Near-optimal sample complexity for noisy or 1-bit tensor completion

by

Navid Ghadermarzy

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Mathematics)

The University of British Columbia
(Vancouver)

August 2018

© Navid Ghadermarzy, 2018
The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis/dissertation entitled:

**NEAR-OPTIMAL SAMPLE COMPLEXITY FOR NOISY OR 1-BIT TENSOR COMPLETION**

submitted by **NAVID GHADERMARZY** in partial fulfilment of the requirements of the degree of Doctor of Philosophy in **Mathematics**

**Examining Committee:**

Ozgur Yilmaz, Faculty of Science  
*Supervisor*

Yaniv Plan, Faculty of Science  
*Supervisor*

Felix Herrmann, Faculty of Science  
*Supervisory Committee Member*

Chen Greif, Faculty of Science  
*University Examiner*

Brian Wetton, Faculty of Science  
*University Examiner*

Richard Lockhart, Simon Fraser University  
*External Examiner*
Abstract

Tensor completion is the problem of recovering a low-rank tensor from a partial subset of its entries. Assuming a rank-$r$, order-$d$ tensor in $\mathbb{R}^{N \times N \times \cdots}$, the best sampling complexity achieved is $O(rN^d)$ which can be obtained by a tensor nuclear-norm minimization problem. This bound is significantly larger than $O(rdN)$, the number of free variables in a rank-$r$ tensor. In this thesis we prove that when $r = O(1)$, we can achieve optimal sample complexity by constraining either one of two proxies for tensor rank, the convex M-norm or the non-convex max-qnorm. The max-qnorm is the generalization of matrix max-norm to tensors which is non-convex. The M-norm on the other hand is a convex atomic norm whose atoms are rank-1 sign tensors. We prove that both max-qnorm and M-norm of a bounded rank-$r$ tensor are bounded by quantities that are independent of $N$. We also prove that the unit balls of both these norms have small Rademacher complexity.

We analyze max-qnorm and M-norm constrained least squares minimization problems for tensor completion, proving that when $r = O(1)$, $m = O(Nd)$ measurements are sufficient for efficient estimation of the original tensor. We also use an information theoretic technique to prove that the dependence on $N$ is optimal. Moreover, we design an alternating method for approximating the solution of max-qnorm tensor completion and do a thorough investigation of its performance on synthetic and real-world data.

We also generalize the 1-bit matrix completion problem to higher-order tensors. We prove that when $r = O(1)$ a bounded rank-$r$, order-$d$ tensor $T$ in $\mathbb{R}^N \times \mathbb{R}^N \times \cdots \times \mathbb{R}^N$ can be estimated efficiently by only $m = O(Nd)$ binary measurements by regularizing either its max-qnorm or its M-norm. We prove that the sample complexity of recovering a low-rank tensor from 1-bit measurements of a subset of its entries is the same as recovering it from unquantized measurements. Moreover, we show the advantage of using 1-bit tensor completion over matricization both theoretically and numerically. Specifically, we show how the 1-bit measurement model can be used for context-aware recommender systems.
Lay summary

Many real-world data sets can be arranged as multi-dimensional arrays or tensors. Tensor completion is an important problem which is applicable whenever the data has missing or corrupted entries. Missing or corrupted entries can be the result of a faulty sensor or when taking measurements is too expensive. Another important application of tensor completion is for recommendation systems such as the movie recommendation used in Netflix or Hulu. Due to the rich understanding of matrices and powerful tools available for matrices, a common approach for completing missing data (tensor completion) is to first rearrange them as a 2-dimensional matrix and then predict the missing entries (matricizing). In this thesis, we prove the advantage of tensor completion over matricizing when we have access to noisy measurements of a subset of the entries of a tensor and also when we have access to partial 1-bit measurements (0 or 1 measurements) of the tensor.
Preface

This thesis consists of my original research, conducted at the Department of Mathematics at the University of British Columbia, Vancouver, Canada, under the supervision of Dr. Ozgur Yilmaz and Dr. Yaniv Plan. The following chapters contain previously published or submitted work for which I was the principal investigator and author.

Chapter 3 and some parts of Chapter 2 are published in [45] which is joint work with Ozgur Yilmaz and Yaniv Plan and has been posted on arXiv and is submitted for publication.

Chapter 4 and some parts of Chapters 2 and 5 are published in [46] which is joint work with Ozgur Yilmaz and Yaniv Plan and has been posted on arXiv and is submitted for publication.
# Table of Contents

Abstract ......................................................... iii
Lay summary ...................................................... iv
Preface .......................................................... v
Table of Contents ............................................... vi
List of Tables ................................................... ix
List of Figures .................................................... x
List of Algorithms .............................................. xiii
Acknowledgments ................................................. xiv

1 Introduction ................................................... 1
   1.1 Matrix completion ..................................... 2
   1.2 1-bit matrix completion ............................... 3
   1.3 Rademacher complexity ............................... 4
   1.4 Notations ............................................... 5
   1.5 Fibers, slices and matricization ..................... 5
   1.6 Rank of a tensor ..................................... 6
   1.7 Outline of the results and organization ............ 8
      1.7.1 Simplified results ............................. 10
      1.7.2 Organization ................................. 12

2 M-norm and max-qnorm of tensors ......................... 15
   2.1 Notions of Orthogonality ............................ 15
4.4 Nuclear norm constrained 1-bit tensor completion ........................................ 70
4.5 1-bit tensor completion under exact rank constraint ................................. 72
4.6 Information-theoretic lower bound ............................................................ 74
4.7 Numerical results ......................................................................................... 75
    4.7.1 Max-qnorm constrained 1-bit tensor completion .................................. 75
    4.7.2 1-bit tensor completion with exact rank constraints ......................... 78
    4.7.3 Synthetic experiments ....................................................................... 79
4.8 Proofs .......................................................................................................... 84
    4.8.1 Proof of Theorem 39 .......................................................................... 84
    4.8.2 Proof of Theorem 47 .......................................................................... 85
    4.8.3 Proof of Theorem 51 .......................................................................... 89
    4.8.4 Proof of Theorem 54 .......................................................................... 94
    4.8.5 Proof of Lemma 65 ............................................................................ 95
5 Applications .................................................................................................. 98
    5.1 Applications of tensor completion ......................................................... 98
    5.1.1 Color image inpainting ....................................................................... 98
    5.1.2 Video completion ................................................................................. 102
    5.1.3 Hyperspectral imaging ....................................................................... 104
    5.1.4 MRI data recovery ............................................................................. 105
    5.2 Applications of 1-bit tensor completion ................................................. 107
    5.2.1 Score prediction ................................................................................ 108
    5.2.2 In-car music recommender system ................................................... 109
    5.2.3 A Restaurant Context-Aware Recommender System .................... 111
6 Conclusion ..................................................................................................... 114
Bibliography .................................................................................................... 117
List of Tables

Table 5.1  Results of a comparison between 1-bit and multi-bit matrix and tensor completion algorithms on incar-music data [6] for predicting whether the unobserved ratings were above or below average. The multi-bit methods use the original ratings from 0 to 5 and the context-aware methods include the context information such as time of the day, weather, location of driving and mood.  

Table 5.2  Results of a comparison between 1-bit and multi-bit matrix tensor completion algorithms on Tijuana restaurant data [102] for predicting whether the unobserved ratings were above or below average. TF refers to using tensor factorization without the max-qnorm regularization. For the 1-bit results we use $f(x) = \Phi(\frac{x}{\sigma})$.  

Table 5.3  Results of a comparison between 1-bit and multi-bit matrix tensor completion algorithms on Tijuana restaurant data [102] showing the mean absolute error.
List of Figures

Figure 1.1 The figure in left shows a third order tensor and its slices and the figure in the right shows its mode-1 matricization. .................................................. 6

Figure 3.1 Log-log plot of average max-qnorm of 3 and 4-dimensional low-rank tensors obtained by Algorithm [1] for various rank and sizes, averaged over 15 draws for each rank and size. ................................................................. 46

Figure 3.2 All the possible situations in the five-point search algorithm. The leftmost and the rightmost red dots are the previous lower bound and upper bound on the max-qnorm $R$, respectively. The green intervals show the new lower bound and upper bound based on the RMSE. ................................................................. 48

Figure 3.3 \textbf{3-dimensions, no noise} Average relative recovery error $\frac{\|T_{\text{recovered}} - T\|_F^2}{\|T\|_F^2}$ for 3-dimensional tensor $T \in \mathbb{R}^{50 \times 50 \times 50}$ and different ranks and samples. ...................... 50

Figure 3.4 \textbf{3-dimensions, 10-dB noise} Average relative recovery error $\frac{\|T_{\text{recovered}} - T\|_F^2}{\|T\|_F^2}$ for 3-dimensional tensor $T \in \mathbb{R}^{50 \times 50 \times 50}$ and different ranks and samples. ...................... 51

Figure 3.5 \textbf{4-dimensions, no noise} Average relative recovery error $\frac{\|T_{\text{recovered}} - T\|_F^2}{\|T\|_F^2}$ for 4-dimensional tensor $T \in \mathbb{R}^{20 \times 20 \times 20 \times 20}$ and different ranks and samples. The plot on the left shows the results for the 400 $\times$ 400 matricized case. ...................... 53

Figure 3.6 \textbf{4-dimensions, 10-dB noise} Average relative recovery error $\frac{\|T_{\text{recovered}} - T\|_F^2}{\|T\|_F^2}$ for 4-dimensional tensor $T \in \mathbb{R}^{20 \times 20 \times 20 \times 20}$ and different ranks and samples. The plot on the left shows the results for the 400 $\times$ 400 matricized case. ...................... 53

Figure 3.7 \textbf{3-dimensions, no noise} Average relative recovery error $\frac{\|T_{\text{recovered}} - T\|_F^2}{\|T\|_F^2}$ for e-dimensional tensors with fixed number of measurements in the left and fixed rank in the right. ...................... 54
Figure 4.1 Average relative squared error, \( \frac{\| \mathbf{T}^{\text{recovered}} - \mathbf{T}^\# \|_F^2}{\| \mathbf{T}^\# \|_F^2} \), obtained from matricized 1bit-TC, max-qnorm, factor-frob norm, and CGD over a range of different value of \( \sigma \). From left to right the original tensor is 2, 3 and 4-dimensional. Results show that the noise variance should not be too small or too big.

Figure 4.2 Average relative squared errors, \( \frac{\| \mathbf{T}^{\text{recovered}} - \mathbf{T}^\# \|_F^2}{\| \mathbf{T}^\# \|_F^2} \), for recovering a 200 \times 200 matrix with various ranks, using \( m \) 1-bit measurements sample for a range of different value of \( m \).

Figure 4.3 Average relative squared errors, \( \frac{\| \mathbf{T}^{\text{recovered}} - \mathbf{T}^\# \|_F^2}{\| \mathbf{T}^\# \|_F^2} \), for recovering a 30 \times 30 \times 30 tensor with various ranks, using \( m \) 1-bit measurements for a range of different value of \( m \). Top left is the result of matricization, top right is CGD for exact-rank constraints, bottom left is the results using the factor-frob function as an estimate for nuclear-norm constraint and bottom right is the result of max-qnorm.

Figure 4.4 Average relative squared errors, \( \frac{\| \mathbf{T}^{\text{recovered}} - \mathbf{T}^\# \|_F^2}{\| \mathbf{T}^\# \|_F^2} \), for recovering a 15 \times 15 \times 15 \times 15 tensor with various ranks, using \( m \) 1-bit measurements for a range of different value of \( m \). Left is the result of matricization, center is the result of using the factor-frob function as an estimate for nuclear-norm constraint and right is the result of max-qnorm.

Figure 5.1 The Baboon image and the RSE of best rank-\( r \) approximations of the 128 \times 128 \times 3 tensor and its 128 \times 384 matricization.

(a) Original Baboon image

(b) RSE of best rank-\( r \) approximations

Figure 5.2 Reconstructed Baboon images with 50\% observed entries. The first row shows the results with missing pixels and the second row shows the results with missing random entries. The results of matricization are worse than using tensors and the results of missing pixels are worse than missing random entries of the tensor.

Figure 5.3 Results for the Peppers image with only 20\% sampling rate.

Figure 5.4 PSNR results for a larger natural image in \( \mathbb{R}^{321 \times 481 \times 3} \). The second row is the results of LRTC [83], TCTF [134] and t-SVD [132].

Figure 5.5 Completing a black and white video from 20\% and 5\% of the entries.

Figure 5.6 Frames 50 and 100 of the tomato video [83]. The subsampling mask includes blocks of pixels as well.
Figure 5.7  An example of a hyperspectral image with 90% missing entries. In the second row, we show the result with added white noise. ............................................ 104

Figure 5.8  The recovery results for four slices of a brain MR image. The columns are images of different slices and the rows show the original image and different sampling rates. First row shows the original images. Each image in the figure shows the sampled image (top) and the recovered image (bottom). .......................................... 106

Figure 5.9  The left figure shows the average relative error of recovering a rank-5, $30 \times 30 \times 30$ tensor by max-qnorm constrained 1-bit tensor completion with different values of $\sigma$ when the 1-bit measurements were obtained by using $\sigma = 0.15$. The right figure shows the percentage of correct sign predictions. .......................... 109

Figure 5.10 Results of applying 1-bit matrix and tensor completion to partially observed verbal scores tensor [87] to determine whether unobserved scores are above or below average. The left figure shows the percentage of correct predictions and the right figure shows the absolute mean error. The scores are in the range $[-40, 40]$ ................................................................. 110
List of Algorithms

1. Estimating max-qnorm of a tensor $T$ ........................................... 45
2. Tensor completion, with cross-validation ..................................... 47
3. 1-bit TC with rank-constraints via a Conditional Gradient Descent (CGD) .... 79
Acknowledgments

First and foremost, I would like to thank my supervisor Dr. Ozgur Yilmaz for his guidance during my PhD. I have been very fortunate to have your valuable mentorship in my mathematical and academic life. I will be forever grateful for teaching me compressed sensing. I appreciate all the time you spent explaining concepts and sharing your insightful suggestions. I am also grateful for your careful evaluation of all my writings and I am sorry about all the grammatical and spelling errors during the years, especially the hundreds of commas you added and removed from my writings. I couldn’t have asked for a better supervisor and thank you for all the support and feedback thorough the years.

This thesis wouldn’t have been possible without Dr. Yaniv Plan. I started working on tensor completion when we had a discussion about it in our first meeting and since then I have learned new concepts or encountered new questions every time we have had a meeting which is more than anyone can ask from a supervisor. Your knowledge and enthusiasm for math in general has always motivated me. Thank you for you guidance during my PhD.

I should also thank Dr. Felix Herrmann and all the members of the SLIM group who I have bugged during my PhD. It was a huge opportunity to learn from a world-known expert in seismic imaging. I also want to thank the mathematics community at UBC for teaching me and helping me with my academic life.

Finally I want to thank all my friends who are always supportive and kind to me. Especially I want to thank Rebeca for her love and support. Thank you for understanding, supporting, and believing in me. Last but not least I would want to thank my family who I owe where I am right now to them.
Chapter 1

Introduction

Representing data as multi-dimensional arrays, i.e., tensors, arises naturally in many modern applications such as collaborative filtering [65], image inpainting [51], interpolating large-scale seismic data [29, 71], medical images [88], computer vision [58], data mining [1], image compression [83, 109], hyperspectral image analysis [79], and radar signal processing [94].

There are many reasons where one may want to work with a subset of the tensor entries; (i) these data sets are usually large, and we wish to store only a small number of the entries (compression). For example, color videos can be arranged in 4-dimensional tensors that can be compressed significantly; (ii) in some applications, the acquisition of each entry can be expensive, e.g., each entry may be obtained by solving a large PDE [123]. One example of such applications is seismic imaging where data are usually 3 to 6 dimensions; (iii) some of the entries might get lost due to physical constraints while gathering them. For example, parts of Hyperspectral images taken by satellites can be blocked by clouds. These restrictions result in situations where one has access only to a subset of the tensor entries. The problem of tensor completion entails recovering a tensor from a subset of its entries.

The general framework of the tensor completion problems can be formulated as the following. Given a $d$-dimensional tensor $T^d \in \mathbb{R}^{N_1 \times \cdots \times N_d}$ we have access to measurements taken from a subset of its entries. In particular, assuming $\Omega \subset [N_1] \times [N_2] \times \cdots \times [N_d]$ to be a subset of the indices of the tensor and $f : \mathbb{R} \to \mathbb{R}$ to be a real function, the most basic form of a tensor completion problem is

$$\text{find } X \in \mathbb{R}^{N_1 \times \cdots \times N_d} \text{ s.t. } f(X(\omega)) = f(T^d(\omega)) \forall \omega \in \Omega.$$ 

It is easy to observe that if the number of observations, $m := |\Omega|$, is smaller than the dimension of
the tensor, $\Pi_{i=1}^{d} N_i$, the above problem does not have a unique result. Hence, the tensor completion problem focuses on two main goals; (i) Given the measurement function $f$, what assumptions on $T^d$ and $\Omega$ can guarantee efficient recovery of $T^d$ from $\{f(T^d(\omega))\}_{\omega \in \Omega}$; and (ii) designing stable and tractable methods that recover the tensor from measurements taken from a subset of its entries.

1.1 Matrix completion

Recovering an object with a number of measurements that is smaller than its ambient dimension is generally impossible without some form of prior information about the object. One example of using such prior information is in Compressed Sensing (CS) which has revolutionized the field of signal acquisition and reconstruction [20, 36]. The main idea of CS is based on using inherent properties of signals, such as sparsity in a specific domain, to decrease the number of necessary measurements for recovery.

In recent years, the framework of recovering an object using measurements far fewer than its ambient dimension by exploiting its inherent properties has been extended in a few directions. One of these extensions is the problem of matrix completion which was popularized by the famous Netflix prize contest. Given a rank-$r$ matrix $A^\sharp \in \mathbb{R}^{N \times N}$ and a random subset of its entries $\Omega = \{\omega_1, \omega_2, \cdots, \omega_m\}$, $\omega_i \in [N] \times [N]$, we observe $m$ noisy samples $\{Y_{\omega_i}\}_{i=1}^{m}$ of the matrix $A^\sharp$, $Y_{\omega_i} = A^\sharp_{\omega_i} + \sigma \xi_i$, where $\xi_i$ is a zero-mean noise. The purpose of matrix completion is to recover $A^\sharp$ from the measurements $Y$. This problem has been extensively studied in the literature [22, 38, 66, 118].

The straightforward solution is finding the matrix with the lowest rank that is consistent with the measurements. In the noiseless case, this translates to

$$\text{arg min} \text{rank}(M) \text{ s.t. } M_\Omega = A^\sharp_\Omega,$$

where $M_\Omega(\omega) = A^\sharp(\omega)$ for $\omega \in \Omega$ and zero otherwise. Rank minimization is NP-hard; therefore, extensive research has been done to find tractable alternatives. The most common approach is using nuclear norm. In [23], the authors showed that solving the nuclear-norm minimization problem

$$\text{arg min} \|M\|_* \text{ s.t. } M_\Omega = A^\sharp_\Omega,$$

would recover a rank-$r$, $N \times N$ matrix from only $O(rN \text{polylog}(N))$ samples under mild incoherence conditions on the matrix. Extensive research has been done in analyzing variants of nuclear-norm
minimization and designing efficient algorithms to solve it as in [17, 21, 23, 126].

The nuclear-norm of a matrix is the sum of its singular values and is the best convex approximation for the rank function. An alternative interpretation of the rank and the nuclear-norm of a matrix is based on the minimum number of columns of its factorizations. In particular, the rank of a matrix $M$ is the minimum number of columns of the factors $U, V$ where $M = UV'$; the nuclear norm is the minimum product of the Frobenius norms of the factors, i.e., $\|M\|_* = \min \|U\|_F \|V\|_F$ subject to $M = UV'$ [118]. An alternative proxy for the rank of a matrix is its max-norm defined as $\|M\|_{\max} = \min \|U\|_{2,\infty} \|V\|_{2,\infty}$ subject to $M = UV'$ [118]. The max-norm bounds the norm of the rows of the factors $U$ and $V$, and was used for matrix completion in [40]. There, the authors studied both max-norm and trace-norm matrix completion by analyzing the Rademacher complexity of the unit balls of these norms. They proved that under uniformly random sampling, either with or without replacement, $m = O(\frac{rN}{\varepsilon} \log^{3}(\frac{1}{\varepsilon}))$ samples are sufficient for achieving mean squared recovery error $\varepsilon$ using max-norm constrained estimation and $m = O(\frac{rN \log(N)}{\varepsilon} \log^{3}(\frac{1}{\varepsilon}))$ samples are sufficient for achieving mean squared recovery error $\varepsilon$ using nuclear-norm constrained estimation.

### 1.2 1-bit matrix completion

In some applications, the observations are highly quantized. For example, in some recommendation systems, the available data are the result of yes/no questions asked from the users, e.g., whether they think an advertisement is relevant to them or not. Here the problem is recovering a low-rank tensor from 1-bit measurements of a subset of its entries. To be precise in tensor completion we have access to (possibly noisy) samples of the tensor, which we also refer to as multi-bit tensor completion and in 1-bit tensor completion we have access to 1-bit information of a subset of the entries of the tensor.

1-bit compressed sensing [16, 59, 99] investigates the possibility of recovering a sparse signal from its 1-bit-quantized measurements and it is proved that robust recovery is possible with $m = O(s \log(n)^2)$ measurements which is surprisingly the same sample complexity one can achieve with unquantized measurements. The generalization of 1-bit CS to matrix recovery has been studied extensively in the literature. Here, the 1-bit samples are obtained by dithering a subset of the entries of the matrix and taking their sign. Given a rank-$r$ matrix $M^\sharp$ and a random subset of its
entries $\Omega$, observations can be modeled as
\[
Y_\omega = \begin{cases}
  +1 & \text{with probability } f(M^\sharp_\omega), \\
  -1 & \text{with probability } 1 - f(M^\sharp_\omega)
\end{cases}
\quad \text{for } \omega \in \Omega.
\] (1.1)

Here $f : \mathbb{R} \to [0, 1]$ is a (previously known) monotone differentiable function which can be interpreted as the cumulative distribution of the dithering noise function. Given these measurements the negative log-likelihood of observing $Y$ from a matrix $M$ is
\[
\mathcal{L}_{\Omega, Y}(M) = \sum_{\omega \in \Omega} \left( \mathbb{1}_{[Y_\omega = 1]} \log \left( \frac{1}{f(M_\omega)} \right) + \mathbb{1}_{[Y_\omega = -1]} \log \left( \frac{1}{1 - f(M_\omega)} \right) \right).
\]

Considering a constraint set $D$ that is preferably low-dimensional such that $M^\sharp \in D$, we can find an estimate for $M^\sharp$ by
\[
\arg\min \mathcal{L}_{\Omega, Y}(M) \text{ s.t. } M \in D.
\]

Using exact-rank constrains has been analyzed in [11, 93] and convex constraints such as nuclear norm [31] and max-norm [18] have been analyzed as well and prove that with high probability, $m = O(rN \log(N))$ measurements are sufficient for efficient recovery of a rank-$r$ matrix in $\mathbb{R}^{N \times N}$.

## 1.3 Rademacher complexity

An important technical tool used in this thesis is treating tensors as functions from indices to values. Using this, we can analyze the complexity of sets of tensors as sets of functions. To be precise, we regard a tensor $T$ as a function $f : [N_1] \times \cdots [N_d] \to \mathbb{R}$ defined as $f(\omega_1, \cdots, \omega_d) := T(\omega_1, \cdots, \omega_d)$.

