value of entry $U(\bar{x})$ from $y$ to $y'$ results in $U'$, which is a rank-$r$ tensor. Then, there are infinitely many scalars $y''$ such that changing the value of entry $U(\bar{x})$ 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 $\tilde{U}_{(d-1)}$. Assume that only one entry of the tensor is changed and consider the column of $\tilde{U}_{(d-1)}$ 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 $\tilde{U}_{(d-1)}$. Note that a column $u \in \mathbb{R}^{N_{d-1}}$ with only a single nonzero entry $U(\tilde{M}_{d-1}(x_1, \ldots, x_{d-1})) = u \neq 0$, is a structured column, as we can rewrite $u$ as the vectorization of the outer product of $[0, \ldots, 0, 1, 0, \ldots, 0]^\top \in \mathbb{R}^{n_i}$ for $1 \leq i \leq d-2$ and $[0, \ldots, 0, u, 0, \ldots, 0]^\top \in \mathbb{R}^{N_{d-1}}$. 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 \( U(\vec{x}) \) 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 \( U(\vec{x}) \) from \( y' \) to \( y'' \) the rank is changed by 2.

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

\[
U(\vec{x}) = \sum_{l=1}^{r-1} a_{1}^{l}(x_{1})a_{2}^{l}(x_{2}) \cdots a_{d}^{l}(x_{d}),
\]

results in a system of \(|\Omega|\) polynomial equations \( p \) in terms of the variables \( a_{i}^{l}, i = 1,...,d, l = 1,...,r-1 \). By assumption, when the value of the entry \( U(\vec{x}) \) is changed from \( y \) to \( y' \), the rank is changed from \( r - 1 \) to \( r \). This means that when \( U(\vec{x}) = y, p \) is feasible; and when \( U(\vec{x}) = y', p \) is not feasible. On the other hand, since changing \( U(\vec{x}) \) to any \( y'' \) results in either rank-\((r-1)\) or rank-\(r\) tensor, we conclude that if changing \( U(\vec{x}) \) to \( y' \) leads to infeasible \( p \), then rank is \( r \). The rest of the proof then follows from Lemma 40. 

Next, we prove Lemma 40 which was used in the above lemma. We first need the following definition.

**Definition 20.** 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 39.** Consider an arbitrary system of polynomial equations \( p = (p_{1},\ldots,p_{m}) = 0 \) in terms of \( n \) variables. The feasible region of \( p \) 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 \( p_i(\vec{x}_i) = 0 \). Note that \( p_i \) is continuous with respect to the vector of variables \( \vec{x} = (x_1, \ldots, x_n) \) and therefore, \( \lim_{t \to \infty} p_i(\vec{x}_i) = 0 \). As a result, \( p_i(\vec{x}_0) = 0 \) and hence, \( \vec{x}_0 \in F_i \). Hence, \( F_i \) is closed.

Define \( F = \cap_{1 \leq i \leq m} F_i \), which denotes the feasible region of \( p \) 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, \( F \) is a closed set in \( \mathbb{R}^n \).

The following lemma is used in the proof of Lemma 38.

**Lemma 40.** Consider an arbitrary system of polynomial equations \( p = (p_1, \ldots, p_{m-1}, p_m - c) = 0 \) 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 \( p \) 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 \( p \) 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 \( p \) 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 \( p \) 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 \( p \) is non-empty.

Let \( F' \subseteq \mathbb{R}^n \) denote the feasible region of the system of polynomial equations \( p' = (p_1, \ldots, p_{m-1}) = 0 \). Note that \( c = y'' \) results that the feasible region of \( p \) is non-empty if and only if there exists a vector \( \vec{x''} \in F' \) such that \( p_m(\vec{x''}) = y'' \). In other words, \( y'' \) belongs to the image of \( p_m \) with \( F' \) as the domain \((p_m(F'))\) if and only if \( c = y'' \) results that the feasible region of \( p \) is non-empty. Consequently, for any \( y'' \in \mathcal{N}(y') \) and \( y'' \neq y' \), there exist at least one vector \( \vec{x''} \in F' \) 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 \( F' \) as the domain, there exist \( \vec{x'1} \in F' \) and \( \vec{x'2} \in F' \) 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 39, \( F' \) 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 $\vec{x}_i$ in $\mathcal{F}''$ that $\lim_{t \to \infty} p_m(\vec{x}_i) = y'$. 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 41.** Let $\mathbf{U}_\Omega$ 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 $\mathbf{U}_\Omega$.

**Proof.** As mentioned in the proof of Lemma 38, changing the value of only one entry of a tensor results in changing the rank of the tensor by at most one. Let $\mathbf{U}_1$ and $\mathbf{U}_2$ denote the rank-$(r-i)$ and rank-$r$ completions, respectively. $\mathbf{U}_1$ and $\mathbf{U}_2$ are the same over the sampled entries, i.e., $(\mathbf{U}_1)_\Omega = (\mathbf{U}_2)_\Omega$, and their difference is only over some of the non-sampled entries. We change the values of non-sampled entries of $\mathbf{U}_1$ one by one to the values of the corresponding non-sampled entries of $\mathbf{U}_2$, which will eventually result in $\mathbf{U}_2$ 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 $\mathbf{U}_3$ of the sampled tensor $\mathbf{U}_\Omega$ such that changing the value of an entry $\mathbf{U}_3(\vec{x})$ from $y$ to $y'$ increases the rank to $r$ for some scalars $y$ and $y'$. Hence, according to Lemma 38, there exists infinitely many rank-$r$ completions of $\mathbf{U}_\Omega$. □

### 5.4.2 Conditions on Tensor Space Clustering

**Theorem 16.** For the tensor space clustering problem discussed in Section 5.3.1 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 $r_{\max} \leq \frac{\min_{1 \leq i \leq d-1} n_i}{6}$, $c_k \geq K(r_{\max} + 2)(\sum_{i=1}^{d-1} n_i)$, for $k = 1, \ldots, K$. ([5.10])
and also, each column of $\tilde{U}_{(d-1)}$ includes at least $l$ sampled entries where

$$l > \max\left\{27 \log \left(\frac{2\max_{1 \leq i \leq d-1} n_i}{\epsilon}\right) + 9 \log \left(\frac{8r_{\max}(d-1)}{\epsilon}\right) + 18, 6r_{\max}\right\}. \quad (5.11)$$

Let $S$ be a set of $r_1$ structured columns and $\bar{T}$ denote the unfolded tensor space generated by its columns span that fits exactly $\bar{c}$ columns of $\tilde{U}_{(d-1)}$ (i.e., $\bar{c}$ is the maximum number of columns of $\tilde{U}_{(d-1)}$ that can be covered by $S$) and assume that $\bar{c} \geq K(r_{\max} + 2)(\sum_{i=1}^{d-1} n_i)$. Then, with probability at least $1 - \epsilon$ the following statement holds: All the $\bar{c}$ columns of $\tilde{U}_{(d-1)}$ covered by $S$ belong to one source $I_{k_0}$ 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 $\tilde{U}_{(d-1)}$ do not belong to $I_{k_0}$ and moreover, $\bar{c} = c_{k_0}$ and $\bar{T} = T_{k_0}$.

**Proof.** According to the pigeonhole principle, at least $\left\lceil \frac{\bar{c}}{K} \right\rceil \geq (r_{\max} + 2)(\sum_{i=1}^{d-1} n_i)$ columns of the $\bar{c}$ covered columns by $S$ are chosen from one source $I_{k_0}$. 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 $S$ that are chosen from one source $I_{k_0}$. Then, according to Lemma 37, 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 $\bar{T}$ 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 41, 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 $\bar{T}$ and $T_{k_0}$ are $r_1$-dimensional (since $r_{k_0} = r_1$), we simply conclude $\bar{T} = T_{k_0}$. Consequently, $\bar{T}$ covers all $c_{k_0}$ columns of $\tilde{U}_{(d-1)}$ that belong to $I_{k_0}$. In order to complete the proof, it suffices to show that $\bar{c} = c_{k_0}$, i.e., $\bar{T}$ does not cover any other column of $\tilde{U}_{(d-1)}$ that belongs to other sources $I_k$ 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 $T_k \subsetneq T_{k_0}$ 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 $\tilde{u}_{\Omega(d-1)}$ of $\tilde{U}_{\Omega(d-1)}$ is chosen from $\mathcal{I}_{k_1}$ for some $k_1 \neq k_0$, and it can be covered by $\mathcal{S}$. Consider $r_1$ random columns of $\tilde{U}_{\Omega(d-1)}$ that belong to $\mathcal{I}_{k_0}$ and denote it by $\tilde{U}_{\Omega(d-1)}^0$, and let $\tilde{U}_{(d-1)}^0$ be the unique rank-$r_1$ completion of $\tilde{U}_{\Omega(d-1)}^0$. Define $\tilde{U}_{\Omega(d-1)}^1 = [\tilde{U}_{(d-1)}^0 \tilde{u}_{\Omega(d-1)}] \in \mathbb{R}^{N_d \times (r_1+1)}$ (only the last column of $\tilde{U}_{\Omega(d-1)}^1$ is incomplete). Note that (5.11) ensures that $\tilde{u}_{\Omega(d-1)}$ includes at least $r_1 + 1$ sampled entries and therefore, $\tilde{U}_{\Omega(d-1)}^1$ includes an $(r_1 + 1) \times (r_1 + 1)$ submatrix such that the $r_1 + 1$ of the sampled entries of $\tilde{u}_{\Omega(d-1)}$ are included and denote such submatrix by $\tilde{U}_{(d-1)}^1$. Note that $\tilde{u}_{(d-1)}$ 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 $\tilde{U}_{(d-1)}^1 \in \mathbb{R}^{(r_1+1) \times (r_1+1)}$ is $r_1 + 1$. Hence, any completion of $\tilde{U}_{\Omega(d-1)}^1$ has matrix rank of at least $r_1 + 1$ and as a result, any completion of the tensor corresponding to the unfolding $\tilde{U}_{\Omega(d-1)}^1$ has CP-rank of at least $r_1 + 1$. Therefore, $\mathcal{S}$ cannot fit $\tilde{u}_{\Omega(d-1)}$ with probability one since $\mathcal{S}$ is an $r_1$-dimensional tensor space, and the proof is complete due to this contradiction.

Remark 19. The required number of subtensors chosen from each source in the matrix analysis approach given by (5.5) is $O(r_{\max} n_1 \ldots n_{d-1})$. However, this number reduces to $O(r_{\max} n_1 + \cdots + n_{d-1})$ in the tensor approach according to Theorem 16. This huge improvement is a consequence of taking advantage of tensor analysis as opposed to matrix analysis. Moreover, Theorem 16 does not require the assumptions of $r_1 = r_2 = \cdots = r_K$ and independently chosen tensor spaces.

Example 6. 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 5.1, given by (5.5) and (5.10), respectively. Note that the bound obtained by matrix analysis is valid only if $r_1 = r_2 = \cdots = r_{10}$, 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 (5.6) and (5.11), respectively, in Figure 5.2. 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.

According to Theorem 16, we identify all columns of $\tilde{U}_{(d-1)}$ chosen from a source (generated
Figure 5.1: The required number of subtensors $c_k$ for correct clustering.

Figure 5.2: The required number of samples per subtensor for correct clustering.
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 16 is applicable again. In particular, let \( \tilde{S}_1, \ldots, \tilde{S}_{K'} \) (for some \( 1 \leq K' < K \)) denote different sets of \( r_1, \ldots, r_{K'} \) structured columns that fit exactly \( \tilde{c}_1, \ldots, \tilde{c}_{K'} \) columns of \( \tilde{U}_{(d-1)} \), respectively, and assume that \( \tilde{c}_k \geq K(r_{\text{max}} + 2)(\sum_{i=1}^{d-1} n_i) \), \( k = 1, \ldots, K' \). Let \( \tilde{T}_1, \ldots, \tilde{T}_{K'} \) denote the column span of \( \tilde{S}_1, \ldots, \tilde{S}_{K'} \), respectively. Moreover, assume that there exist \( K(r_{\text{max}} + 2)(\sum_{i=1}^{d-1} n_i) \) columns of \( \tilde{U}_{(d-1)} \) covered by \( \tilde{S}_k \) that cannot be covered by any of \( \tilde{S}_1, \ldots, \tilde{S}_{k-1} \), \( k = 1, \ldots, K' \). Then, using Theorem 16, we have \( \tilde{T}_1 = T_{k_1} \) and \( \tilde{c}_1 = c_{k_1} \) 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 \( T_{k_1} \). Therefore, similarly, we can apply Theorem 16 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 \( \tilde{U}_{(d-1)} \) of the first source, there still exist \( K(r_{\text{max}} + 2)(\sum_{i=1}^{d-1} n_i) \) columns covered by \( \tilde{S}_2 \) that cannot be covered by \( \tilde{S}_1 \). Hence, we apply Theorem 16 again and therefore, with probability at least \( (1 - \epsilon)^2 \) the following statement holds: All the \( \tilde{c}_k \) columns of \( \tilde{U}_{(d-1)} \) covered by \( \tilde{S}_k \) belong to one source \( I_{k'} \) such that \( r_{k'} = r_k \) and the rest of the columns of \( \tilde{U}_{(d-1)} \) do not belong to \( I_{k'} \) and moreover, \( \tilde{c}_k = c_{k'} \) and \( \tilde{T}_k = T_{k'} \), \( k = 1, 2 \). By simply repeating this procedure, we conclude the following corollary.

**Corollary 21.** 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 \( r_{\text{max}} = \max_{1 \leq k \leq K} r_k = r_K \). Assume further that \( r_{\text{max}} \leq \frac{\min_{1 \leq i \leq d-1} n_i}{6} \), (5.10) holds, and also, each column of \( \tilde{U}_{(d-1)} \) includes at least \( l \) sampled entries such that (5.11) holds. Let \( \tilde{S}_1, \ldots, \tilde{S}_{K'} \) (for some \( 1 \leq K' < K \)) denote different sets of \( r_1, \ldots, r_{K'} \) structured columns that fit exactly \( \tilde{c}_1, \ldots, \tilde{c}_{K'} \) columns of \( \tilde{U}_{(d-1)} \), respectively, and assume that \( \tilde{c}_k \geq K(r_{\text{max}} + 2)(\sum_{i=1}^{d-1} n_i) \), \( k = 1, \ldots, K' \). Let \( \tilde{T}_1, \ldots, \tilde{T}_{K'} \) denote the column span of \( \tilde{S}_1, \ldots, \tilde{S}_{K'} \), respectively. Moreover, assume that there exist \( K(r_{\text{max}} + 2)(\sum_{i=1}^{d-1} n_i) \) columns of \( \tilde{U}_{(d-1)} \) covered by \( \tilde{S}_k \) that cannot be covered by any of \( \tilde{S}_1, \ldots, \tilde{S}_{k-1} \), \( k = 1, \ldots, K' \). Then, with probability at least \( (1 - \epsilon)^{K'} \) the following statement holds: All the \( \tilde{c}_k \) columns of \( \tilde{U}_{(d-1)} \) covered by \( \tilde{S}_k \) belong to one source \( I_{k'} \) such that
$r_{k'} = r_k$ and the rest of the columns of $\tilde{U}_{(d-1)}$ do not belong to $I_{k'}$ and moreover, $\bar{c}_k = c_{k'}$ and $\bar{T}_k = T_{k'}$, $k = 1, \ldots, K'$.

5.5 Summary

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.
Part II

Efficient Algorithms
Chapter 6

Algorithm For Low-Rank Data Completion With Very Low Sampling Rate

Newton’s method is a widely applicable and empirically efficient method for finding the solution to a set of equations. The recently developed algebraic geometric analyses provide information-theoretic bounds on the sampling rate to ensure the existence of a unique completion with a high probability. We study a remained open question from these works, which is whether we can retrieve the sampled data when the sampling rate is very close to the mentioned information-theoretic bounds. This work is concerned with proposing algorithms to retrieve the sampled data when the sampling rate is too small and close to the mentioned information-theoretic bounds. Hence, we propose a new approach for recovering a partially sampled low-rank matrix or tensor when the number of samples is only slightly more than the dimension of the corresponding manifold, by solving a set of polynomial equations using Newton’s method. In particular, we consider low-rank matrix completion, matrix sensing and tensor completion. Each observed entry contributes one polynomial equation in terms of the factors in the rank factorization of the data. By exploiting the specific structures of the resulting set of polynomial equations, we analytically characterize the convergence regions of the Newton’s method for matrix completion and matrix sensing. Through extensive numerical results, we show that the proposed approach outperforms the well known methods such as nuclear norm
minimization and alternating minimization in terms of the success rate of data recovery (noiseless case) and peak signal-to-noise ratio (noisy case), especially when the sampling rate is very low. In terms of the running time, the proposed Newton’s method is significantly faster than the nuclear norm minimization method, but it is slower than the alternating minimization method.

6.1 Introduction

This chapter is concerned with the problem of reconstructing a partially sampled low-rank matrix or tensor, i.e., the low-rank data completion problem, which has many applications in various areas of engineering and applied science including image or signal processing [Candès et al., 2013; Ji et al., 2010], network coding [Harvey et al., 2005], control [Mesbahi and Papavassilopoulos, 1997], data mining [Elđén, 2007], data clustering [Ashraphijuo and Wang, 2019a], recommender systems and collaborative filtering [Rennie and Srebro, 2005], etc. There are various techniques for tackling the low-rank data completion problem, including convex relaxations of rank [Candès and Tao, 2009; Candès and Tao, 2010; Cai et al., 2010; Candès et al., 2013; Gandy et al., 2011], alternating minimization [Jain et al., 2013; Jain and Oh, 2014; Wang et al., 2008; Wang et al., 2016], augmented Lagrangian method [Yang and Yuan, 2013], algebraic geometric analyses [Pimentel-Alarcón et al., 2016b; Ashraphijuo and Wang, 2017a; Ashraphijuo et al., 2019a; Ashraphijuo and Wang, 2020a; Ashraphijuo et al., 2017b; Pimentel-Alarcón et al., 2017b] and other heuristics [Liu et al., 2016; Kressner et al., 2014; Krishnamurthy and Singh, 2013]. Among them, the nuclear norm minimization method is the most effective and robust solution, although its complexity is very high [Candès and Tao, 2010]. On the other hand, the alternating minimization method is very fast, but with less satisfactory performance. The matrix sensing problem is a generalization of the matrix completion problem, where a set of linear matrix equations instead of a set of entries are given. A review on matrix sensing problem and its applications can be found in [Recht et al., 2010].