For a function class $\mathcal{F}$, the empirical Rademacher complexity of $\mathcal{F}$ with respect to a sample set $\Omega$ is defined as
\[
\hat{R}_\Omega(\mathcal{F}) = \frac{2}{|\Omega|} \mathbb{E}_{\epsilon} \left[ sup_{f \in \mathcal{F}} \sum_{\omega \in \Omega} \epsilon_i f(\omega) \right],
\]
where $\epsilon_i$ is a Rademacher random variable. The Rademacher complexity expresses how well a function class correlates with random noise on average and is a powerful tool for proving generalization bounds. In chapter 2, we prove that the set of bounded low-rank tensors has low Rademacher complexity.
1.4 Notations

In this section, we introduce some notations that will be used throughout this thesis. We adopt the notation of Kolda and Bader’s review on tensor decompositions [70]. Below, \( \lambda, \sigma, \) and \( \alpha \) are used to denote scalars, and \( C \) and \( c \) are used to denote universal constants. Vectors are denoted by lower case letters, e.g., \( u, \) and \( v. \) Both matrices and tensors are represented by upper case letters, usually using \( A \) and \( M \) for matrices, and \( T \) and \( X \) for tensors. Tensors are a generalization of matrices to higher order, also called multi-dimensional arrays. For example, a first-order tensor is a vector and a second-order tensor is a matrix. \( X \in \otimes_{i=1}^{d} \mathbb{R}^{N_i} \) is a \( d \)-th order tensor whose \( i \)-th size is \( N_i. \) We also denote \( \otimes_{i=1}^{d} \mathbb{R}^{N} \) as \( \mathbb{R}^{Nd}. \) Elements of a tensor are either specified as \( X_{i_1,i_2,\cdots,i_d} \) or \( X(i_1,i_2,\cdots,i_d), \) where \( 1 \leq i_j \leq N_j \) for \( 1 \leq j \leq d. \) We also use \( X_\omega \) as a shorthand to refer the index \( \omega \) of a tensor, where \( \omega = (i_1,i_2,\cdots,i_d) \) is an \( n \)-tuple determining the index \( X(i_1,i_2,\cdots,i_d). \)

Inner products are denoted by \( \langle \cdot, \cdot \rangle. \) The symbol \( \circ \) represents both matrix and vector outer products where \( T = U_1 \circ U_2 \circ \cdots \circ U_d \) means \( T(i_1,i_2,\cdots,i_d) = \sum_k U_1(i_1,k)U_2(i_2,k) \cdots U_d(i_d,k), \) where \( k \) ranges over the columns of the factors. In the special case of vectors, \( T = u_1 \circ u_2 \circ \cdots \circ u_d \) means \( T(i_1,i_2,\cdots,i_d) = u_1(i_1)u_2(i_2) \cdots u_d(i_d). \) Finally \([N] := \{1,\cdots,N\}\) is the shorthand notation we use for the set of integers from 1 to \( N. \)

1.5 Fibers, slices and matricization

Matricization is a fundamental step in many algorithms that have been designed for tensor and 1-bit tensor completion. This method rearranges the tensor as a matrix by assigning one dimension of the tensor as the rows and all the other dimensions as the columns. Fibers are the higher-order analogue of matrix rows and columns. For a tensor \( X \in \otimes_{i=1}^{d} \mathbb{R}^{N_i}, \) mode-\( i \) fibers of the tensor are \( \Pi_{j\neq i}N_j \) vectors obtained by fixing all indices of \( X \) except for the \( i \)-th one. For example mode-1 fibers of a third order tensor \( X \) is denoted by \( X(:,j,k) \) for fixed \( j \) and \( k. \) Slices of a tensor, on the other hand, are obtained by fixing all the indices of a tensor except for two. The mode-\( i \) matricization of \( X, \) denoted by \( X(i) \in \mathbb{R}^{N_i \times \Pi_{j\neq i}N_j} \) is obtained by arranging all the mode-\( i \) fibers of \( X \) along columns of the matrix \( X(i). \) More precisely, \( X(i)(i,j) = X(i_1,i_2,\cdots,i_d), \) where

\[
    j = 1 + \sum_{k=1,k\neq i}^{d} (i_k - 1)J_k \quad \text{with} \quad J_k = \Pi_{m=1,m\neq i}^{k-1}N_m.
\]

A detailed illustration of these definitions can be found in [69, 70]. Figure [1.1] shows an example of mode-1 matricization of an order-3 tensor.
A generalization of these unfoldings was proposed by [90] that rearranges $X^{(1)}$ into a more balanced matrix: For $j \in \{1, \cdots, d\}$, $X^{[j]}$ is obtained by arranging the first $j$ dimensions along the rows and the rest along the columns. In particular, using Matlab notations $X^{[j]} = \text{reshape}(X^{(1)}, \prod_{i=1}^{j} N_i, \prod_{i=j+1}^{d} N_i)$.

\section{1.6 Rank of a tensor}

A unit tensor is a tensor $U \in \bigotimes_{j=1}^{d} \mathbb{R}^{N_j}$ that can be written as

$$U = u^{(1)} \circ u^{(2)} \circ \cdots \circ u^{(d)},$$

where $u^{(j)} \in \mathbb{R}^{N_j}$ is a unit-norm vector and $\circ$ denotes the vector outer product. The vectors $u^{(j)}$ are called the components of $U$. A rank-1 tensor is a scalar multiple of a unit tensor. Define $\mathbb{U}_d$ to be the set of unit tensors of order $d$.

The rank of a tensor $T$, denoted by $\text{rank}(T)$, is defined as the smallest number of rank-1 tensors that generate $T$ as their sum, i.e.,

$$T = \sum_{i=1}^{r} \lambda_i U_i = \sum_{i=1}^{r} \lambda_i u^{(1)}_i \circ u^{(2)}_i \circ \cdots \circ u^{(d)}_i,$$

where $U_i \in \mathbb{U}_d$ is a unit tensor. This low-rank decomposition is also known as CANDECOMP-P/PARAFAC (CP) decomposition [24, 53]. In this thesis, we concentrate on CP decompositions which generalizes the outer product machinery of matrix SVD. However, unlike matrix
SVD, there is no restriction on orthogonality of the set of vectors \( \{ u_i^{(j)} \}_{i=1}^{N_j} \). More importantly, for a rank-\( r \) tensor \( T = \sum_{i=1}^{r} \lambda_i u_i^{(1)} \odot u_i^{(2)} \odot \cdots \odot u_i^{(d)}, \) mode-\( i \) matricization is \( T_{(j)} = \sum_{i=1}^{r} \lambda_i u_i^{(j)} \odot (u_i^{(1)} \otimes \cdots \otimes u_i^{(j-1)} \otimes u_i^{(j+1)} \otimes \cdots \odot u_i^{(d)}) \) and the balanced matricization introduced in [90] is \( T_{[j]} = \sum_{i=1}^{r} \lambda_i (u_i^{(1)} \otimes u_i^{(2)} \otimes \cdots \odot u_i^{(j)}) \odot (u_i^{(j+1)} \otimes \cdots \odot u_i^{(d)}). \) Here, the symbol \( \otimes \), represents the Kronecker product. The rank of all matricizations defined above are less than or equal to the rank of the tensor.

Even though tensor-rank is a direct generalization of matrix-rank it does not have a lot of properties of matrix rank. For example, tensor-rank can be different over \( \mathbb{R} \) and \( \mathbb{C} \) [28]. Moreover, there is no straightforward algorithm to compute the rank of a tensor [70]. Furthermore, the set of rank-\( r \) tensors is not necessarily closed, e.g., there exists a sequence of rank-2 tensors that converge to a rank-3 tensor. An example of such tensors is presented in [97]. Consider a rank-3 tensor defined as

\[
X = a_1 \odot b_1 \odot b_2 + a_1 \odot b_2 \odot c_2 + a_2 \odot b_1 \odot c_1,
\]

where the matrices \( A = [a_1 a_2], B = [b_1 b_2], \) and \( C = [c_1 c_2] \) have linearly independent columns. This tensor can be approximated arbitrarily closely by a rank-two tensor of the following form:

\[
X_t = t(a_1 + \frac{1}{t} a_2) \odot (b_1 + \frac{1}{t} b_2) \odot (c_1 + \frac{1}{t} c_2) - ta_1 \odot b_1 \odot c_1.
\]

Theoretical and numerical difficulties of approximating a tensor by its CP decomposition has resulted in defining other decompositions that are used in the literature. The Tucker decomposition [122] is a form of higher-order PCA, which decomposes a tensor into a core tensor that gets multiplied by a matrix in each mode [70]. The core tensor can be thought of as a multi-dimensional tensor of singular values. For a tensor \( T \), define \( r_i \) to be the rank of its mode-\( i \) matricization, i.e., assume \( \text{rank}(T_{(i)}) = r_i \). Then

\[
T = G \times_1 A_1 \times_2 A_2 \cdots \times_d A_d
\]

is its Tucker decomposition. Here \( G \in \mathbb{R}^{r_1 \times \cdots \times r_d} \) is the core tensor, \( A_i \in \mathbb{R}^{N_i \times r_i} \) is a factor of the decomposition and \( (r_1, \ldots, r_d) \) is its multi-linear rank. The notation \( G \times_i A_i \) is the multiplication along mode-\( i \) which is the result of multiplication of \( A_i \) with the matricization of \( G \) along its \( i \)-th mode. The generalization of rank to multi-rank removes some of the difficulties associated with CP-decomposition. Most notable is that the set of tensors with multi-linear rank \( (r_1, \ldots, r_d) \) is closed. As a result, every tensor has a best multi-rank \( (r_1, \ldots, r_d) \)-approximation which can be obtained by using HOSVD [33]. Notice that the CP decomposition is a special case of the Tucker decomposition when \( G \) is diagonal. For a tensor with multi-linear rank \( (r, \ldots, r) \), the number of free
variables is exponential in $r$. This has motivated the use of hierarchical tensor decompositions such as Hierarchical Tucker decomposition\[48, 52\], its subclass tensor train decomposition [27, 96] and many more variants whose number of free variables is linear with $r$. In this thesis, we concentrate on CP-decomposition because of its simplicity and similarity to a certain nuclear decomposition of low max-qnorm tensors that we will define later.

1.7 Outline of the results and organization

In this thesis, we study the generalization of matrix completion and 1-bit matrix completion to higher-order tensors, i.e., considering an order-$d$ tensor $T$ with size $N_1 \times N_2 \times \cdots \times N_d$, we want to reconstruct $T$ by sampling a subset of its entries (or some function of the samples). Similar to its lower-dimensional counterparts, without assuming further structure on the underlying tensor, there is no hope of recovering the missing entries as they are independent of the observed entries. Therefore, here (and in many applications) tensors of interest are the ones that can be expressed approximately as a lower-dimensional object. In particular, we consider tensors that have low CP-rank [24, 53]. The low-rank assumption makes tensor completion a feasible problem. For example, an order-$d$, rank-$r$ tensor, which has size $N_1 \times N_2 \times \cdots N_d$ where $N_i = O(N)$, has $O(rN^d)$ free variables, which is much smaller than $N^d$, the ambient dimension of the tensor.

Generalizing the techniques and machinery used in matrix completion to the tensor completion problem is generally hard. Maybe the most important complication is that tensors do not have a unique generalization of SVD and most extensions manage to capture just some of the properties of the matrix SVD. For example, generalizing the outer product machinery of SVD leads to the CP-decomposition, which does not have orthogonality condition of the matrices. Therefore, despite all the powerful tools and algorithms developed for matrix completion, tensor completion problem is still fairly open and not as well understood. For instance, there is a large gap between theoretical guarantees and what is observed in numerical simulations. This is mainly due to the lack of efficient orthogonal decompositions, low-rank approximations, and limited knowledge of the structure of low-rank tensors compared to matrices. These limitations have resulted in numerous works which are based on connecting the general tensor completion problem to matrix completion by rearranging the tensor as a matrix, including the sum of nuclear norms (SNN) model that minimizes the sum of the nuclear norm of matricizations of the tensor along all its dimensions, leading to sufficient recovery with $m = O(rN^{d-1})$ samples [44, 83]. More balanced matricizations, such as the one introduced in [90], can result in a better bound of $m = O(rN^\lceil\frac{d}{2}\rceil)$ samples.
Once we move from matrices to higher-order tensors, many of the well-known facts of matrix algebra cease to be true. For example, even a best rank-$k$ approximation may not exist for some tensors, illustrated in [70, Section 3.3], showing that the space of tensors with rank at most 2 is not closed. Interestingly there is a paper titled “Most tensor problems are NP-hard” [55] which proves that many common algebraic tasks are NP-hard for tensors with $d \geq 3$, including computing the rank, spectral norm, and nuclear norm. The computational complexity of directly solving tensor completion and the inferior results of matricization make tensor completion challenging.

Having all these complications in mind, on the theoretical side, a low-rank tensor has $O(rdN)$ free variables, but an upper-bound of $O(rN^{\lceil \frac{d}{2} \rceil})$ on the sample complexity. When $d > 2$, the polynomial dependence on $N$ seems to have a lot of room for improvement. Moreover, it is well-known that empirical recovery results are much better when the tensor is not rearranged as a matrix, even though these results are attempting to solve an NP-hard problem. This has resulted in efforts towards narrowing this gap, including heuristic algorithms [9, 83]. In spite of good empirical results and reasonable justifications, a theoretical study filling this gap was not presented in these cases.

In the second part of the thesis, we do an extensive investigation of the recovery guarantees one can obtain for the 1-bit tensor completion problem. Unlike the tensor completion problem, the 1-bit tensor completion has not been well studied in the literature. Matricizing the tensor along different modes was suggested in [2]. Our analysis is based on considering three constraint sets for a low-rank tensor. The set of low-rank tensors (fixed rank) and two convex approximations of this set: the set of low nuclear-norm tensor and the set of low M-norm tensors. Our investigation is based on generalizing three different results on 1-bit matrix completion. Along the way, we discuss the differences between these methods and the complications of generalizing the 1-bit matrix completion methods to tensors which answers some of the open questions and highlights some other open problems in the field.

In this thesis, we analyze both 1-bit and multi-bit tensor completion and provide near-optimal sample complexity guarantees for both of them. Nuclear norm has been extended to tensors in [35, 43, 129]. In an effort to obtain linear dependence on $N$, we analyze 1-bit and multi-bit tensor completion using a max-qnorm (max-quasi-norm) constrained algorithm where the max-qnorm is a direct generalization of the matrix max-norm to the case of tensors. Unfortunately, max-qnorm is non-convex. However, analyzing the unit-ball of the dual of the dual of the max-qnorm (which is a convex norm) led us to define and analyze a convex atomic-norm (which we call M-norm)
constrained least squares problem, where we obtained optimal recovery bounds on the size of the tensor.

1.7.1 Simplified results

Without going into details, we briefly state the upper bounds we establish (in Section 3.3) on the recovery errors associated with M-norm and max-qnorm constrained tensor completion. Given a rank $r$, order $d$ tensor $T^z \in \otimes_{i=1}^{d} \mathbb{R}^N$, and a random subset of its entries with indices in $S = \{\omega_1, \omega_2, \ldots, \omega_m\}$, $\omega_i \in [N]^d$, we observe $m$ noisy entries $\{Y_{\omega_t}\}_{t=1}^{m}$ of $\{T^z(\omega_t)\}_{t=1}^{m}$, where each observation is perturbed by i.i.d. noise with mean zero and variance $\sigma^2$. To give a simple version of the result we assume that indices in $S$ are drawn independently at random with the same probability for each observation, i.e., we assume uniform sampling to give a simple theorem.

**Theorem 1.** Consider a rank-$r$, order-$d$ tensor $T^z \in \otimes_{i=1}^{d} \mathbb{R}^N$ with $\|T^z\|_\infty \leq \alpha$. Assume that we are given a collection of noisy observations

$$Y_{\omega_t} = T^z(\omega_t) + \sigma \xi_t, \quad t = 1, \ldots, m,$$

where the noise sequence $\xi_t$ are i.i.d. standard normal random variables and each index $\omega_t$ is chosen uniformly random over all the indices of the tensor. Then, there exists a constant $C < 20$ such that the solution of

$$\hat{T}_M = \arg\min_X \frac{1}{m} \sum_{t=1}^{m} (X_{\omega_t} - Y_{\omega_t})^2 \quad \text{subject to} \quad \|X\|_\infty \leq \alpha, \|X\|_{M} \leq (r \sqrt{r})^{d-1} \alpha, \quad (1.3)$$

satisfies

$$\frac{\|T^z - \hat{T}_M\|^2_F}{N^d} \leq C(\alpha + \sigma)\alpha (r \sqrt{r})^{d-1} \sqrt{\frac{dN}{m}},$$

with probability greater than $1 - e^{-\frac{N}{\min(N)}} - e^{-\frac{dN}{2}}$. Moreover, the solution of

$$\hat{T}_{\max} = \arg\min_X \frac{1}{m} \sum_{t=1}^{m} (X_{\omega_t} - Y_{\omega_t})^2 \quad \text{subject to} \quad \|X\|_\infty \leq \alpha, \|X\|_{\max} \leq (\sqrt{r^2 - d}) \alpha, \quad (1.4)$$

satisfies

$$\frac{\|T^z - \hat{T}_{\max}\|^2_F}{N^d} \leq C_d(\alpha + \sigma)\sqrt{r^2 - d} \sqrt{\frac{dN}{m}},$$

with probability greater than $1 - e^{-\frac{N}{\min(N)}} - e^{-\frac{dN}{2}}$. 

10
Above, $\|X\|_M$ is the M-norm of tensor $X$ which is an atomic norm whose atoms are rank-1 sign tensors defined in Section 2.3.2 (2.12) and $\|X\|_{\text{max}}$ is max-qnorm defined in (2.11).

Without going into details, the following theorem summarizes the results in Chapter 4 on recovering a rank-$r$ tensor from partial 1-bit observations.

**Theorem 2.** Consider a rank-$r$, order-$d$ tensor $T^\sharp \in \otimes_{i=1}^d \mathbb{R}^N$ with $\|T^\sharp\|_\infty \leq \alpha$ and a random subset of its indices $\Omega \subset [N]^d$ where $|\Omega| = m$. Suppose that we only observe 1-bit measurements of $T^\sharp$ on $\Omega$ according to

$$Y_{\omega} = \begin{cases} +1 & \text{with probability } f(T^\sharp_{\omega}), \\ -1 & \text{with probability } 1 - f(T^\sharp_{\omega}) \end{cases} \quad \text{for } \omega \in \Omega. \quad (1.5)$$

Here $f : \mathbb{R} \to [0, 1]$ is a previously-known well-defined monotone differentiable function. Define the negative log-likelihood function

$$L_{\Omega, Y}(X) = \sum_{\omega \in \Omega} \left( \mathbb{1}_{[Y_\omega = 1]} \log \left( \frac{1}{f(X_\omega)} \right) + \mathbb{1}_{[Y_\omega = -1]} \log \left( \frac{1}{1 - f(X_\omega)} \right) \right).$$

Then the minimizer of the M-norm constrained optimization problem

$$\hat{T}_M = \arg \min_X L_{\Omega, Y}(X) \quad \text{subject to} \quad \|T\|_\infty \leq \alpha, \|T\|_M \leq (r\sqrt{r})^{d-1} \alpha,$$

satisfies

$$\frac{\|T^\sharp - \hat{T}_M\|_F^2}{Nd} \leq C_{\beta} \{ L_{\alpha}(r\sqrt{r})^{d-1} \alpha \sqrt{\frac{dN}{m}} + U_{\alpha} \sqrt{\frac{\log(\frac{4}{\delta})}{m}} \},$$

with probability greater than $1 - \delta$. Here $L_{\alpha}$, $\beta_{\alpha}$ and $U_{\alpha}$ are constants that just depend on $\alpha$. Moreover, the minimizer of the max-qnorm constrained optimization problem

$$\hat{T}_{\text{max}} = \arg \min_X L_{\Omega, Y}(X) \quad \text{subject to} \quad \|T\|_\infty \leq \alpha, \|T\|_{\text{max}} \leq \sqrt{r^{d-1}} \alpha,$$

satisfies

$$\frac{\|T^\sharp - \hat{T}_{\text{max}}\|_F^2}{Nd} \leq C_d \beta_{\alpha} \{ L_{\alpha} r^{d-2} \alpha \sqrt{\frac{dN}{m}} + U_{\alpha} \sqrt{\frac{\log(\frac{4}{\delta})}{m}} \},$$

with probability greater than $1 - \delta$. Furthermore, the minimizer of the nuclear-norm constrained...
optimization

\[ \hat{T}_{\text{nuc}} = \arg \min_X \mathcal{L}_{\Omega, Y}(X) \quad \text{subject to} \quad \|X\|_{\infty} \leq \alpha, \|X\|_* \leq \sqrt{r^d - 1} N^d \alpha, \]

satisfies

\[ \frac{\|T^* - \hat{T}_{\text{nuc}}\|_F^2}{N^d} \leq C_{d, \alpha} \sqrt{r^d - 1} \left( \frac{\log^d(N) \sqrt{N}}{\sqrt{m}} + \frac{\log(N) \sqrt{N^d}}{m} \right), \]

with probability at least \(1 - \frac{C}{N^d}\) and the minimizer of the exact-rank constrained optimization

\[ \hat{T}_r = \arg \min_X \mathcal{L}_{\Omega, Y}(X) \quad \text{subject to} \quad \text{rank}(X) \leq r, \|X\|_{\infty} \leq \alpha, \]

satisfies

\[ \frac{\|T^* - \hat{T}_r\|_F^2}{N^d} \leq \max \left( 4 \alpha^2 N^d \frac{m}{m}, C_{\alpha, \gamma} (2r)^{-\frac{d}{2}} \frac{\sqrt{N^d}}{m} \right). \]

with probability at least \(1 - \exp(-cN^d)\).

1.7.2 Organization

Next, we briefly explain the contents of each chapter. To explain the main contributions we consider an order-\(d\) tensor \(T^* \in \mathbb{R}^{N_1 \times \cdots \times N_d}\) where \(N_i = O(N)\) for \(1 \leq i \leq d\).

Chapter 2

In this chapter, we review some of the basic definitions related to tensors. Specifically, we find a bound on the nuclear norm of bounded rank-\(r\) tensors which is tight in \(N\) by connecting the CP-decomposition with an orthogonal decomposition. Next, we define the M-norm and max-qnorm of tensors as a robust proxy for the rank of a tensor. We prove that both M-norm and max-qnorm of a bounded low-rank tensor are bounded above by quantities that just depend on the tensor’s rank and its infinity norm and are independent of \(N\). We use a generalization of Grothendieck’s theorem to connect the max-qnorm of a tensor to its nuclear decomposition with unit infinity-norm factors. Using this, we bound the Rademacher complexity of the set of bounded tensors with low max-qnorm. This also establishes a theoretical framework for further investigation of low max-qnorm tensors.
Chapter 3

In this section, we investigate the problem of tensor completion. In particular, we prove that, with high probability, \( m = O(r^{3d}dN) \) (or \( m = O(R^2N) \) if M-norm is bounded by \( R \)) samples are sufficient to estimate a rank-\( r \) bounded tensor using a convex least squares algorithm. Moreover, we derive an information-theoretic lower bound that proves \( m = O(R^2N) \) measurements are necessary for recovering tensors with M-norm less than \( R \). This proves that our bound is optimal both in its dependence on \( N \) and the M-norm bound \( R \). It is worth mentioning though, that the bound we prove in this chapter is not necessarily optimal in \( r \), the rank of the tensor.

We also propose an alternating method to approximate the solution of the max-qnorm tensor completion. In an effort to get a better understanding of the behavior of the max-qnorm, we use this algorithm to investigate the max-qnorm of low-rank tensors whose factors are random Gaussian and random binary values. We also do a thorough investigation of the performance of the algorithm for different values of rank, \( r \), the maximum length of the tensor dimensions, \( N \), and the number of measurements, \( m \). Through synthetic numerical examples, we illustrate the advantage of using algorithms designed for max-qnorm constrained tensor completion instead of matricization. These algorithms significantly outperform algorithms based on matricization and alternating least squares (ALS).

Chapter 4

In this section, we investigate the problem of 1-bit tensor completion. First, we analyze 1-bit tensor completion using M-norm constraints on the underlying tensor (this is a convex constraint). We prove that, with high probability, the mean squared error (MSE) of recovering a rank-\( r \) tensor \( T^\# \) from \( m \) 1-bit measurements by solving an M-norm constrained optimization is \( O(\sqrt{r^{3d-3}} \sqrt[3]{N^d/m}) \). Moreover, we analyze a related non-convex function, max-qnorm, and prove that MSE of optimizing a log-likelihood function constrained by max-qnorm is \( O(\sqrt{r^{d^2-d} \sqrt{Nd/m}}) \).

Next, we use nuclear norm as a convex constraint. The theoretical sample complexity we obtain by using nuclear norm is not as tight as using M-norm and we show that with the tools we use here (or in the matrix-case) there is no hope of getting tighter results. We believe that achieving a tighter upper bound is only possible if we use nuclear norm as a regularizer with carefully chosen regularization parameters (and not as a constraint). We also analyze 1-bit tensor completion using exact-rank constraints. The estimation error relies on bounding the M-norm of rank-\( r \) tensors and on upper bound on the dual-M-norm of a random tensor which might be of independent interest.
We also derive an information-theoretic lower bound that proves the MSE of any arbitrary algorithm is at least \( \Omega(r \sqrt{\frac{N}{m}}) \) for a rank-\( r \) tensor and \( \Omega(R \sqrt{\frac{N}{m}}) \) for a tensor with M-norm less than \( R \). This proves that our upper bound is optimal in \( N \) and the M-norm bound \( R \) (but not necessarily in \( r \)).

Finally, we propose a numerical method to approximate the solution of max-qnorm constrained 1-bit tensor completion and show its advantage over 1-bit matrix completion using synthetic data.

**Chapter 5**

In this chapter, we show some of the real-world applications of multi-bit and 1-bit tensor completion. We first show the application of the tensor completion algorithm in Chapter 3 to image inpainting, video completion, hyperspectral imaging and completing MRI images.

In the second part of the chapter, we show some applications of the 1-bit tensor completion algorithm. In particular, we concentrate on context-aware recommendation systems.
Chapter 2

M-norm and max-qnorm of tensors

In this chapter, we generalize some of the well-known norms defined for matrices to tensors. Some of the definitions are straightforward generalizations of matrix-norms. First, we explain some notions of orthogonality in tensors and then we define Frobenius and nuclear norm of tensors. The rest of the chapter is devoted to defining and analyzing the max-qnorm and atomic M-norm which is an important pillar of the analysis in the rest of the thesis.

2.1 Notions of Orthogonality

In this section, we define some notions of orthogonality of tensors which we later use to bound nuclear norm of a bounded low-rank tensor. Before that though, notice that for \( X, T \in \bigotimes_{j=1}^{d} \mathbb{R}^{N_j} \), the inner product of \( X \) and \( T \) is defined as:

\[
\langle X, T \rangle := \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \cdots \sum_{i_d=1}^{N_d} X(i_1, i_2, \ldots, i_d)T(i_1, i_2, \ldots, i_d).
\]  

which is a direct generalization of matrix inner product. For the sake of simplicity, in the rest of this section, we assume \( N_j = N \) for \( 1 \leq j \leq d \).

**Definition 3.** Two tensors \( X, T \in \mathbb{R}^{N^d} \) are orthogonal if \( \langle X, T \rangle = 0 \).

Consider \( U := u^{(1)} \circ u^{(2)} \circ \cdots \circ u^{(d)} \) and \( V := v^{(1)} \circ v^{(2)} \circ \cdots \circ v^{(d)} \). In this case, \( \langle U, V \rangle = \prod_{j=1}^{d} \langle u^{(j)}, v^{(j)} \rangle \). In particular, two rank-1 tensors are orthogonal if and only if any two of their corresponding components are orthogonal.
Definition 4. \( U \) and \( V \) are completely orthogonal, \( U \perp V \), if \( u_j \perp v_j \) for all \( j \in \{1, \ldots, d\} \). Moreover, \( U \) and \( V \) are strongly orthogonal, \( U \perp^s V \), if \( U \perp V \) and for all \( 1 \leq j \leq d \), either \( u^{(j)} = \pm v^{(j)} \) or \( u^{(j)} \perp v^{(j)} \).

Definition 5. [Orthogonal rank decomposition [68]] Define rank\(_{\perp}(T)\) to be the minimal \( r \), such that \( T \) can be written as the sum of \( r \) orthogonal unit tensors as follows:

\[
T = \sum_{i=1}^{r} \lambda_i U_i, \quad \langle U_i, U_j \rangle = 0 \text{ for } i \neq j
\]

where \( U_i \in \mathbb{R}^{N_d} \) are unit tensors.

Lemma 6. For any rank-\( r \) tensor \( T = \sum_{i=1}^{r} \lambda_i u^{(1)}_i \circ u^{(2)}_i \circ \cdots \circ u^{(d)}_i \), we have \( \text{rank}_{\perp}(T) \leq (r^d - 1) \)

Proof of this lemma is provided in Section 2.4. In the next section we use the orthogonal decomposition to bound the nuclear and Frobenius norm of a tensor whose rank\(_{\perp}(T) \leq (r^d - 1)\).

### 2.2 Frobenius and nuclear norm

The Frobenius norm of a tensor is defined as

\[
\|T\|_F^2 := \langle T, T \rangle = \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \cdots \sum_{i_d=1}^{N_d} T_{i_1,i_2,\ldots,i_d}^2.
\]

Remember that a unit tensor \( U \in \bigotimes_{j=1}^{d} \mathbb{R}^{N} \) is a tensor that can be written as an outer product of \( d \) unit vectors. Using this definition, we define the spectral norm of a tensor as

\[
\|T\| := \max_{U \in U_d} \langle T, U \rangle,
\]

where \( U_d \) is the set of all unit tensors in \( \mathbb{R}^{N_d} \). The nuclear norm of a tensor, defined as the dual of the spectral norm, was originally formulated in [50, 107] and has been revisited in greater depth in the past few years, e.g., in [35, 42, 57, 81]. To be precise,

\[
\|T\|_* := \max_{\|X\| \leq 1} \langle T, X \rangle.
\]

In the matrix case, the nuclear norm is equal to the sum of the singular values of the matrix. However, a similar singular value decomposition is not available for tensors in general. What we
know so far is that calculating the nuclear norm of $d$-tensors with $d > 2$ is NP-hard [43] and similar to the matrix case, we can prove that

$$\|T\|_* := \min \{ \sum \lambda_i : T = \sum \lambda_i u_i^{(1)} \circ u_i^{(2)} \circ \cdots \circ u_i^{(d)}, \|u_i\| = 1, \forall i \}.$$  \hspace{1cm} (2.6)

Using the $\ell_1$-norm and $\ell_2$-norm of the vector of singular values of a matrix, we can prove that for any rank-$r$ matrix $M$, $\|M\|_* \leq \sqrt{r} \|M\|_F$. Using a combinatorial orthogonal decomposition, we can find a similar bound for a general rank-$r$ tensor.

**Theorem 7.** For any rank-$r$ tensor $T$, $\|T\|_F \leq \sqrt{r^{d-1}} \|T\|$ and $\|T\|_* \leq \sqrt{r^{d-1}} \|T\|_F$.

We prove this theorem by connecting the CP-decomposition with the orthogonal rank decomposition. The details of the proof are in Section 2.4.2.

### 2.3 Max-qnorm and atomic M-norm

In this section, we introduce the max-qnorm and M-norm of tensors and characterize the unit ball of these norms as tensors that have a specific decomposition with bounded factors. This then helps us to prove a bound on the max-qnorm and M-norm of low-rank tensors that is independent of $N$. We give an overview of properties of max-qnorm and M-norm and prove them in Section 2.4.

#### 2.3.1 Matrix max-norm

First, we define the max-norm of matrices which was first defined in [82] as $\gamma_2$ norm. We also mention some of the properties of the matrix max-norm which we generalize later on in this section. Recall that the max-norm of a matrix is defined as

$$\|M\|_{\max} = \min_{M = U \circ V} \{ \|U\|_{2,\infty} \|V\|_{2,\infty} \},$$  \hspace{1cm} (2.7)

where $\|U\|_{2,\infty} = \sup_{\|x\|_\infty = 1} \|Ux\|_\infty$ is the maximum $\ell_2$-norm of the rows of $U$ [82, 116].

Considering all the possible factorizations of a matrix $M = U \circ V$, the rank of $M$ is the minimum number of columns in the factors and nuclear norm of $M$ is the minimum product of the Frobenius norms of the factors. The max-norm, on the other hand, finds the factors with the smallest row-
norm as $\|U\|_{2,\infty}$ is the maximum $\ell_2$-norm of the rows of the matrix $U$. Furthermore, it was proved in [78] that max-norm is comparable with nuclear norm in the following sense:

$$\|M\|_{\text{max}} \approx \inf \left\{ \sum_j |\sigma_j| : M = \sum_j \sigma_j u_j v_j^T, \|u_j\|_{\infty} = \|v_j\|_{\infty} = 1 \right\}.$$  \hfill (2.8)

Here, the factor of equivalence is the Grothendieck’s constant $K_G \in (1.67, 1.79)$. To be precise,

$$\frac{\inf \sum_j |\sigma_j|}{K_G} \leq \|M\|_{\text{max}} \leq \inf \sum_j |\sigma_j|,$$  \hfill (2.9)

where the infimum is taken over all nuclear decompositions $M = \sum_j \sigma_j u_j v_j^T, \|u_j\|_{\infty} = \|v_j\|_{\infty} = 1$. Moreover, in connection with element-wise $\ell_\infty$ norm we have:

$$\|M\|_{\infty} \leq \|M\|_{\text{max}} \leq \sqrt{\text{rank}(M)} \|M\|_{1,\infty} \leq \sqrt{\text{rank}(M)} \|M\|_{\infty}.$$  \hfill (2.10)

This is an interesting result that shows that we can bound the max-norm of a low-rank matrix by an upper bound that is independent of $N$.

### 2.3.2 Tensor max-qnorm and atomic M-norm

We generalize the definition of max-norm to tensors as follows. Let $T$ be an order-$d$ tensor. Then

$$\|T\|_{\text{max}} := \min_{T = U^{(1)} \circ U^{(2)} \circ \cdots \circ U^{(d)}} \left\{ \prod_{j=1}^d \|U^{(j)}\|_{2,\infty} \right\}.$$  \hfill (2.11)

Notice that this definition agrees with the definition of max-norm for matrices when $d = 2$. As in the matrix case, the rank of the tensor is the minimum possible number of columns in the low-rank factorization of $T = U^{(1)} \circ U^{(2)} \circ \cdots \circ U^{(d)}$ and the max-qnorm is the minimum product of the row norms of the factors over all such decompositions.

**Theorem 8.** For $d \geq 3$, the max-qnorm (2.11) does not satisfy the triangle inequality. However, it satisfies a quasi-triangle inequality

$$\|X + T\|_{\text{max}} \leq 2^{d-1} (\|X\|_{\text{max}} + \|T\|_{\text{max}}),$$

and, therefore, is a quasi-norm.

The proof of this theorem is in Section [2.4.3] Later on, in Section [3.3], we prove that certain
max-qnorm constrained minimization problems, break the $O(N_d^2)$ limitation on the number of measurements required for tensor completion and 1-bit tensor completion mainly because of two main properties:

- Max-qnorm of a bounded low-rank tensor does not depend on the size of the tensor.
- Defining $T_{\pm} := \{ T \in \{\pm 1\}^{N_1 \times N_2 \times \cdots \times N_d} \mid \text{rank}(T) = 1 \}$, the unit ball of the tensor max-qnorm is a subset of $C_d \text{conv}(T_{\pm})$ which is a convex combination of $2^{Nd}$ rank-1 sign tensors. Here $C_d$ is a constant that only depends on $d$ and $\text{conv}(S)$ is the convex envelope of the set $S$.

However, the max-qnorm is non-convex. To obtain a convex alternative that still satisfies the properties mentioned above, we consider the norm induced by the set $T_{\pm}$ directly; this is an atomic norm as discussed in [25]. The atomic M-norm of a tensor $T$ is then defined as the gauge of $T_{\pm}$ given by

$$\|T\|_M := \inf\{t > 0 : T \in t\ \text{conv}(T_{\pm})\}. \quad (2.12)$$

As $T_{\pm}$ is centrally symmetric around the origin and spans $\bigotimes_{j=1}^d \mathbb{R}^{N_j}$, this atomic norm is a convex norm and the gauge function can be rewritten as

$$\|T\|_M = \inf\{ \sum_{X \in T_{\pm}} c_X : T = \sum_{X \in T_{\pm}} c_X X, c_X \geq 0, X \in T_{\pm} \}. \quad (2.13)$$

### 2.3.3 Unit max-qnorm ball of tensors

In the next lemma, we prove that, similar to the matrix case, the tensor unit max-qnorm ball is comparable to the set $T_{\pm}$. First define $\mathcal{B}^T_{\text{max}}(1) := \{ T \in \mathbb{R}^{N_1 \times \cdots \times N_d} \mid \|T\|_{\text{max}} \leq 1 \}$ and $\mathcal{B}_M(1) := \{ T : \|T\|_M \leq 1 \}$.

**Lemma 9.** The unit ball of the max-qnorm, unit ball of atomic M-norm, and $\text{conv}(T_{\pm})$ satisfy the following:

- $\mathcal{B}_M(1) = \text{conv}(T_{\pm})$,
- $\mathcal{B}^T_{\text{max}}(1) \subset c_1 c_d^2 \text{conv}(T_{\pm})$.

Here $c_1$ and $c_2$ are derived from the generalized Grothendieck theorem [15, 121] which is explained thoroughly in Section 2.4.3.
Using Lemma [9], it is easy to analyze the Rademacher complexity of the unit ball of these two norms. In fact, noticing that \( T_\pm \) is a finite class with \( |T_\pm| < 2^{dN} \) and some basic properties of Rademacher complexity we can prove the following lemma. Below, \( \hat{R}_S(X) \) denotes the empirical Rademacher complexity of \( X \). To keep this section simple, we refer to Section 2.4.5 for the definition of Rademacher complexity and proof of Lemma [10].

**Lemma 10.** The empirical Rademacher complexities of the M-norm and max-qnorm unit balls are bounded by

- \( \sup_{S: |S| = m} \hat{R}_S(\mathbb{B}_M(1)) < 6\sqrt{\frac{dN}{m}} \),

- \( \sup_{S: |S| = m} \hat{R}_S(\mathbb{B}_{\max}^T(1)) < 6c_1d^2\sqrt{\frac{dN}{m}} \).

### 2.3.4 Max-qnorm and M-norm of bounded low-rank tensors

Next, we bound the max-qnorm and M-norm of a rank-\( r \) tensor whose (entry-wise) infinity norm is less than \( \alpha \). First, we bound the max-qnorm and a similar proof can be used to obtain a bound on the M-norm as well which we explain in the Section 2.4.6. As mentioned before, for \( d = 2 \), i.e., the matrix case, an interesting inequality has been proved which does not depend on the size of the matrix, i.e., \( \|M\|_{\max} \leq \sqrt{\text{rank}(M) \alpha} \) [116]. In what follows, we bound the max-qnorm and M-norm of a rank-\( r \) tensor \( T \) with \( \|T\|_\infty \leq \alpha \).

**Theorem 11.** Assume \( T \in \mathbb{R}^{N_1 \times \cdots \times N_d} \) is a rank-\( r \) tensor with \( \|T\|_\infty = \alpha \). Then

- \( \alpha \leq \|T\|_M \leq (r\sqrt{r})^{d-1} \alpha \).

- \( \alpha \leq \|T\|_{\max} \leq \sqrt{r^{d-1}} \alpha \).

The proofs of these two bounds are similar and both of them can be found in Section 2.4.6. Notice the discrepancy of Theorem 11 when \( d = 2 \). This is an artifact of the proof which hints at the fact that Theorem 11 might be not optimal in \( r \) for general \( d \) as well.

**Remark 12.** When \( d = 2 \), the atoms of the M-norm are rank-1 sign matrices. Therefore, flat low-rank matrices have small M-norm. Moreover, M-norm and max-norm are equivalent and the M-norm of a matrix \( M \) with \( \text{rank}(M) = r \) and \( \|M\|_\infty \leq \alpha \) is bounded by \( K_G \alpha \sqrt{r} \), where \( K_G \in \{1.671.79\} \) is the Grothendieck's constant [78].
2.4 Proofs

2.4.1 Proof of Lemma 6

To find an orthogonal rank decomposition we find an orthonormal basis for the sets spanned by corresponding components of $U_1, \ldots, U_r$. In particular, define $S_j := \{u^{(j)}_1, u^{(j)}_2, \ldots, u^{(j)}_r\}$ for $1 \leq j \leq d$. Next, define $B_j := \{b^{(j)}_1, b^{(j)}_2, \ldots, b^{(j)}_r\}$ to be an orthonormal basis for the set span{$S_j$}. Hence, for $1 \leq i \leq r$ and $1 \leq j \leq d$, we can write $u^{(j)}_i = \sum_{k=1}^d \sigma^{(j)}_{i,k} b^{(j)}_k$. Now by expanding each $u^{(j)}_i$ in $T = \sum_{i=1}^r \lambda_i u^{(1)}_i \circ u^{(2)}_i \circ \cdots \circ u^{(d)}_i$, we can get $T = \sum_{i=1}^r \sum_{j=1}^r \cdots \sum_{j_d=1}^r \sigma^{(1)}_{i,j_1,\ldots,j_d} b^{(1)}_{j_1} \circ b^{(2)}_{j_2} \circ \cdots \circ b^{(d)}_{j_d}$ where $\sigma^{(1)}_{i,j_1,\ldots,j_d} = \sum_{i=1}^r \lambda_i \sigma^{(1)}_{i,j_1} \sigma^{(2)}_{i,j_2} \cdots \sigma^{(d)}_{i,j_d}$. Hence we can write $T = \sum_{i=1}^r \lambda_i T_i$ where $T_i$’s are strongly orthogonal.

Notice that finding an orthogonal decomposition suffices and we can improve the result slightly by dropping the strong orthogonality of the sub-tensors. To do this notice that we can combine all the $T_i$’s that only differ in the first factor, i.e., defining $X_{j_2,\ldots,j_d} = (\sum_{i=1}^r \sigma^{(1)}_{i,j_2,\ldots,j_d} b^{(1)}_{i}) \circ b^{(2)}_{j_2} \circ \cdots \circ b^{(d)}_{j_d}$, we can observe that

$$T = \sum_{j_2=1}^r \sum_{j_3=1}^r \cdots \sum_{j_d=1}^r X_{j_2,\ldots,j_d} = \sum_{i=1}^r \lambda_i X_i,$$

such that any two different $X_i$ and $X_j$ differ at least in one of the factors 2 to $d$ and therefore, are orthogonal to each other.

2.4.2 Proof of Theorem 7

To prove this theorem, we first bound the Frobenius norm and nuclear norm of a tensor with rank$_\perp(T) = r$ and then use Lemma 6 to finish the proof.

Lemma 13. For a tensor $T$ with rank$_\perp(T) = r$, we have $\|T\|_F \leq \sqrt{r} \|T\|$ and $\|T\|_* \leq \sqrt{r} \|T\|_F$.

Proof: Assume that $T$ has the strongly orthogonal decomposition $T = \sum_{i=1}^r \lambda_i T_i$. Notice that for $1 \leq i \leq r$, $T_i$ is a rank-1 unit tensor with $\|T_i\| = \|T_i\|_F = \|T_i\|_* = 1$. Tensors $T_i$ are mutually orthogonal. Therefore, $\langle T, T_i \rangle = \lambda_i$ and we can conclude that

$$\|T\| \geq \max_{i=1}^r |\lambda_i|. \quad (2.14)$$

On the other hand, $\|T\|_F^2 = \langle T, T \rangle = \sum_{i=1}^r \lambda_i T_i, \sum_{i=1}^r \lambda_i T_i \rangle$ and as $\langle T_i, T_j \rangle = 0$ for $i \neq j$ we con-
clude that
\[ \|T\|_F = \sqrt{\sum_{i=1}^{r} \lambda_i^2}. \]  

(2.15)

Next, we bound the nuclear norm. Notice that
\[ \|T\|_* = \|\sum_{i=1}^{r} \lambda_i T_i\|_* \leq \sum_{i=1}^{r} |\lambda_i| \|T_i\|_* \]  which means that
\[ \|T\|_* \leq \sum_{i=1}^{r} |\lambda_i|. \]  

(2.16)

The lemma can be verified using (2.14), (2.15), and (2.16).

Combining Lemma 6 and Lemma 13 finishes the proof.

2.4.3 Proof of Theorem 8

Notice that for any tensors \( T \) and \( X \), there exist decompositions \( T = \bigcirc_{i=1}^{d} U^{(i)} \) and \( X = \bigcirc_{i=1}^{d} V^{(i)} \), where \( \|U^{(j)}\|_{2,\infty} \leq (\|T\|_{\text{max}})^{\frac{1}{2}} \) and \( \|V^{(j)}\|_{2,\infty} \leq (\|X\|_{\text{max}})^{\frac{1}{2}} \). Moreover, one way to factorize the tensor \( X + T \) is by concatenating the factors of \( X \) and \( T \) as \( X + T = \bigcirc_{i=1}^{d} [U^{(i)}, V^{(i)}] \) and therefore,
\[ \|X + T\|_{\text{max}} \leq \prod_{j=1}^{d} \|U^{(j)}, V^{(j)}\|_{2,\infty} \leq (\sqrt{\|X\|_{2,\infty}^2 + \|T\|_{2,\infty}^2})^d \leq 2^{d-1} (\|X\|_{\text{max}} + \|T\|_{\text{max}}), \]
which proves that max-qnorm is a quasi-norm. Notice that the last inequality follows from the inequality \( |a + b|^p \leq 2^{p-1}(|a|^p + |b|^p) \) for \( p > 1 \). It is easy to check that the max-qnorm satisfies the triangle inequality, when \( d = 2 \). However, this is not true for \( d > 2 \). Next, we prove this for \( d = 3 \) and higher-order cases can be proven similarly.

The main challenge in proving that the max-qnorm does not satisfy the triangle-inequality when \( d = 3 \) is that the size of the factors is not fixed. However, it can be observed from the following simple counter-example. Let \( T = T_1 + T_2 \), where \( T_1 = \left[ \begin{array}{ccc} 1 & 1 & 1 \\ 0 & 0 & 1 \\ \end{array} \right] \) and \( T_2 = \left[ \begin{array}{ccc} 1 & 1 & 1 \\ 1 & 1 & 1 \\ \end{array} \right] \), and note that \( T \) is a rank-2, \( 2 \times 2 \times 2 \) tensor. Here, \( T_1 \) and \( T_2 \) are rank-1 tensors with \( \|T_1\|_{\text{max}} = 1 \) and \( \|T_2\|_{\text{max}} = 1 \) (notice that for any \( T \), \( \|T\|_{\text{max}} \geq \|T\|_{\infty} \)). Therefore, if max-qnorm satisfies triangle-inequality, then \( \|T\|_{\text{max}} \) cannot exceed 2. In what follows we prove that this is not possible. Notice that even though \( T \) is a small tensor with only 8 elements we don’t have a straightforward way to calculate its max-qnorm. However, we can bound it from above and below. If \( \|T\|_{\text{max}} \leq 2 \), then there exists a decomposition \( T = U^{(1)} \circ U^{(2)} \circ U^{(3)} \) such that \( \|T\|_{\text{max}} = \prod_{j=1}^{3} \|U^{(j)}\|_{2,\infty} \leq 2 \), and
with a simple rescaling of the factors,

\[ \|U^{(1)}\|_{2,\infty} \leq \sqrt{2}, \quad \|U^{(2)}\|_{2,\infty} \leq \sqrt{2}, \quad \|U^{(3)}\|_{2,\infty} \leq 1. \]  \hspace{1cm} (2.17)

First, notice that \( T \) is an all-ones tensor except for one entry where \( T(1,1,1) = 2 \). Defining the generalized inner product as

\[ \langle x_1, \cdots, x_d \rangle := \sum_{i=1}^{k} \prod_{j=1}^{d} x_j(i), \]  \hspace{1cm} (2.18)

this means that

\[ \langle U^{(1)}(1,:), U^{(2)}(1,:), U^{(3)}(1,:) \rangle = 2. \]  \hspace{1cm} (2.19)

Using Cauchy-Schwarz

\[ \langle U^{(1)}(1,:), U^{(2)}(1,:), U^{(3)}(1,:) \rangle \leq \|U^{(1)}(1,:)\| \|U^{(2)}(1,:)\| \|U^{(3)}(1,:)\|_{\infty}. \]  \hspace{1cm} (2.20)

Combining (2.17), (2.19), and (2.20), we get

\[ 2 \leq \|U^{(1)}(1,:)\| \|U^{(2)}(1,:)\| \leq \|U^{(1)}\|_{2,\infty} \|U^{(2)}\|_{2,\infty} \leq 2, \]

which together with (2.17) proves that

\[ \|U^{(1)}(1,:)\| = \sqrt{2}, \text{ and } \|U^{(2)}(1,:)\| = \sqrt{2}. \]  \hspace{1cm} (2.21)

Moreover, similarly

\[ 2 \leq 2\|U^{(3)}(1,:)\|_{\infty} \leq 2 \Rightarrow \|U^{(3)}(1,:)\|_{\infty} = 1. \]

Notice that \( \|U^{(3)}(1,:)\| \leq 1 \), and \( \|U^{(3)}(1,:)\|_{\infty} = 1 \) which proves that \( U^{(3)}(1,:) \) is an all-zeros vector with a single non-zero entry of one. Remember that the number of columns of \( U^{(3)} \) is arbitrary.

Without loss of generality, we can assume

\[ U^{(3)}(1,:) = (1,0,\cdots,0). \]  \hspace{1cm} (2.22)

Combining this with (2.19), and (2.21) we can also prove that

\[ U^{(1)}(1,:) = U^{(2)}(1,:) = (\sqrt{2},0,\cdots,0). \]  \hspace{1cm} (2.23)
Now from $T(1,1,2) = 1$ and the above two equations we have to have $U(3)(2,1) = \frac{1}{2}$ and similarly $U(2)(2,1) = \frac{1}{\sqrt{2}}$. Finally $T(1,2,2) = U(1)(1,1) U(2)(2,1) U(3)(2,1) = \sqrt{2} \frac{1}{\sqrt{2}} \frac{1}{2} = \frac{1}{2}$ which is a contradiction. \hfill \Box

### 2.4.4 Proof of Lemma 9

Characterization of the unit ball of the atomic M-norm follows directly from (2.12). By definition, any tensor $T$ with $\|T\|_M \leq 1$ is a convex combination of the atoms of $T_\pm$, $T = \sum_{X \in T_\pm} c_X X$, $c_X > 0$ with $\sum_{X \in T_\pm} c_X = 1$. This proves that $\mathbb{B}_M(1) = \text{conv}(T_\pm)$.

To characterize the unit ball of max-qnorm, we use a generalization of Grothendieck’s theorem to higher-order tensors [15, 121]. First, we generalize the matrix $\|\cdot\|_{\infty,1}$ norm ($\|M\|_{\infty,1} := \sup_{\|x\|_\infty = 1} \|Mx\|_1$) as:

**Definition 14.** $\|T\|_{\infty,1} := \sup_{\|x\|_\infty = 1} |\sum_{i_1=1}^N \cdots \sum_{i_d=1}^N T(i_1, \ldots, i_d)x_1(i_1) \cdots x_d(i_d)|$.