The recently developed algebraic geometric analyses in [Pimentel-Alarcón et al., 2016b; Ashraphijuo and Wang, 2017a; Ashraphijuo et al., 2019a; Ashraphijuo et al., 2020; Ashraphijuo and Wang, 2020a] make use of the rank decomposition and transform the data completion problem to the problem of solving a system of polynomial equations. In particular, each observed entry corre-
Chapter 6. Algorithm for Low-Rank Data Completion with Very Low Sampling Rate

sponds to a polynomial in terms of the entries of the factors in the rank factorization of the data. Therefore, any solution to the set of polynomial equations results in a completion of the data of the given rank constraint. As a result, the problem of low-rank matrix completion can be translated to finding a root for a system of semi-homogeneous polynomials. In addition, the mentioned works provide information-theoretic bounds on the sampling rate to ensure the existence of a unique completion with a very high probability. A remained open question from the above-mentioned works is whether we can retrieve the sampled data when the sampling rate is very close to the obtained information-theoretic bounds on the sampling rate. This work is concerned with proposing algorithms to retrieve the sampled data when the sampling rate is too small and close to the mentioned information-theoretic bounds.

In this chapter, we propose to use Newton’s method to solve the set of polynomial equations and to obtain the completion of the data. This can be done for any well-known data structure by using the rank factorization as we will show. By exploiting the structure of the polynomials, we can analytically characterize the convergence region of the proposed Newton’s method for matrix completion and matrix sensing. In particular, our method is extremely efficient in comparison with the existing methods in the literature, when the sampling rate is too low. This can be also easily extended to homotopy methods. In the numerical experiments, we use the successful recovery rate and the peak signal-to-noise ratio as performance metrics for noiseless and noisy cases, respectively. We show that the proposed Newton’s method outperforms the existing nuclear norm minimization and alternating minimization methods especially when the sampling rate is very low, i.e., when the number of samples is close to the dimension of the corresponding manifold. Note that alternating minimization is much faster than nuclear norm minimization, but nuclear norm minimization requires fewer samples to successfully recover the data. Our proposed approach outperforms nuclear norm minimization in terms of both speed and the number of samples for successful completion and it outperforms alternating minimization in terms of the number of required samples. We also provide numerical experiments on real-world datasets such as the MovieLens dataset and show that the proposed Newton’s method outperforms both existing methods.

The remainder of this chapter is organized as follows. In Section 6.2, the Newton’s method for matrix completion is proposed. In Section 6.3, we characterize the convergence region of the proposed Newton’s matrix completion algorithm. In Section 6.4, we provide numerical results
to compare the proposed method with existing matrix completion algorithms. In Section 6.5, we extend the proposed framework and propose Newton’s method for solving matrix sensing and tensor completion problems. Finally, Section 6.6 summarizes the chapter.

6.2 Matrix Completion

6.2.1 Background

Let $U \in \mathbb{R}^{n_1 \times n_2}$ denote a rank-$r$ matrix that is partially observed. Denote $\Omega$ as the set of indices $(i, j)$ such that $U(i, j)$ is observed. Moreover, define $U_\Omega$ as the matrix obtained from sampling $U$ according to $\Omega$, i.e.,

$$U_\Omega(i, j) = \begin{cases} U(i, j) & \text{if } (i, j) \in \Omega, \\ 0 & \text{if } (i, j) \notin \Omega. \end{cases}$$ (6.1)

The matrix completion problem is to recover the original rank-$r$ matrix $U$ given its sampled version $U_\Omega$. Alternating minimization [Jain et al., 2013] and nuclear norm minimization [Candès and Recht, 2009] are two well-known methods for matrix completion.

In the alternating minimization approach, we write $U = XY$ such that $X \in \mathbb{R}^{n_1 \times r}$ and $Y \in \mathbb{R}^{r \times n_2}$. Starting with some initial $X_0$ and $Y_0$, at the $k$-th iteration, given $X_{k-1}$ and $Y_{k-1}$, we first update $X_k$ by solving the following convex program

$$\text{minimize}_{X_k \in \mathbb{R}^{n_1 \times r}} \quad \|U_\Omega - (X_k Y_{k-1})_\Omega\|_F,$$ (6.2)

and then update $Y_k$ by solving

$$\text{minimize}_{Y_k \in \mathbb{R}^{r \times n_2}} \quad \|U_\Omega - (X_k Y_k)_\Omega\|_F,$$ (6.3)

where $\|\cdot\|_F$ denotes the Frobenius norm. The iteration continues until both errors are below certain threshold.

On the other hand, in the nuclear norm minimization approach, we do not make use of the rank constraint and solve the following relaxed problem which is a convex program

$$\text{minimize}_{U' \in \mathbb{R}^{n_1 \times n_2}} \quad \|U'\|_*,$$ (6.4)

subject to $U'_\Omega = U_\Omega$. 
where $\|\cdot\|_*$ denotes the matrix nuclear norm, i.e., the sum of singular values.

In this chapter, we compare our approach for low-rank data completion with two well-known approaches, i.e., nuclear norm minimization and alternating minimization. All these three approaches are different only in terms of formulation, which results in different performances. For example, among the three of them alternating minimization is the fastest and our approach is slightly slower but nuclear norm minimization (the only convex formulation) is very slow. However, our formulation requires much less number of samples to recover the original sampled data in comparisons with the two other formulations. Note that one of the advantages of nuclear norm minimization approach in comparison with the non-convex approaches like alternating minimization and our proposed method is the fact that the only required input is the observed partial matrix and not the value of the rank. Whereas, in the mentioned non-convex approaches, the value of the rank is required. On the other hand, generally, these non-convex approaches are much faster than the convex approaches like nuclear norm minimization. For the alternating minimization formulation, we use the numerical method in [Jain et al., 2013] and for the nuclear norm minimization we used both CVX toolbox and also the online implementation [Shabat, 2015].

Our method and the described alternating minimization method are both solving the same problem, i.e., Low Rank Matrix Completion (LRMC) problem, and they are only two different numerical methods.

6.2.2 Matrix Completion via Solving Polynomial Equations

The dimension of the manifold of $n_1 \times n_2$ matrices of rank $r$ is $r(n_1 + n_2 - r)$. $X \in \mathbb{R}^{n_1 \times r}$ is called a basis of $U$ if each column of $U$ can be written as a linear combination of the columns of $X$. The following lemma gives a unique canonical decomposition of $U$.

**Lemma 42.** Let $U$ denote a generically chosen matrix from the manifold of $n_1 \times n_2$ matrices of rank $r$. Then, there exists a unique decomposition $U = XY$ such that $X \in \mathbb{R}^{n_1 \times r}$, $Y \in \mathbb{R}^{r \times n_2}$ and $X(1 : r, 1 : r) = I_r$, where $X(1 : r, 1 : r)$ represents the submatrix of $X$ consisting of the first $r$ columns and the first $r$ rows and $I_r$ denotes the $r \times r$ identity matrix.

**Proof.** We show that there exists exactly one decomposition $U = XY$ such that $X(1 : r, 1 : r) = I_r$ with probability one. Considering the first $r$ rows of $U = XY$, we conclude $U(1 : r,:) = I_r Y = Y$. 
Therefore, we need to show that there exists exactly one $X(r + 1 : n_1, :)$ such that $U(r + 1 : n_1, :) = X(r + 1 : n_1, :) Y$ or equivalently $U(r + 1 : n_1, :)^T = U(1 : r, :)^T X(r + 1 : n_1, :)^T$. It suffices to show that each column of $X(r + 1 : n_1, :)$ can be determined uniquely having $u = U(1 : r, :)^T x$ where $u \in \mathbb{R}^{n_2 \times 1}$ and $x \in \mathbb{R}^{r \times 1}$. As $U$ is a generically chosen $n_1 \times n_2$ matrix of rank $r$, we have $\text{rank}(U(1 : r, :)) = r$ with probability one. Hence, $u(1 : r) = U(1 : r, 1 : r)^T x$ results in $r$ independent degree-1 equations in terms of the $r$ variables (entries of $x$), and therefore $x$ has exactly one solution with probability one.

Remark 20. The above lemma holds true for any permutation of rows of the identity matrix among all $n_1$ rows of the basis matrix $X$.

Remark 21. This structure is only for simplifying the obtained system of polynomial equations. Note that we only use this structure when $U$ is a generically chosen matrix from the manifold of $n_1 \times n_2$ matrices of rank $r$. Otherwise, we do not use this structure.

Consider a rank decomposition of the original matrix, i.e., $U = XY$, where $X \in \mathbb{R}^{n_1 \times r}$ and $Y \in \mathbb{R}^{r \times n_2}$. Note that each observed entry of $U$ results in a polynomial in terms of the entries of $X$ and $Y$ as follows

$$\sum_{k=1}^{r} X(i, k) Y(k, j) - U(i, j) = 0, \quad \text{for } (i, j) \in \Omega. \quad (6.5)$$

Given the canonical structure, to recover $U$ we need to solve for the $D = r(n_1 + n_2 - r)$ unknown entries of $(X, Y)$ from the set of second-order polynomial equations in (6.5). An illustrative example is provided next to show how to obtain such an equation set.

Example 7. In this example, we want to show the process of obtaining the mentioned polynomials based on the given sampled entries. Consider a $4 \times 3$ matrix $U$ of rank 2 with the following observed entries

$$U_\Omega = \begin{bmatrix} 4 & 7.3 & 0 \\ 0 & 0 & 0 \\ 1 & 8.3 & 22.1 \\ 0 & 12.1 & 0 \end{bmatrix}.$$
Since \( \text{rank}(U) = 2 \), the canonical decomposition is

\[
U = \begin{bmatrix}
1 & 0 \\
0 & 1 \\
x_1 & x_2 \\
x_3 & x_4
\end{bmatrix}
\begin{bmatrix}
y_1 & y_3 & y_5 \\
y_2 & y_4 & y_6
\end{bmatrix}.
\]

Denote \( z = [x_1, \ldots, x_4, y_1, \ldots, y_6]^\top \in \mathbb{R}^{10} \) and then the corresponding system of second-order polynomial equations is

\[
p(z) = 0 \Rightarrow \begin{cases}
y_1 - 4 = 0, \\
y_3 - 7.3 = 0, \\
x_1 y_1 + x_2 y_2 - 1 = 0, \\
x_1 y_3 + x_2 y_4 - 8.3 = 0, \\
x_1 y_5 + x_2 y_6 - 22.1 = 0, \\
x_3 y_3 + x_4 y_4 - 12.1 = 0.
\end{cases}
\]

Moreover, we can simplify the polynomials by permuting the rows of the identity matrix mentioned in Remark 21 as follows

\[
U = \begin{bmatrix}
1 & 0 \\
x_1 & x_2 \\
x_3 & x_4
\end{bmatrix}
\begin{bmatrix}
y_1 & y_3 & y_5 \\
y_2 & y_4 & y_6
\end{bmatrix}.
\]

Then, the corresponding system of second-order polynomial equations becomes

\[
\bar{p}(z) = 0 \Rightarrow \begin{cases}
y_1 - 4 = 0, \\
y_3 - 7.3 = 0, \\
y_2 - 1 = 0, \\
y_4 - 8.3 = 0, \\
y_6 - 22.1 = 0, \\
x_3 y_3 + x_4 y_4 - 12.1 = 0.
\end{cases}
\]

It is easy to see that compared with \( p(z) \), \( \bar{p}(z) \) has more first-order equations and therefore it is easier to solve.
CHAPTER 6. ALGORITHM FOR LOW-RANK DATA COMPLETION WITH VERY LOW SAMPLING RATE

6.2.3 The Algorithm

6.2.3.1 Locations of Unknowns in X

We choose the \( r \) rows with the most number of samples in \( U_\Omega \) and permute the rows of the identity matrix in \( X \) to those rows to have as many linear (i.e., first-order) polynomials as possible.

6.2.3.2 Newton’s Method for Solving \( p(z) = 0 \)

Define the vector \( z \in \mathbb{R}^{D \times 1} \) such that it contains all the \( D \) unknown elements of the decomposition. For example, if we choose \( X(1:r, 1:r) = I_r \), then we have \( z = \text{vec}(X(r+1:n_1,:), Y) \), where \( \text{vec}() \) denotes the vectorization operator. Then, it follows from (6.5) that each observed entry results in a polynomial that involves \( r \) entries of \( X \) and \( r \) entries of \( Y \). Hence, we have a set of second-order polynomial equations \( p_i(z) = 0, \ i = 1, \ldots, |\Omega| \), where \( |\Omega| \) denotes the number of observed entries. Denote \( p(z) = [p_1(z), \ldots, p_{|\Omega|}(z)]^\top \).

In order to solve \( p(z) = 0 \), we use the simple Newton’s method. In particular, we start with some initial \( z_0 \in \mathbb{R}^{D \times 1} \), and perform the following iteration

\[
z_n = z_{n-1} - (\nabla p(z_{n-1}))^\dagger p(z_{n-1}),
\]

where \( \nabla p(z) \in \mathbb{R}^{|\Omega| \times D} \) and its \((i,j)\)-th element denotes the partial derivate of \( p_i(z) \) with respect to \( z_j \), i.e., \( \frac{\partial p_i(z)}{\partial z_j} \) and the operator \( ^\dagger \) denotes pseudoinverse. Equivalently,

\[
\nabla p(z_n) = [\nabla p_1(z_n)^\top \ldots \nabla p_{|\Omega|}(z_n)^\top]^\top \in \mathbb{R}^{|\Omega| \times D},
\]

where

\[
\nabla p_i(z_n) = [\frac{\partial p_i}{\partial z_1}, \ldots, \frac{\partial p_i}{\partial z_D}] \in \mathbb{R}^{1 \times D},
\]

for \( i = 1, \ldots, |\Omega| \). We assume that \( |\Omega| \geq D \) and \( \nabla p(z_0) \) is full column-rank. Corollary 22 in Section 6.3 states that under mild assumptions if \( \nabla p(z_0) \) is full column-rank and \( z_0 \) falls into a certain ball around \( z^* \), then \( \nabla p(z_n) \) is also full column-rank, \( n = 1, 2, \ldots \).

Note that the matrix \( \nabla p(z) \) has a very sparse structure. This is because the number of involved variables in each polynomial is either \( r \) or \( 2r \) (depending on the location of canonical pattern). Such sparsity enables a fast computation of its inverse or pseudoinverse, e.g., the command
sparse(∇p(z)) \ p(z) in Matlab is an efficient way to calculate (∇p(z_{n-1}))^t p(z_{n-1}). For instance, for the example \( \hat{p}(z) \) described earlier we have

\[
\nabla \hat{p}(z) = \begin{bmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & y_3 & y_4 & 0 & 0 & x_3 & x_4 & 0 & 0
\end{bmatrix}.
\]

6.2.3.3 Initialization

We can use the initialization scheme for alternating minimization as in [Jain et al., 2013]. Consider the singular value decomposition \( U_\Omega = U_0 S_0 V_0^\top \), where the number of nonzero entries of \( S_0 \) can be more than \( r \). Now, we define a decomposition corresponding to the \( r \) largest singular values, i.e.,

\[
U_0(:, 1 : r) S_0(1 : r, 1 : r) V_0(:, 1 : r)^\top = LR,
\]

where \( L = U_0(:, 1 : r) S_0(1 : r, 1 : r) \in \mathbb{R}^{n_1 \times r} \) and \( R = V_0(:, 1 : r)^\top \in \mathbb{R}^{r \times n_2} \).

We choose the \( r \) rows of \( U_\Omega \) with the most number of samples and permute the rows of the identity matrix in the basis matrix \( X \) to those rows to increase the number of linear polynomials as much as possible. Let \( I = \{i_1, \ldots, i_r\} \) denote the indices corresponding to these \( r \) rows. Let \( L' = L (L(I, 1 : r))^{-1} \) and \( R' = L(I, 1 : r) R \). Finally, we set \( (X_0, Y_0) = (L', R') \) as the initial value in Newton’s method. As we will observe in the numerical results, this initialization significantly increases the efficiency and decreases the number of required observations for successful recovery. We denote the initialization procedure described above as \( \text{Init}(U_\Omega) \).

As an alternative initialization method, we start by applying the nuclear norm minimization to \( U_\Omega \) and obtain a completion \( \tilde{U} \). Then, we apply the above-mentioned initialization procedure using \( \tilde{U} \) to obtain \( (X_0, Y_0) = \text{Init}(\tilde{U}) \).

6.2.3.4 Stopping Criterion

For the non-convex approaches (Newton and alternating minimization), we stop the program if either \( \|z_n\| \) converges or \( \|z_n\| \) becomes larger than \( \max\{10^6, 10^6 \|z_0\|\} \). When \( \|z_n\| \) becomes larger
than $\max \{10^6, 10^6 \|z_0\|\}$, we count it as failure of the algorithm for recovering data (divergence).

### 6.3 Convergence Analysis

Since the equations $p(z) = 0$ involve only polynomials which are differentiable, the following classical convergence result on Newton’s method holds for our proposed algorithm. Let $z^*$ denote the solution to $p(z) = 0$. Moreover, define the error vector at iteration $n$ as $e_n = z_n - z^* \in \mathbb{R}^D$.

**Proposition 3.** [Luenberger, 1973] Consider a system of polynomial equations $p(z) = 0$. There exists $\epsilon > 0$ such that if $\|e_0\|_2 < \epsilon$, then we have $\|e_n\|_2 \to 0$ as $n \to \infty$.

However, in the above classical result, the exact characterization of the convergence neighborhood, i.e., $\epsilon$, is not specified. In this section, we will give an explicit expression for $\epsilon$ by exploiting the specific structure of the polynomial set $p$.

We will first derive a relationship between $\|e_n\|_2$ and $\|e_{n+1}\|_2$ through the smallest singular value of a submatrix of $\nabla p(z^*)$. Then, by bounding $\|e_0\|_2$ in terms of this smallest singular value, we can ensure $\|e_n\|_2 \to 0$ as $n \to \infty$.

From Newton’s update given in (6.6) we can write

$$p(z_n) + \nabla p(z_n)(z_{n+1} - z_n) = 0. \quad (6.9)$$

On the other hand, using Taylor’s expansion and the fact that the polynomials in $p$ are twice differentiable at any point, we have

$$p(z_n) + \nabla p(z_n)(z^* - z_n) + h(t_n) = p(z^*) = 0, \quad (6.10)$$

for some $t_n$ belonging to a ball centered at $z^*$ and of radius $|z^* - z_n|$, where

$$h(t_n) = \left[ \frac{1}{2}(z^* - z_n)^\top \nabla^2 p_1(t_n)(z^* - z_n), \ldots, \frac{1}{2}(z^* - z_n)^\top \nabla^2 p_{|\Omega|}(t_n)(z^* - z_n) \right]^\top \in \mathbb{R}^{|\Omega| \times 1}, \quad (6.11)$$

with $\nabla^2 p_i = \left[ \frac{\partial^2 p_i}{\partial z_j \partial z_{j'}} \right]_{j,j'} \in \mathbb{R}^{D \times D}$, for $1 \leq j, j' \leq D$ and $i = 1, \ldots, |\Omega|$.

Using (6.9) and (6.10), we conclude that

$$\nabla p(z_n)(z^* - z_{n+1}) + h(t_n) = 0. \quad (6.12)$$
Note that for the proposed algorithm all polynomials are of the form (6.5). Consider polynomial \( p_i \) and assume that it corresponds to the observed entry \((i', j') \in \Omega \) of \( U \). Then, \( p_i \) involves \( X(i', \ell'), Y(\ell', j') \) for \( \ell' = 1, \ldots, r \). Note that \( X(i', :) \) either corresponds to \( r \) variables in \( z \), or it is a constant vector corresponding to a row of the identity submatrix in the canonical structure. Hence, \( \nabla^2 p_i \) is a binary matrix that is independent of \( t_n \), where there are 1’s at entries corresponding to the variables \( X(i', \ell') \) and \( Y(\ell', j') \), and zeros everywhere else.

Denote \( e_{X,n}^i = e_{X,n}^{\ell_1}, \ldots, e_{X,n}^{\ell_r} \) as the components of \( e_n \) that corresponds to the \( r \) variables \( Y(\ell', j') \) in \( p_i \), \( \ell' = 1, \ldots, r \). Moreover, define \( e_{X,n}^{\ell'} = e_{X,n}^{\ell_1}, \ldots, e_{X,n}^{\ell_r} \) as follows: if \( X(i', \ell') \) is a variable in \( z \), then \( e_{X,n}^{\ell'}(\ell') = 0, \ell' = 1, \ldots, r \).

Then it is easily verified that
\[
\frac{1}{2} (z^* - z_n)^\top \nabla^2 p_i(t_n)(z^* - z_n) = (e_{X,n}^i)^\top e_{Y,n}^i.
\]
Hence
\[
h(t_n) = \left[ (e_{X,n}^1)^\top e_{Y,n}^1, \ldots, (e_{X,n}^{[0]})^\top e_{Y,n}^{[0]} \right]^\top \in \mathbb{R}^{[0]}.
\]
Therefore, (6.12) can be simplified as
\[
\nabla p (z_n) e_{n+1} = \left[ (e_{X,n}^1)^\top e_{Y,n}^1, \ldots, (e_{X,n}^{[0]})^\top e_{Y,n}^{[0]} \right]^\top.
\]

Assume that \( \nabla p (z_n) \in \mathbb{R}^{[0] \times D} \) is full column-rank. Hence, there exists a set of \( D \) polynomials \( \tilde{p} = \{p_1, \ldots, p_D\} \) such that
\[
\nabla \tilde{p} (z_n) e_{n+1} = \left[ (e_{X,n}^{i_1})^\top e_{Y,n}^{i_1}, \ldots, (e_{X,n}^{i_D})^\top e_{Y,n}^{i_D} \right]^\top,
\]
or
\[
e_{n+1} = (\nabla \tilde{p} (z_n))^{-1} \left[ (e_{X,n}^{i_1})^\top e_{Y,n}^{i_1}, \ldots, (e_{X,n}^{i_D})^\top e_{Y,n}^{i_D} \right]^\top,
\]
where \( \nabla \tilde{p} (z_n) \in \mathbb{R}^{D \times D} \) is a full-rank matrix.
It then follows that
\[
\|e_{n+1}\|_2 = \left\| \left( \nabla \tilde{p}(z_n) \right)^{-1} \left[ (e_{X,n}^i)^\top e_{Y,n}^i, \ldots, (e_{X,n}^D)^\top e_{Y,n}^i \right]^\top \right\|_2 \\
\leq \sigma_{\text{max}}(\left( \nabla \tilde{p}(z_n) \right)^{-1}) \times \left\| \left[ (e_{X,n}^i)^\top e_{Y,n}^i, \ldots, (e_{X,n}^D)^\top e_{Y,n}^i \right]^\top \right\|_2 \\
= \frac{1}{\sigma_{\text{min}}(\left( \nabla \tilde{p}(z_n) \right))} \times \left\| \left[ (e_{X,n}^i)^\top e_{Y,n}^i, \ldots, (e_{X,n}^D)^\top e_{Y,n}^i \right]^\top \right\|_2. \tag{6.17}
\]

We can now state our convergence result.

**Theorem 17.** Consider the matrix $T$ such that $T^{-1} \nabla \tilde{p}(z^*) = \Sigma^*$ is the diagonal matrix consisting of the singular values of the full-rank matrix $\nabla \tilde{p}(z^*)$ and define $\delta = \|T^{-1}\|_F\|T\|_F$. Assume that $\|e_0\|_2 \leq \min\{\frac{\sqrt{\sigma_{\text{min}}(\nabla \tilde{p}(z^*))}}{2\delta \sqrt{D}}, \frac{\sigma_{\text{min}}(\nabla \tilde{p}(z^*))}{8\delta \sqrt{D}}\}$. Then, we have $\|e_n\|_2 \to 0$ as $n \to \infty$.

**Proof.** Since each polynomial $p_i(z)$ is second-order as given in (6.5), $\nabla p_i(z)$ is a vector with entries either equal to zero or one of the elements of $z$. For example, if $z \in \mathbb{R}^6$ and $p_i(z) = z_1 z_4 + z_2 z_5 - 4 = 0$, then we have $\nabla p_i(z) = [z_4, z_5, 0, z_1, z_2, 0]$. Define $E_n = \nabla \tilde{p}(z_n) - \nabla \tilde{p}(z^*)$. It is easy to observe that each element of $E_n$ is either zero or one element of the error vector $e_n$. As a result, $\|\nabla p_i(z_n) - \nabla p_i(z^*)\|_2 \leq \|e_n\|_2$, and therefore $\|E_n\|_F \leq \sqrt{D}\|e_n\|_2$. Hence, $\nabla \tilde{p}(z_n) = \nabla \tilde{p}(z^*) + E_n$, where $\|E_n\|_F \leq \sqrt{D}\|e_n\|_2$.

The perturbation analysis [Deif, 1995] results that for any eigenvalue $\lambda$ of $\nabla \tilde{p}(z_n)$, there exists $i \in \{1, \ldots, D\}$ such that
\[
|\lambda - \lambda_i| \leq \|T^{-1}\|_F\|T\|_F\|E_n\|_F, \tag{6.18}
\]
where $\lambda_i$'s are the eigenvalues of $\nabla \tilde{p}(z^*)$ and $T^{-1} \nabla \tilde{p}(z^*) T = \Sigma^*$ is the diagonal matrix consisting of the singular values of $\nabla \tilde{p}(z^*)$. Recall that $\|E_n\|_F \leq \sqrt{D}\|e_n\|_2$, and according to the assumption we have $\|e_n\|_2 \leq \frac{\sqrt{\sigma_{\text{min}}(\nabla \tilde{p}(z^*))}}{2\delta \sqrt{D}}$, with $\delta = \|T^{-1}\|_F\|T\|_F$. Therefore, for any eigenvalue $\lambda$ of $\nabla \tilde{p}(z_n)$, there exists $i \in \{1, \ldots, D\}$ such that
\[
|\lambda - \lambda_i| \leq \frac{\sqrt{\sigma_{\text{min}}(\nabla \tilde{p}(z^*))}}{2} = \frac{1}{2} \min_{1 \leq j \leq D} |\lambda_j|. \tag{6.19}
\]
It simply follows from (6.19) that for any eigenvalue $\lambda$ of $\nabla \hat{p}(z_n)$ we have $|\lambda| \geq \frac{1}{2} \min_{1 \leq j \leq D} |\lambda_j|$, and therefore

$$\sigma_{\text{min}}((\nabla \hat{p}(z_n))) \geq \frac{\sigma_{\text{min}}((\nabla \hat{p}(z^*)))}{4}. \tag{6.20}$$

Therefore, $\nabla \hat{p}(z_n)$ is full-rank, i.e., $\nabla p(z_n)$ is full column-rank. Then, using (6.17) we conclude

$$\|e_{n+1}\|_2 \leq \frac{4}{\sigma_{\text{min}}((\nabla \hat{p}(z^*)) \times \left\|\left(\begin{array}{c} e_{i,n}^1 \top e_{i,n}^1, \ldots, (e_{i,D,n}^i \top e_{i,D,n}^i) \top \right)\right\|_2. \tag{6.21}$$

On the other hand, by definition, each element of $e_{i,n}^1$ and $e_{i,n}^i$ is either 0 or an element of $e_n$. Hence we can write

$$\left\|\left(\begin{array}{c} e_{i,n}^1 \top e_{i,n}^1, \ldots, (e_{i,D,n}^i \top e_{i,D,n}^i) \top \right)\right\|_2 \leq \sqrt{Dr^2 (\max_{1 \leq j \leq D} \{\|e_n(j)\|\})^4}, \tag{6.22}$$

where $e_n(j)$ denotes the $j$-th element of the vector $e_n$. As a result, we have

$$\|e_{n+1}\|_2 \leq \frac{4r \sqrt{D} \|e_n\|_2^2}{\sigma_{\text{min}}((\nabla \hat{p}(z^*))} (b) \|e_n\|_2^2, \tag{6.23}$$

where (a) follows from the fact that $(\max_{1 \leq j \leq D} \{\|e_n(j)\|\})^2 \leq \|e_n\|_2^2$ and (b) follows from the assumption that $\|e_n\|_2 \leq \frac{\sigma_{\text{min}}((\nabla \hat{p}(z^*))}{8r \sqrt{D}}$.

Starting from $e_0$ that satisfies the condition in the statement of the theorem, we have $\|e_1\|_2 \leq \frac{\|e_0\|_2^2}{2}$. Hence $e_1$ also satisfies the condition and $\|e_2\|_2 \leq \frac{\|e_1\|_2^2}{2}$, and so on. Therefore $\|e_n\|_2 \to 0$ as $n \to \infty$. \qed

**Corollary 22.** Consider the matrix $T$ such that $T^{-1} \nabla \hat{p}(z^*) T = \Sigma^*$ is the diagonal matrix consisting of the singular values of the full-rank matrix $\nabla \hat{p}(z^*)$ and define $\delta = \|T^{-1} \|_F \|T\|_F$. Assume that $\|e_0\|_2 \leq \min\{\frac{\sigma_{\text{min}}((\nabla \hat{p}(z^*))}{2r \sqrt{D}}, \frac{\sigma_{\text{min}}((\nabla \hat{p}(z^*))}{8r \sqrt{D}}\}$. Then, $\nabla \hat{p}(z_n)$ is full-rank, $i = 1, 2, \ldots$. Hence, $\nabla p(z_n)$ is full column-rank, $i = 1, 2, \ldots$.

**Proof.** The proof is straightforward since (6.20) simply results that $\sigma_{\text{min}}((\nabla \hat{p}(z_n))) > 0$. \qed

### 6.4 Numerical Results

#### 6.4.1 Noiseless Matrix

Here, we are interested in generating a random low-rank matrix to randomly sample it by a sampling probability and compare the proposed algorithms with the existing algorithms in the literature. To
this end, we first generate $X \in \mathbb{R}^{n_1 \times r}$ and $Y \in \mathbb{R}^{r \times n_2}$ by choosing each entry of them uniformly from some interval on real numbers, e.g., $[1, 10]$. Then, we generate a random $n_1 \times n_2$ matrix of rank $r$, i.e., $U = XY$. We sample each entry independently with probability $0 < p < 1$. We say the sampled matrix $U$ is recovered if $\frac{\|\hat{U} - U\|_F}{\|U\|_F} < 0.01$, where $\hat{U}$ denotes the obtained matrix through the corresponding completion algorithm.

For each value of $p$, we complete 200 matrices and calculate the average recovery rate of each algorithm.

For the first example, we consider a $500 \times 500$ matrix of rank 3. The dimension of the manifold is $D = n_1 r + n_2 r - r^2 = 2991$. Since there are $D$ unknowns, at least $D$ samples are needed and therefore $\frac{D}{n_1 n_2} = \frac{2991}{250000} = 0.011964$ is the absolute lower bound on the sampling rate. Figure 6.1 shows the recovery rate of different completion algorithms in terms of the normalized sampling rate $\frac{p}{D/n_1 n_2}$. We have the following observations:

- Our proposed Newton’s method with initialization $\text{Init}(\hat{U})$ outperforms all other methods in the sense of requiring the least number of samples to recover the original matrix $U$. Newton’s method with initialization $\text{Init}(U_\Omega)$ is the second best and outperforms nuclear norm minimization and alternating minimization.

- With $2D$ samples the proposed Newton’s method can achieve recovery rates of 0.9 and 0.65 using the initializations $\text{Init}(\hat{U})$ and $\text{Init}(U_\Omega)$, respectively; whereas the recovery rate is zero for the conventional methods (i.e., nuclear norm minimization and alternating minimization).

- The average number of samples required for a recovery rate of at least 90% is $2.3D$, $3.25D$ and $4.25D$ using Newton’s method with initialization $\text{Init}(U_\Omega)$, nuclear norm minimization and alternating minimization, respectively. This leads to $\frac{3.25 - 2.3}{3.25} \times 100\% = 29.2\%$ and $\frac{4.25 - 2.3}{4.25} = 45.8\%$ reductions in the required number of samples, respectively.

We also show the average running time for each algorithm in Figure 6.2. It is seen that alternating minimization is the fastest method with running time almost independent of the sample size. The proposed Newton’s method with initialization $\text{Init}(U_\Omega)$ is significantly faster than nuclear norm minimization especially when the sample size is large.

For the second example, we consider a $200 \times 200$ matrix of rank 10. The dimension is $D = n_1 r + n_2 r - r^2 = 3900$ and the lower bound on the sampling rate is $\frac{D}{n_1 n_2} = \frac{3900}{400000} = 0.0975$. Figure 6.3 shows the recovery rates for different completion algorithms. We have the following observations:
Newton’s method with Init($U\Omega$) outperforms nuclear norm minimization and alternating minimization. The average number of samples required for a recovery rate of at least 90% is $1.9D$, $2.125D$ and $2.2D$ using Newton’s method with Init($U\Omega$), nuclear norm minimization and alternating minimization, respectively. This leads to $\frac{2.125-1.9}{2.125} \times 100% = 10.5\%$ and $\frac{2.2-1.9}{2.2} = 13.6\%$ reductions in the required number of samples, respectively.

Using Newton’s method with Init($\tilde{U}$), we are able to recover the data with $1.3D$ samples with probability of almost one, whereas $2D$, $2.15D$, $2.25D$ samples are required for Newton’s method with Init($U\Omega$), nuclear norm minimization and alternating minimization, respectively. This leads to $\frac{2-1.3}{2} \times 100% = 35\%$, $\frac{2.15-1.3}{2.15} \times 100% = 39.5\%$ and $\frac{2.25-1.3}{2.25} \times 100% = 42.2\%$ reductions in the required number of samples, respectively.

In these comparisons, we have compared the non-convex approaches that are able to incorporate and take advantage of the rank value with nuclear norm minimization. Note that the calculated time for Newton’s method with Init($\tilde{U}$) includes the time spent for solving the nuclear norm mini-
6.4.2 Noisy Matrix

6.4.2.1 Additive Gaussian Noise

We first generate a random matrix \( U \in \mathbb{R}^{n_1 \times n_2} \) of rank \( r \) and then we add a noise matrix \( E \) to \( U \), where \( E \) consists of i.i.d. \( \mathcal{N}(0, \sigma^2) \) entries. We sample each entry of \( E + U \) with probability \( p \).

We use the peak signal-to-noise ratio (PSNR) as the completion performance metric, defined as

\[
\text{PSNR} \left( \hat{U}, U \right) = 20 \log_{10} (M) - 10 \log_{10} \left( \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \left( \hat{U}(i,j) - U(i,j) \right)^2 \right),
\]

(6.24)

where \( M \) denotes upper bound of the entries in \( U \) (since \( U = XY \) and each element of \( X \) and \( Y \) is uniform in \([1, 10]\), we have \( M = 100r \) here).