**Theorem 15 (Generalized Grothendieck theorem).** *Let $T$ be an order-$d$ tensor such that

\[
\|T\|_{\infty,1} = \sup_{\|x\|_\infty = 1} |\sum_{i_1=1}^N \cdots \sum_{i_d=1}^N T(i_1, \ldots, i_d)x_1(i_1) \cdots x_d(i_d)| \leq 1,
\]

and let $u_{ij}^l \in \mathbb{R}^k$, $1 \leq j \leq d$, $1 \leq i \leq N$ be $N \times d$ vectors such that $\|u_{ij}^l\| \leq 1$. Then

\[
|\sum_{i_1=1}^N \cdots \sum_{i_d=1}^N T(i_1, \ldots, i_d) \langle u_{i_1}^1, u_{i_2}^2, \ldots, u_{i_d}^d \rangle| \leq c_1 c_2^d,
\]

where $\langle u_{i_1}^1, u_{i_2}^2, \ldots, u_{i_d}^d \rangle$ is the generalized inner product of $u_{i_1}^1, u_{i_2}^2, \ldots, u_{i_d}^d$ as defined in (2.18). Here, $c_1 \leq \frac{K_d}{2}$ and $c_2 \leq 2.83$.*

Now we use Theorem 15 to prove Lemma 9.

**Proof of Lemma 9** The dual norm of the max-qnorm is

\[
\|T\|_\infty^* = \max_{\|U\|_\infty \leq 1} \langle T, U \rangle = \max_{\|u_{ij}^l\| \leq 1} \sum_{i_1=1}^N \cdots \sum_{i_d=1}^N T(i_1, \ldots, i_d) \langle u_{i_1}^1, u_{i_2}^2, \ldots, u_{i_d}^d \rangle.
\]

Above, the length of the vectors $u_{i_1}^1, \ldots, u_{i_d}^d$ is not constrained (to clarify, notice that the length of these vectors are larger or equal to the rank of $U$). Using Theorem 15, $\|T\|_\infty^* \leq c_1 c_2^d \|T\|_{\infty,1}$. On
the other hand, in the special case when \( u_{i_1}^1, \ldots, u_{i_d}^d \in \mathbb{R} \) the right-hand side of (2.25) is equal to \( \|T\|_{\infty,1} \). Therefore, \( \|T\|_{\infty,1} \leq \|T\|_{\max}^* \). Taking the dual:

\[
\frac{\|T\|_{\infty,1}^*}{c_1c_2^d} \leq (\|T\|_{\max}^*)^* \leq \|T\|_{\infty,1}^*
\]

(2.26)

Notice that the max-qnorm, defined in (2.11) is a quasi-norm and therefore, \((\|T\|_{\max}^*)^* \) is not equal to \( \|T\|_{\max} \). However, notice that the max-qnorm is absolutely homogeneous and therefore,

\[
(\|T\|_{\max}^*)^* = \max_{\|Z\|_{\max} \leq 1} \langle T, Z \rangle \leq \|T\|_{\max}.
\]

which implies that

\[
\frac{\|T\|_{\infty,1}^*}{c_1c_2^d} \leq \|T\|_{\max}.
\]

(2.27)

To calculate the unit ball of \( \|\cdot\|_{\infty,1}^* \), notice that the argument of the supremum in Definition 14 is linear in each variable \( x_j(i_j) \) and as \(-1 \leq x_j(i_j) \leq 1 \), the supremum is achieved when \( x_j(i_j) = \pm 1 \) which means that \( \|T\|_{\infty,1} = \sup_{U \in T_{\pm}} |\langle T, U \rangle| \). Therefore, \( \text{conv}(T_{\pm}) \) is the unit ball of \( \|\cdot\|_{\infty,1}^* \) and Lemma 9(ii) follows from (2.27).

### 2.4.5 Proof of Lemma 10

A technical tool that we use in the proof of our main results involves data-dependent estimates of the Rademacher and Gaussian complexities of a function class. We refer to [8, 116] for a detailed introduction of these concepts.

Two important properties that will be used in the following lemma is: First, if \( \mathbb{F} \subset \mathbb{G} \), then \( \hat{R}_S(\mathbb{F}) \leq \hat{R}_S(\mathbb{G}) \) and second is \( \hat{R}_S(\mathbb{F}) = \hat{R}_S(\text{conv}(\mathbb{F})) \).

**Lemma 16.** \( \sup_{S: |S|=m} \hat{R}_S(\mathbb{B}_M(1)) < 6\sqrt{\frac{dN}{m}} \)

**Proof:** By definition, \( \mathbb{B}_M(1) = \text{conv}(T_{\pm}) \), and \( T_{\pm} \) is a finite class with \( |T_{\pm}| < 2^{dN} \). Therefore, \( \hat{R}_S(T_{\pm}) < \sqrt{\frac{72dN+\log|S|}{|S|}} \) [114] which concludes the proof.

**Lemma 17.** \( \sup_{S: |S|=m} \hat{R}_S(\mathbb{B}_{\max}^T(1)) < 6c_1c_2^d \sqrt{\frac{dN}{m}} \)

**Proof:** By Lemma 9, \( \mathbb{B}_{\max}^T(1) \subset c_1c_2^d \text{conv}(T_{\pm}) \) and we have \( \hat{R}_S(T_{\pm}) < \sqrt{\frac{72dN+\log|S|}{|S|}} \). Taking
the convex hull of this class and using \( |S| = m \leq N^d \) and scaling by \( c_1 c_2^d \) we get
\[
\hat{R}_S(\mathbb{B}_\infty^T(1)) \leq 6 c_1 c_2^d \sqrt{\frac{dN}{m}}.
\]

### 2.4.6 Proof of Theorem 11

In order to prove the tensor max-qnorm bound, we first sketch the proof of \([103]\) for the matrix case. That is, assuming that \( M \) is a matrix with rank(\( M \)) = \( r \) and \( \| M \|_\infty \leq \alpha \), we show that there exists a decomposition \( M = U \circ V \) where \( U \in \mathbb{R}^{N_1 \times r} \), \( V \in \mathbb{R}^{N_2 \times r} \) and \( \| U \|_\infty \leq \sqrt{r} \), \( \| V \|_\infty \leq \alpha \). To prove this, we first state a version of the John’s theorem \([103]\).

**Theorem 18** (John’s theorem \([64]\)). For any full-dimensional symmetric convex set \( K \subseteq \mathbb{R}^r \) (dim(\( K \)) = \( r \)) and any ellipsoid \( E \subseteq \mathbb{R}^r \) (defined with respect to \( \ell_2 \) norm) that is centered at the origin, there exists an invertible linear map \( S \) so that \( E \subseteq S(\mathbb{R}^r) \subseteq \sqrt{r}E \).

**Theorem 19.** \([103]\) Corollary 2.2] For any rank-\( r \) matrix \( M \in \mathbb{R}^{N_1 \times N_2} \) with \( \| M \|_\infty \leq \alpha \) there exist vectors \( u_1, \cdots, u_{N_1}, v_1, \cdots, v_{N_2} \in \mathbb{R}^r \) such that \( \langle u_i, v_j \rangle = M_{i,j} \) and \( \| u_i \| \leq \sqrt{r} \) and \( \| v_j \| \leq \alpha \).

The proof is based on considering any rank-\( r \) decomposition of \( M = X \circ Y \) where \( X \in \mathbb{R}^{N_1 \times r} \) and \( Y \in \mathbb{R}^{N_2 \times r} \) and \( M_{i,j} = \langle x_i, y_j \rangle \). Defining \( K \) to be the convex hull of the set \( \{ \pm x_i : i \in [N_1] \} \). Then using the linear map \( S \) in John’s Theorem for the set \( K \) with the ellipsoid \( E = \mathbb{B}_r := \{ x \in \mathbb{R}^r : \| x \|_2 \leq 1 \} \), the decomposition \( M = (XS) \circ (YS^{-1}) \) satisfies the conditions of Theorem 19 \([103]\).

The following lemma proves the existence of a nuclear decomposition for bounded rank-\( r \) tensors, which can be used directly to bound the M-norm of a bounded rank-\( r \) tensor.

**Lemma 20.** Any order-\( d \), rank-\( r \) tensor \( T \), with \( \| T \|_\infty \leq \alpha \) can be decomposed into \( r^{d-1} \) rank-one tensors whose components have unit infinity norm such that
\[
T = \sum_{j=1}^{r^{d-1}} \sigma_j u_j^1 \circ u_j^2 \circ \cdots \circ u_j^d, \quad \| u_j^1 \|_\infty, \cdots, \| u_j^d \|_\infty \leq 1, \quad \text{with} \quad \sum_{j=1}^{r^{d-1}} |\sigma_j| \leq (r\sqrt{r})^{d-1} \alpha. \tag{2.28}
\]

**Proof:** We prove this lemma by induction. The proof for \( d = 2 \) follows directly from applying John’s theorem to a rank-\( r \) decomposition of \( T \), i.e., \( T = XS \circ YS^{-1} \) where \( T = X \circ Y \) \([103]\). This is summarized in Theorem 19 above as well.

Now assume an order-\( d \) tensor which can be written as \( T = \sum_{j=1}^{r^d} \lambda_j v_j^1 \circ v_j^2 \circ \cdots \circ v_j^d \) and \( \| T \|_\infty \leq \alpha \). Matricizing along the first dimension results in \( T_{[1]} = \sum_{j=1}^{r} (\lambda_j u_i^{(1)}) \circ (u_i^{(2)} \otimes \cdots \otimes u_i^{(d)}) \). Using
Matlab notations we can write $T_{[1]} = U \circ V$ where $U(:,i) = \lambda_i u_i^{(1)} \in \mathbb{R}^{N_1}$, and $V(:,i) = u_i^{(2)} \otimes \cdots \otimes u_i^{(d)} \in \mathbb{R}^{\Pi_{k=2}^{d} N_k}$.

Using John’s theorem, there exists an $S \in \mathbb{R}^{r \times r}$ where $T_{[1]} = X \circ Y$ where $X = US$, $Y = VS^{-1}$, $\|X\|_\infty \leq \|X\|_2 \leq \sqrt{r}$, and $\|Y\|_\infty \leq \|Y\|_2 \leq \alpha$. Furthermore, each column of $Y$ is a linear combination of the columns of $V$, i.e., there exist $\zeta_1, \cdots, \zeta_r$ such that $Y(:,i) = \sum_{j=1}^r \zeta_j (u_j^{(2)} \otimes \cdots \otimes u_j^{(d)})$. Therefore, unfolding $i$-th column of $Y$ into a $(d-1)$-dimensional tensor $E_i \in \mathbb{R}^{N_2 \times \cdots \times N_d}$ would result in a rank-$r$, $(d-1)$-dimensional tensor with $\|E_i\|_\infty \leq \|Y\|_\infty \leq \alpha$. By induction, $E_i$ can be decomposed into $r^{d-2}$ rank-one tensors with bounded factors, i.e.,

$$E_i = \sum_{j=1}^{r^{d-2}} \sigma_{i,j} v_{i,j}^1 \circ v_{i,j}^2 \circ \cdots \circ v_{i,j}^d, \quad \|v_{i,j}\|_\infty \leq 1, \quad \sum |\sigma_{i,j}| \leq (r\sqrt{r})^{d-2} \alpha.$$

Going back to the original tensor, as $T_{[1]} = X \circ Y$, we also have $T = \sum_{i=1}^r X(:,i) \circ (\sum_{j=1}^{r^{d-2}} \sigma_{i,j} v_{i,j}^1 \circ v_{i,j}^2 \circ \cdots \circ v_{i,j}^d)$. Notice that $\|X(:,i)\|_\infty \leq \sqrt{r}$. Therefore, we can rewrite

$$T = \sum_{i=1}^r \sum_{j=1}^{r^{d-2}} (\sigma_{i,j} \|X(:,i)\|_\infty) \frac{X(:,i)}{\|X(:,i)\|_\infty} \circ v_{i,j}^1 \circ v_{i,j}^2 \circ \cdots \circ v_{i,j}^d.$$

By rearranging, we get $T = \sum_{k=1}^{r^{d-1}} \sigma_k u_k^1 \circ u_k^2 \circ \cdots \circ u_k^d$, $\|u_k^1\|_\infty, \cdots, \|u_k^d\|_\infty \leq 1$ and

$$\sum_{k=1}^{r^{d-1}} |\sigma_k| = \sum_{i=1}^r \sum_{j=1}^{r^{d-2}} |\sigma_{i,j}| \|X(:,i)\|_\infty \leq \sum_{i=1}^r \sqrt{r} \sum_{j=1}^{r^{d-2}} |\sigma_{i,j}| \leq \sum_{i=1}^r \sqrt{r} (r\sqrt{r})^{d-2} \alpha = (r\sqrt{r})^{d-1} \alpha,$$

which concludes the proof of Lemma [20]. This lemma can be used directly to bound the M-norm of a bounded rank-$r$ tensor.

Next, we bound the max-qnorm of a bounded rank-$r$ tensor. The following lemma proves the existence of a nuclear decomposition for bounded rank-$r$ tensors, which can be used directly to bound their max-qnorm. As the max norm is one-homogeneous, without loss of generality we assume $\|T\|_\infty \leq 1$.

**Lemma 21.** Any order-$d$, rank-$r$ tensor $T \in \bigotimes_{i=1}^d \mathbb{R}^{N_i}$, with $\|T\|_\infty \leq 1$ can be decomposed into
$r^{d-1}$ rank-one tensors, $T = \sum_{j=1}^{r^{d-1}} u_j^1 \circ u_j^2 \circ \cdots \circ u_j^d$, where:

$$\sum_{j=1}^{r^{d-1}} (u_j^k(t))^2 \leq r^{d-1} \text{ for any } 1 \leq k \leq d, \ 1 \leq t \leq N_k. \quad (2.29)$$

Notice that $\sqrt{\sum_{j=1}^{r^{d-1}} (u_j^k(t))^2}$ is the spectral norm of $j$-th row of $k$-th factor of $T$, i.e., $\sum_{j=1}^{r^{d-1}} (u_j^k(t))^2 \leq r^{d-1}$ means that the two-infinity norm of the factors is bounded by $\sqrt{r^{d-1}}$.

**Remark 22.** [Proof by Lemma 20] At the end of this subsection, we provide a proof for the lemma as stated above. However, using the decomposition obtained in Lemma 20, we can find a decomposition with $\sum_{j=1}^{r^{d-1}} (u_j^k(t))^2 \leq r^d$. To do this notice that by Lemma 20, and defining $\bar{\sigma} := \{\sigma_1, \ldots, \sigma_{r^{d-1}}\}$ we can write

$$T = \sum_{j=1}^{r^{d-1}} \sigma_j v_j^1 \circ v_j^2 \circ \cdots \circ v_j^d, \parallel v_j^1 \parallel_\infty, \cdots, \parallel v_j^d \parallel_\infty \leq 1, \text{ with } \parallel \bar{\sigma} \parallel_1 \leq (r\sqrt{r})^{d-1}.$$  

Now define

$$u_j^k := (\sigma_j)^{\frac{1}{d}} v_j^k \text{ for any } 1 \leq k \leq d, \ 1 \leq t \leq N_k.$$  

It is easy to check that $T = \sum_{j=1}^{r^{d-1}} u_j^1 \circ u_j^2 \circ \cdots \circ u_j^d$ and

$$\sum_{j=1}^{r^{d-1}} (u_j^k(t))^2 = \sum_{j=1}^{r^{d-1}} \sigma_j^2 (v_j^k(t))^2 \leq \sum_{j=1}^{r^{d-1}} \sigma_j^2 = \parallel \bar{\sigma} \parallel_2^2.$$  

Using Holder’s inequality, when $d \geq 2$ we have

$$\sum_{j=1}^{r^{d-1}} (u_j^k(t))^2 = \parallel \bar{\sigma} \parallel_2^2 \leq \parallel \bar{\sigma} \parallel_2^2 (r^{d-1})^{1-\frac{2}{d}} \leq r^{(d-1)\frac{d}{2}} \leq r^{\frac{(d-1)(d+1)}{d}}.$$  

This proves an upper bound which is close to the one in the lemma. To get a more optimal upper bound (the one stated in the Lemma 21) we need to go over the induction steps as explained below.

**Proof of Lemma 21** We prove this lemma by induction. The proof for $d = 2$ follows directly from applying John’s theorem to a rank-$r$ decomposition of $T$, i.e., $T = XS \circ YS^{-1}$ where $X = Y \circ Y$. Now assume an order-$d$ tensor which can be written as $T = \sum'_{j=1} u_j^1 \circ u_j^2 \circ \cdots \circ u_j^d$ and $\|T\|_\infty \leq 1$. Matricizing along the first dimension results in $T_{[1]} = \sum'_{j=1} (u_j^{(1)}) \circ (u_j^{(2)}) \otimes \cdots \otimes (u_j^{(d)})$. Using matrix notation we can write $T_{[1]} = U \circ V$ where $U(\cdot, i) = u_i^{(1)} \in \mathbb{R}^{N_1}$, and $V(\cdot, i) = u_i^{(2)} \otimes \cdots \otimes u_i^{(d)} \in \mathbb{R}^{\Pi_{k=2} N_k}$. 

28
Using John’s theorem, there exists an \( S \in \mathbb{R}^{r \times r} \) where \( T[1] = X \odot Y \) where \( X = US \), \( Y = VS^{-1} \), \( \|X\|_{2,\infty} \leq \sqrt{r} \), and \( \|Y\|_{\infty} \leq \|Y\|_{2,\infty} \leq 1 \). More importantly, each column of \( Y \) is a linear combination of the columns of \( V \). More precisely, there exist \( \zeta_1, \cdots, \zeta_r \) such that \( Y(:,i) = \sum_{j=1}^{r} \zeta_j (u_j^{(2)} \otimes \cdots \otimes u_j^{(d)}) \). Therefore, unfolding \( i \)-th column of \( Y \) into a \((d-1)\)-dimensional tensor \( E_i \in \mathbb{R}^{N_1 \times \cdots \times N_d} \) would result in a rank-\( r \), \((d-1)\)-dimensional tensor with \( \|E_i\|_{\infty} \leq \|Y\|_{\infty} \leq 1 \). By induction, \( E_i \) can be decomposed into \( r^{d-2} \) rank-one tensors where \( E_i = \sum_{j=1}^{r^{d-2}} v_{i,j}^2 \circ v_{i,j}^3 \circ \cdots \circ v_{i,j}^d \), where \( \sum_{j=1}^{r^{d-2}} (v_{i,j}(t))^2 \leq r^{d-2} \) for any \( 2 \leq k \leq d \) and any \( 1 \leq t \leq N_k \). Notice that the factors start from \( v_{i,j}^2 \) to emphasize that \( E_i \) is generated from the indices in the dimensions 2 to \( d \).

Going back to the original tensor, as \( T[1] = X \odot Y \), we can write

\[
T = \sum_{i=1}^{r} X(:,i) \odot (\sum_{j=1}^{r^{d-2}} v_{i,j}^2 \circ v_{i,j}^3 \circ \cdots \circ v_{i,j}^d).
\]

By distributing the outer product we get

\[
T = \sum_{i=1}^{r} \sum_{j=1}^{r^{d-2}} X(:,i) \odot v_{i,j}^2 \circ v_{i,j}^3 \circ \cdots \circ v_{i,j}^d.
\]

Renaming the vectors in the factors we get

\[
T = \sum_{k=1}^{r^{d-1}} u_k^1 \circ u_k^2 \circ \cdots \circ u_k^d.
\]

Now we bound the max norm of \( T \) using this decomposition by considering each factor separately using the information we have about \( X \) and \( E_i \)s.

Starting from the first factor, notice that \( \|X\|_{2,\infty} \leq \sqrt{r} \) or more precisely \( \sum_{i=1}^{r} X(t,i)^2 \leq r \) for any \( 1 \leq t \leq N_1 \). By careful examining of the two decompositions of \( T \) stated above, we get

\[
u_k^1 = X(:, \text{mod}(k-1,r)+1)
\]

and therefore

\[
\sum_{k=1}^{r^{d-1}} (u_k^1(t))^2 = r^{d-2} \sum_{i=1}^{r} X(t,i)^2 \leq r^{d-2} r = r^{d-1}, \text{ for any } 1 \leq t \leq N_1, \tag{2.30}
\]

which proves the lemma for the vectors in the first dimension of the decomposition.
For the second dimension, define $j := \text{mod}(k - 1, r^{d-2}) + 1$, and $j := \frac{k - j}{r^{d-2}} + 1$. Then

$$u_k^2 = v_{i,j}^k;$$

and therefore,

$$\sum_{k=1}^{r^d-1} (u_k^2(t))^2 = \sum_{i=1}^r \sum_{j=1}^{r^{d-2}} (v_{i,j}^2(t))^2 \leq \sum_{i=1}^r r^{d-2} \leq r^{d-1}, \text{ for any } 1 \leq t \leq N_2,$$

which finishes the proof of the lemma for the vectors in the second dimension. All the other dimensions can be bounded in an exactly similar way to the second dimension.

The bound on the max-qnorm of a bounded rank-$r$ tensor follows directly from Lemma 21 and definition of tensor max-qnorm.

Remark 23. In both lemmas [20] and [21], we start by decomposing a tensor $T = U_1 \circ U_2 \circ \cdots \circ U_d$ into $T_{[1]} = U_1 \circ V$ and generating $K$ (in the John’s theorem) by the rows of the factor $U_1$. Notice that John’s theorem requires the set $K$ to be full-dimensional. This condition is satisfied in the matrix case as the low-rank decomposition of a matrix (with the smallest rank) is full-dimensional. However, this is not necessarily the case for tensors. In other words, the matricization along a dimension might have smaller rank than the original tensor. To take care of this issue, consider a factor $U_{add}$ with the same size of $U_1$ such that $U_1 + U_{add}$ is full-dimensional. Now the tensor $T_\varepsilon = T + \varepsilon U_{add} \circ U_2 \circ \cdots \circ U_d \varepsilon$ satisfies the conditions of the John’s theorem and by taking $\varepsilon$ to zero we can prove that $\|T\|_M = \|T_\varepsilon\|_M$ and $\|T\|_{\max} = \|T_\varepsilon\|_{\max}$. Notice that M-norm is convex and max-qnorm satisfies $\|X + T\|_{\max} \leq (\sqrt[3]{\|X\|_{\max}^2 + \|T\|_{\max}^2})^d$. 

30
Chapter 3

Tensor completion

In many applications, acquiring all the indices of a tensor is either impossible or too expensive, which results in having access to just a subset of the entries of the tensor. In this section, we consider the problem of tensor completion from noisy measurements of a subset of its entries. As explained before, we assume that the indices of the entries that are measured are drawn independently at random with replacement. Also, the tensor of interest is low-rank and has bounded entries. Instead of constraining the problem to the set of low-rank bounded tensors, we consider a more general case and consider the set of bounded tensors with bounded M-norm which includes the set of low-rank bounded tensors. We minimize a constrained least squares (LS) problem given in (3.5) below. Similar results can be obtained for a max-qnorm constrained LS. We only provide the final result of the max-qnorm constrained problem in Theorem 29 as the steps are exactly similar to the M-norm constrained one. When $d = 2$, i.e., the matrix case, max-norm constrained matrix completion has been thoroughly studied in [19], so we will not discuss the lemmas and theorems that can be directly used in the tensor case; see [19] for more details.

3.1 Introduction and past results

The problem of matrix completion, i.e., recovering a matrix from partial noisy measurements of a subset of its entries arises in a wide variety of practical applications, including collaborative filtering [47], sensor localization [14, 113], system identification [84] and seismic data interpolation [3]. The low-rank structure of a matrix makes it possible to complete it by subsampling a number of entries which is much smaller than the ambient dimension of the matrix. Matrix completion is useful for applications where acquiring the full data exactly is either expensive or impossible due to physical limitations.
Assuming that we have access to \( m \) random entries of a matrix \( M^\# \), the set of indices of which is denoted by \( \Omega \), where \(|\Omega| = m\), the matrix completion problem entails finding a matrix \( \hat{M} \) with smallest rank that agrees with the \( m \) obtained samples, i.e., defining \( M_\Omega \) to be the projection of \( M \) onto the set of matrices whose all non-zero entries are in \( \Omega \),

\[
\hat{M} := \arg \min \text{rank}(M) \text{ s.t. } M_\Omega = M_\Omega^\#.
\]

This problem has been extensively studied in the literature \([17, 22, 23, 66]\). In general, rank minimization is NP-hard. Therefore, one of the most common approaches is minimizing convex surrogates of the rank, such as nuclear norm \([17, 21, 23]\), and max-norm \([19, 40]\). Both these convex regularizers can be used to recover the underlying matrix with as few as \( O(rN\log(N)) \) measurements which is close to the number of free variables of a rank-\( r \) matrix in \( \mathbb{R}^{N \times N} \) which is \( O(Nr) \).

Using max-norm for learning low-rank matrices was pioneered in \([118]\) where max-norm was used for collaborative prediction. In this chapter, we use max-qnorm for tensor completion which is a generalization of a recent result on matrix completion using max-norm constrained optimization \([19]\). First, we review some of the results which are related to M-norm and max-qnorm tensor completion. In particular, we first go over some of the matrix completion results including using nuclear norm and max-norm and then review some of the results on tensor completion.