We consider \( U \in \mathbb{R}^{500 \times 500} \) of rank 3 with \( D = n_1 r + n_2 r - r^2 = 2991 \) and \( \frac{D}{n_1 n_2} = 0.011964 \). The PSNR performances of different completion algorithms are shown in Figures 6.4(a) and 6.4(b) for two SNR (defined as \( 10 \log_{10} \left( \frac{\sigma_U}{\sigma} \right) \), where \( \sigma_U \) denotes the variance of each entry of \( U \)) values 17 dB and 12.5 dB, respectively. As we can observe, for the non-convex approaches (Newton’s method and alternating minimization) performance becomes less stable as we decrease SNR. However, for moderate SNR, Newton’s method mainly outperforms alternating minimization and nuclear norm minimization with small number of samples.
Figure 6.4: Comparison of PSNR for $U \in \mathbb{R}^{500 \times 500}$ of rank 3.
The “MovieLens-1M” dataset found online at “grouplens” website includes 6040 rows representing the users and 3706 columns representing the movies and each entry is between 1 and 5, which represents the rating of a particular user for a particular movie from Netflix rating table. For each single experiment we choose a random $2000 \times 2000$ submatrix of this dataset and sample each entry with probability $p$. Note that for the non-convex approaches, i.e., Newton’s method and alternating minimization, we need a rank value. We start with a small number, e.g., $r = 1$ and if the algorithm diverges, we increase $r$ and repeat this trial a few times.

We use the root-mean-square error (RMSE) for the unobserved entries as the performance metric, given by

$$\text{RMSE}(\hat{U}, U) = \sqrt{\frac{1}{|\Omega|} \sum_{(i,j) \in \Omega} (\hat{U}(i,j) - U(i,j))^2},$$

(6.25)

where $\Omega$ denotes the set of unobserved entries. For each value of $p$, we repeat the experiment 20 times and obtain the average RMSE. For this example, Newton’s method with the two different initializations give the same solution. The results are summarized in Table 6.1. It is seen that the proposed Newton’s method outperforms both alternating minimization and nuclear norm minimization.

In this example, the average running times of one experiment for Newton’s method with $\text{Init}(U_{\Omega})$, alternating minimization, nuclear norm minimization and Newton’s method with $\text{Init}(\hat{U})$ are 12, 9, 457 and 470 seconds, respectively.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$p = 0.05$</th>
<th>$p = 0.07$</th>
<th>$p = 0.09$</th>
<th>$p = 0.11$</th>
<th>$p = 0.13$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newton’s method</td>
<td>1.4019</td>
<td>1.2401</td>
<td>1.1395</td>
<td>1.0871</td>
<td>1.0498</td>
</tr>
<tr>
<td>Alternating minimization</td>
<td>24.2515</td>
<td>8.2742</td>
<td>6.9766</td>
<td>2.7103</td>
<td>1.9240</td>
</tr>
<tr>
<td>Nuclear norm minimization</td>
<td>1.6027</td>
<td>1.4373</td>
<td>1.3282</td>
<td>1.2648</td>
<td>1.2139</td>
</tr>
</tbody>
</table>

Table 6.1: Comparison of RMSE for MovieLens data.
6.5 Extensions

6.5.1 Matrix Sensing

Denote the inner product between two \( n_1 \times n_2 \) matrices \( U \) and \( M \) as \( \langle U, M \rangle = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} U(i,j)M(i,j) \). The matrix sensing problem is to recover a low-rank matrix \( U \) with \( \text{rank}(U) = r \), given a set of linear measurements \( \langle U, M_i \rangle = b_i, i = 1, \ldots, L \), where \( L \) denotes the total number of measurements. Hence we have a set of \( L \) polynomial equations \( p(z) = 0 \), whose elements are \( p_i(z) = \langle XY, M_i \rangle - b_i \).

In the alternating minimization approach, we again write \( U = XY \) such that \( X \in \mathbb{R}^{n_1 \times r} \) and \( Y \in \mathbb{R}^{r \times n_2} \). Starting with some initial \( X_0 \) and \( Y_0 \) (\( \text{Init}(U_M) \) described later), at the \( k \)-th iteration, given \( X_{k-1} \) and \( Y_{k-1} \), we first update \( X_k \) by solving the following convex program

\[
\text{minimize}_{X_k \in \mathbb{R}^{n_1 \times r}} \quad \sqrt{\sum_{i=1}^{L} \left( \langle X_kY_{k-1}, M_i \rangle - b_i \right)^2},
\]

and then update \( Y_k \) by solving

\[
\text{minimize}_{Y_k \in \mathbb{R}^{r \times n_2}} \quad \sqrt{\sum_{i=1}^{L} \left( \langle X_kY_k, M_i \rangle - b_i \right)^2}.
\]

The iteration continues until both errors are below certain threshold. In the nuclear norm minimization approach, we solve the following relaxed convex problem

\[
\text{minimize}_{U' \in \mathbb{R}^{n_1 \times n_2}} \quad \|U'\|_*
\text{subject to} \quad \langle U', M_i \rangle = b_i, \quad i = 1, \ldots, L.
\]

Similar to the algorithm described in Section 6.2.3, we can use Newton’s method to solve the system of polynomial equations \( p(z) = 0 \). We use the same initialization for alternating minimization as in [Jain et al., 2013] for the matrix sensing problem. We define the matrix \( U_M = \sum_{i=1}^{L} b_i M_i \). Then, we apply the initialization procedure described in Section 6.2.3 using \( U_M \) instead of \( U_\Omega \) and denote this initialization by \( \text{Init}(U_M) \).

Or we can first obtain a completion \( \tilde{U} \) by using the nuclear norm minimization method, and then apply the initialization procedure in Section 6.2.3 using \( \tilde{U} \) to obtain \( (X_0, Y_0) = \text{Init}(\tilde{U}) \).

Similar to Theorem 17, we can characterize the convergence region of the Newton’s algorithm for matrix sensing as follows.
Theorem 18. Consider the matrix $T$ such that $T^{-1}\nabla \tilde{p}(z^*)T = \Sigma^*$ is the diagonal matrix consisting of the singular values of $\nabla \tilde{p}(z^*)$ and define $\delta = \|T^{-1}\|F\|T\|F$. Suppose that the number of non-zero elements of $M_i$ (denoted by $K_i$), i.e., $\|M_i\|_0$, is upper bounded by a constant number $K$, $i = 1, \ldots, L$. Assume that $\|e_0\|_2 \leq \min \{\frac{\sqrt{\sigma_{\min}(\nabla \tilde{p}(z^*))}}{2\sqrt{D}}, \frac{\sigma_{\min}(\nabla \tilde{p}(z^*))}{8rK^2\sqrt{D}}\}$. Then, we have $\|e_n\|_2 \to 0$ as $n \to \infty$.

Proof. The proof is similar to the proof of Theorem 17 with the only difference that each polynomial in this scenario can be written as a linear combination of at most $K$ polynomials in the matrix completion scenario. It is easily verified that for matrix sensing (6.12) holds true. Moreover, instead of (6.13) we have the following

\[
h(t_n) = \begin{bmatrix} (w_{1X,n}^1)^\top & w_{1Y,n}^1 & \cdots & (w_{LX,n}^L)^\top & w_{LY,n}^L \end{bmatrix}^\top \in \mathbb{R}^L,
\]

where $w_{iX,n}^i = \sum_{j=1}^{K_i} e_{iX,n}^j \in \mathbb{R}^r$ for each of the $K_i$ involved polynomials in the $i$-th linear measurement. Recall that we have $K_i \leq K$. Therefore, (6.21) becomes as follows

\[
\|e_{n+1}\|_2 \leq \frac{4}{\sigma_{\min}(\nabla \tilde{p}(z_n))} \times \left\|\begin{bmatrix} (w_{1X,n}^1)^\top & w_{1Y,n}^1 & \cdots & (w_{LX,n}^L)^\top & w_{LY,n}^L \end{bmatrix}^\top \right\|_2.
\]

(6.30)

Note that since $K_i \leq K$, each element of $h(t_n)$ changes to the summation of at most $K$ terms in the form of $\begin{bmatrix} e_{iX,n}^1 \end{bmatrix}^\top e_{iY,n}^1$. As a result, the RHS of (6.22) becomes $\sqrt{Dr^2K^4 \max_{1 \leq j \leq D}(\|e_n(j)\|)^4}$, i.e., we have

\[
\left\|\begin{bmatrix} (w_{iX,n}^{i_1})^\top & w_{iY,n}^{i_1} & \cdots & (w_{LX,n}^{i_D})^\top & w_{LY,n}^{i_D} \end{bmatrix}^\top \right\|_2 \leq \sqrt{Dr^2K^4 \max_{1 \leq j \leq D}(\|e_n(j)\|)^4}.
\]

(6.31)

Therefore (6.23) changes to

\[
\|e_{n+1}\|_2 \leq K^2 \frac{4r\sqrt{D}\|e_n\|_2^2}{\sigma_{\min}(\nabla \tilde{p}(z^*))} \leq \frac{\|e_n\|_2}{2},
\]

(6.32)

where (a) follows from the modified assumption on $\|e_0\|_2$. 

6.5.1.1 Column-wise Measurements

In the first experiment, we assume that each matrix $M_i$ has only one non-zero column, which is chosen uniformly from the $n_2$ columns. The $n_1$ elements of the non-zero column are i.i.d. uniformly
chosen from the interval $[-1, 1]$. We say the matrix $\mathbf{U}$ is recovered if $\frac{\|\hat{\mathbf{U}} - \mathbf{U}\|_F}{\|\mathbf{U}\|_F} < 0.01$, where $\hat{\mathbf{U}}$ denotes the output of the corresponding algorithm. For each value of $L$, we repeat the experiment 200 times and calculate the average recovery rates of different algorithms.

For the first example, we consider a $50 \times 50$ matrix $\mathbf{U}$ of rank 3. The dimension of corresponding manifold $D = n_1r + n_2r - r^2 = 291$. Figure 6.5 shows the recovery rates of different algorithms in terms of the normalized sample size $\frac{L}{D}$. It is seen that the Newton’s method with both initialization schemes outperforms the nuclear norm minimization and alternating minimization.

![Figure 6.5: Comparison of recovery rates for $\mathbf{U} \in \mathbb{R}^{50 \times 50}$ of rank 3.](image)

In these comparisons, we have compared the non-convex approaches that are able to incorporate and take advantage of the rank value with nuclear norm minimization. Note that the calculated time for Newton’s method with Init($\tilde{\mathbf{U}}$) includes the time spent for solving the nuclear norm minimization.

For the second numerical experiment, we consider randomly generated $\mathbf{U} \in \mathbb{R}^{400 \times 400}$ of rank 3 and hence $D = n_1r + n_2r - r^2 = 2391$. Since nuclear norm minimization is significantly slower than Newton’s method with Init($\mathbf{U}_M$) and alternating minimization (it almost gets stuck using a regular laptop) for this experiment, we only compare the two non-convex algorithms in Figure 6.6. Again the Newton’s method significantly outperforms alternating minimization.

In this example, the average running times of one experiment for Newton’s method with Init($\mathbf{U}_M$) and alternating minimization are 702 and 129 seconds, respectively (we stopped nuclear norm minimization after 2 hours, as it takes a very long time to run on a regular computer).

Then we add a noise matrix $\mathbf{E} \in \mathbb{R}^{400 \times 400}$ to $\mathbf{U}$, where $\mathbf{E}$ consists of i.i.d. $\mathcal{N}(0, \sigma^2)$ entries. Again, we use the PSNR as the performance metric for matrix sensing problem. The PSNR performances of different algorithms are shown in Figure 6.7 for SNR = 18 dB.
6.5.1.2 Random Measurements

In this experiment, each entry of $M_i$ is non-zero with probability $p'$ independent of other entries. And nonzero entries are chosen uniformly from the interval $[-1, 1]$.

We consider $U \in \mathbb{R}^{400 \times 400}$ of rank 3. Note that $D = n_1 r + n_2 r - r^2 = 1191$. Observe that $p' = \frac{1}{200}$ results in an average $C = p' n_1 n_2 = 200$ nonzero entries, which is the number of entries of each column. In order to observe the impact of $p'$ on the recovery rate, we provide the results for $p' = 2 \times \frac{1}{200}$ and $p' = 0.5 \times \frac{1}{200}$ for different values of $L$. Since nuclear norm minimization is significantly slower than Newton’s method and alternating minimization for this example, we only compare the latter two methods in Figures 6.8(a) and 6.8(b).
Figure 6.8: Recovery rate for $\mathbf{U} \in \mathbb{R}^{200 \times 200}$ of rank 3.
6.5.2 Tensor Completion

Assume that a $d$-way tensor $\mathcal{U} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is sampled. The CP rank of a tensor $\mathcal{U}$ is defined as the minimum number $r$ such that there exist $a^l_i \in \mathbb{R}^{n_i}$ for $1 \leq i \leq d$ and $1 \leq l \leq r$, such that

$$\mathcal{U} = \sum_{l=1}^{r} a^1_1 \otimes a^2_2 \otimes \cdots \otimes a^d_d,$$

or equivalently,

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

where $\otimes$ denotes the tensor product (outer product) and $a^l_i(x_i)$ denotes the $x_i$-th entry of vector $a^l_i$. Note that $a^l_i \otimes a^l_2 \otimes \cdots \otimes a^l_d \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is a rank-1 tensor, $l = 1, 2, \ldots, r$. Denote $\Omega$ as the set of indices such that $\vec{x} \in \Omega$ if $\mathcal{U}(\vec{x})$ is observed, where $\mathcal{U}(\vec{x})$ represents an entry of tensor $\mathcal{U}$ with coordinate $\vec{x} = (x_1, \ldots, x_d)$. Moreover, define $\mathcal{U}_\Omega$ as the tensor obtained from sampling $\mathcal{U}$ according to $\Omega$, i.e.,

$$\mathcal{U}_\Omega(\vec{x}) = \begin{cases} \mathcal{U}(\vec{x}) & \text{if } \vec{x} \in \Omega, \\ 0 & \text{if } \vec{x} \notin \Omega. \end{cases}$$

In the alternating minimization approach, we write $\mathcal{U} = \sum_{l=1}^{r} a^l_1 \otimes a^l_2 \otimes \cdots \otimes a^l_d$ such that $a^l_i \in \mathbb{R}^{n_i}$ for $1 \leq i \leq d$ and $1 \leq l \leq r$. Starting with some initial $a^l_i_{0} \in \mathbb{R}^{n_i}$ for $1 \leq i \leq d$ and $1 \leq l \leq r$ (described later), at the $k$-th iteration, we update all $a^l_i'_{k}$'s by solving the following convex programs

$$\min_{a^l_i'_{k} \in \mathbb{R}^{n_i}} \left\| \mathcal{U}_\Omega - \left( \sum_{l=1}^{r} a^l_1_{k(1,l)} \otimes a^l_2_{k(2,l)} \otimes \cdots \otimes a^l_d_{k(d,l)} \right) \right\|_F,$$

where $k(i,l) = k$ if $i \leq i'$ and $l \leq l'$ and $k(i,l) = k - 1$ otherwise. The iteration continues until both errors are below certain threshold.

Define $\mathbf{z} \in \mathbb{R}^{(n_1 + \cdots + n_d)r}$ as the vector that contains $a^l_i$ for $1 \leq i \leq d$ and $1 \leq l \leq r$. Then the set of observed entries result in a set of $|\Omega|$ $r$-th order polynomial equations $\mathbf{p}$ such that each polynomial has the form

$$p_i(\mathbf{z}) = \sum_{l=1}^{r} a^l_1(x_1) a^l_2(x_2) \cdots a^l_d(x_d) - \mathcal{U}(\vec{x}), \quad \vec{x} \in \Omega.$$
Again we can apply Newton’s method described in Section 6.2.3 to solve the set of equations \( p(z) = 0 \). In particular, for initialization, we use the Matlab toolbox “Tensorlab” found online to calculate the CP decomposition of \( \mathcal{U}_\Omega \) and the leading \( r \) rank-1 components to obtain \( z_0 \).

### 6.5.2.1 Noiseless Data

For the numerical experiments, we consider \( \mathcal{U} \in \mathbb{R}^{100 \times 100 \times 100} \) of rank 3 (by generating random vectors \( \mathbf{a}_l^i \in \mathbb{R}^{100} \) for \( 1 \leq i \leq 3 \) and \( 1 \leq l \leq 3 \)) and hence the number of unknowns is \( D = n_1 r + n_2 r + n_3 r = 900 \). The lower bound on the sampling probability is \( \frac{D}{n_1 n_2 n_3} = \frac{900}{10^6} = 0.0009 \). Figure 6.9 shows the recovery rate comparison between Newton’s method and alternating minimization, as a function of the normalized sampling rate \( \frac{p D}{D/(n_1 n_2 n_3)} \).

Note that the average number of samples required for a recovery rate of at least 50% is 14.9\( D \) and 17\( D \) using Newton’s method with \( \text{Init}(\mathcal{U}_\Omega) \) and alternating minimization, respectively. This leads to \( \frac{17 - 14.9}{17} \times 100\% = 12.3\% \) reduction in the required number of samples. Moreover, the average running times of one experiment for Newton’s method with \( \text{Init}(\mathcal{U}_\Omega) \) and alternating minimization are 152 and 72 seconds, respectively.