Inspired by the result of \([38]\), which proved that the nuclear norm is the convex envelope of the rank function, most of the research on matrix completion has focused on using nuclear-norm minimization. Assuming \( M^\# \) to be a rank-\( r \), \( N \times N \) matrix, and \( M_\Omega^\# \) to be the set of \( m \) independent samples of this matrix, in \([22, 114]\), it was proved that solving

\[
\hat{M} := \arg \min \|M\|_* \text{ subject to } M_\Omega = M_\Omega^\#,
\]  

(3.1)

recovers the matrix \( M^\# \) exactly if \( |\Omega| > C N^{1.2} r \log(N) \), provided that the row and column space of the matrix is “incoherent”. This result was later improved in \([66]\) to \( |\Omega| = O(Nr \log(N)) \). There has been significant research in this area since then, either in sharpening the theoretical bound, e.g., \([12, 23, 104]\) or designing efficient algorithms to solve (3.1), e.g., \([17, 61]\).

More relevant to noisy tensor completion are the results of \([19, 21, 67]\) which consider recov-
ering $M^\sharp$ from measurements $Y_\Omega$, where $Y = M^\sharp + Z$, and $|\Omega| = m$; here $Z$ is a noise matrix. It was proved in [21] that if $\|Z_\Omega\|_F \leq \delta$, by solving the nuclear-norm minimization problem

$$\arg\min \|M\|_* \text{ subject to } \|(M - Y)_\Omega\|_F \leq \delta,$$

we can recover $\hat{M}$ where

$$\frac{1}{N} \|M^\sharp - \hat{M}\|_F \leq C \sqrt{\frac{N}{m} \delta + 2 \delta N},$$

provided that there are sufficiently many measurements for perfect recovery in the noiseless case.

Another approach was taken by [67] where the authors assume that $\|M^\sharp\|_\infty \leq \alpha$, and $Z$ is a zero-mean random matrix whose entries are i.i.d. with subgaussian-norm $\sigma$. They then suggest initializing the left and right-hand singular vectors ($L$ and $R$) from the observations $Y_\Omega$ and prove that by solving

$$\min_{L,S,R} \frac{1}{2} \|M^\sharp - LSR\|_F^2 \text{ subject to } L'L = I_r, R'R = I_r,$$

one can recover a rank-$r$ matrix $\hat{M}$ where

$$\frac{1}{N} \|M^\sharp - \hat{M}\|_F \leq C\alpha \sqrt{\frac{Nr}{m}} + C'\sigma \sqrt{\frac{Nr\alpha \log(N)}{m}}.$$

Inspired by promising results of using max-norm for collaborative filtering [118], a max-norm constrained optimization was employed in [40] to solve the noisy matrix completion problem under the uniform sampling assumption. Nuclear norm minimization has been proven to be rate-optimal for matrix completion. However, it is not entirely clear if it is the best approach for non-uniform sampling. In many applications, such as collaborative filtering, the uniform sampling assumption is not a reasonable assumption. For example, in the Netflix problem, some movies get much more attention and therefore have more chance of being rated compared to others. To tackle the issue of non-uniform samples, using a weighted nuclear norm was suggested in [91], imposing probability distributions on samples belonging to each row or column. Due to similar considerations, [19] generalized the max-norm matrix completion to the case of non-uniform sampling and proved that, with high probability, $m = O\left(\frac{N^r \log^3(\frac{1}{\epsilon})}{\epsilon^3}\right)$ samples are sufficient for achieving mean squared recovery error $\epsilon$, where the mean squared error is dependent on the distribution of the observations. To be more precise, in their error bound, indices that have higher probability of being observed are recovered more accurately compared to the entries that have less probability of being observed. In particular, [19] assumed a general sampling distribution as explained in Section 3.2 (when $d = 2$).
that includes both uniform and non-uniform sampling. Assuming that each entry of the noise matrix is an independent zero-mean Gaussian random variable with variance $\sigma$, and $\|M\|_\infty \leq \alpha$, they proved that the solution $\hat{M}_{\text{max}}$ of

$$\min_{\|M\|_{\text{max}} \leq \sqrt{r\alpha}} \|(M^\dagger - M)\Omega\|_F^2,$$

assuming each entry gets observed with probability larger than $\frac{1}{\mu N^2}$ on average, satisfies

$$\frac{1}{N^2}\|\hat{M}_{\text{max}} - M^\dagger\|_F^2 \leq C\mu(\alpha + \sigma)\alpha \sqrt{\frac{rN}{n}},$$

with probability greater than $1 - 2e^{-dN}$. This section is a generalization of the above result to tensor completion.

Finally, we explain some of the past results on tensor completion. To our knowledge, this chapter provides the first result that proves linear dependence of the sufficient number of random samples on $N$ when the measurements are taken independently at random. It is worth mentioning, though, that [73] proves that $O(Nr^{d-0.5}d \log(r))$ adaptively chosen samples is sufficient for exact recovery of tensors. However, the result is heavily dependent on the samples being adaptive.

There is a long list of heuristic algorithms that attempt to solve the tensor completion problem by using different decompositions or matricizations which, in spite of showing good empirical results, do not do a theoretical comparison of tensor completion and matrix completion [49, 83]. The most popular approach is minimizing the sum of nuclear norms of all the matricizations of the tensor along all modes. To be precise one solves

$$\min_{X} \sum_{i=1}^{d} \beta_i \|X_{(i)}\|_*, \text{ subject to } X_{\Omega} = T^\dagger_{\Omega}, \quad (3.2)$$

where $X_{(i)}$ is the mode-$i$ matricization of the tensor (see [44, 83, 111, 120]). The result obtained by solving (3.2) is highly sensitive on the choice of the weights $\beta_i$ and an exact recovery result is not available. At least in the special case of tensor sensing, where the measurements of the tensor are its inner products with random Gaussian tensors, [90] proves that $m = O(rN^{d-1})$ is necessary for (3.2), whereas a more balanced matricization such as $X_{\left\lfloor \frac{d}{2} \right\rfloor}$ (as explained in Section 1.5) can achieve successful recovery with $m = O(r^{\left\lfloor \frac{d}{2} \right\rfloor}N^{\left\lfloor \frac{d}{2} \right\rfloor})$ Gaussian measurements.
Assuming $T$ is symmetric and has an orthogonal CP decomposition, it was proved in [60] that when $d = 3$, an alternating minimization algorithm can achieve exact recovery from $O(r^5 N^3 \log(N)^4)$ random samples. However, the empirical results of this work show good results for non-symmetric tensors as well if a good initial point can be found.

Decompositions such as Tucker [122], Hierarchical Tucker [48], higher-order svd [32] and tensor train [95] which have more expressive power (i.e., more free variable) than CP decomposition has been considered in various settings and has been applied to various applications. In [72], tensors with fixed multi-linear rank are estimated by performing Riemannian optimization techniques on the manifold of tensors of fixed multi-linear rank. An algorithm based on Tucker decomposition was proposed in [39] which is numerically shown to perform well even when the multi-linear ranks are unknown. In [131], a generalization of the singular value decomposition for tensors, called t-SVD, is used to prove that a third order tensor ($d = 3$) can be recovered from $O(rN^2 \log(N)^2)$ measurements, provided that the tensor satisfies some incoherence conditions called tensor incoherence conditions. In [5], a heuristic algorithm was proposed that can find an approximation to the tensor in the hierarchical Tucker format by inspecting $O(dr^3 + d \log(d)rN)$ entries of the tensor adaptively. An efficient optimization framework for problems whose solutions are well-approximated by Hierarchical Tucker (HT) tensors was developed in [29]. Finally, the tensor train decomposition was used for tensor completion in [56, 106].

In short, any method that is based on naive matricization can not improve the $O(N^{\lceil d/2 \rceil})$ required measurements bottleneck. The large gap between the theoretical guarantees and the number of free variables of the set of low-rank tensors ($O(rNd)$) motivated a lot of researchers to consider formulations that are not based on matricization (including the work in this chapter).

A noteworthy paper is an interesting theoretical result that minimizes the nuclear norm of a tensor defined as the dual of the spectral norm and avoids any kind of matricization [129]. They show that using the nuclear norm, the sample size requirement for a tensor with low coherence using nuclear norm is $m = O(\sqrt{rN^d} \log(N))$ which is still far away from the number of free variables. Comparing our result with the result of [129], an important question that needs to be investigated is whether max-qnorm is a better measure of the complexity of low-rank tensors compared to nuclear norm or whether the difference is just an artifact of the proofs. We investigate this question in more detail in Chapter 4. Another difficulty of using tensor nuclear norm is the lack of sophisticated or even approximate algorithms that can minimize the nuclear norm of a tensor. We propose a factor-
frob function in Chapter 4 which can be considered as a proxy for nuclear norm but needs more investigation.

Using polynomial-time algorithms was considered in [7] with two important results. First, for third-order tensors they show that if \( m = O(N^{3r/2}) \) then there is a polynomial time algorithm for completing it. Further, they show a connection between noisy tensor completion and strongly refuting random 3-SAT and 3-XOR formulas which hints that polynomial-time algorithms might not be able to achieve a sample complexity better than \( O(N^{3r/2}) \). We believe that max-qnorm and M-norm cannot be calculated in polynomial-time but the results in this section might shed some light on the above claim.

Finally, the sum of squares method has been considered for exact tensor completion in [101] and has better sample complexity \( (O(rN^{d/2})) \) than matricization (for odd values of \( d \)). Similar to the sum of squares algorithm in [130], this algorithm does not scale well to large problem instances. Therefore spectral algorithms were considered in [89] where they obtain sample complexity of \( O(rN^{d/2}) \) (when \( d = 3 \)) which although not proven in [89] or in this thesis, seems to be optimal when \( r \gg N \).

### 3.2 Observation model

Given an order-\( d \) tensor \( T^\sharp \in \mathbb{R}^{N^d} \) and a random subset of indices \( S = \{\omega_1, \omega_2, \ldots, \omega_m\} \), \( \omega_t \in [N]^d \), we observe \( m \) noisy entries \( \{Y_{\omega_t}\}_{t=1}^m \):

\[
Y_{\omega_t} = T^\sharp(\omega_t) + \sigma \xi_t, \quad t = 1, \ldots, m,
\]

for some \( \sigma > 0 \). The variables \( \xi_t \) are zero-mean i.i.d. random variables with \( \mathbb{E}(\xi_t^2) = 1 \). The indices in \( S \) are drawn randomly with replacement from a predefined probability distribution \( \Pi = \{\pi_\omega\} \), for \( \omega \in [N]^d \), such that \( \sum_\omega \pi_\omega = 1 \). Obviously, \( \max \pi_\omega \geq \frac{1}{N^d} \). Although it is not a necessary condition for our proof, it is natural to assume that there exists \( \mu \geq 1 \) such that

\[
\pi_\omega \geq \frac{1}{\mu N^d} \quad \forall \omega \in [N]^d,
\]

which ensures that each entry is observed with some positive probability. This observation model includes both uniform and non-uniform sampling and is a better fit than uniform sampling in many
practical applications.

### 3.3 M-norm constrained least squares estimation

Given a collection of noisy observations \( \{ Y_{\omega t} \}_{t=1}^m \) of a low-rank tensor \( T^\sharp \), following the observation model (3.3), we solve a least squares problem to find an estimate of \( T^\sharp \). Consider the set of bounded M-norm tensors with bounded infinity norm

\[
K^T_M(\alpha, R) := \{ T \in \mathbb{R}^{N^d} : \| T \|_\infty \leq \alpha, \| T \|_M \leq R \}.
\]

Notice that assuming that \( T^\sharp \) has rank \( r \) and \( \| T^\sharp \|_\infty \leq \alpha \), Theorem 11 ensures that a choice of \( R = (r\sqrt{r})^{d-1} \alpha \) is sufficient to include \( T^\sharp \) in \( K^T_M(\alpha, R) \). Defining

\[
\mathcal{L}_m(X, Y) := \frac{1}{m} \sum_{t=1}^m (X_{\omega t} - Y_{\omega t})^2,
\]

we bound the recovery error for the estimate \( \hat{T}_M \) obtained by solving the optimization problem

\[
\hat{T}_M = \arg\min_X \mathcal{L}_m(X, Y) \quad \text{subject to} \quad X \in K^T_M(\alpha, R).
\]

In words, \( \hat{T}_M \) is a tensor with entries bounded by \( \alpha \) and M-norm less than \( R \) that is closest to the sampled tensor in Frobenius norm. Moreover, as for any tensor \( T \), \( \| T \|_M \) and \( \| T \|_{\text{max}} \) are greater than or equal to \( \| T \|_\infty \), we assume \( R \geq \alpha \).

We now state the main result on the performance of M-norm constrained tensor completion as in (3.5) for recovering a bounded low-rank tensor.

**Theorem 24.** Consider an order-\( d \) tensor \( T^\sharp \in \bigotimes_{i=1}^d \mathbb{R}^N \) with \( \| T^\sharp \|_\infty \leq \alpha \) and \( \| T^\sharp \|_M \leq R \). Given a collection of noisy observations \( \{ Y_{\omega t} \}_{t=1}^m \) following the observation model (3.3) where the noise sequence \( \xi_t \) are i.i.d. standard normal random variables, there exists a constant \( C < 20 \) such that the minimizer \( \hat{T}_M \) of (3.5) satisfies:

\[
\| \hat{T}_M - T^\sharp \|_{\Pi}^2 := \sum_{\omega} \pi_\omega \left( \hat{T}_M(\omega) - T^\sharp(\omega) \right)^2 \leq C \left( \sigma (R + \alpha) + R\alpha \right) \sqrt{dN \frac{dN}{m}},
\]

with probability greater than \( 1 - e^{-\frac{N}{mN}} - e^{-\frac{dN}{2}} \).
Corollary 25. If we assume each entry of the tensor is sampled with some positive probability, \( \pi_\omega \geq \frac{1}{\mu N^d} \) \( \forall \omega \in [N]^d \), then

\[
\frac{1}{N^d} \| \hat{T}_M - T^\sharp \|_F^2 \leq C\mu(\alpha + \sigma)R \sqrt{\frac{dN}{m}}.
\]

with probability greater than \( 1 - e^{\frac{-N}{\ln(N)}} - e^{-\frac{dN}{2}} \).

Remark 26. In Section 1.7.1, we presented a simplified version of the above theorem when \( \mu = 1 \) and \( T^\sharp \) is a rank-\( r \) tensor which uses the bound \( \| T^\sharp \|_M < (r \sqrt{r})^{d-1} \alpha \) proved in Theorem 11.

Remark 27. The upper bound (3.6) is general and does not impose any restrictions on the sampling distribution \( \pi \). However, the recovery error depends on the distribution. In particular, the entries that have a bigger probability of being sampled have a better recovery guarantee compared to the ones that are sampled with a smaller probability (notice that in Corollary 25 we assume \( \pi_\omega \geq \frac{1}{\mu N^d} \) \( \forall \omega \in [N]^d \)).

Corollary 28. Under the same assumptions as in Theorem 24 but assuming instead that \( \xi_t \) are independent sub-exponential random variables with sub-exponential norm \( K \) such that

\[
\max_{n=1,\ldots,n} \mathbb{E}[\exp\left(\frac{|\xi_n|}{K}\right)] \leq e,
\]

then

\[
\frac{1}{N^d} \| \hat{T}_M - T^\sharp \|_F^2 \leq C\mu(\alpha + \sigma K)R \sqrt{\frac{dN}{m}}.
\]

with probability greater than \( 1 - 2e^{\frac{-N}{\ln(N)}} \).

Although equation (3.8) proves linear dependence of sample complexity with \( N \), we are not aware of a polynomial-time method for estimating (or even attempting to estimate) the solution of (3.5). However, we later propose an algorithm that is inspired by max-qnorm constrained tensor completion and illustrate its efficiency numerically. Therefore, now we analyze the error bound of max-qnorm constrained tensor completion which is very similar to the error bound of (3.5). To this end, we define the set of low max-qnorm tensors as

\[
K^T_{\max}(\alpha, R) := \{ T \in \mathbb{R}^{N^d} : \| T \|_\infty \leq \alpha, \| T \|_{\max} \leq R \}.
\]

Note that, Theorem 11 ensures that a choice of \( R = \sqrt{r^d - d} \alpha \) is sufficient to include \( T^\sharp \) in \( K^T_{\max}(\alpha, R) \). The following theorem provides the bound on max-qnorm constrained LS estimation.
Theorem 29. Consider an order-$d$ tensor $T^\sharp \in \bigotimes_{i=1}^d \mathbb{R}^N$ with $\|T^\sharp\|_\infty \leq \alpha$ and $\|T^\sharp\|_{\max} \leq R$. Given a collection of noisy observations $\{Y_{\omega t}\}_{t=1}^m$ following the observation model (3.3) where the noise sequence $\xi_t$ are i.i.d. standard normal random variables, define

$$\hat{T}_{\max} = \arg\min_X \mathcal{L}_m(X, Y) \quad \text{subject to} \quad X \in K_{\max}^T(\alpha, R).$$

Then there exists a constant $C_d$ such that the minimizer $\hat{T}_M$ of (3.5) satisfies:

$$\|\hat{T}_{\max} - T^\sharp\|_F^2 = \sum_{\omega} \pi_\omega (\hat{T}_{\max}(\omega) - T^\sharp(\omega))^2 \leq C_d (\sigma(R + \alpha) + R\alpha) \sqrt{\frac{dN}{m}},$$

with probability greater than $1 - e^{-\frac{m\sigma^2}{2N}} - e^{-\frac{mN}{2}}$.

Corollary 30. Moreover, if we assume each entry of the tensor is sampled with some positive probability, $\pi_\omega \geq \frac{1}{\mu N^d} \forall \omega \in [N]^d$, then

$$\frac{1}{N^d} \|\hat{T}_{\max} - T^\sharp\|_F^2 \leq C_d \mu (\alpha + \sigma) R \sqrt{\frac{dN}{m}}.$$
Then with probability $1 - \frac{2}{m}$ over the choice of sample set,

$$|\hat{T}_M - T^\sharp\|_\Pi \leq C\left(\sigma \left(\sqrt{\log(m)^3 \frac{R^2 N}{m}} + \sqrt{\log(m)^3 \frac{\alpha^2}{m}}\right) + \log(m)^3 \frac{R^2 N}{m} + \log(m)^3 \frac{\alpha^2}{m}\right).$$  (3.12)

Theorem 32 proves decay rate $O(\frac{1}{m})$ when $\sigma = 0$. The proof is similar to the proof in [19, Theorem 6.2] which is based on a general theorem in [119] and is included in Section 3.7.2 for completeness.

### 3.4 Comparison to past results

As discussed in Section 3.1, there are several works that have considered max-norm for matrix completion [37, 41, 78, 110, 118]. However, the closest work to our result is [19], where the authors study max-norm constrained matrix completion, which is a special case of max-qnorm constrained tensor completion with $d = 2$. Here, we have generalized the framework of [19] to the problem of tensor completion. Although the main ideas of the proof are similar, the new ingredients include building a machinery for analyzing the max-qnorm and M-norm of low-rank tensors, as explained in Section 2.3. As expected, our result reduces to the one in the [19] when $d = 2$. More interestingly when $d > 2$, compared to the matrix error bound, the only values in upper bound (3.6) that change is the upper bound on the max-qnorm of the $d$-th order tensor (which is independent of $N$) and the order $d$, which changes the constants slightly.

As can be seen from Theorem 11, for a rank-$r$ tensor $T$ with $\|T\|_\infty \leq \alpha$, we have $\|T\|_M \leq (r\sqrt{r})^{d-1} \alpha$. Therefore, assuming $\alpha = O(1)$, to obtain an error bound of $\frac{1}{N^d} \|\hat{T}_M - T^\sharp\|_F \leq \epsilon$, it is sufficient to have $m > C\frac{(r\sqrt{r})^{d-1} dN}{\epsilon^2}$ samples. Similarly, using the max-qnorm, for an approximation error bounded by $\epsilon$, it is sufficient to obtain $m > C_d \frac{r^{d^2-d} dN}{\epsilon^2}$ samples. In contrast, the sufficient number of measurements with the best possible matricization is $m > C_d \frac{r N^{\frac{d}{2}}}{\epsilon^2}$, which is significantly larger for higher-order tensors.

Tensor completion using nuclear norm gives significantly inferior bounds as well. In particular, fixing $r$, and $d$, compared to latest results on tensor completion using nuclear norm [130], using M-norm lowers the theoretical sufficient number of measurements from $O(N^2 \frac{d}{4})$ to $O(dN)$.  

40
3.5 Information-theoretic lower bound

To prove a lower bound on the performance of (3.5), we employ a classical information-theoretic technique to establish a minimax lower bound for non-uniform sampling of random tensor completion on the max-qnorm ball. A similar strategy in the matrix case has been used in [19, 31]. In order to derive a lower bound on the performance of (3.5), we find a set of tensors in the set $K^T_M$ that are sufficiently far away from each other. Fano’s inequality implies that with the finite amount of information that we have, there is no method that can differentiate between all the elements of a set with too many elements and therefore any method will fail to recover at least one of them with a large probability. The main idea and the techniques closely follow [19, Section 6.2]; therefore we only explain the main steps we take to generalize this approach from matrices to tensors.

Similar to the upper bound case, we analyze a general restriction on the max-qnorm of the tensors instead of concentrating on low-rank tensors. Plugging the upper bound of the max-qnorm of low-rank tensors as a special case provides a lower bound for low-rank tensors as well.

Restating the set of bounded low M-norm tensors given by

$$K^T_M(\alpha, R) := \{ T \in \mathbb{R}^{N \times d} : \| T \|_{\infty} \leq \alpha, \| T \|_M \leq R \},$$

(3.13)

We will find a lower bound on the recovery error of any method that takes $\{ Y_{\omega_t} \}_{t=1}^m$ as input and outputs an estimate $\hat{T}$. This includes $\hat{T}_M$ that is obtained by

$$\hat{T}_M = \arg\min_{X} \mathcal{L}_m(X, Y) \quad \text{subject to} \quad X \in K^T_M(\alpha, R).$$

(3.14)

In particular, we show that when the sampling distribution satisfies

$$\frac{\mu}{N^d} \leq \min_{\omega} \pi_\omega \leq \max_{\omega} \pi_\omega \leq \frac{L}{N^d},$$

the M-norm constrained least squares estimator is rate optimal on $K^T_M(\alpha, R)$. The following theorem is proved in Section 3.7.3.

Theorem 33. Assume that the noise sequence $\xi_t$ are i.i.d. standard normal random variables and the sampling distribution $\Pi$ satisfies $\max_{\omega} \pi_\omega \leq \frac{L}{N^d}$. Fix $\alpha, R, N, $ and $m$ such that

$$R^2 \geq \frac{48 \alpha^2 K^2_G}{N},$$

(3.15)
where $K_G$ is the Grothendieck’s constant. Then the minimax recovery error is lower bounded by

$$
\inf_{\hat{T}_M} \sup_{T \in K_G^T(\alpha, R)} \frac{1}{N^d} \mathbb{E}\|\hat{T}_M - T\|_F^2 \geq \min \left\{ \frac{\alpha^2}{16}, \frac{\sigma R}{128\sqrt{2}K_G} \sqrt{\frac{N}{mL}} \right\}.
$$

(3.16)

Remark 34. Comparing the above theorem with (3.8), we observe that as long as

$$\frac{\sigma R}{128\sqrt{2}K_G} \sqrt{\frac{N}{mL}} < \frac{\alpha^2}{16},$$

M-norm constrained tensor completion is optimal in both $N$ and $R$.

### 3.6 Experiments

In this section, we present algorithms that we use to solve (3.9) and experiments concerning max-qnorm of specific classes of tensors and max-qnorm constrained tensor completion. As mentioned before most of the typical procedures such as calculating the nuclear norm or even calculating the rank of a tensor are NP-hard. The situation seems even more hopeless if we consider the results of [7] which connects 3-dimensional tensor completion with refuting 3-SATs, which has a long line of research behind it. In short, if we assume that either max-qnorm or M-norm is computable in polynomial time, a conjecture of [30] for refuting 3-SATs will be disproved. All these being said, the current chapter is the first work that considers max-qnorm for tensor completion and the preliminary results we show in this section are promising, outperforming matricization in every experiment we ran, and even outperforming the TenALS algorithm of [60].

In this section, we concentrate on (3.9) instead of (3.5) as we are not aware of any algorithm that can even attempt to solve (3.5) and simple heuristic algorithms we designed for (3.9) give promising results even though we do not know of any algorithm that is known to converge due to the non-convexity of the optimization problem (3.9).

There are two questions that need to be answered while solving (3.9). First, how to choose the max-qnorm bound $R$, and second, how to solve the least squares problem once $R$ is fixed. We address both these question in the next sections. We also run some experiments to estimate the tensor max-qnorm of some specific classes of tensors to get an idea of the dependency of the max-qnorm of a tensor on its size and rank. Finally, we compare the results of max-qnorm constrained tensor completion with TenALS and matricizing.
3.6.1 Algorithms for max-qnorm constrained least squares estimation

In this section, we introduce a few algorithms that attempt to solve (or approximate the solution of) (3.9). Defining \( f(V_1, \cdots, V_d, Y) := \hat{L}_m((V_1 \circ \cdots \circ V_d), Y) \), we minimize

\[
\min f(V_1, \cdots, V_d, Y) \text{ subject to } \max_i (\|V_i\|_{2,\infty}) \leq \sqrt[d]{R},
\]

where \( R \) is the max-qnorm constraint. In the definition of the max-qnorm, there is no limitation on the column size of the factors \( V_i \).

In the experiments we run in this section, we limit the factor sizes to \( N \times 2N \). Although this is an arbitrary value and we haven’t derived an error bound in the max-qnorm of tensors with this limitation, we believe (and our experiments also confirm) that this choice is large enough when \( r \ll N \). We defer the exact details of the effect of this choice on the error bounds to future work.

All the algorithms mentioned in this section are first-order methods that are scalable for higher dimensions and just require access to the first derivative of the loss function.

Projected gradient

The first algorithm is the projected gradient algorithm that for each factor, fixes all the other factors and takes a step according to the gradient of the loss function. Next, we project back all the factors on the set \( C := \{X | \|X\|_{2,\infty} \leq \sqrt[d]{R} \} \). To be precise, for each factor \( V_i \), define the matricization of \( T = V_1 \circ \cdots \circ V_d \) along the \( i \)-th dimension, \( T_i \), to be \( T_i = V_i \circ R_i \) (notice that \( R_i \) is obtained from the Kronecker product of the factors \( V_1, \cdots, V_{i-1}, V_{i+1}, \cdots, V_d \)) and define \( f_i(X) := \hat{L}((X \circ R_i), Y_i) \), where \( Y_i \) is the matricization of \( Y \) along its \( i \)-th dimension. Fixing a step size \( \gamma \), the algorithm updates all the factors in parallel via

\[
[V_i] \leftarrow P_C([V_i - \gamma \nabla (f_i)R_i]).
\]

where \( P_C \) simply projects the factor onto the set of matrices with \( \ell_2 \)-infinity norm less than \( \sqrt[d]{R} \). This projection looks at each row of the matrix and if the norm of a row is bigger than \( \sqrt[d]{R} \), it scales that row back down to \( \sqrt[d]{R} \) and leaves other rows unchanged.

This algorithm is a well-known algorithm with a lot of efficient implementations and modifications. Furthermore, using Armijo line search rule to guarantee sufficient decrease of the loss function, it is guaranteed to find a stationary point of (3.17).
Projected quasi-Newton

Stacking all the factors in a matrix $X$,

$$X = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_d \end{bmatrix}$$

and defining $f(X) := \hat{\mathcal{L}}_m((V_1 \circ \cdots \circ V_d), Y)$, this algorithm uses BFGS quasi-Newton method to form a quadratic approximation to the function at the current estimate and then uses spectral projected gradient (SPG) method to minimize this quadratic function, constrained to $X \in C$. We use the implementation of [108] which uses limited memory BFGS and uses a Barzilai-Borwein scaling of the gradient, and use a non-monotone Armijo line search along the feasible direction to find the next iterate in the SPG step.

Stochastic gradient

The loss function

$$\hat{\mathcal{L}}_m(X, Y) = \frac{1}{m} \sum_{t=1}^{m} (X_{\omega_t} - Y_{\omega_t})^2,$$

is decomposable into the sum of $m$ loss functions, each concerning one observed entry. This makes it very easy to use stochastic gradient methods that at each iteration take one or more of the entries, and find the feasible direction according to this subset of observations. In particular, at each iteration, we take a subset of the $m$ entries, $S \subset \Omega$, and minimize the loss function

$$\hat{\mathcal{L}}_S(X, Y) = \frac{1}{|S|} \sum_{\omega_t \in S} (X_{\omega_t} - Y_{\omega_t})^2.$$

This approach is useful when we are dealing with very high dimension sizes and accessing all the measurements at once is not an option or very costly. There has been plenty of research on the efficiency of this method and its recovery guarantees [74]. The projection part is done as before with the advantage that we just need to project the rows in the factors that correspond to the subset of entries chosen in this iteration and not necessarily all of them which saves time in large applications.
3.6.2 Experiment on max-qnorm of tensors

In this section, we run an experiment to find the dependency of the max-qnorm on its rank and size. To this end, we consider tensors whose low-rank factors come from Gaussian distribution. We also mention the results for tensors coming from random sign factors. Although in comparison with other ways of generating low-rank tensors, these specific classes of tensors do not necessarily represent tensors with highest possible max-qnorm, they can be helpful in giving us an idea of how does the max-qnorm scale with size and rank.

In order to estimate the max-qnorm of a tensor, we employ a max-qnorm constrained tensor completion while accessing all the entries of the tensor and find the smallest constraint that successfully recovers the tensor. Using the bisection method to estimate the max-qnorm of the tensor, starting from a lower bound and an upper bound for the max-qnorm of the tensor we first check if the tensor can be recovered with max-qnorm bound equal to the average of the upper bound and the lower bound. Next, we increase the lower bound if the max-qnorm constraint is too small for full recovery and reduce the upper bound if the max-qnorm bound is large enough. Algorithm 1 explains this algorithm in more details. For small ranks, we get to the approximate max-qnorm very fast, for example, in less than \(\log(\text{rank}^{\text{dim}-1}) + k\) iterations we can estimate the max-qnorm with an error less than \(2^{-k}\). Moreover, we assume successful recovery is achieved once the root means squared error (RMSE) is less than a small predefined value. This algorithm becomes faster after the first iteration, as we use the factors found in the previous iteration as a good initial point for the next iteration.

\begin{algorithm}
\caption{Estimating max-qnorm of a tensor \(T\)}
\begin{algorithmic}[1]
\State \textbf{Input} \(T, \Omega = [N_1] \times [N_2] \times \cdots \times [N_d], \text{lowerbound}, \text{upperbound}\)
\State \textbf{Output} \(\|T\|_{\max}\) with an estimation error of at most \(0.01\)
\For {iteration = 1 to \(\lceil \log_2(\text{upperbound} - \text{lowerbound}) \rceil + 6\)}
\State \(\hat{T} = \arg\min_{X \in \mathbb{R}^{N_1} \times \cdots \times \mathbb{R}^{N_d}} \|X_{\Omega} - T_{\Omega}\|_F^2 \text{ subject to } \|X\|_{\max} \leq \frac{\text{lowerbound} + \text{upperbound}}{2}\)
\EndFor
\State \(\text{RMSE} = \frac{\|X - T\|_F}{\sqrt{\prod_{i=1}^d N_i}}\)
\If {\(\text{RMSE} \leq 1 \times 10^{-3}\)}
\State \(\text{upperbound} = \frac{\text{lowerbound} + \text{upperbound}}{2}\)
\Else
\State \(\text{lowerbound} = \frac{\text{lowerbound} + \text{upperbound}}{2}\)
\EndIf
\State \textbf{return} \(\frac{\text{lowerbound} + \text{upperbound}}{2}\)
\end{algorithmic}
\end{algorithm}
Figure 3.1 shows the results for both 3 and 4-dimensional tensors when low-rank factors are drawn either from Gaussian distribution 3.1.a, and 3.1.b or from Bernoulli distribution 3.1.c. In both cases, we have considered \( r \in \{1, 2, \cdots, 10\} \), and \( N \in \{5, 10, 15, 20\} \) for the 3-dimensional case and \( N \in \{5, 10\} \) for the 4-dimensional case. In all the cases the average max-qnorm is similar for different value of \( N \) when rank and order is fixed. These results confirm that max-qnorm of a tensor just depends on its rank and its order and is independent of the size of the tensor.

The results are averaged over 15 experiments. Because of the linear effect of infinity norm of a tensor on its max-qnorm, we rescale all tensors to have \( \|T\|_\infty = 1 \) before estimating their max-qnorm. Comparing Figure 3.1.a and 3.1.b shows that the max-qnorm is around \( \sqrt{r} \) for \( d = 3 \), and \( r \) for \( d = 4 \). These values are attained exactly when the factors are drawn from Bernoulli random variables. Moreover, the dependence is constant when \( d = 2 \). This suggests a multiplicative increase of \( \sqrt{r} \) when the order is increased by one. However, whether or not the actual bound in general case is \( O(\sqrt{r^{d-1}}) \) is an interesting open question.

![Figure 3.1](image)

\( \text{(a) 3-dimensional, Gaussian factors} \quad \text{(b) 4-dimensional, Gaussian factors} \quad \text{(c) 3 and 4-dimensional, sign factors} \)

**Figure 3.1:** Log-log plot of average max-qnorm of 3 and 4-dimensional low-rank tensors obtained by Algorithm 1, for various rank and sizes, averaged over 15 draws for each rank and size.

### 3.6.3 Automatic algorithm for choosing optimal max-qnorm bound \( R \)

As explained, before, other than designing algorithms for solving constrained max-qnorm minimization, we need to design a procedure to find good bounds on the max-qnorm. The theoretical bounds found in this thesis might not be tight and even if they are proven to be tight, such theoretical upper bounds usually capture the worst-case scenarios which might not be optimal for a
general problem. This issue is very important as the result of the tensor completion is very dependent on choosing the right upper bound. Other than this in many practical applications, we don’t have access to the actual rank of the underlying tensor which shows the importance of finding the upper bound automatically and not as an input to the optimization problem.

Algorithm 2 Tensor completion, with cross-validation

1: **Input** possibly noisy measurements \( Y_\Omega = T_\Omega + \sigma N(0, 1) \), observed entries \( \Omega \), **lowerbound**, **upperbound**
2: **Output** \( \hat{T} \)
3: Divide the observations into \( \Omega_{\text{train}} \) and \( \Omega_{\text{validate}} \)
4: for iteration = 1 to \( \lceil \log_2(\text{upperbound} - \text{lowerbound}) \rceil + 6 \) do
5: for \( \text{iter check} = 0 \) to 4 do
6: \( \text{bound}(\text{iter check}) = \frac{\text{iter check}}{4} \ast \text{upperbound} + \frac{4 - \text{iter check}}{4} \ast \text{upperbounds} \)
7: \( \hat{T}(\text{iter check}) = \arg\min_{X \in \mathbb{R}^{N_1 \times \cdots \times N_d}} \|X_\Omega - Y_\Omega\|_F^2 \) s.t. \( \|X\|_{\text{max}} \leq \text{bound}(\text{iter check}) \)
8: \( \text{RMSE}(\text{iter check}) = \frac{\|\hat{T}(\text{iter check})_{\text{validate}} - (Y)_{\text{validate}}\|_F}{\sqrt{|\Omega_{\text{validate}}|}} \)
9: end for
10: \( \text{min index} = \arg\min \text{RMSE} \)
11: \( \text{lowerbound} = \text{bound}(\text{min index} - 1) \)
12: \( \text{upperbound} = \text{bound}(\text{min index} + 1) \)
13: \( \hat{T} = \hat{T}(\text{min index}) \)
14: end for
15: return \( \hat{T} \)

The first approach is modifying algorithm 1 to find a good upper bound. There are two complications with generalizing this approach. First, in tensor completion, we don’t have access to the full tensor and we have to estimate the recovery error on the indices we have not observed as otherwise choosing a large upper bound can result in over-training, i.e, fitting the observations exactly and losing the low-rank (and low max-qnorm) structure of the tensor. The other complication, is that in algorithm 1 we use \( \text{RMSE} < 1e - 3 \) as an approximation for full recovery. Doing such a thing in a noisy problem is not possible and using the RMSE of the noise is an independent problem that depends on the noise-type and is usually not optimal in a general case. Other than this, even if we know the RMSE of the noise, we won’t be bale to decide if a bound is larger than the optimal max-qnorm bound or smaller due to the possibility of over-fitting.

To address these two issues, we propose algorithm 2 for tensor completion that uses cross-validation for estimating the RMSE and uses a five-point search for the optimal max-qnorm bound.
Figure 3.2: All the possible situations in the five-point search algorithm. The leftmost and the rightmost red dots are the previous lower bound and upper bound on the max-qnorm $R$, respectively. The green intervals show the new lower bound and upper bound based on the RMSE.

As explained above we need to find a way to estimate the true RMSE to avoid over-fitting and also as a measure of how good the approximation is. In order to do this, before starting the optimization process, we randomly divide the observed samples, $\Omega$, into two sets: $\Omega_{\text{train}}$ and $\Omega_{\text{validate}}$ and we solve each max-qnorm constrained sub-problem just using the samples in $\Omega_{\text{train}}$ which contains 80% of the total samples and use the reserved samples $\Omega_{\text{validate}}$ (20% of total samples) to estimate the RMSE. We are aware of more sophisticated cross-validation (such as k-fold cross validation) algorithms that try to remove the bias in the samples as much as possible. However, considering that each max-qnorm constrained tensor completion problem is expensive, our numerical results show that the simple cross-validation used in algorithm 2 is good enough for finding an approximately optimal upper bound.

Now we explain algorithm 2 more thoroughly. To find the optimal upper bound we input a large enough upper bound and a small enough lower bound that are bigger and smaller than the optimal max-qnorm bound and iteratively refine these bounds until the two bounds become close to each other. To determine the next upper and lower bounds, checking the middle point is not enough because unlike algorithm 1 the RMSE is not going to be zero for bounds bigger than the optimal bound. Therefore, other than the lower bound and the upper bound we calculate the RMSE using three points in the interval between the lower bound and the upper bound as well and consider the
best bound among those three points to be the center of the new upper bound and the new lower bound. The derivation of this approach is in Algorithm 2. A hand-wavy reasoning behind this is assuming that there is an optimal bound $R^\#$ that gives the best RMSE, any bound larger than $R^\#$ results in over-training and any bound smaller than $R^\#$ results in under training. This over-training or under-training becomes more severe when the bound is further from the optimal max-qnorm bound and therefore the problem becomes finding the minimum of a function where we don’t know the derivative of the function. Deriving a provably exact algorithm to find the optimal upper bound in such a situation is an interesting question that we postpone to future work. Assuming all the justification above is roughly correct, Figure 3.2 shows that the five-point method explained in Algorithm 2 finds the optimal upper bound. Considering all the assumptions above to be true, the figure plots the RMSE against the max-qnorm upper bound $R$, and shows all the possible situations that the five points (red stars) can have in such a curve. The green lines show the new interval bounded by the new lower and upper bound. Notice that the optimal value always stays in the middle of these two bounds. Our numerical experiments in the next section show that although this justification is not mathematically rigorous, the final outcome is very promising. However, providing a better explanation or designing a more rigorous method is an interesting future work.

3.6.4 Numerical results of low-rank tensor completion

In this section, we present the results of max-qnorm constrained tensor completion, and compare it with those of matricization and alternating least squares (tenALS) [60]. As explained at the beginning of Section 3.6, we pick the low-rank factors to have size $N \times 2N$. Although the choice of $2N$ is arbitrary, we believe it is large enough for small ranks and does not result in large errors. This also has the additional benefit of not requiring the knowledge of the exact rank of the tensors. As explained above, we just assume the knowledge of an upper bound and a lower bound on the max-qnorm of the tensor and use cross-validation to the find the optimal max-qnorm bound. Obviously, the algorithm becomes faster if these bounds are closer to the actual max-qnorm of the tensor.

Figures 3.3 and 3.4 show the results of completing a $50 \times 50 \times 50$ tensor with $m$ random samples of it taken uniformly at random (without and with 10-dB noise respectively). The results are averaged over 15 experiments, with various ranks ranging from 3 to 30 and sampling rate ranging from 0.01 to 0.1. The row $i$ and column $j$ of each subplot shows the average squared relative recovery error $\|T - \hat{T}\|_F^2 / \|T\|_F^2$ for a random tensor with rank $i$, where different columns represent different number of samples observed according to (3.3). We rescale the true tensors to have infinity norm equal to 1 before adding the noise and the RMSE of the noise is around 0.1. As expected in all
experiments we get better results with higher number of measurements $m$, and smaller rank $r$. The matricized results are obtained by applying the Fixed Point Continuation with Approximate SVD algorithm (FPCA), introduced in [85] to flattened $900 \times 30$ matrices (This algorithm results in the best outcomes in our experiments, compared to other noisy matrix completion algorithms). Next, for a more fair comparison we have included the results of tensor completion using alternating least squares (ALS) where the code is provided online [60]. The two plots in the second row show the max-qnorm constrained (MNC) results in two scenarios. One with exact low-rank factors and second, as the theory suggests, with factors with larger number of columns. Notice that the exact max-qnorm formulation does not put any limitation on the number of columns of the factors and we chose $2N$ to balance the computational cost and the accuracy of the computed max-qnorms. The results unanimously show the advantage of using $N \times 2N$ factors instead of $N \times r$ factors. Although we are dealing with higher-dimensional factors and this makes the algorithm a little slower,
using larger factors has the additional benefit of not getting stuck in local minima compared to exact low-rank factors. The ALS algorithm, show some discrepancies in the results, i.e., there are cases that the average error increases with smaller rank or higher sampling rate. We believe this is because of high non-convexity in this algorithm. We expected to see similar discrepancies in the max-qnorm results as well but at least for these dimensions and setup, our algorithm seems to be able to run away from local minima which is surprising. We can see the importance of this issue when we compare the results of using $N \times r$ factors and $N \times 2N$ factors. It is worth mentioning here that the ALS algorithm uses a number of initial vectors to use in the initialization step and larger values give a better estimate at the expense of longer processing time. We use 50 vectors so that the time spent by the algorithms is comparable to each other. However, the result can be slightly improved if we use more initial vectors.
The results of matricization is always inferior to those of tensor completion with both ALS and MNC which is expected, especially as here the tensors have an odd order. The difference between matrix completion and MNC with $N \times 2N$ factors is significant. For example, when $m/N^2 = 0.1$, Max-qnorm constrained TC (MNC) recovers all the tensor with rank less 10, whereas matrix completion starts to fails for ranks bigger than 1.

In Figures 3.5 and 3.6 we have included the results for completing a 4-dimensional $20 \times 20 \times 20 \times 20$ tensor. The results are similar to the 3-dimensional case and therefore we are not including the exact low-rank factor results which are inferior to using $N \times 2N$ factors. The available online code of ALS algorithm is not suitable for 4-dimensional cases as well. Similar to the 3-dimensional case MNC TC always beats matrix completion. Notice that because of the even dimension, the matricization can be done in a more balanced way and therefore, results of matrix completion are a little better compared to the 3-dimensional case but still we can see the huge advantage of using tensor completion instead of matricizing.

It is worth mentioning that although the algorithms take reasonable time for small dimensions used here (less than 15 seconds in each case), it is still slower than matricizing. In other words using tensors enjoys better sample complexity and can be used to save very large data but it comes at the cost of computational complexity. Deriving scalable and efficient algorithms is an interesting and necessary future work for practical applications.

In Figure 3.7 we do a thorough investigation of the recovery error of our algorithm with respect to $r$, $N$, and the number of measurements $m$ when we use max-qnorm constrained 1-bit TC (3.9). In the figures, dark blue shows full recovery. In the left image in Figure 3.7 we fix $m = 1500$ and show the average squared error for varying values of $N$ and $r$. As expected, we observe an increase in the average error for larger values of $N$ and $r$. In the right image of figure 3.7 we fix $r = 5$ and show the results for varying $N$ and $m$. The low-rank tensors are generated by random sign factors. In Figure 3.1 we showed that for such tensors the max-qnorm is equal to $\sqrt{r}$ and therefore, the error bound (in the noiseless case) is $O(\log(m)^3 rN/m)$. By observing the rows of the left image, we can approximately see the dependence on $r$ is linear. However, for $m$ and $N$ the dependence of of error is much more complicated than what we expect theoretically. We think this is due to a few factors. First, our algorithm is not exactly solving the max-qnorm constrained problem due to the limited number of columns in the factors and inexact infinity-norm projections. Second and more important, the theoretical bound is a worst-case error bound that is true with high probability over
Figure 3.5: 4-dimensions, no noise Average relative recovery error \( \frac{\| \mathbf{T}_{\text{recovered}} - \mathbf{T}^\sharp \|_F^2}{\| \mathbf{T}^\sharp \|_F^2} \) for 4-dimensional tensor \( \mathbf{T} \in \mathbb{R}^{20 \times 20 \times 20 \times 20} \) and different ranks and samples. The plot on the left shows the results for the 400 \( \times \) 400 matricized case.

Figure 3.6: 4-dimensions, 10-dB noise Average relative recovery error \( \frac{\| \mathbf{T}_{\text{recovered}} - \mathbf{T}^\sharp \|_F^2}{\| \mathbf{T}^\sharp \|_F^2} \) for 4-dimensional tensor \( \mathbf{T} \in \mathbb{R}^{20 \times 20 \times 20 \times 20} \) and different ranks and samples. The plot on the left shows the results for the 400 \( \times \) 400 matricized case.

the sampling scheme. Whereas, the figures show the average error.
Figure 3.7: 3-dimensions, no noise Average relative recovery error \( \left\| \frac{T_{\text{recovered}} - T^z}{T^z} \right\|_F \) for e-dimensional tensors with fixed number of measurements in the left and fixed rank in the right.

3.7 Proof

3.7.1 Proof of Theorem 24

In this section, we prove Theorem 24. We make use of Lemma 9, Lemma 10 and Theorem 11 repeatedly. However, some parts of the other calculations are simple manipulations of the proof in [19, Section 6].

For ease of notation define \( \hat{T} := \hat{T}_M \). Notice that \( T^z \) is feasible for (3.5) and therefore,

\[
\frac{1}{m} \sum_{t=1}^{m} (\hat{T}_{\omega_t} - Y_{\omega_t})^2 \leq \frac{1}{m} \sum_{t=1}^{m} (T^z_{\omega_t} - Y_{\omega_t})^2.
\]

Plugging in \( Y_{\omega_t} = T^z(\omega_t) + \sigma \xi_t \) and defining \( \Delta = \hat{T} - T^z \in K_M^T(2\alpha, 2R) \) we get

\[
\frac{1}{m} \sum_{t=1}^{m} \Delta(\omega_t)^2 \leq \frac{2\sigma}{m} \sum_{t=1}^{m} \xi_t \Delta(\omega_t). \tag{3.19}
\]

The proof is based on a lower bound on the left-hand side of (3.19) and an upper bound on its right-hand side.
Upper bound on right-hand side of (3.19)

First, we bound \( \hat{R}_m(\alpha, R) := \sup_{\Delta \in K^T_{\alpha, R}} |\frac{1}{m} \sum_{t=1}^{m} \xi_t \Delta(\omega_t)| \) where \( \xi_t \) is a sequence of \( \mathcal{N}(0,1) \) random variables. With probability at least \( 1 - \delta \) over \( \xi = \{ \xi_t \} \), we can relate this value to a Gaussian maxima as follows [98, Theorem 4.7]:

\[
\sup_{\Delta \in K^T_{\alpha, R}} |\frac{1}{m} \sum_{t=1}^{m} \xi_t \Delta(\omega_t)| \leq R \mathbb{E}_{\xi} \left[ \sup_{\Delta \in K^T_{\alpha, R}} \left| \frac{1}{m} \sum_{t=1}^{m} \xi_t \Delta(\omega_t) \right| \right] + \pi \alpha \sqrt{\frac{\log \left( \frac{1}{\delta} \right)}{2m}}.
\]

where \( T \in T_{\pm} \) is the set of rank-one sign tensors with \( |T_{\pm}| < 2^{Nd} \). Notice that in the second inequality we use \( \mathbb{B}_M(1) = \text{conv}(T_{\pm}) \) (Lemma 9) and remove the convex hull from the supremum.

Since for each \( T \), \( \sum_{t=1}^{m} \xi_t T(\omega_t) \) is a Gaussian with mean zero and variance \( m \), the expected maxima is bounded by \( \sqrt{2m \log(|T_{\pm}|)} \). Gathering all the above information, we end up with the following upper bound with probability larger than \( 1 - \delta \):

\[
\sup_{T \in K^T_{\alpha, R}} \left| \frac{1}{m} \sum_{t=1}^{m} \xi_t T(\omega_t) \right| \leq R \mathbb{E}_{\xi} \left[ \sup_{T \in T_{\pm}} \left| \frac{1}{m} \sum_{t=1}^{m} \xi_t T(\omega_t) \right| \right] + \pi \alpha \sqrt{\frac{\log \left( \frac{1}{\delta} \right)}{2m}}.
\]

Choosing \( \delta = e^{-\frac{Nd}{2}} \), we get that with probability at least \( 1 - e^{-\frac{Nd}{2}} \)

\[
\sup_{T \in K^T_{\alpha, R}} \left| \frac{1}{m} \sum_{t=1}^{m} \xi_t T(\omega_t) \right| \leq 2(R + \alpha) \sqrt{\frac{N \delta}{m}}.
\]

Lower bound on left-hand side of (3.19)

In this section, we prove that with high probability, \( \frac{1}{m} \sum_{t=1}^{m} \Delta(\omega_t)^2 \) does not deviate much from its expectation \( \| \Delta \|_{\Pi}^2 \). For ease of notation, define \( T_S = (T(\omega_1), T(\omega_2), \cdots, T(\omega_m)) \) to be the set of chosen samples drawn from \( \Pi \) where

\[
\| T \|_{\Pi}^2 = \frac{1}{m} \mathbb{E}_{S \sim \Pi} \| T_S \|_2^2 = \sum_{\omega} \pi_\omega T(\omega)^2.
\]
We prove that with high probability over the samples,

\[ \frac{1}{m} \|TS\|_2^2 \geq \|T\|_\Pi^2 - C\beta \sqrt{\frac{Nd}{m}}, \quad (3.22) \]

holds uniformly for all tensors \( T \in K_M^T(1, \beta) \).