![Figure 6.9: Comparison of recovery rates for \( \mathcal{U} \in \mathbb{R}^{100 \times 100 \times 100} \) of rank 3.](image)

### 6.5.2.2 Noisy Data

We again consider a tensor \( \mathcal{U} \in \mathbb{R}^{100 \times 100 \times 100} \) of rank 3. Then, we add a noise tensor \( \mathcal{E} \) to \( \mathcal{U} \), where \( \mathcal{E} \) consists of i.i.d. \( \mathcal{N}(0, \sigma^2) \) entries. We sample each entry of \( \mathcal{E} + \mathcal{U} \) with probability \( p \) and use the PSNR as the performance metric of each algorithm in recovering the data \( \mathcal{U} \). Figure 6.10 shows the PSNR of different algorithms in terms of the normalized sampling probability \( \frac{p D}{D/(n_1 n_2 n_3)} = \frac{p}{0.0009} \).
CHAPTER 6. ALGORITHM FOR LOW-RANK DATA COMPLETION WITH VERY LOW SAMPLENG RATE

for SNR = 22.4 dB. As we can observe, Newton’s method outperforms alternating minimization.

![Bar chart showing comparison of PSNR for different methods](image.png)

Figure 6.10: Comparison of PSNR for \( U \in \mathbb{R}^{100 \times 100 \times 100} \) of rank 3 for SNR = 22.4 dB.

6.6 Summary

We have studied the problem of retrieving a partially sampled low-rank data, when the sampling rate is very close to the information-theoretic bounds on the sampling rate for existence of a unique solution, i.e., we have considered the problem of low-rank data completion where the number of samples is comparable to the dimension of the corresponding manifold. By using rank factorization, each observed entry gives a polynomial equation of the factor entries and the solution to the set of such polynomial equations constitutes a completion of the data. We have proposed to employ Newton’s method to solve the set of polynomial equations. The convergence regions of the proposed Newton’s methods for matrix completion and matrix sensing are analytically characterized. Extensive numerical results have been provided to demonstrate that the proposed Newton’s method for data completion outperforms the well-known existing methods, such as nuclear norm minimization and alternating minimization, especially when the sampling rate is very low, i.e., comparable to the normalized dimension of the corresponding manifold. Our numerical experiments include both noiseless and noisy made up low-rank data (both two-dimensional data and higher dimensions) and also real-world low-rank data. Moreover, the proposed method is significantly faster than nuclear norm minimization but slower than alternating minimization. Hence, our proposed method is the most efficient when the sampling rate is very low and almost close to the information-theoretic bounds for existence of a unique completion of the sampled data.
Chapter 7

Structured Alternating Minimization
For Union of Nested Low-Rank Subspaces Data Completion

In this chapter, we consider a particular data structure consisting of a union of several nested low-rank subspaces with missing data entries. Given the rank of each subspace, we treat the data completion problem, i.e., to estimate the missing entries. Starting from the case of two-dimensional data, i.e., matrices, we show that the union of nested subspaces data structure leads to a structured decomposition $U = XY$ where the factor $Y$ has blocks of zeros that are determined by the rank values. Moreover, for high-dimensional data, i.e., tensors, we show that a similar structured CP decomposition also exists, $\mathcal{U} = \sum_{l=1}^{r} a_{1}^{l} \otimes a_{2}^{l} \otimes \ldots \otimes a_{d}^{l}$, where $A_{d} = [a_{1}^{1} \ldots a_{r}^{r}]$ contains blocks of zeros determined by the rank values. Based on such structured decompositions, we develop efficient alternating minimization algorithms for both matrix and tensor completions, by enforcing the above structures in each iteration including the initialization. Compared with naive approaches where either the additional rank constraints are ignored, or data completion is performed part by part, the proposed structured alternating minimization approaches exhibit faster convergence and higher recovery accuracy.
CHAPTER 7. STRUCTURED ALTERNATING MINIMIZATION FOR UNION OF NESTED LOW-RANK SUBSPACES DATA COMPLETION

7.1 Introduction

Given the ubiquitousness of multi-perspective, multi-dimensional big data in our day-to-day lives, a common feature shared by such datasets is the inherent sparsity or low-rank property. On the other hand, missing and faulty data are the norm rather than the exception. Hence a fundamental task in many big data applications is data completion, i.e., to recover the missing data points by exploiting the underlying sparsity structure. In particular, the low-rank matrix completion problem [Nguyen et al., 2019] is a classical problem that finds applications in various areas including compressed sensing [Lim and Comon, 2010; Sidiropoulos and Kyrillidis, 2012; Gandy et al., 2011], image inpainting [Yu et al., 2019], network coding [Harvey et al., 2005], image processing [Candès et al., 2013; Ji et al., 2010], data mining [Eldén, 2007], etc. The low-rank tensor completion problem has received more attention in the past decade and plays a vital role in multilinear data analysis [Kressner et al., 2014; Zhang et al., 2015], 3D image reconstruction [Sauve et al., 1999], state estimation [Madbhavi et al., 2020], color image inpainting [Qin et al., 2020], video inpainting [Patwardhan et al., 2007], hyperspectral data recovery [Li and Li, 2010], higher-order web link analysis [Kolda et al., 2005], etc.

In general, low-rank data completion techniques can be classified into convex and non-convex approaches, and a recent survey can be found in [Song et al., 2019]. Specifically, convex approaches to matrix completion are typically based on nuclear norm minimization with theoretical optimality [Candes and Recht, 2012; Candès and Recht, 2009]. Moreover, non-convex approaches such as alternating minimization are much faster than convex methods and empirically observed to always converge to the optimum, which has also been shown theoretically [Jain et al., 2013; Sun and Luo, 2016]. Similarly, for the low-rank tensor completion, various nuclear norm minimization methods with theoretical performance guarantees have been introduced [Liu et al., 2013b; Song et al., 2020; Liu et al., 2020; Zhang and Aeron, 2016], as well as the non-convex approaches such as alternating minimization [Zheng et al., 2019; Liu et al., 2019; Wang et al., 2016], that make use of various tensor decompositions and are much faster than convex methods.

Related to data completion is the problem of data clustering in a union of subspaces with missing data. For example, let $S_k$ be a subspace of $\mathbb{R}^N$ with rank $r_k$, $k = 1, \ldots, K$. Given a data matrix $U \in \mathbb{R}^{N \times T}$ possibly with missing entries, the problem is to assign each column $u_t$ of $U$ to a particular subspace $S_k$. For example, in face recognition, the $K$ subspaces represent
K persons and each vector $u_t$ corresponds to the photo of a person. The clustering problem is then to assign each photo to one of the $K$ persons [Liu and Yan, 2011]. Other applications of clustering a union of low-rank data structures include motion recognition [Rao et al., 2009], texture analysis [Ma et al., 2007], MIMO channel estimation [Zhang and Bitmead, 2005], image analysis [Cheng et al., 2011], etc. Classical approaches to subspace clustering include maximum likelihood methods [Rao et al., 2008; Yang et al., 2006], algebraic algorithms [Costeira and Kanade, 1998; Gear, 1998; Wu et al., 2001] and their iterative implementations [Zhang et al., 2009], and spectral clustering of high-dimensional data based on low-rank representation [Liu et al., 2010]. Moreover, most of these techniques can be extended to handle missing data. For example, nuclear norm minimization is employed in dictionary learning for spectral clustering in [Liu and Yan, 2011]; and in algebraic methods, subspaces where the sampled columns belong to are identified by analyzing a set of homogeneous polynomials [Vidal et al., 2005].

In the union of subspace data structure mentioned above, the subspaces $S_1, \ldots, S_K$ are unrelated to each other. In this chapter, we consider a union of nested low-rank subspaces, i.e., we assume that the subspaces are related according to $S_1 \subset S_2 \subset \ldots \subset S_K$ which reflects the hierarchical data structure. For example, $S_1$ can correspond to pictures of German Shepherd dogs, $S_2$ to pictures of dogs, and $S_3$ to pictures of animals. Theoretical aspects of clustering and completion of such union of nested subspaces data are studied in [Ashraphijuo and Wang, 2019a; Ashraphijuo and Wang, 2020b; Ashraphijuo and Wang, 2019b]. In particular, the fundamental conditions on the sampling patterns for correctly clustering are characterized for matrices and tensors in [Ashraphijuo and Wang, 2019a] and [Ashraphijuo and Wang, 2020b], respectively. Moreover, conditions on the sampling patterns for unique completability of the correctly clustered data are given for matrices and tensors in [Ashraphijuo and Wang, 2019b] and [Ashraphijuo and Wang, 2020b], respectively. However, to date there is no algorithmic study on union of nested subspaces data.

In this chapter, we develop efficient alternating minimization based completion algorithms for union of nested subspaces two-dimensional (matrix) and higher-dimensional (tensor) data. For both matrix and tensor cases, first, we show that the union of nested subspaces structure and the corresponding rank constraints lead to a structured decomposition where certain factor has blocks of zeros. Then we develop an alternating minimization algorithm that alternatively updates each factor in the structured decomposition. Since initialization plays an important role in non-
convex optimization, we also propose two structured initialization methods, one is based on random initialization and the other is based on solving several smaller least-squares problems. Extensive simulation results are provided to compare our proposed structured approaches to several naive methods that are also based on alternating minimization. Extensive simulation results show that the proposed structured approaches offer both faster convergence speed and higher data recovery accuracy, for both matrix and tensor data, with or without noise.

The remainder of the chapter is organized as follows. In Section 7.2, we formulate the problem of union of nested subspaces data completion for the matrix case, outline three naive methods for solving it based on alternating minimization, and then propose our structured alternating minimization algorithm. In Section 7.3, we consider the same problem for the tensor case and develop the corresponding structured alternating minimization algorithm. Simulation results are presented in Section 7.4. And finally summary is drawn in Section 7.5.

### 7.2 Completion of Union of Nested Low-Rank Matrices

#### 7.2.1 Problem Statement

Assume that $K \geq 2$, $n_1 < n_2 < \cdots < n_K$ and $m$ are given integers. Let $\mathbf{U} \in \mathbb{R}^{m \times n_K}$ be a sampled matrix and denote the matrix consisting of the first $n_k$ columns of $\mathbf{U}$ by $\mathbf{U}_k \in \mathbb{R}^{m \times n_k}$ and also define $\mathbf{M}_k \in \mathbb{R}^{m \times (n_k - n_{k-1})}$ as the matrix consisting of the $(n_k - n_{k-1})$ columns of $\mathbf{U}_k$ that does not belong to $\mathbf{U}_{k-1}$, $k = 1, \ldots, K$. This is shown in Fig. 7.1 and note that $\mathbf{U} = \mathbf{U}_K$. Moreover, assume that $\text{rank}(\mathbf{U}_k) = r_k$, $k = 1, \ldots, K$. Hence, we have $r_1 \leq r_2 \leq \ldots \leq r_K$.

Let $\Omega$ denote the set of indices corresponding to the sampled entries, i.e., $\Omega = \{(i,j) : \mathbf{U}(i,j) \text{ is sampled}\}$. Moreover, define $\mathbf{U}_\Omega$ as the matrix obtained from sampling $\mathbf{U}$ according to $\Omega$, i.e.,

$$
\mathbf{U}_\Omega(i,j) = \begin{cases} 
\mathbf{U}(i,j) & \text{if } (i,j) \in \Omega, \\
0 & \text{if } (i,j) \notin \Omega.
\end{cases}
$$

We are interested in retrieving the missing entries of $\mathbf{U}$ using an efficient alternating minimization-based method. The challenge is to take advantage of all $K$ rank constraints simultaneously.
CHAPTER 7. STRUCTURED ALTERNATING MINIMIZATION FOR UNION OF NESTED LOW-RANK SUBSPACES DATA COMPLETION

Figure 7.1: The union of nested subspaces data structure, rank($U_k$) = $r_k$.

7.2.2 Naive Approaches

7.2.2.1 Naive Initialization Methods

Our goal is to find $X \in \mathbb{R}^{m \times r_K}$ and $Y \in \mathbb{R}^{r_K \times n}$ such that $U = XY$. We consider two simple methods for setting the initial values of $X$ and $Y$—SVD-based initialization and random initialization. In the SVD-based method, we first compute the singular value decomposition (SVD) of $U_\Omega$ and pick the $r_K$ largest eigenvalues and their corresponding eigenvectors to construct the initial matrices $X_0$ and $Y_0$. In particular, if $U_\Omega = U_0 S_0 V_0^T$, where the number of nonzero diagonal entries of $S_0$ can be more than $r_K$. Then, we define a decomposition corresponding to the $r_K$ largest singular values, i.e., $U_0(:, 1: r_K) S_0(1: r_K, 1: r_K) V_0(:, 1: r_K)^T = X_0 Y_0$, where $X_0 = U_0(:, 1: r_K) S_0(1: r_K, 1: r_K) \in \mathbb{R}^{m \times r_K}$ and $Y_0 = V_0(:, 1: r_K)^T \in \mathbb{R}^{r_K \times n_K}$.

On the other hand, for random initialization, we simply set $X_0$ and $Y_0$ as matrices that contain i.i.d. $\mathcal{N}(0, 1)$ samples. Note that the initial matrix satisfies only one rank constraint $r_K$ for both methods.
7.2.2.2 Naive Alternating Minimization Methods

(i) Naive approach 1: In this approach we simply discard the rank constraints \( r_1, \ldots, r_{K-1} \) and complete \( U \) using only the constraint \( \text{rank}(U) = r_K \). Starting from the above initial \( X_0 \) and \( Y_0 \), in each iteration, we alternatively optimize \( X_i \) and \( Y_i \) until convergence.

In particular, at the \( i \)-th iteration, given \( X_{i-1} \) and \( Y_{i-1} \), we first update \( X_i \) by solving the following regularized least-squares problem

\[
\min_{X_i \in \mathbb{R}^{m \times r_K}} \| U_{\Omega} - (X_i Y_{i-1})_{\Omega} \|_F + \lambda \| X_i \|_F,
\]

and then update \( Y_i \) by solving

\[
\min_{Y_i \in \mathbb{R}^{r_K \times n_K}} \| U_{\Omega} - (X_i Y_i)_{\Omega} \|_F + \lambda \| Y_i \|_F,
\]

where \( \| \cdot \|_F \) denotes the Frobenius norm and \( \lambda \) is a small constant. The purpose of the regularization term is to avoid singularity in solving the least-squares problems. The iteration continues until it reaches convergence or until the algorithm diverges. The solutions to (7.2) and (7.3) can be obtained row by row and column by column, respectively. In particular, denote \( E \) as an all-one \( m \times n_K \) matrix, then (7.2) can be solved as

\[
X_i(j, :) = \arg\min_{X_i(j, :)} \| U_{\Omega}(j, :) - X_i(j, :)Y_i-1\text{Diag}[E_{\Omega}(j, :)] \|_F + \lambda \| X_i(j, :) \|_F,
\]

where \( \text{Diag}[v] \) denotes a diagonal matrix with the diagonal entries being the entries of \( v \). Similarly, (7.3) can be solved as

\[
Y_i(:, j) = \arg\min_{Y_i(:, j)} \| U_{\Omega}(:, j) - \text{Diag}[E_{\Omega}(:, j)] X_i Y_i(:, j) \|_F + \lambda \| Y_i(:, j) \|_F,
\]

Note that the output of this simple approach, \( \hat{U} = X_N Y_N \) for some \( N \), satisfies only the rank constraint \( \text{rank}(\hat{U}) = r_K \), but \( \hat{U} \) may not satisfy other \( K-1 \) rank constraints.

(ii) Naive approach 2: In this approach, we break the original problem into \( K \) independent completion problems, i.e., completing \( M_k \in \mathbb{R}^{m \times (n_k-n_{k-1})} \) with \( \text{rank}(M_k) = r_k \). This method may
be fast as each subproblem has a smaller dimension. However, it may result in a solution that does not satisfy any of the rank constraints except for the first one (for $M_1$), since rank ($M_k$) = $r_k$ does not necessarily result in rank ($U_k$) = $r_k$ (except for $k = 1$).

(iii) Naive approach 3: In this approach, we first complete $U_1$ with constraint rank($U_1$) = $r_1$ using the above alternating minimization method. Then, we complete $U_2$ = [$U_1$|$M_2$] with the constraint rank($U_2$) = $r_2$. Note that the $U_1$ part of $U_2$ is already complete and all missing entries are in the $M_2$ part of $U_2$. This is repeated and in the $k$-th step, we complete the $M_k$ part of $U_k$ = [$U_{k-1}$|$M_k$] with the constraint rank($U_k$) = $r_k$. One important issue with this method is the error propagation when the sampling rate is low, i.e., the erroneously recovered entries at any step will lead to further errors in subsequent steps. However, the output of this method satisfies all rank constraints.

7.2.3 Structured Decomposition

In this chapter, we propose a structured alternating minimization method for completing $U$ such that: (1) all $K$ rank constraints are satisfied at each iteration and, (2) it converges faster than the conventional alternating minimization for matrix completion with a single constraint rank($U$) = $r_K$, by exploiting the additional $K - 1$ rank constraints. To this end, we make use of a structured decomposition of $U$ that is determined by the $K$ rank constraints.

**Definition 21.** Consider a decomposition $U = XY$ such that $X \in \mathbb{R}^{m \times r_K}$, $Y \in \mathbb{R}^{r_K \times n_K}$, and $Y(r_1 + 1 : r_K, 1 : n_1) = 0_{(r_K-r_1)\times n_1}$, $Y(r_2 + 1 : r_K, n_1 + 1 : n_2) = 0_{(r_K-r_2)\times (n_2-n_1)}$, ... and $Y(r_{K-1} + 1 : r_K, n_{K-2} + 1 : n_{K-1}) = 0_{(r_K-r_{K-1})\times (n_{K-1}-n_{K-2})}$. This structure is shown in Fig. 7.2 and we call such decomposition $U = XY$ a structured decomposition.

**Lemma 43.** Consider a matrix $U \in \mathbb{R}^{m \times n_K}$ that has a structured decomposition $U = XY$. Then, rank($U(:,1:n_k)$) $\leq r_k$, $k = 1, \ldots, K$.

**Proof.** Note that $U(:,1:n_k) = XY(:,1:n_k)$. Hence, under the structured decomposition, we conclude that $U(:,1:n_k) = X(:,1:r_k)Y(1:r_k,1:n_k)$ (because $Y(r_k + 1 : r_K, 1 : n_k) = 0_{(r_K-r_k)\times n_k}$), $k = 1, \ldots, K$. Then, $U(:,1:n_k) = X(:,1:r_k)Y(1:r_k,1:n_k)$ results that rank($U(:,1:n_k)$) $\leq r_k$, $k = 1, \ldots, K$. \qed
CHAPTER 7. STRUCTURED ALTERNATING MINIMIZATION FOR UNION OF NESTED LOW-RANK SUBSPACES DATA COMPLETION

Figure 7.2: A matrix \( Y \) that satisfies the properties of a structured decomposition given in Definition 21.

Lemma 44. If the matrix \( U \in \mathbb{R}^{m \times nK} \) has the union of nested subspaces structure shown in Fig. 7.1, then there exists a structured decomposition \( U = XY \).

Proof. We need to show that there exists a basis \( X \) for \( U \) such that the first \( n_k \) columns of \( U \) belong to the subspace span of the first \( r_k \) columns of \( X \), \( k = 1, \ldots, K \). Note that it is easily verified that this statement is equivalent with the existence of a decomposition \( U = XY \) such that \( Y \) satisfies the structure given in Definition 21. We show the mentioned statement by induction on \( k \). In the \( k \)-th step, we construct \( X^k \) such that \( U_k' \) belongs to the column span of the first \( r_k' \) columns of \( X^k \), \( k' = 1, \ldots, k \). Note that for \( k = 1 \) it is straightforward to construct \( X^1 \), which is simply a basis for \( U_1 \). Induction hypothesis results in the matrix \( X^k \) with the mentioned properties and in order to complete the induction, we need to show the existence of a matrix \( X^{k+1} \) such that \( U_{k+1}' \) belongs to the column span of the first \( r_{k+1}' \) columns of \( X^{k+1} \), \( k' = 1, \ldots, k+1 \).

We first note that \( X^k \) belongs to the column span of \( U_{k+1} \), because according to the induction hypothesis, \( X^k \) is a basis for \( U_k \), and \( U_k \) is a subset of columns of \( U_{k+1} \). Let \( S_k \) denote the column span of \( X^k \), which is an \( r_k \)-dimensional space and \( S_{k+1}' \) denote the column span of \( U_{k+1}' \), which is an \( r_{k+1}' \)-dimensional space. As a result of our earlier claim, \( S_k \) is a subspace of \( S_{k+1}' \). Let \( S_k'' \) denote the \((r_{k+1} - r_k)\)-dimensional subspace of \( S_{k+1}' \) such that the union of \( S_k \) and \( S_k'' \) is \( S_{k+1}' \).

Consider an arbitrary basis \( X^{k'} \in \mathbb{R}^{m \times (r_{k+1} - r_k)} \) for the space \( S_k'' \). Observe that by putting together the columns of \( X^k \) and \( X^{k'} \), i.e., \( X^{k+1} = [X^k \ X^{k'}] \), the new matrix \( X^{k+1} \in \mathbb{R}^{m \times r_{k+1}} \) is a basis for the space \( S_{k+1}' \). Therefore, \( U_{k+1} \) belongs to the column span of the first \( r_{k+1} \) columns of \( X^{k+1} \) since \( X^k \) has exactly \( r_{k+1} \) columns. Given the induction hypothesis, the proof is complete as
\( U_{k'} \) belongs to the column span of the first \( r_{k'} \) columns of \( X^{k+1}, k' = 1, \ldots, k + 1. \)

### 7.2.4 Proposed Structured Alternating Minimization Algorithm

We are interested in imposing the structured decomposition in the alternating minimization procedure for two reasons: (i) we know that according to Lemma 44 there exists such a decomposition and also, according to Lemma 43 the \( K \) rank constraints on the original data will hold in such decomposition and therefore, it is more likely that such decomposition results in the recovery of the original data. (ii) A structured decomposition has many zeros and convergence may be much faster than an unstructured decomposition.

There are two main challenges to impose a structured decomposition in alternating minimization: (i) an efficient initialization, and (ii) an efficient update of \( Y \) at each iteration.

#### 7.2.4.1 Structured Initialization Methods

**SVD-based structured initialization:** We first discuss the SVD-based method.

In order to obtain an initialization \( X_0 \in \mathbb{R}^{m \times r_K} \) and \( Y_0 \in \mathbb{R}^{r_K \times n_K} \), we will obtain \( X^k \in \mathbb{R}^{m \times (r_k - r_{k-1})} \) and \( Y^k \in \mathbb{R}^{r_K \times (n_k - n_{k-1})} \) for \( k = 1, \ldots, K \), where \( n_0 = r_0 = 0 \). Then \( X_0 = [X^1 \ldots X^K] \) and \( Y_0 = [Y^1 \ldots Y^K] \).

In other words, in the \( k \)-th step, we obtain \( r_k - r_{k-1} \) columns of the basis, i.e., \( X^k \), and the corresponding coefficients of these \( r_k - r_{k-1} \) columns of the basis in \( M_{k'} \)'s, i.e., \( Y^k \). Note that we set the coefficients corresponding to \( X^k \) in \( M_{k'} \) for \( k' < k \) as zeros to meet the structured decomposition.

As the first step, we compute the SVD of \( M_{10} \in \mathbb{R}^{m \times n_1} \) and pick the \( r_1 \) largest eigenvalues and their corresponding eigenvectors to construct matrices \( X^1 \in \mathbb{R}^{m \times r_1} \) and \( Z^1 \in \mathbb{R}^{r_1 \times n_1} \), similar to the initialization explained in Sec. 7.2.1 for the naive approaches. Then, we define \( Y^1 = [Z^1^\top 0_{n_1 \times (r_K - r_1)}] \in \mathbb{R}^{r_K \times n_1} \), as shown in Fig. 7.3.

In the second step, we first obtain the SVD of \( M_{20} \in \mathbb{R}^{m \times n_2} \) and pick the \( r_2 - r_1 \) largest eigenvalues and their corresponding eigenvectors to construct matrices \( X^2 \in \mathbb{R}^{m \times (r_2 - r_1)} \) and \( Z^2 \in \mathbb{R}^{(r_2 - r_1) \times (n_2 - n_1)} \). Next, we want to obtain \( Y^2 = [K^2^\top Z^2^\top 0_{(n_2 - n_1) \times (r_K - r_2)}] \in \mathbb{R}^{r_K \times (n_2 - n_1)} \), as shown in Fig. 7.3, where \( K^2 \in \mathbb{R}^{r_1 \times (n_2 - n_1)} \) represents the coefficients of \( X^2 \) in \( M_1 \), which is based
on the projection of matrix \((M_2 - X^2Z^2)_\Omega\) on \(X^1\). Specifically, we have

\[
K^2 = \arg\min_{K^2 \in \mathbb{R}^{r_1 \times (n_2-n_1)}} \|((M_2 - X^2Z^2) - X^1K^2)_\Omega\|_F + \lambda\|K^2\|_F, \tag{7.6}
\]

which can be solved column by column similar to (7.5)

\[
K^2(:, j) = \left( X^1^\top \text{Diag} [E_{\Omega}(\cdot, j)] X^1 + \lambda I \right)^{-1} X^1^\top \text{Diag} [E_{\Omega}(\cdot, j)] T^2_{\Omega}(\cdot, j), \quad j = 1, \ldots, n_2 - n_1, \tag{7.7}
\]

where \(T^2 = M_2 - X^2Z^2\).

Similarly, in the \(k\)-th step, we first obtain \(X^k \in \mathbb{R}^{m \times (r_k-r_{k-1})}\) and \(Z^k \in \mathbb{R}^{(r_k-r_{k-1}) \times (n_k-n_{k-1})}\) from the SVD of \(M_k \in \mathbb{R}^{m \times n_k}\). Then, we construct

\[
Y^k = \begin{bmatrix} K^k^\top & Z^k^\top & 0_{(n_k-n_{k-1}) \times (r_k-r_k)} \end{bmatrix}^\top \in \mathbb{R}^{r_K \times (n_k-n_{k-1})}, \tag{7.8}
\]

as shown in Fig. 7.3, where \(K^k \in \mathbb{R}^{r_k \times (n_k-n_{k-1})}\) is the coefficient of \(\bar{X}^{k-1} = [X^1 \ldots X^{k-1}]\) in \(M_k\), which is obtained based on the projection of \((M_k - X^kZ^k)_\Omega\) on \(\bar{X}^k\), i.e.,

\[
K^k = \arg\min_{K^k \in \mathbb{R}^{r_k \times (n_k-n_{k-1})}} \|((M_k - X^kZ^k) - \bar{X}^{k-1}K^k)_\Omega\|_F + \lambda\|K^k\|_F. \tag{7.9}
\]

The solution is given by

\[
K^k(:, j) = \left( \bar{X}^{k-1}^\top \text{Diag} [E_{\Omega}(\cdot, j)] \bar{X}^{k-1} + \lambda I \right)^{-1} \bar{X}^{k-1}^\top \text{Diag} [E_{\Omega}(\cdot, j)] T^k_{\Omega}(\cdot, j), \quad j = 1, \ldots, n_k - n_{k-1}, \tag{7.10}
\]

where \(T^k = M_k - X^kZ^k\).

**Remark 22.** In the above SVD-based structured initialization the choice of \(K^k\) in (7.9) plays a critical role. In particular, numerical experiments show that simply setting \(K^k = 0_{r_K \times (n_k-n_{k-1})}\), for \(k = 2, \ldots, K\), will result in a poor initialization which significantly reduces the convergence speed of the algorithm.

Random structured initialization: We also consider random initialization for the proposed structured alternating minimization where \(X_0\) contains i.i.d. \(\mathcal{N}(0, 1)\) samples and the non-zero entries of \(Y_0\) in Definition 21 are also i.i.d. \(\mathcal{N}(0, 1)\) samples.
7.2.4.2 Structured Alternating Minimization

Note that the initialization satisfies the structured decomposition. Now, we need to make sure that at each iteration of the algorithm this property still holds. In particular, in the $i$-th iteration of the structured alternating minimization procedure, given $X_{i-1}$ and $Y_{i-1}$, we first update $X_i$ according to (7.4). Then, in (7.5) we only need to update the non-zero entries of $Y_i(:,j)$ in the structured decomposition. That is, for $1 \leq k \leq K$ and $n_{k-1} + 1 \leq j \leq n_k$ we have

$$Y_i(1 : r_k, j) = \left(X_i(:, 1 : r_k)^\top \text{Diag} [E_{\Omega(:, j)}] X_i(:, 1 : r_k) + \lambda I\right)^{-1} X_i(:, 1 : r_k)^\top \text{Diag} [E_{\Omega(:, j)}] U_{\Omega(:, j)}.$$  

(7.11)

Finally, we summarize the proposed structured alternating minimization algorithm for union of nested low-rank matrices completion in Algorithm 1. Note that at each iteration of this algorithm, including the initialization, the structured decomposition holds and therefore all $K$ rank constraints hold.

7.3 Completion of Union of Nested Low-Rank Tensors

In this section, we generalize the structured alternating minimization approach to a union of nested low-rank tensor spaces.
Algorithm 1 Structured Alternating Minimization - Matrix Case

1: Input $\mathbf{U}_\Omega$, $r_1, \ldots, r_K$ and $n_1, \ldots, n_K$.
2: Initializing $\mathbf{X}_0 \in \mathbb{R}^{m \times r_K}$ and $\mathbf{Y}_0 \in \mathbb{R}^{r_K \times n_K}$ using either the SVD-based or random structured initialization.
3: repeat
   4:   for $j = 1 : m$ do
   5:       Compute Eq. (7.4).
   6:   end for
   7:   for $k = 1 : K$ do
      8:       for $j = n_{k-1} + 1 : n_k$ do
      9:           Compute Eq. (7.11).
     10:       end for
    11:   end for
   12: until convergence/divergence

7.3.1 Background

Recall that the CP-rank of a tensor $\mathbf{U} \in \mathbb{R}^{m_1 \times m_2 \times \ldots \times m_{d-1} \times m_d}$ is the minimum number $r$ such that there exist $\mathbf{a}^l_j \in \mathbb{R}^{m_j}$ for $1 \leq j \leq d$ and $1 \leq l \leq r$ and

$$
\mathbf{U} = \sum_{l=1}^r \mathbf{a}^l_1 \otimes \mathbf{a}^l_2 \otimes \ldots \otimes \mathbf{a}^l_d,
$$

or equivalently,

$$
\mathbf{U}(x_1, x_2, \ldots, x_d) = \sum_{l=1}^r \mathbf{a}^l_1(x_1) \mathbf{a}^l_2(x_2) \ldots \mathbf{a}^l_d(x_d),
$$

where $\otimes$ denotes the tensor product (outer product) and $\mathbf{U}(x_1, x_2, \ldots, x_d)$ denotes the entry of tensor $\mathbf{U}$ with coordinate $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ and $\mathbf{a}^l_j(x_j)$ denotes the $x_j$-th entry of vector $\mathbf{a}^l_j$. In other words, the CP-rank of a tensor $\mathbf{U}$ is the minimum number of rank-1 tensors that $\mathbf{U}$ can be decomposed to.

For notational convenience, define $M_{d-1} \triangleq m_1 m_2 \ldots m_{d-1}$. Moreover, define the matrix $\overline{\mathbf{U}} \in \mathbb{R}^{M_{d-1} \times m_d}$ as the $(d-1)$-th unfolding of tensor $\mathbf{U}$, such that $\mathbf{U}(\mathbf{x}) = \overline{\mathbf{U}}(v(x_1, \ldots, x_{d-1}, x_d))$, where $v : (x_1, \ldots, x_{d-1}) \to \{1, 2, \ldots, M_{d-1}\}$ is a bijective mapping. Note that this is a vectorization
mapping that merges the first \((d-1)\) dimensions and therefore, there is a corresponding inverse mapping \(v^{-1} : \{1, 2, \ldots, M_{d-1}\} \rightarrow (x_1, \ldots, x_{d-1})\). Moreover, for a \((d-1)\)-dimensional tensor \(\mathcal{V} \in \mathbb{R}^{m_1 \times \cdots \times m_{d-1}}\) we can define a vectorization operator \(\text{vec} : \mathbb{R}^{m_1 \times \cdots \times m_{d-1}} \rightarrow \mathbb{R}^{m_{d-1}}\) using the mapping \(v(\cdot)\) such that \(\mathcal{V}(x_1, ..., x_{d-1}) = \text{vec}(\mathcal{V})(v(x_1, ..., x_{d-1}))\). We call a vector \(\mathbf{u} \in \mathbb{R}^{m_{d-1}}\) an “structured column” if \(\text{vec}^{-1}(\mathbf{u}) \in \mathbb{R}^{m_1 \times \cdots \times m_{d-1}}\) is a rank-1 tensor, i.e., there exist \(\mathbf{u}_j \in \mathbb{R}^{m_j}\) for \(j = 1, \ldots, d-1\), such that \(\mathbf{u} = \text{vec}(\mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d-1})\).

**Lemma 45.** The CP-rank of a tensor \(\mathcal{U}\) is equal to the minimum number of structured columns that span all columns of \(\mathcal{U}\).

**Proof.** First we show that there exist \(\mathbf{b}_1^l \in \mathbb{R}^{m_{d-1}}\) and \(\mathbf{b}_2^l \in \mathbb{R}^{m_d}\) for \(1 \leq l \leq r\) such that

\[
\mathcal{U} = \sum_{l=1}^{r} \mathbf{b}_1^l \otimes \mathbf{b}_2^l. \tag{7.14}
\]

Recall the CP decomposition in (7.12). Then, we define \(\mathcal{A}_1^l = \mathbf{a}_1^l \otimes \cdots \otimes \mathbf{a}_{d-1}^l\) and \(\mathbf{b}_2^l = \mathbf{a}_d^l\) for \(1 \leq l \leq l\) and define \(\mathbf{b}_1^l = \text{vec}(\mathcal{A}_1^l)\). Hence, there exist \(\mathbf{b}_1^l \in \mathbb{R}^{m_{d-1}}\) and \(\mathbf{b}_2^l \in \mathbb{R}^{m_d}\) for \(1 \leq l \leq r\) such that (7.14) holds. Therefore, there exist \(r\) structured columns that span all columns of \(\mathcal{U}\). Similarly, if there exists \((r-1)\) structured columns that span all columns of \(\mathcal{U}\) we can use \(\text{vec}^{-1}\) and obtain a CP-decomposition of rank \((r-1)\) for \(\mathcal{U}\). Therefore, \(\text{rank}(\mathcal{U}) = r\) means that \(r\) is the minimum number of structured columns that span all columns of \(\mathcal{U}\).

**Definition 22.** According to the above lemma, \(\text{rank}(\mathcal{U}) = r\) concludes that there exists a set \(\mathcal{S}\) consisting of \(r\) structured columns whose column span (denoted by \(\mathcal{T}\)) includes any column of \(\mathcal{U}\). In other words, the column span of these \(r\) structured columns, i.e., \(\mathcal{T}\), is an unfolded tensor space of rank \(r\). We call such \(r\) structured columns a tensor basis for \(\mathcal{U}\).