**Lemma 35.** Defining \( \delta(S) := \sup_{T \in K_M^T(1, \beta)} \frac{1}{m} \|TS\|_2^2 - \|T\|_\Pi^2 \), and assuming \( N, d > 2 \) and \( m \leq Nd \), there exists a constant \( C > 20 \) such that

\[ \Pr(\delta(S) > C\beta \sqrt{\frac{Nd}{m}}) \leq e^{-\frac{N}{m(N)}}. \]

To prove this lemma, we show that we can bound the \( t \)-th moment of \( \delta(S) \) as

\[ \mathbb{E}_{S \sim \Pi}[\delta(S)^t] \leq \left( \frac{8\beta \sqrt{Nd + t \ln(m)}}{\sqrt{m}} \right)^t. \quad (3.23) \]

Before stating the proof of this bound, we show how we can use it to prove Lemma 35 by using Markov’s inequality.

\[ \Pr(\delta(S) > C\beta \sqrt{\frac{Nd}{m}}) = \Pr((\delta(S))^t > (C\beta \sqrt{\frac{Nd}{m}})^t) \leq \frac{\mathbb{E}_{S \sim \Pi}[\delta(S)^t]}{(C\beta \sqrt{\frac{Nd}{m}})^t}. \quad (3.24) \]

Using (3.23) and simplifying we get

\[ \Pr(\delta(S) > C\beta \sqrt{\frac{Nd}{m}}) \leq \left( \frac{4\sqrt{Nd + t\ln(m)}}{C\sqrt{Nd}} \right)^t. \]

Taking \( t = \frac{Nd}{\ln(m)} \) and for \( C > 12 \),

\[ \Pr(\delta(S) > C\beta \sqrt{\frac{Nd}{m}}) \leq e^{-\frac{Nd}{m}} \leq e^{-\frac{N}{m(N)}}. \]

Now we prove (3.23) by using some basic techniques of probability in Banach space, including symmetrization and contraction inequality [31, 77]. Regarding the tensor \( T \in \bigotimes_{i=1}^d \mathbb{R}^N \) as a function from \([N]^d \to R\), we define \( f_T(\omega_1, \omega_2, \cdots, \omega_d) := T(\omega_1, \omega_2, \cdots, \omega_d)^2 \). We are interested in

56
bounding $\delta(S) := \sup_{f_T : T \in K_T^\beta}(1, \beta) |\frac{1}{m} \sum_{i=1}^{m} f_T(\omega_i) − \mathbb{E}(f_T(\omega_i))|$. A standard symmetrization argument and using contraction principle yields

$$\mathbb{E}_{S^{-}\Pi}[\delta(S)^{\prime}] \leq \mathbb{E}_{S^{-}\Pi}\{2\mathbb{E}_{\varepsilon}\sup_{T \in K_T^\beta}(1, \beta) \frac{1}{m} \sum_{i=1}^{m} \varepsilon_i T(\omega_i)^2\}^{\prime} \leq \mathbb{E}_{S^{-}\Pi}\{4\mathbb{E}_{\varepsilon}\sup_{T \in K_T^\beta}(1, \beta) \frac{1}{m} \sum_{i=1}^{m} \varepsilon_i T(\omega_i)\}^{\prime},$$

where $\varepsilon_i$’s are Rademacher random variables. Notice that if $T_M \leq \beta$, then $T \in \beta \text{conv}(T_{\pm})$ and therefore

$$\mathbb{E}_{S^{-}\Pi}[\delta(S)^{\prime}] \leq \mathbb{E}_{S^{-}\Pi}\{4\beta\mathbb{E}_{\varepsilon}\sup_{T \in T_{\pm}} \frac{1}{m} \sum_{i=1}^{m} \varepsilon_i T(\omega_i)\}^{\prime} = \beta^{\prime}\mathbb{E}_{S^{-}\Pi}\{\mathbb{E}_{\varepsilon}\sup_{T \in T_{\pm}} \frac{4}{m} \sum_{i=1}^{m} \varepsilon_i\}^{\prime}.$$

To bound the right-hand side above notice that for any $\alpha > 0$ [114, Theorem 36]

$$\mathbb{P}(\frac{4}{m} \sum_{i=1}^{m} \varepsilon_i \geq \frac{\alpha}{\sqrt{m}}) = \mathbb{P}\left(\text{Binom}(m, \frac{1}{2}) \geq \frac{m}{2} + \frac{\alpha \sqrt{m}}{8}\right) \leq e^{-\frac{\alpha^2}{16}}.$$

Taking a union bound over $T_{\pm}$, where $|T_{\pm}| \leq 2^{Nd}$ we get

$$\mathbb{P}(\sup_{T \in T_{\pm}} (\frac{4}{m} \sum_{i=1}^{m} \varepsilon_i T(\omega_i))) \geq \frac{\alpha}{\sqrt{m}} \leq 2^{Nd+1}e^{-\frac{\alpha^2}{16}}.$$

Combining the above result and using Jensen’s inequality, when $i > 1$

$$\beta^{\prime}\mathbb{E}_{S^{-}\Pi}\{\mathbb{E}_{\varepsilon}\sup_{T \in T_{\pm}} \frac{4}{m} \sum_{i=1}^{m} \varepsilon_i\}^{\prime} \leq \beta^{\prime}\mathbb{E}_{S^{-}\Pi}\{\mathbb{E}_{\varepsilon}\sup_{T \in T_{\pm}} \frac{4}{m} \sum_{i=1}^{m} \varepsilon_i\}^{\prime} \leq \beta^{\prime} \left(\frac{\alpha}{\sqrt{m}}\right)^{\prime} + 4^{\prime} 2^{Nd+1}e^{-\frac{\alpha^2}{16}}$$

Choosing $\alpha = \sqrt{16 \ln(4 \times 2^{Nd}1) + 4^{\prime} \ln(m)}$ and simplifying proves (3.23).  

**Gathering the results of (3.19), (3.21), and (4.49)**

Now we combine the upper and lower bounds in the last two sections to prove Theorem 24.

On one hand from (3.21), as $\Delta \in K_T(2\alpha, 2R)$, we get

$$\frac{1}{m} \|\Delta_S\|^2 \leq 8\sigma(R + \alpha)\sqrt{\frac{Nd}{m}},$$

57
with probability greater than $1 - e^{-Nd^2/2}$. On the other hand, using Lemma 35 and rescaling, we get
\[ \|\Delta\|^2_{\Pi} \leq \frac{1}{m} \|\Delta_S\|^2_2 + CR\alpha \sqrt{\frac{Nd}{m}}, \]
with probability greater than $1 - e^{-N \ln(N)}$. The above two inequalities finishes the proof of Theorem 24.

**Remark 36.** There are only two differences in the proof of Theorem 24 and Theorem 29. First is an extra constant, $c_1 c_2^d$, which shows up in Rademacher complexity of unit max-qnorm ball which changes the constant $C$ in Theorem 24 to $Cc_1 c_2^d$ in Theorem 29 and second is the max-qnorm of the error tensor $\Delta = \hat{T} - T^\sharp$ (refer to equation (3.19)) which belongs to $K_{\max}^T(2\alpha, 2^{d-1}R)$ instead of $K_{\max}^T(2\alpha, 2R)$.

### 3.7.2 Proof of Theorem 32

The proof is based on [119] which establishes excess risk bounds of risk minimization with smooth loss functions and a hypothesis class with bounded empirical Rademacher complexity. To that end, remember that $Y(\omega_t) = T^\sharp(\omega_t) + \sigma \xi_t$. We define the loss function
\[ \mathcal{L}(X) := \mathbb{E}_{\omega_t \sim \Pi, \xi_t \sim \mathcal{N}(0,1)} (X(\omega_t) - Y(\omega_t))^2 = \|X - T^\sharp\|^2_{\Pi} + \sigma^2 \]
and its corresponding empirical loss function
\[ \hat{\mathcal{L}}_m(X, Y) := \frac{1}{m} \sum_{t=1}^{m} (X_{\omega_t} - Y_{\omega_t})^2, \]
where $\{Y_{\omega_t}\}_{t=1}^{m}$ is i.i.d. noisy samples generated according to $\Pi$. Our goal is to bound $\|\hat{T}_M - T^\sharp\|^2_{\Pi}$ where $\hat{T}_M := \arg\min_{X \in K_{\max}^T(\alpha, R)} \hat{\mathcal{L}}_m(X, Y)$. First step is proving that with high probability the noise is bounded. In particular, with probability greater than $1 - \frac{1}{m}$ [19, equation (6.16)],
\[ \max_{1 \leq t \leq m} |\xi_t| < c \sqrt{\log(m)}. \]
Using this and \cite[Theorem 1]{119}, we get than with probability greater than $1 - \delta$

$$ \mathcal{L}(\hat{T}_M) - \min_{X \in K_T(\alpha, R)} \mathcal{L}(X) \leq \left[ C \left( \sigma \left( \log(m)^3 R_m(K_T(\alpha, R)) + \frac{b \log(\frac{1}{\delta})}{m} \right) + \log(m)^3 R_m(K_T(\alpha, R))^2 + \frac{b \log(\frac{1}{\delta})}{m} \right) \right], $$

(3.25)

where $b = 5\alpha^2 + 4\alpha\sigma \sqrt{\log(m)}$, and $R_m(K_T(\alpha, R))$ is bounded by $6\sqrt{\frac{dN}{m}}$ (Lemma 10). Moreover, note that $\min_{X \in K_T(\alpha, R)} \mathcal{L}(X) = \mathcal{L}(T^\#) = \sigma^2$ and $\mathcal{L}(\hat{T}_M) = \|\hat{T}_M - T^\#\|_F^2 + \sigma^2$. The proof is finished by plugging these two into the left-hand side of (3.25).

### 3.7.3 Proof of Theorem 33

**Packing set construction**

In this section, we construct a packing for the set $K^T_M(\alpha, R)$.

**Lemma 37.** Let $r = \lfloor \frac{R}{\alpha K_G} \rfloor^2$ and let $K^T_M(\alpha, R)$ be defined as in (3.13) and let $\gamma \leq 1$ be such that $\frac{r}{\gamma^2}$ is an integer and suppose $\frac{r}{\gamma^2} \leq N$. Then there exists a set $\chi^T \subset K^T_M(\alpha, R)$ with

$$ |\chi^T| \geq \exp(\frac{rN}{16\gamma^2}) $$

such that

- For $T \in \chi^T$, $|T(\omega)| = \alpha \gamma$ for $\omega \in [N]^d$.
- For any $T^{(i)}, T^{(j)} \in \chi^T$, $T^{(i)} \neq T^{(j)}$

$$ \|T^{(i)} - T^{(j)}\|_F^2 \geq \frac{\alpha^2 \gamma^2 N^d}{2}. $$

**Proof:** This packing is a tensor version of the packing set generated in \cite[Lemma 3]{31} with similar properties and our construction is based on the packing set generated there for low-rank matrices with bounded entries. In particular, we know that there exists a set $\chi \subset \{M \in \mathbb{R}^{N \times N} : \|M\|_\infty \leq \alpha, \text{ rank}(M) = r\}$ with $|\chi| \geq \exp(\frac{rN}{16\gamma^2})$ and for any $M^{(i)}, M^{(j)} \in \chi$, $\|M^{(i)} - M^{(j)}\|_F^2 \geq \frac{\alpha^2 \gamma^2 N^d}{2}$ when $i \neq j.$

59
Take any $M^{(k)} \in \chi$. $M^{(k)}$ is a rank-$r$ matrix with $\|M^{(k)}\|_\infty = \alpha \gamma \leq \alpha$ and therefore $\|M^{(k)}\|_\text{max} \leq \sqrt{r} \alpha$. By using (2.9), there exists a nuclear decomposition of $M^{(k)}$ with bounded singular vectors $M^{(k)} = \sum_i \sigma_i u_i \circ v_i$, $\|u_i\|_\infty, \|v_i\|_\infty \leq 1$, such that $\sum_{i=1} |\sigma_i| \leq K_G \sqrt{r} \alpha$. Define $T^{(k)} = \sum_i \sigma_i u_i \circ v_i \circ \mathbf{1}_i \cdots \circ \mathbf{1}$ where $\mathbf{1}_i \in \mathbb{R}^N$ is the vector of all ones. Notice that $\|u_i\|_\infty, \|v_i\|_\infty, \|\mathbf{1}_i\|_\infty \leq 1$ and therefore by definition, $\|T^{(k)}\|_M \leq K_G \sqrt{r} \alpha \leq R$ by Lemma 9. The tensor is basically generated by stacking the matrix $M^{(k)}$ along all the other $d - 2$ dimensions and therefore $|T^{(k)}(\omega)| = \alpha \gamma$ for $\omega \in [N]^d$, and $\|T^{(k)}\|_\infty \leq \alpha$. Hence we build $\chi^T$ from $\chi$ by taking the outer product of the matrices in $\chi$ by the vector $\mathbf{1}$ along all the other dimensions. Obviously, $|\chi^T| = |\chi| \geq \exp\left(\frac{rN}{16 \gamma^2}\right)$. It just remains to prove

$$\|T^{(i)} - T^{(j)}\|^2_F \geq \frac{\alpha^2 \gamma^2 N d^2}{2}.$$

Assuming that $T^{(i)}$ is generated from $M^{(i)}$ and $T^{(j)}$ is generated from $M^{(j)}$, since $T^{(i)}(i_1, i_2, \cdots, i_d) = M^{(i)}(i_1, i_2)$, $\|T^{(i)} - T^{(j)}\|^2_F = \sum_{i_1=1}^N \sum_{i_2=1}^N \sum_{i_d=1}^N (T^{(i)}(i_1, i_2, \cdots, i_d) - T^{(j)}(i_1, i_2, \cdots, i_d))^2 = N^d - 2 \|M^{(i)} - M^{(j)}\|^2_F \geq \frac{\alpha^2 \gamma^2 N d^2}{2}$ which concludes proof of the lemma.

**Proof of Theorem 33**

Now we use the construction in Lemma 66 to obtain a $\delta$-packing set, $\chi^T$ of $K_M^T$, with $\delta = \alpha \gamma \sqrt{\frac{N d}{2}}$. For the lower bound, we assume that the sampling distribution satisfies

$$\max_{\omega} \pi_\omega \leq \frac{L}{N d}. \quad (3.26)$$

The proof is based on the proof in [19, Section 6.2] which we will rewrite the main parts and refer to [19] for more details. The proof is based on two main arguments. First is a lower bound on the $\|\cdot\|_F$-risk in terms of the error in a multi-way hypothesis testing problem [127]

$$\inf_{\hat{T}} \sup_{T \in K_M^T(\alpha, R)} \mathbb{E}_T \|\hat{T} - T\|^2_F \geq \frac{\delta^2}{4} \min_{\bar{T}} \mathbb{P}(\bar{T} \neq T^\sharp),$$

where $T^\sharp$ is uniformly distributed over the pacing set $\chi^T$. The left hand side above is the maximum expected error obtained by the best estimator in terms of Frobenius norm. Second argument is a variant of the Fano’s inequality which conditioned on the observations $S = \{\omega_1, \cdots, \omega_m\}$, gives the
lower bound

$$\mathbb{P}(\hat{T} \neq T^2|S) \geq 1 - \frac{(|\mathcal{X}_2|)^{-1} \sum_{k \neq j} K(T^k||T^j) + \log(2)}{\log|\mathcal{X}|}$$

where $K(T^k||T^j)$ is the Kullback-Leibler divergence between distributions $(Y_S|T^k)$ and $(Y_S|T^j)$. Here, $(Y_S|T^k)$ is the probability of observing $Y_S$ given that the measurements are taken from $T^k$.

For our observation model with i.i.d. Gaussian noise, we have

$$K(T^k||T^j) = \frac{1}{2\sigma^2} \sum_{i=1}^{m} (T^k(\omega_i) - T^j(\omega_i))^2,$$

and therefore,

$$\mathbb{E}_S[K(T^k||T^j)] = \frac{m}{2\sigma^2} ||T^k - T^j||_F^2.$$

From the first property of the packing set generated in Lemma 66, $||T^k - T^j||_F^2 \leq 4\gamma^2 N^d$. This combined (3.26) and (3.27) yields

$$\mathbb{P}(\hat{T} \neq T^2) \geq 1 - \frac{32\gamma^2 \alpha^2 m}{\sigma^2 rN} + \frac{12\gamma^2}{\sigma^2 rN} \geq 1 - \frac{32\gamma^2 \alpha^2 m}{\sigma^2 rN} - \frac{12\gamma^2}{\sigma^2 rN} \geq \frac{3}{4} - \frac{32\gamma^2 \alpha^2 m}{\sigma^2 rN},$$

provided $rN > 48$. Now if $\gamma^2 \leq \frac{\sigma^2 rN}{128\gamma^2}$,

$$\inf_{\hat{T}} \sup_{T \in \mathcal{K}_M(\alpha,R)} \frac{1}{N^d} \mathbb{E}||\hat{T} - T||_F^2 \geq \frac{\alpha^2 \gamma^2}{16}.$$

Therefore, if $\frac{\sigma^2 rN}{128\gamma^2} \geq 1$ choosing $\gamma = 1$ finishes the proof. Otherwise choosing $\gamma^2 = \frac{\sigma^2}{8\sqrt{2\alpha}} \sqrt{\frac{rN}{Lm}}$ results in

$$\inf_{\hat{T}} \sup_{T \in \mathcal{K}_M(\alpha,R)} \frac{1}{N^d} \mathbb{E}||\hat{T} - T||_F^2 \geq \frac{\sigma \alpha}{128\sqrt{2L}} \sqrt{\frac{rN}{Lm}} \geq \frac{\sigma R}{128\sqrt{2K_G}} \sqrt{\frac{N}{Lm}}.$$
Chapter 4

1-bit tensor completion

In this chapter, we analyze the problem of recovering a low-rank tensor from 1-bit measurements of a subset of its entries. Namely, we provide recovery guarantees using max-qnorm and M-norm constrained ML estimation in Section 4.3, nuclear-norm constrained ML estimation in section 4.4, and finally exact-rank-constrained ML estimation in Section 4.5. All the proofs are presented separately in Section 4.8. We explain the max-qnorm constrained analysis first, as some of the concepts in its proof will be used in the other two sections. Before that, we explain the observation model and main goal of the problem more concretely.

4.1 Introduction and past results

Compressed sensing, as the name suggests, analyzes the possibility of simultaneously measuring (sensing) a signal and compressing it at the acquisition level. Sometimes hardware limitations or the nature of the measurements force these measurements to be highly quantized which adds an extra complication to this process. 1-bit compressed sensing is a technique that was developed to address this issue [16], where we examine the extreme case where we just store 1-bit measurements of a signal. It is interesting to note that when the signal to noise ratio is low, numerical experiments demonstrate that such extreme quantization can be optimal [76].

Generalizing from vectors to matrices, a framework for 1-bit matrix completion was developed in [31] where the indices of a low-rank matrix is sampled after dithering them by a known noise function and then storing the sign of these dithered samples. Interestingly they showed that the sample complexity of 1-bit matrix completion is similar to the case where the measurements are not quantized.
The extreme quantization framework is especially interesting for recommender systems. There are two reasons for this. First, users’ ratings have to be quantized so that they have qualitative interpretations. The most extreme case is yes/no questions in surveys. Moreover, user ratings are influenced by many factors which can be modeled by considering an additive noise. In [31] it was shown that 1-bit matrix completion is more successful than multi-bit matrix completion in estimating the movie-ratings.

In the 1-bit matrix completion problem, the measurements are in the form of \( Y(\omega) = \text{sign}(M^\dagger(\omega) + Z(\omega)) \) for \( \omega \in \Omega \), where \( Z \) is a noise matrix. Notice that without the noise, there will be no difference in the 1-bit measurements of \( M^\dagger \) and \( \alpha M^\dagger \) for any positive \( \alpha \). However, if we assume that the additive noise comes from a log-concave distribution, the problem can be solved by minimizing a regularized negative log-likelihood function given the measurements. In other words, the noise matrix has a dithering effect which is essential for unique recovery of the matrix. Under these assumptions, a nuclear-norm constrained maximum likelihood (ML) optimization was used in [31] to recover \( M^\dagger \) and it was proved that it is minimax rate-optimal under the uniform sampling model. In particular, in order to recover a rank-\( r \) matrix \( M^\dagger \) with \( \|M^\dagger\|_\infty \leq \alpha \) from 1-bit measurements, they considered the weaker assumption that the matrix belongs to the set \( \{A | \|A\|_* \leq \alpha \sqrt{rN^2}, \|A\|_\infty \leq \alpha \} \) and proved

\[
\frac{1}{N^2} \|M^\dagger - M_{\text{recovered}}\|_F^2 \leq C\alpha \sqrt{\frac{rN}{m}},
\]

provided that \( m > CN \log(N) \). The authors also show that these estimates are near-optimal.

A line of research followed this paper and concentrated on analyzing the recovery of low-rank matrices from 1-bit measurements [11, 18, 93]. Max-norm constrained ML estimation was considered in [18]. The max norm, an alternative convex proxy of the rank of a matrix, bounds the spectral norm of the rows of the low-rank factors of \( M^\dagger \), say \( U \) and \( V \). It is defined as \( \|M^\dagger\|_{\text{max}} := \min \|U\|_2 \|V\|_2 \) s.t. \( M^\dagger = UV^T \) [117]. The max norm has been extensively investigated in the machine learning community after the work of [116] and shown to be empirically superior to nuclear-norm minimization for collaborative filtering problems. A max-norm constrained ML estimation was analyzed in [18] and it was shown to be near optimal. To be precise, it was proved in [18] that under some weak assumptions on the sampling distribution, with probability greater than \( 1 - \frac{4}{N} \),

\[
\frac{1}{N^2} \|M^\dagger - M_{\text{recovered}}\|_F^2 \leq C\alpha \sqrt{\frac{rN}{m}} + U\alpha \sqrt{\frac{\log(N)}{m}}.
\]
Notice that $m = O(rN)$ is sufficient for estimating the true matrix efficiently.

Finally using exact low-rank constraints for 1-bit matrix completion is considered in [11, 93]. Here, they consider a rank-constrained ML estimator and analyze it using a second-order Taylor expansion. The sampling model of [11] corresponds to a $d$-regular bipartite graph and their estimation error bound, $O(\frac{N^3 \sqrt{rN}}{m^2})$, is inferior to that of [93] which is $O(\sqrt{\frac{rN}{m}})$. Although the low-rank constraints are non-convex, an efficient conditional gradient descent algorithm was proposed and analyzed in [93] which will be discussed later.

Other generalizations of this problem have been considered in the literature. Learning from a finite alphabet was considered in [75]. Notice that finite alphabet is different than what we call multi-bit where multi-bit can be thought of as infinite-bit as well. Subspace learning from bits was considered in [26] and more general formulations of the problem were considered in [100, 125].

In this chapter we generalize 1-bit matrix completion to multi-dimensional tensors and analyze recovery guarantees of 1-bit measurements obtained from a rank-$r$ bounded tensor. Similar to the multi-bit tensor completion, we show the advantage of using M-norm and max-qnorm constrained completion instead of matricizing both theoretically and numerically when $r = O(1)$.

Generalizing 1-bit matrix completion to 1-bit tensor completion was considered in [2] as well. There they solve the 1-bit tensor completion problem by a weighted sum of the 1-bit matrix completion problem resulting from matricization of the tensor along each one of the $d$ dimensions. Although this method can be successful at finding the best matricization, its performance is still bounded by the best performance one can get using matricization which inferior to direct tensor completion when $r \ll N$.