### 7.3.2 Problem Statement

Consider a fixed number \(K \geq 2\) and partially sampled \(d\)-way tensors \(\mathcal{M}_k \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_{d-1} \times c_k}\), \(k = 1, 2, \ldots, K\). Define \(n_k = c_1 + \cdots + c_k\) for \(k = 1, \ldots, K\), and \(c_0 = n_0 = 0\). Let \(\mathcal{U}_k \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_{d-1} \times n_k}\), be the concatenation of \(\mathcal{M}_1, \ldots, \mathcal{M}_k\) along the \(d\)-th dimension, and \(r_k\) denote the CP-rank of \(\mathcal{U}_k\), \(k = 1, 2, \ldots, K\). Let \(\Omega\) denote the sampled index set, i.e., \(\Omega = \{\vec{x} = (x_1, \ldots, x_d) : \mathcal{U}(\vec{x}) \text{ is sampled}\}\). Moreover, define \(\mathcal{U}_\Omega\) as the tensor obtained from sampling \(\mathcal{U} = \mathcal{U}_K\) according to
\[ U_{\Omega}(\vec{x}) = \begin{cases} 
U(\vec{x}) & \text{if } \vec{x} \in \Omega, \\
0 & \text{if } \vec{x} \notin \Omega. 
\end{cases} \] (7.15)

Moreover, we assume a union of nested tensor subspaces structure similar to the matrix case. Specifically, assume that there exist structured columns \( u_l \in \mathbb{R}^{M_{d-1}} \), \( l = 1, \ldots, m_j \), such that \( S_k = \{ u_1, \ldots, u_{r_k} \} \) is a tensor basis for \( U_k \), \( k = 1, \ldots, K \). Note that we have \( \text{rank}(U_k) = \text{rank}(M_k) = r_k \), \( k = 1, \ldots, K \). The problem then is to complete the tensor \( U_\Omega \) given the above mentioned union of nested tensor subspaces structure, and the rank values \( r_1, \ldots, r_K \).

### 7.3.3 Alternating Minimization For Tensor Completion

Recall the CP decomposition \( U = \sum_{i=1}^{r_K} a_1^i \otimes a_2^i \otimes \cdots \otimes a_d^i \), where \( a_j^i \in \mathbb{R}^{m_j} \) for \( 1 \leq j \leq d \) and \( 1 \leq l \leq r_K \). Define \( A_j = [a_j^1| \ldots |a_j^{r_K}] \in \mathbb{R}^{m_j \times r_K} \), \( j = 1, \ldots, d \). In alternating minimization, given the result of the \( (i-1) \)-th iteration \( A_j^{(i-1)} \in \mathbb{R}^{m_j \times r_K} \), \( j = 1, \ldots, d \), at the \( i \)-th iteration, we update all \( A_j \)'s one by one in \( d \) steps. In particular, in the \( j \)-th step, we solve for \( A_j^{(i)} \) using the latest values \( A_1^{(i)}, \ldots, A_{j-1}^{(i)} \), and \( A_{j+1}^{(i-1)}, \ldots, A_d^{(i-1)} \) by solving the following regularized least squares problem

\[
\min_{A_j^{(i)} \in \mathbb{R}^{m_j \times r_K}} \left\| U_{\Omega} - \left( \sum_{l=1}^{r_K} A_1^{(i)}(:,l) \otimes \cdots \otimes A_{j-1}^{(i)}(:,l) \otimes A_j^{(i)}(:,l) \otimes A_{j+1}^{(i-1)}(:,l) \otimes \cdots \otimes A_d^{(i-1)}(:,l) \right) \right\|_F \\
+ \lambda \| A_j^{(i)} \|_F, \; j = 1, \ldots, d.
\] (7.16)

To solve (7.16), we first write it in matrix form. To do this, we define an operator that reorders the dimensions of a tensor. Consider the tensor in (7.12) and another tensor

\[ U' = \sum_{l=1}^{r_K} a_1^l \otimes \cdots \otimes a_{j-1}^l \otimes a_{j+1}^l \otimes \cdots \otimes a_d^l \otimes a_j^l. \] (7.17)

Then, it is clear that the only difference between these two tensors is that the order of dimensions has changed from \( 1, 2, \ldots, d \) in \( U \) to \( 1, 2, \ldots, j - 1, j + 1, \ldots, d, j \) in \( U' \). Denote such a dimension reordering operation by \( U' = \sigma_j(U) \in \mathbb{R}^{m_1 \times \cdots \times m_{j-1} \times m_{j+1} \times \cdots \times m_d \times m_j} \) such that

\[ U(x_1, x_2, \ldots, x_d) = \sigma_j(U)(x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_d, x_j). \] (7.18)
Then, (7.16) can be rewritten as

\[
\min_{A^{(i)}_{j} \in \mathbb{R}^{m_j \times r_K}} \left\| \sigma_j(U)_{\sigma_j(\Omega)} - \left( \sum_{l=1}^{r_K} A^{(i)}_{1}(; l) \otimes \cdots \otimes A^{(i)}_{j-1}(; l) \otimes A^{(i-1)}_{j+1}(; l) \otimes \cdots \otimes A^{(i-1)}_{d}(; l) \otimes A^{(i)}_{j}(; l) \right)_{\sigma_j(\Omega)} \right\|_F \\
+ \lambda \|A^{(i)}_{j}\|_F, j = 1, \ldots, d.
\]  

(7.19)

Now we define \( \Gamma^{(i)}_{j} \in \mathbb{R}^{m_1 \ldots m_{j-1} m_{j+1} \ldots m_d \times r_K} \) such that

\[
\Gamma^{(i)}_{j}(; \ell) \triangleq \text{vec}(A^{(i)}_{1}(; \ell) \otimes \cdots \otimes A^{(i-1)}_{j-1}(; \ell) \otimes A^{(i-1)}_{j+1}(; \ell) \otimes \cdots \otimes A^{(i-1)}_{d}(; \ell)) \in \mathbb{R}^{m_1 \ldots m_{j-1} m_{j+1} \ldots m_d}, \ell = 1, \ldots, r_K.
\]  

(7.20)

Then, we can rewrite (7.19) using its \((d-1)\)-th unfolding as

\[
\min_{A^{(i)}_{j} \in \mathbb{R}^{m_j \times r_K}} \left\| \hat{\sigma}_j(U)_{\sigma_j(\Omega)} - \left( \Gamma^{(i)}_{j} A^{(i)}_{j} \right)_{\sigma_j(\Omega)} \right\|_F + \lambda \|A^{(i)}_{j}\|_F, 
\]  

(7.21)

where \( \hat{\sigma}_j(U) \) and \( \hat{\sigma}_j(\Omega) \) denote the \((d-1)\)-th unfolding of \( \sigma_j(U) \) and \( \sigma_j(\Omega) \), respectively. Note that (7.21) is of the same form as (7.3) and hence similar to (7.5), we can write

\[
A^{(i)}_{j}(l, :) = \left( \Gamma^{(i)}_{j} \right)^\top \text{Diag} \left[ E_{\sigma_j(\Omega)}(; l) \right] \Gamma^{(i)}_{j} \left( \left( \Gamma^{(i)}_{j} \right)^\top \text{Diag} \left[ E_{\sigma_j(\Omega)}(; l) \right] \Gamma^{(i)}_{j} + \lambda I \right)^{-1}, l = 1, \ldots, m_j,
\]  

(7.22)

where \( E \) denotes an all-one \((m_1 \ldots m_{j-1} m_{j+1} \ldots m_d) \times m_j \) matrix.

### 7.3.4 Naive Approaches

We can generalize the three naive approaches for the matrix case to the tensor case as follows.

(i) Naive approach 1: We apply the alternating minimization procedure described in Sec. 7.3.3 to tensor \( U \) with the only constraint \( \text{rank}(U) = r_K \).

(ii) Naive approach 2: We break the original problem into \( K \) independent completion problems, i.e., completing \( M_k \in \mathbb{R}^{m_1 \times \cdots \times m_d \times (n_k - n_{k-1})} \) with \( \text{rank}(M_k) = r_k, k = 1, \ldots, K \).

(iii) Naive approach 3: We first complete \( U_1 \) with constraint \( \text{rank}(U_1) = r_1 \) using the above alternating minimization method. Then, we complete \( U_2 \) with the constraint \( \text{rank}(U_2) = r_2 \). Note
that the $U_1$ part of $U_2$ is already complete and all missing entries are in the $M_2$ part of $U_2$. This is repeated and in the $k$-th step, we complete the $M_k$ part of $U_k$ with the constraint $\text{rank}(U_k) = r_k$, $k = 1, \ldots, K$.

Similarly to the matrix case, for each of the above naive methods, either CP-based or random initialization can be employed. Specifically, for CP-based initialization, we first calculate the CP decomposition of $U_{\Omega}$. Then, we normalize each vector $a^l_j$ in (7.12) to have unit norm so that the $l$-th out-product has a weight of $\|a^l_1\| \cdots \|a^l_d\|$. Then, we choose the leading $r_K$ rank-1 components, sorted according to the weights, to obtain a rank-$r_K$ initialization. And for random initialization, we simply set entries of $A^{(0)}$, $j = 1, \ldots, d$ as i.i.d $\mathcal{N}(0,1)$ samples.

### 7.3.5 Structured Decomposition

Similar to the matrix case, we develop a tensor completion method based on alternating minimization that takes into account all $K$ rank constraints. First, similar to Definition 1 and Lemma 2 for the matrix case, we have the following lemma on the existence of a structured CP decomposition for tensor $U$ that has a union of nested subspaces structure defined in Sec. III.B. Recall that $n_k = c_1 + \cdots + c_k$.

**Lemma 46.** If the tensor $U \in \mathbb{R}^{m_1 \times \cdots \times m_{d-1} \times r_K}$ has the union of nested tensor subspaces structure, then there exist $a^l_d \in \mathbb{R}^{m_d}$ for $l = 1, \ldots, r_K$ such that $U = \sum_{l=1}^{r_K} a^l_1 \otimes \cdots \otimes a^l_{d-1} \otimes a^l_d$ and for any $k = 1, \ldots, K$, $x = n_{k-1} + 1, \ldots, n_k$ and $l = r_k + 1, \ldots, r_K$ we have $a^l_d(x) = 0$. In other words, $A_d(n_{k-1} + 1 : n_k, 1 : r_K - r_k) = 0_{c_k \times (r_K - r_k)}$, $k = 1, \ldots, K$. We call such CP-decomposition of $U$ a structured decomposition (shown in Fig. 7.4).

**Proof.** Note that since each column of $\tilde{M}_k$ (the $(d-1)$-th unfolding of $M_k$) is chosen from the column span of $S_k$, there exist $B_k \in \mathbb{R}^{r_k \times c_k}$ such that $\tilde{M}_k = [u_1 \ldots u_{r_k}]B_k$, $k = 1, \ldots, K$. Recall that $\tilde{U} = [\tilde{M}_1 \ldots \tilde{M}_K]$. Therefore, we can write

$$
\tilde{U} = [u_1 \ldots u_{r_K}] \begin{bmatrix} C_1 & \cdots & C_K \end{bmatrix},
$$

where $C_k = [B^T_k 0_{c_k \times (r_K - r_k)}] \in \mathbb{R}^{r_K \times c_K}$ and $[a^l_d \ldots a^l_d] \in [C_1 \ldots C_K]$. Hence, for any $k = 1, \ldots, K$ and $l = r_k + 1, \ldots, r_K$ we have $a^l_d(x) = 0$ if $n_{k-1} + 1 \leq x \leq n_k$. 


7.3.6 Proposed Structured Alternating Minimization for Union of Nested Tensor Subspaces

7.3.6.1 Structured Initialization Methods

CP-based structured initialization: We obtain such structured initialization in K steps: in the k-th step, k = 1, . . . , K, we obtain $B_j^k \in \mathbb{R}^{n_j \times (r_k - r_{k-1})}$ for $j = 1, \ldots, d - 1$, and $B_d^k \in \mathbb{R}^{c_k \times r_K}$, and the initialization is $A_j^{(0)} = [B_1^1 \ldots B_j^K]$, $j = 1, \ldots, d - 1$ and $A_d^{(0)} = [B_d^1 \ldots B_d^K]^\top$. Note that here $A_j^{(0)}$ for $j = 1, \ldots, d - 1$ and $A_d^{(0)}$ correspond to $X_0$ and $Y_0$ in Sec. 7.2.4, respectively.

We first perform the CP-decomposition of $M_{1\Omega} \in \mathbb{R}^{m_1 \times \ldots \times m_{d-1} \times c_1}$ and retain the $r_1$ leading
rank-1 components, to obtain \( B_j^1 \in \mathbb{R}^{m_j \times r_1} \) for \( j = 1, \ldots, d - 1 \), and \( C_d^1 \in \mathbb{R}^{c_1 \times r_1} \), i.e.,

\[
\mathcal{M}_{1\Omega} \approx \sum_{l=1}^{r_1} B_1^1(:,l) \otimes \cdots \otimes B_{d-1}^1(:,l) \otimes C_d^1(:,l). \tag{7.25}
\]

Then we define \( B_d^1 = [C_d^1 \, \mathbf{0}_{c_1 \times (r_K - r_1)}] \in \mathbb{R}^{c_1 \times r_K} \) that meets the structure of the top block row in Fig. 7.4.

In the \( k \)-th step, \( k = 2, \ldots, K \), we perform the CP-decomposition of \( \mathcal{M}_{k\Omega} \in \mathbb{R}^{m_1 \times \cdots \times m_{d-1} \times c_k} \) and retain the \( r_k - r_{k-1} \) leading rank-1 components denoted by \( B_j^k \in \mathbb{R}^{m_j \times (r_k - r_{k-1})} \) for \( j = 1, \ldots, d - 1 \), and \( C_d^k \in \mathbb{R}^{c_k \times (r_k - r_{k-1})} \). Then, we define \( \tilde{B}_d^k = [K^k \, C_d^k \, \mathbf{0}_{c_k \times (r_K - r_k)}] \in \mathbb{R}^{c_k \times r_K} \) that meets the structure of the \( k \)-th block row in Fig. 7.4, where \( K^k \in \mathbb{R}^{c_k \times r_k - 1} \) represents the coefficients of the structured columns 1 to \( r_k - 1 \) in \( \mathcal{M}_k \), that is calculated as follows.

Let \( \tilde{\mathcal{M}}_k \in \mathbb{R}^{m_1 \times \cdots \times m_{d-1} \times c_k} \) denote the rank-(\( r_k - r_{k-1} \)) approximation of \( \mathcal{M}_{k\Omega} \), i.e.,

\[
\tilde{\mathcal{M}}_k \triangleq \sum_{l=1}^{r_k-r_{k-1}} \begin{pmatrix} B_1^k(:,l) \otimes \cdots \otimes B_{d-1}^k(:,l) \otimes C_d^k(:,l) \end{pmatrix}. \tag{7.26}
\]

Define \( \tilde{B}_j^{k-1} = [B_j^1 \ldots B_j^{k-1}] \in \mathbb{R}^{m_j \times r_{k-1}} \), \( j = 1, \ldots, d - 1 \). Then \( K^k \) is the projection of tensor \( (\mathcal{M}_k - \tilde{\mathcal{M}}_k) \) on the structured columns 1 to \( r_{k-1} \), i.e.,

\[
K^k = \arg\min_{K^k \in \mathbb{R}^{c_k \times r_{k-1}}} \left\| (\mathcal{M}_k - \tilde{\mathcal{M}}_k)_{\Omega} - \left( \sum_{l=1}^{r_k-r_{k-1}} \tilde{B}_1^{k-1}(:,l) \otimes \cdots \otimes \tilde{B}_{d-1}^{k-1}(:,l) \otimes K^k(:,l) \right) \right\|_{\mathcal{F}} + \lambda \| K^k \|_{\mathcal{F}}, \tag{7.27}
\]

which is similar to (7.19) and can be rewritten using the \( (d-1) \)-th unfoldings of the corresponding tensors as

\[
K^k = \arg\min_{K^k \in \mathbb{R}^{c_k \times r_{k-1}}} \left\| (\mathcal{M}_k - \tilde{\mathcal{M}}_k)_{\Omega} - \left( \tilde{B}_k K^k^\top \right)_{\Omega} \right\|_{\mathcal{F}} + \lambda \| K^k \|_{\mathcal{F}}, \tag{7.28}
\]

where \( \tilde{B}_k \in \mathbb{R}^{m_1 \cdots m_{d-1} \times r_{k-1}} \) is defined as

\[
\tilde{B}_k(:,\ell) \triangleq \text{vec}(\tilde{B}_1^{k-1}(:,\ell) \otimes \cdots \otimes \tilde{B}_{d-1}^{k-1}(:,\ell)) \in \mathbb{R}^{m_1 \cdots m_{d-1}}, \quad \ell = 1, \ldots, r_{k-1}. \tag{7.29}
\]

And the solution is

\[
K^k(l,:) = \left( \mathcal{M}_k - \tilde{\mathcal{M}}_k \right)_{\Omega}^\top (l,:) \text{Diag} \left[ E_{\Omega}(l,:) \right] \tilde{B}_k \left( \tilde{B}_k^\top \text{Diag} \left[ E_{\Omega}(l,:) \right] \tilde{B}_k + \lambda I \right)^{-1}, \quad l = 1, \ldots, c_k, \tag{7.30}
\]
where \( E \) denotes an all-one \((m_1 \ldots m_{d-1}) \times m_d\) matrix.

**Random structured initialization:** We also consider the random initialization for the proposed structured alternating minimization where \( A^{(0)}_j \) contains i.i.d. \( \mathcal{N}(0,1) \) samples for \( j = 1, \ldots, d-1 \), and the non-zero entries of \( A^{(0)}_d \) in Lemma 46 are also i.i.d. \( \mathcal{N}(0,1) \) samples.

### 7.3.6.2 Structured Alternating Minimization

Note that the initialization satisfies the structured decomposition. Now, we need to make sure that at each iteration of the algorithm this property still holds. In particular, in the \( i \)-th iteration of the structured alternating minimization procedure, given \( A^{(i-1)}_j \in \mathbb{R}^{m_j \times r_K} \) for \( j = 1, \ldots, d-1 \), we first update \( A^{(i)}_j \in \mathbb{R}^{m_j \times r_K} \) for \( j = 1, \ldots, d-1 \) according to (7.22). Then, to update \( A^{(i)}_d \in \mathbb{R}^{m_d \times r_K} \) we only need to update the non-zero entries in the structured decomposition. That is in (7.22), for \( 1 \leq k \leq K \) and \( n_{k-1} + 1 \leq l \leq n_k \) we have

\[
A^{(i)}_d(l, 1 : r_k) = \tilde{U}_{\Omega}^\top(l, :) \text{Diag}[E_{\tilde{\Omega}}(:, l)] \tilde{A}^{(i)}_d(:, 1 : r_k) \left( (\tilde{A}^{(i)}_d(:, 1 : r_k))^\top \text{Diag}[E_{\tilde{\Omega}}(:, l)] \tilde{A}^{(i)}_d(:, 1 : r_k) + \lambda I \right)^{-1},
\]

(7.31)

where \( \tilde{A}^{(i)}_d \in \mathbb{R}^{m_1 \ldots m_{d-1} \times r_K} \) denotes the \((d - 1)\)-th unfolding of \( \tilde{A}^{(i)}_d = A^{(i)}_1 \otimes \ldots \otimes A^{(i)}_{d-1} \in \mathbb{R}^{m_1 \times \ldots \times m_{d-1} \times r_K} \) and \( E \) denotes an all-one \((m_1 \ldots m_{d-1}) \times m_d\) matrix.

Finally, we summarize the proposed structured alternating minimization algorithm for union of nested low-rank tensor subspaces completion in Algorithm 2. Note that at each iteration of this algorithm, including the initialization, the structured decomposition holds and therefore all \( K \) rank constraints hold.

### 7.4 Simulation Results

#### 7.4.1 Matrix Case

We consider an example where \( K = 4 \), \( m = 1000 \), \( n_1 = 300 \), \( n_2 = 500 \), \( n_3 = 700 \), \( n_4 = 900 \), \( r_1 = 50 \), \( r_2 = 60 \), \( r_3 = 70 \) and \( r_4 = 80 \). In order to generate a matrix that is randomly chosen from the manifold corresponding to the given rank constraints, we first generate \( X \in \mathbb{R}^{1000 \times 80} \) (\( r_4 = 80 \) basis columns) with entries being i.i.d. \( \mathcal{N}(0,1) \) samples. Then, we generate \( Y \in \mathbb{R}^{80 \times 900} \) such that
CHAPTER 7. STRUCTURED ALTERNATING MINIMIZATION FOR UNION OF NESTED LOW-RANK SUBSPACES DATA COMPLETION

Algorithm 2 Structured Alternating Minimization - Tensor Case

1: Input $U_\Omega, r_1, \ldots, r_K$ and $n_1, \ldots, n_K$.
2: Initializing $A_j^{(0)} \in \mathbb{R}^{m_j \times r_K}$ for $j = 1, \ldots, d$ using either the CP-based or random structured initialization.
3: repeat
4: for $j = 1 : d - 1$ do
5: for $l = 1 : m_j$ do
6: Compute Eq. (7.22).
7: end for
8: end for
9: for $k = 1 : K$ do
10: for $l = n_{k-1} + 1 : n_k$ do
11: Compute Eq. (7.31).
12: end for
13: end for
14: until convergence/divergence
it satisfies the structured decomposition given in Definition 1, i.e., \( Y(51 : 80, 1 : 300) = 0_{30 \times 300} \), 
\( Y(61 : 80, 301 : 500) = 0_{20 \times 200} \) and \( Y(71 : 80, 501 : 700) = 0_{10 \times 200} \) and the rest of the entries are
i.i.d. \( \mathcal{N}(0,1) \) samples. Then, the matrix \( U = XY \) satisfies all the rank constraints. We sample
the entries of \( U \) independently with probability \( 0 < p < 1 \). The regularization weight is set as
\( \lambda = 0.01 \). We define the convergence metric as
\( \epsilon_i = \frac{\|X_i Y_i\|_F - \|X_{i-1} Y_{i-1}\|_F}{\|X_i Y_i\|_F} \) and convergence is reached
if \( \epsilon_i < 10^{-3} \). On the other hand, divergence is declared if \( \|X_i\|_F > 10^6 \|X_0\|_F \) or \( \|Y_i\|_F > 10^6 \|Y_0\|_F \).

7.4.1.1 Noiseless Matrix

We say the sampled matrix \( U \) is recovered if the algorithm converges and the normalized error
satisfies \( \frac{\|\hat{U} - U\|_F}{\|U\|_F} < 0.01 \), where \( \hat{U} \) denotes the completed matrix. We consider different number of
rank constraints: for \( K = 4 \), we include all rank constraints \( r_1, r_2, r_3 \) and \( r_4 \); for \( K = 3 \), we include
rank constraints \( r_2, r_3 \) and \( r_4 \); and for \( K = 2 \), we include rank constraints \( r_3 \) and \( r_4 \). For each
case, we generate 100 random matrices from the corresponding manifold. Then for each value of
the sampling probability \( p \), we run different completion algorithms on these sampled matrices and
calculate the recovery rates.

First, to see the impact of multiple rank constraints on the convergence, in Fig. 7.5 we illustrate
the convergence behaviors of the Naive method 1 and the structured approach, for \( K = 4 \), \( p = 0.3 \)
and a particular sampled matrix. It is seen that it takes 8 and 10 iterations for the structured
approach and the Naive 1 method, respectively, to reach the convergence condition \( \epsilon_i < 10^{-3} \).

![Figure 7.5: Convergence comparison for noiseless matrices with \( K = 4 \) and \( p = 0.3 \).](image-url)
Next the recovery rate performances of different algorithms are compared in Figs. 7.6(a), 7.6(b), and 7.6(c) for $K = 2, 3, 4$, respectively. A number of observations are in order. First, for all three values of $K$, for both Naive methods 2 and 3, the recovery rate is 1 for $p \geq 0.28$ and it is 0 when $p \leq 0.24$; for both Naive method 1 and the structured approach, the recovery rate is 1 for $p \geq 0.25$; and the recovery rate is 0 for $p \leq 0.23$ for Naive 1. Hence among the three naive methods, Naive 1 has the best recovery performance even though it ignores all additional rank constraints. Second, the structured approach mainly improves the region where the recovery rate is below 1. In particular, the recovery rate is 0 for $p \leq 0.22$ when $K = 2, 3$, whereas it becomes $p \leq 0.21$ when $K = 4$. Moreover, in the region where the recovery rate is below 1, i.e., $p \in (0.22, 0.25)$, its recovery rate is higher than that of Naive 1. Thirdly, for all three naive methods, random initialization leads to better performance than the SVD-based initialization; whereas for the structured approach, SVD-based initialization performs better.

Finally, we show the average running time comparisons among different algorithms in Fig. 7.7 for $K = 4$ and $p = 0.3$. It is seen that the Naive method 3 is much slower than the other methods, since the matrix it processes has more and more samples over the later stages. Moreover, the Naive method 2 is the fastest due to the smaller sizes of the matrices it processes. The structured approach takes only slightly longer than the Naive method 1.

### 7.4.1.2 Noisy Matrix

We now consider the case that the matrix to be completed is noisy, i.e., $Z = U + N = XY + N$, where $X$ and $Y$ are generated the same way as described in Sec. IV.A; and the entries of $N$ are i.i.d. $\mathcal{N}(0, \sigma^2)$ samples. We define the signal-to-noise-ratio as $\text{SNR} = 10 \log_{10} \left( \frac{1}{mnK} \sum_{i=1}^{m} \sum_{j=1}^{nK} U(i,j)^2 }{ \sigma^2 } \right)$. Moreover, we define the signal-to-error-ratio for the recovered matrix $\hat{U}$ as $\text{SER} = 10 \log_{10} \left( \frac{1}{mnK} \sum_{i=1}^{m} \sum_{j=1}^{nK} (\hat{U}(i,j) - U(i,j))^2 }{ \sigma^2 } \right)$. Each result of (SNR, SER) is the average of 100 realizations of $Z$.

First, for $K = 4$, $p = 0.3$, $\text{SNR} = 10\text{dB}$ and a particular sampled matrix, we show the convergence behaviors of the structured approach and the Naive method 1 in Fig. 7.8. By comparing Fig. 7.8 and Fig. 7.5, we observe that for all methods, it takes more iterations to converge in the noisy case, but still the structured approach converges faster than the Naive method 1. Moreover, for the structured approach, the SVD initialization leads to faster convergence, whereas for the Naive
Figure 7.6: Recovery rate performances for noiseless matrices with $K = 2, 3, 4$. 
CHAPTER 7. STRUCTURED ALTERNATING MINIMIZATION FOR UNION OF NESTED LOW-RANK SUBSPACES DATA COMPLETION

Figure 7.7: Running time comparisons for noiseless matrices with $K = 4$ and $p = 0.3$.

The Naive 1 method, random initialization converges faster.

Figure 7.8: Convergence comparison for noisy matrices with $K = 4$, $p = 0.15$, and SNR = 10dB.

Next, the SER performance results are shown in Figs. 7.9(a) and 7.9(b), for $p = 0.15$ and $p = 0.25$, respectively. It is seen that among the naive methods, the Naive method 1 still performs the best in the noisy case. But now there is a significant gain in SER by the proposed structured approach over the naive methods. For example, at SNR = 12dB, for $p = 0.15$ and $p = 0.25$, the SER gains over the Naive 1 method is 3.8dB and 1.9dB, respectively. Moreover, similar to the noiseless case, the SVD-based initialization performs better for the structured approach whereas random initialization performs better for the naive methods.
7.4.2 Tensor Case

For the tensor case, we consider an example where $d = 4$, $K = 4$, $m_1 = m_2 = m_3 = 40$, $n_1 = 25$, $n_2 = 30$, $n_3 = 35$, $n_4 = 40$, $r_1 = 50$, $r_2 = 60$, $r_3 = 70$ and $r_4 = 80$. In order to generate a tensor that is randomly chosen from the manifold corresponding to the given rank constraints, we first generate $a^l_j \in \mathbb{R}^{40}$ (structured columns) with entries being i.i.d. $\mathcal{N}(0, 1)$ samples for $1 \leq j \leq (d - 1)$ and $1 \leq l \leq r_K$. Then, we generate $a^l_d \in \mathbb{R}^{40}$ such that it satisfies the structured decomposition given in Lemma 4, i.e., for any $k = 1, \ldots, K$, $x = n_{k-1} + 1, \ldots, n_k$ and $l = r_k + 1, \ldots, r_K$ we have $a^l_d(x) = 0$, and the rest of the entries are i.i.d. $\mathcal{N}(0, 1)$ samples. Therefore, the tensor $U = \sum_{l=1}^{r_1} a^l_1 \otimes a^l_2 \otimes \cdots \otimes a^l_d$ satisfies all the rank constraints. Then, we sample the entries of $U$ independently with probability $0 < p < 1$. The regularization weight is set as $\lambda = 0.01$. We define the convergence metric as $\epsilon_i = \frac{\|U_i\|_F - \|U_{i-1}\|_F}{\|U_i\|_F}$ (where $U_i = \sum_{l=1}^{r_K} A^{(i)}_1(:, l) \otimes \cdots \otimes A^{(i)}_d(:, l)$) and convergence is reached if $\epsilon_i < 10^{-3}$. On the other hand, divergence is declared if $\|A^{(i)}_j\|_F > 10^6\|A^{(0)}_j\|_F$, for any $j \in \{1, \ldots, d\}$.
7.4.2.1 Noiseless Tensor

We say the sampled tensor $\mathcal{U}$ is recovered if the algorithm converges and the normalized error satisfies $\frac{\|\hat{\mathcal{U}} - \mathcal{U}\|_F}{\|\mathcal{U}\|_F} < 0.01$, where $\hat{\mathcal{U}}$ denotes the completed tensor. Similar to the matrix case, we consider different number of rank constraints: $K = 2, 3$ and $4$. For each case and a given sampling probability $p$, we run different completion algorithms on 100 random tensors from the corresponding manifold and calculate the recovery rates.

In Fig. 7.10 we illustrate the convergence behaviors of the Naive method 1 and the structured approach, for $K = 4$, $p = 0.2$ and a particular sampled tensor. It is seen that the convergence condition $\epsilon_i < 10^{-3}$ is reached after 17 and 21 iterations for the structured approach and the Naive 1 method, respectively.

![Figure 7.10: Convergence comparison for noiseless tensors with $K = 4$ and $p = 0.2$.](image)

Figs. 7.11(a), 7.11(b), and 7.11(c) show the recovery rate performances of different algorithms for $K = 2, 3$, and 4, respectively. Similarly as in the matrix case, Naive 1 has the best recovery performance among the three naive methods. Compared with Naive 1, the structured approach mainly improves the region where the recovery rate is below 1, i.e., $p \in (0.12, 0.15)$. And, for all three naive methods, random initialization perform better than the CP-based initialization; whereas for the structured approach, CP-based initialization is better.

Fig. 7.12 shows the average running time comparisons among different algorithms for $K = 4$ and $p = 0.2$. Similar to the matrix case, the Naive method 3 is the slowest and the Naive method 2 is the fastest. The structured approach is slightly slower than the Naive method 1.
Figure 7.11: Recovery rate performances for noiseless tensors with $K = 2, 3, 4$. 

(a) $K = 2$.  
(b) $K = 3$.  
(c) $K = 4$. 

CHAPTER 7. STRUCTURED ALTERNATING MINIMIZATION FOR UNION OF NESTED LOW-RANK SUBSPACES DATA COMPLETION
CHAPTER 7. STRUCTURED ALTERNATING MINIMIZATION FOR UNION OF NESTED LOW-RANK SUBSPACES DATA COMPLETION

7.4.2.2 Noisy Tensor

We now consider the noisy case, i.e., \( Z = U + N \), where \( U \) is generated the same way as described in Sec. IV.B; and the entries of \( N \) are i.i.d. \( \mathcal{N}(0, \sigma^2) \) samples. We define the signal-to-noise-ratio as

\[
\text{SNR} = 10 \log_{10} \left( \frac{1}{m_1 \cdots m_d} \sum_{x_1=1}^{m_1} \cdots \sum_{x_d=1}^{m_d} U(x_1, \ldots, x_d)^2 \right) \sigma^2.
\]

Moreover, we define the signal-to-error-ratio for the recovered tensor \( \hat{U} \) as

\[
\text{SER} = 10 \log_{10} \left( \frac{1}{m_1 \cdots m_d} \sum_{x_1=1}^{m_1} \cdots \sum_{x_d=1}^{m_d} (\hat{U}(x_1, \ldots, x_d) - U(x_1, \ldots, x_d))^2 \right). \]

Each result of (SNR, SER) is the average of 100 realizations of \( Z \).

Fig. 7.13 shows the convergence behaviors of the structured approach and the Naive method 1 for \( K = 4, p = 0.2 \), SNR = 10dB and a particular sampled tensor. Similar to the matrix case, it takes more iterations to converge in the noisy case for all methods, and the structured approach converges faster. Moreover, for the structured approach, the CP initialization leads to faster convergence, whereas for the Naive method 1, random initialization converges faster.

Figs. 7.14(a) and 7.14(b) show the SER performances for \( p = 0.05 \) and \( p = 0.15 \), respectively. It is seen that there is a significant gain in SER by the proposed structured approach over the naive methods. For example, at SNR = 12dB, for \( p = 0.05 \) and \( p = 0.15 \), the SER gains over the Naive (which performs the best among naive methods) method 1 is 2.9dB and 2dB, respectively. Moreover, similar to the noiseless case, the CP-based initialization performs better for the structured approach whereas random initialization performs better for the naive methods.
CHAPTER 7. STRUCTURED ALTERNATING MINIMIZATION FOR UNION OF NESTED LOW-RANK SUBSPACES DATA COMPLETION

Figure 7.13: Convergence comparison for noisy tensors with $K = 4$, $p = 0.15$, and SNR = 10dB.

Figure 7.14: SER performances for noisy tensors with $K = 4$. 

(a) $p = 0.05$.  

(b) $p = 0.15$.  

7.5 Summary

In this chapter, we have developed a structured alternating minimization approach to data completion where the data has a union of nested subspaces structure with multiple known rank constraints. Both matrix and tensor cases are studied. Our key observation is that the union of nested subspaces structure leads to a structured decomposition where some factors (\(Y\) for matrix case and \(A_d\) for tensor case) contain blocks of zeros determined by the rank values. The proposed structured alternating minimization algorithms for both matrix and tensor completion enforce such structures in each iteration including the initialization. Simulation results show that compared with naive methods, the proposed structured approaches achieve faster convergence and higher recovery accuracy, especially for noisy data completion.
Chapter 8

Conclusions

In this dissertation, we have made progress on addressing the following questions: (1) What is the condition on sampling pattern or the fundamental limit on the sampling rate to ensure that the data can be retrieved uniquely? (2) What is the condition on sampling pattern or the fundamental limit on the sampling rate to estimate the rank of a partially sampled data? (3) What is the fundamental limit on the sampling rate to correctly cluster a union of matrices or tensors? (4) What is an efficient way to approach such fundamental limits on the sampling rate to retrieve the sampled data? For each of the mentioned problems, we have developed a novel method and we have studied different data structures such as low-rank matrices, low-rank multi-view matrices, and low-rank tensors. We believe that still the achievability of the proposed limits with efficient algorithms is a challenging and important problem that can be studied further and this research will give insights into cracking these general problems.
Bibliography