The framework of 1-bit tensor completion has not been analyzed before. However, tensor factorization for context-aware recommender systems has been investigated in [54, 65, 128, 133] with two distinct differences. First, analyses in these works are different from the 1-bit machinery introduced in [31] and second is the lack of theoretical analysis. However, these works show the power of using tensor factorization for context-aware recommender systems.
4.2 Problem formulation and observation model

In this section, we explain the 1-bit tensor completion problem. Given a \(d\)-dimensional tensor \(T^\# \in \bigotimes_{i=1}^d \mathbb{R}^{N_i}\) (for simplicity, we assume \(N_i = N\)) and a random subset of its indices \(\Omega \subset [N] \times [N] \times \cdots [N]\), suppose that we observe 1-bit measurements of the indices in the subset \(\Omega\) according to the following rule:

\[
Y_{\omega} = \begin{cases} 
+1 & \text{with probability } f(T_{\omega}^\#), \\
-1 & \text{with probability } 1 - f(T_{\omega}^\#) 
\end{cases}
\quad \text{for } \omega \in \Omega. \tag{4.1}
\]

Here \(f : \mathbb{R} \to [0, 1]\) is a monotone differentiable function which can be interpreted as the cumulative distribution of some noise function \(Z_{\omega}\) where the observation then becomes [31]:

\[
Y_{\omega} = \begin{cases} 
+1 & \text{with probability } T_{\omega} + Z_{\omega} \geq 0, \\
-1 & \text{with probability } T_{\omega} + Z_{\omega} < 0 
\end{cases}
\quad \text{for } \omega \in \Omega.
\]

Suppose that we observe 1-bit measurements of a random subset of the entries of the tensor. A standard assumption is that observed indices are chosen uniformly in random, i.e., we randomly pick \(|\Omega|\) indices out of all the \(N^d\) indices and then take 1-bit measurements from this set. Here, \(\Omega\) is chosen at random such that \(\mathbb{P}(\omega \in \Omega) = \frac{m}{N^d}\). Alternatively, the measurements may be picked uniformly at random with replacement so each index is sampled independently of the other measurements. We take a generalization of the later model, i.e., we assume a general sampling distribution \(\Pi = \{\pi_\omega\}\) for \(\omega \in [N] \times [N] \times \cdots \times [N]\) such that \(\sum_\omega \pi_\omega = 1\) and then each index is sampled using the distribution \(\Pi\). Although in this method it is possible to choose a single index multiple times, it has been proven in various contexts that sampling with replacement can be as powerful theoretically as sampling without replacement[18, 19]. Other than theoretical considerations, assuming non-uniform sampling is a better model for various applications of 1-bit matrix and tensor completion, including the Netflix problem. The challenges and benefits of using non-uniform sampling are thoroughly discussed in [18, 115].

We assume that \(f\) is such that the following quantities are well defined:

\[
L_\alpha := \sup_{|x| \leq \alpha} \frac{|f'(x)|}{f(x)(1 - f(x))}, \\
\beta_\alpha := \sup_{|x| \leq \alpha} \frac{f(x)(1 - f(x))}{(f'(x))^2}, \tag{4.2}
\]
where $\alpha$ is the upper bound of the absolute values of entries of $T$. Here, $L_\alpha$ controls the steepness of $f$, $\beta_\alpha$ controls its flatness. Furthermore, we assume $f$ and $f'$ are non-zero in $[-\alpha, \alpha]$ and the quantity:

$$U_\alpha := \sup_{|x| \leq \alpha} \log\left( \frac{1}{f(x)(1-f(x))} \right),$$

is well-defined. A few well-known examples are [31]:

- **Logistic regression/Logistic noise** defined as $f(x) = \frac{e^x}{1+e^x}$,
  $L_\alpha = 1$, $\beta_\alpha = \frac{(1+e^\alpha)^2}{e^\alpha}$, and $U_\alpha = 2\log(e^\alpha + e^{-\alpha})$.

- **Probit regression/Gaussian noise** defined as $f(x) = \Phi\left(\frac{x}{\sigma}\right)$,
  $L_\alpha \leq \frac{4}{\sigma} \left( \frac{\alpha}{\sigma} + 1 \right)$, $\beta_\alpha \leq \pi \sigma^2 \exp\left(\frac{\alpha^2}{2\sigma^2}\right)$, and $U_\alpha \leq \frac{\alpha}{\sigma} + 1$.

For the exact-rank constrained proof we use quadratic lower and upper bounds on the second order Taylor expansion of $f$. Therefore we also define $\gamma_\alpha$, and $\mu_\alpha$ to be

$$\gamma_\alpha := \min\left( \inf_{|x| \leq \alpha} \left\{ \frac{f''(x)}{f'(x)} \right\}, \inf_{|x| \leq \alpha} \left\{ \frac{f''(x)}{(1-f(x))^2} \right\} \right),$$

$$\mu_\alpha := \max\left( \sup_{|x| \leq \alpha} \left\{ \frac{f''(x)}{f'(x)} \right\}, \sup_{|x| \leq \alpha} \left\{ \frac{f''(x)}{(1-f(x))^2} \right\} \right).$$

(4.3)

$\gamma_\alpha$ and $\mu_\alpha$ are fixed constants that do just depend on $\alpha$. For example for the Logistic regression model, we have $\gamma_\alpha = \frac{e^\alpha}{(1+e^\alpha)^2}$ and $\mu_\alpha = \frac{1}{4}$.

The problem of 1-bit tensor completion entails recovering $T^\sharp$ from the measurements $Y_\omega$ given the function $f$. To do this, similar to [31], we minimize the negative log-likelihood function

$$\mathcal{L}_{\Omega, f}(X) = \sum_{\omega \in \Omega} \left( \mathbb{1}_{[Y_\omega = 1]} \log\left( \frac{1}{f(X_\omega)} \right) + \mathbb{1}_{[Y_\omega = -1]} \log\left( \frac{1}{1-f(X_\omega)} \right) \right),$$

given our observations and subject to certain constraints to promote low-rank solutions.

### 4.2.1 Discrepancy

In this section, we briefly describe some more notation which are direct generalizations of corresponding definitions for matrix distances. First, the Hellinger distance for two scalars $0 \leq p, q \leq 1$ is defined by:

$$d_H^2(p, q) = (\sqrt{p} - \sqrt{q})^2 + ((\sqrt{1-p} - \sqrt{1-q})^2,$$
which gives a standard notation on the distance between two probability distributions. We generalize this definition to define Hellinger between two tensors \( T, U \in [0, 1]^{N_1 \times \cdots \times N_d} \) as:

\[
\mathcal{d}_H^2(T, U) = \frac{1}{N_1 \cdots N_d} \sum_{i_1, \ldots, i_d} d_H^2(S(i_1, \ldots, i_d), U(i_1, \ldots, i_d)).
\]

The Kullback-Leibler divergence between two tensors \( T, U \in [0, 1]^{N_1 \times \cdots \times N_d} \) is generalized as:

\[
\mathcal{K}(T \| U) := \frac{1}{N_1 \cdots N_d} \sum_{i_1, \ldots, i_d} \mathcal{K}(S(i_1, \ldots, i_d) \| U(i_1, \ldots, i_d)),
\]

where \( \mathcal{K}(p \| q) := p \log \left( \frac{p}{q} \right) + (1 - p) \log \left( \frac{1 - p}{1 - q} \right) \), for \( 0 \leq p, q \leq 1 \). It is easy to prove that for any two scalars \( 0 \leq p, q \leq 1 \), \( d_H^2(p, q) \leq \mathcal{K}(p \| q) \) and hence:

\[
d_H^2(T, U) \leq \mathcal{K}(T \| U). \tag{4.4}
\]

### 4.3 Max-qnorm and M-norm constrained 1-bit tensor completion

In this section, we analyze the problem of 1-bit tensor recovery using max-qnorm constrained ML estimation as well as M-norm constrained ML estimation. These two optimization problems are closely related, and our analysis and results are similar. Therefore, we explain the corresponding results together. To be precise, we recover \( T^\sharp \) from the 1-bit measurements \( Y_\Omega \), acquired following the model (4.1), via minimizing the negative log-likelihood function

\[
\mathcal{L}_{\Omega, Y}(X) = \sum_{\omega \in \Omega} \left( I_{[Y_\omega = 1]} \log \left( \frac{1}{f(X_\omega)} \right) + I_{[Y_\omega = -1]} \log \left( \frac{1}{1 - f(X_\omega)} \right) \right), \tag{4.5}
\]

subject to \( X \) having low max-qnorm or M-norm and small infinity-norm. Defining

\[
K_{\max}^T(\alpha, R) := \{ T \in \mathbb{R}^{N_1 \times N_2 \times \cdots \times N_d} : \| T \|_\infty \leq \alpha, \| T \|_{\max} \leq R \},
\]

\[
K_{M}^T(\alpha, R) := \{ T \in \mathbb{R}^{N_1 \times N_2 \times \cdots \times N_d} : \| T \|_\infty \leq \alpha, \| T \|_M \leq R \},
\]

67
we analyze the following optimization problems:

\[
\hat{T}_{\text{max}} = \arg \min_X \mathcal{L}_{\Omega,Y}(X) \quad \text{subject to} \quad X \in K^T_{\text{max}}(\alpha, R_{\text{max}}). \tag{4.6}
\]

\[
\hat{T}_M = \arg \min_X \mathcal{L}_{\Omega,Y}(X) \quad \text{subject to} \quad X \in K^T_M(\alpha, R_M). \tag{4.7}
\]

**Remark 38.** Flat low-rank tensors are contained in \(K^T_{\text{max}}(\alpha, R_{\text{max}})\) and \(K^T_M(\alpha, R_M)\) \((R_{\text{max}}\) and \(R_M\) are quantities that just depend on \(\alpha\) and the rank of the tensor) \([45, \text{Theorem 7}]\). M-norm, max-qnorm and infinity norm are all continuous functions and therefore, both \(K^T_{\text{max}}(\alpha, R_{\text{max}})\) and \(K^T_M(\alpha, R_M)\) contain approximately low-rank tensors. We assume the data is generated from one of these sets of approximate low-rank tensors.

**Theorem 39.** Suppose that we have \(m\) 1-bit measurements of a subset of the entries of a tensor \(T^\sharp \in \mathbb{R}^{N^d}\) following the probability distribution explained above where \(f\) is chosen such that \(L_\alpha, \beta_\alpha, \) and \(U_\alpha\) are well defined. Moreover, the indices of the tensor are sampled according to a probability distribution \(\Pi\). Then assuming \(\|T^\sharp\|_{\text{max}} \leq R_{\text{max}}\), there exist absolute constants \(C_{\text{max}}\) and \(C_M\) such that for a sample size \(2 \leq m \leq N^d\) and for any \(\delta > 0\) the maximizer \(\hat{T}_{\text{max}}\) of (4.6) satisfies:

\[
\|T^\sharp - \hat{T}_{\text{max}}\|_{\Pi}^2 := \sum_\omega \pi_\omega (\hat{T}_{\text{max}}(\omega) - T^\sharp(\omega))^2 \leq C_{\text{max}} c_d^2 \beta_\alpha \left( L_\alpha R_{\text{max}} \sqrt{\frac{dN}{m}} + U_\alpha \sqrt{\frac{\log\left(\frac{4}{\delta}\right)}{m}} \right)
\]

with probability at least \(1 - \delta\). Similarly, assuming \(\|T^\sharp\|_M \leq R_M\), for any \(\delta > 0\), the maximizer \(\hat{T}_M\) of (4.7) satisfies:

\[
\|T^\sharp - \hat{T}_M\|_{\Pi}^2 := \sum_\omega \pi_\omega (\hat{T}_M(\omega) - T^\sharp(\omega))^2 \leq C_M \beta_\alpha \left( L_\alpha R_M \sqrt{\frac{dN}{m}} + U_\alpha \sqrt{\frac{\log\left(\frac{4}{\delta}\right)}{m}} \right)
\]

with probability at least \(1 - \delta\).

**Corollary 40.** A rank-\(r\) tensor \(T_r^\sharp\) with \(\|T_r^\sharp\|_\infty \leq \alpha\) satisfies \(\|T_r^\sharp\|_{\text{max}} \leq \sqrt{r^{d-2}} \alpha\) and therefore, the maximizer \(\hat{T}_{\text{max}}\) of (4.6) with \(R_{\text{max}} = \sqrt{r^{d-2}} \alpha\) satisfies:

\[
\|T_r^\sharp - \hat{T}_{\text{max}}\|_{\Pi}^2 \leq C_{\text{max}} c_d^2 \beta_\alpha \left( L_\alpha \sqrt{r^{d-2}} \alpha \sqrt{\frac{dN}{m}} + U_\alpha \sqrt{\frac{\log\left(\frac{4}{\delta}\right)}{m}} \right),
\]

with probability greater than \(1 - \delta\). Moreover, \(\|T_r^\sharp\|_M \leq (r \sqrt{r})^{d-1} \alpha\) and the maximizer \(\hat{T}_M\) of (4.7)
with \( R_M = (r\sqrt{r})^{d-1} \alpha \) satisfies

\[
\|T^\sharp_r - \hat{T}_M\|_\Omega^2 \leq C_M \beta \alpha \{L_\alpha (r\sqrt{r})^{d-1} \sqrt{\frac{dN}{m}} + U_\alpha \sqrt{\frac{\log(\frac{1}{\delta})}{m}} \},
\]

with probability greater than \( 1 - \delta \).

**Remark 41.** There are two differences in the right-hand sides of (4.8) and (4.9). First is the difference in \( R_{\text{max}} \) and \( R_M \) which is due to the different upper bounds on max-qnorm and M-norm of a bounded rank-\( r \) tensor. The second difference is in the constants \( C_{\text{max}} \) and \( C_M \) where \( C_M < \frac{C_{\text{max}}}{c_1} \) which implies that (4.9) is tighter than (4.8). Moreover, for \( d \geq 2 \), the M-norm bound of a low-rank tensor is smaller than its max-qnorm bound [45].

**Remark 42.** Assuming that all entries have a nonzero substantial probability of being observed, i.e., in the sense that there exists a constant \( \eta \geq 1 \) such that \( \pi_\omega \geq \frac{1}{\eta N^d} \), for all \( \omega \in [N_1] \times \cdots \times [N_d] \), we can simplify (4.8) as

\[
\frac{1}{N^d} \|T^\sharp_r - \hat{T}_{\text{max}}\|_\Omega^2 \leq C_\eta \beta \alpha \{L_\alpha R_{\text{max}} \sqrt{\frac{dN}{m}} + U_\alpha \sqrt{\frac{\log(dN)}{m}} \}.
\]

A similar bound can be obtained for M-norm constrained ML estimation (4.9) as well.

**Remark 43.** The terms \( r^{3d-3} \) and \( r^{d^2-d} \) in the upper bounds come from the M-norm and max-qnorm of low-rank tensors. We believe this is sub-optimal in \( r \). Indeed, when \( d = 2 \), we know that both these upper bounds can be improved to \( \sqrt{r} \) instead of \( r^{\frac{3}{2}} \) and \( r^2 \).

**Remark 44.** For a rank-\( r \) tensor \( T^\sharp_r \) with \( \|T^\sharp_r\|_\infty \leq \alpha \), it is sufficient to choose \( R_{\text{max}} = r^{\frac{d^2-d}{2}} \alpha \) in (4.6) and \( R_M = r^{\frac{3d-3}{2}} \alpha \) in (4.7) for \( T^\sharp_r \) to be feasible [45]. This proves that efficient recovery can be achieved when \( m = O(r^{d^2-d}N) \) using (4.6) and \( m = O(r^{3d-3}N) \) using (4.7). These are significant improvements over matricizing, which would require \( \Omega(rN^2) \) measurements when \( r \ll N \).

**Remark 45.** 1-bit tensor completion can recover the magnitude of the tensor because of the dithering effect of the noise we add before sampling. If the SNR is too low, we might end up estimating the noise by a low-rank tensor and if the SNR is too high, we risk losing magnitude information. In Section 4.7 we conduct some experiments and discuss the optimal noise level using synthetic data.

**Remark 46.** 1-bit matrix completion using max-norm is analyzed in [18]. A direct generalization resulted in the error bound (4.8) which recovers their result when \( d = 2 \). However, the direct
generalization of max-norm to tensors is a quasi-norm and the resulting constraint is non-convex. The closely related upper bound using M-norm (4.9) is governed by a convex optimization and gives similar results when \( d = 2 \). In this case, the only difference is the Grothendieck’s constant \( 1.67 < c_1 c_2^2 < 1.79 \) [62].

### 4.4 Nuclear norm constrained 1-bit tensor completion

In this section, we analyze tensor recovery from 1-bit measurements under nuclear-norm constraints. Instead of considering the set of tensors with bounded rank, we consider the set of tensors with bounded nuclear norm. This constraint is more relaxed compared to the non-convex set of low-rank tensors. Using Theorem 7, the nuclear norm of a rank-\( r \) tensor \( T \) with \( \| T \|_\infty \leq \alpha \) is bounded by \( \| T \|_* \leq \alpha \sqrt{r^d - 1} N^d \). To be consistent with the results obtained in [31], in this section, we assume uniform sampling, i.e., we assume \( \mathbb{P}(\omega \in \Omega) = \frac{m}{N^d} \) (the exact same results are also true for (4.1) with \( \pi_\omega = \frac{m}{N^d} \)). Considering all these assumptions, in this section, we use the following optimization problem to recover \( T^\# \).

\[
\hat{T}_{\text{nuc}} = \arg \max_X \mathcal{L}_{\Omega,Y}(X) \quad \text{subject to} \quad \| X \|_* \leq \sqrt{r^d - 1} N^d \alpha. \tag{4.10}
\]

**Theorem 47.** Assume that \( \| T^\# \|_* \leq \sqrt{r^d - 1} N^d \alpha \) and \( \| T^\# \|_\infty \leq \alpha \) and suppose the sample set \( \Omega \) is chosen at random such that \( \mathbb{P}(\omega \in \Omega) = \frac{m}{N^d} \) and 1-bit measurements of these samples are taken via (4.1). If \( \hat{T}_{\text{nuc}} \) is the solution to (4.10), then with probability at least \( 1 - \frac{C}{N^d} \)

\[
\frac{\| T^\# - \hat{T}_{\text{nuc}} \|_F^2}{N^d} \leq C d, \alpha \sqrt{r^d - 1} \left( \frac{\log^d(N) \sqrt{N}}{\sqrt{m}} + \frac{\log(N) N^d}{m} \right).
\]

**Remark 48.** Theorem 47 is a direct generalization of the 1-bit matrix completion using nuclear-norm constrained ML in [31] and the results are the same when \( d = 2 \).

**Remark 49.** Comparing Theorem 39 and Theorem 47 shows that the nuclear-norm constrained algorithm gives much weaker results compared to the M-norm constrained one. In this remark, we attempt to clarify this difference. The main step in the proof of Theorem 47 which results in the \( O(N^2) \) requirement in the number of measurements, is the second inequality of (4.28) below which is similar to the Rademacher complexity of low nuclear-norm tensors. In the following remark we show that with the assumptions in Theorem 47 we can’t hope to get a better result and if one hopes to achieve tighter upper bounds, further assumptions should be considered. For simplicity, we assume \( \alpha = 1 \).
Define \( \vec{1}_N \) to be the vector of all ones with length \( N \). Then for the \( d \)-dimensional tensor \( T = \vec{1}_N \circ \vec{1}_N \circ \cdots \circ \vec{1}_N \), rank(\( T \)) = 1, \( \|T\|_\infty = 1 \), and \( \|T\|_* = \sqrt{Nd} \). This shows that the bound on the nuclear norm of bounded rank-\( r \) tensors is tight (at least) in \( N \).

Considering the set \( G = \{ X \in \bigotimes_{i=1}^d \mathbb{R}^N : \|X\|_* \leq \sqrt{rd-1}Nd \} \), the empirical Rademacher complexity of set \( G \) is defined as \( E \left[ \sup_{X \in G} |<\Delta_\Omega \ast E, X>| \right] \), where \( \ast \) denotes the Hadamard product, \( \Delta_\Omega \) is the sampling operator which is one on \( \Omega \) and zero otherwise and \( E \) is a tensor of Rademacher random variables. This expression shows up in (4.28) and is at least \( \sqrt{rd-1}Nd \) which can be achieved by considering a tensor \( T \in G \) which is zero everywhere except for one single entry with the value of \( \sqrt{rd-1}Nd \). For such a tensor \( T \), \( \|T\|_* = \sqrt{rd-1}Nd \) and obviously, for any random tensor \( \Delta_\Omega \ast E \) there exists such a tensor that makes \( \left[ \sup_{X \in G} |<\Delta_\Omega \ast E, X>| \right] \geq \sqrt{rd-1}Nd \). Therefore, it seems that imposing an explicit constraint on the infinity norm is necessary if one hopes to get a better error bound.

Even considering the set \( G = \{ X \in \bigotimes_{i=1}^d \mathbb{R}^N : \|X\|_* \leq \sqrt{rd-1}Nd, \|X\|_\infty \leq 1 \} \) doesn’t result in better results. For any sampling set \( \Omega \) and tensor \( \Delta_\Omega \), consider the tensor \( T \) with

\[
T(\omega) = \begin{cases} \min(1, \sqrt{\frac{rd-1}N}|\Omega|) \cdot \text{sign}(\Delta_\Omega) & \text{for } \omega \in \Omega, \\
0 & \text{for } \omega \notin \Omega \end{cases}
\]

Then \( \|T\|_\infty \leq 1 \), \( \|T\|_* \leq \min(|\Omega|, \sqrt{rd-1}Nd) \) and more importantly \( |<\Delta_\Omega \ast E, X>| = \min(|\Omega|, \sqrt{rd-1}Nd) \).

Tracking back the steps of the proof we get

\[
\frac{1}{Nd} \|T - \hat{T}_{\text{nuc}}\|_F^2 \leq C_{d,\alpha} \left( \log^d(N) \min(\frac{1}{Nd-1}, \frac{\sqrt{rd-1}N}{\sqrt{m}}) + \frac{\log(N) \min(m, \sqrt{rd-1}Nd)}{m} \right),
\]

which is essentially the same result as before and the theoretical bound is meaningful when \( m > \sqrt{rd-1}Nd \).

It is worth mentioning that the above remark just proves that with the approach taken in our proof (or any nuclear-norm constrained approach that depends on the Rademacher complexity of low nuclear-norm tensors) we can not expect to get a better result. We believe that without some extra incoherence assumptions on the tensor, it’s impossible to get better results. The specifics of
such assumptions depend on the considered applications. An example of such assumptions which is widely used in matrix completion literature is an incoherence condition on the low-rank factors. In particular, assume a rank-$r$ tensor $T = U_1 \circ U_2 \circ \cdots \circ U_d$, where $U_j := [u_1^{(j)} \ u_2^{(j)} \ \cdots \ u_r^{(j)}]$ is the low-rank factor whose columns are the low-rank vectors corresponding to the $j$-th dimension. Then assuming $\mathbb{P}_{U_j}$ to be the orthogonal projection onto the columns of $U_j$, the coherence of $U_j$ is defined as

$$\mu(U_j) = \frac{N_j}{r} \max_{1 \leq i \leq N_j} \|\mathbb{P}_{U_j} e_i\|_2^2.$$ 

The incoherence assumption $\max_j(\mu(U_j)) \leq \mu_0$, for some positive constant $\mu_0$ forces the elements of the low-rank factors to be of the same order. Once we make this assumption, we can observe that we can bound the entries of the tensor by (using Matlab notations below for the rows):

$$|T(i_1, \cdots, i_d)| = |\sum_{1 \leq k \leq r} U_1(i_1, k)U_2(i_2, k)\cdots U_d(i_d, k)|$$

$$\leq \|U_1(i_1,:\|_2\|U_2(i_2,:\|_2\|U_3(i_3,:\|_\infty\cdots\|U_d(i_d,:\|_\infty\leq (\frac{\mu_0 r}{N})^d,$$

which removes the $N^d$ factor in the upper bound of Theorem 47 with the expense of an extra $r^d$. However, one can argue that such an argument is implicitly assuming that the $\|T\|_\infty \sim O(\frac{1}{N^{\frac{d}{3}}})$ which is much stronger than the assumptions in Theorem 47. We believe that applicability of such extra assumptions depends on specific applications and needs to be explored separately. We keep the results as they are to be consistent with the results of 1-bit matrix completion in the literature.

**Remark 50.** Remark 49 studies the nuclear-norm constrained ML. We do not know if bringing the nuclear norm into the optimization as a regularizer, i.e., minimizing $\mathcal{L}_{\Omega,Y}(X) + \lambda \|X\|_*$ can achieve better sample complexity for a suitable $\lambda$. We postpone the study of this problem to future work. It is worth mentioning though that finding the tensor with the least nuclear norm that satisfies measurements have been studied in [129] for the problem of tensor completion; they prove the required number of samples for exact recovery is $O(rN^3)$ when $d = 3$ which is similar to our result for the 1-bit case.

### 4.5 1-bit tensor completion under exact rank constraint

Next, we consider recovery with exact rank constraints. Such formulations in the matrix case were first proposed and analyzed in [11], where the authors considered a rank-constrained ML estimation for recovering a matrix where the sampling is drawn from a $p$-regular bipartite graph. Here, $p$ is the number of indices observed in each row. Even though this observation model is
restrictive, for an \( N \times N \) matrix, their estimation error bound, \( O(\frac{N^3 \sqrt{rN}}{|\Omega|^2}) \), is not minimax optimal. Later [93] improved this result to \( O(\frac{rN \log(N)}{|\Omega|}) \) using the fact that under mild assumptions, low-rank matrices follow a restricted strong convexity condition [91].

The exact rank constraint adds some complications to the problem of 1-bit tensor completion. First, unlike the M-norm and nuclear-norm constraints, the rank constraint is non-convex and second, we need to know the rank of the true matrix (or an accurate estimation of the rank). For the sake of simplicity we assume that the samples are drawn randomly and independently such that \( \mathbb{P}(\omega \in \Omega) = \frac{m}{N^d} \). Notice that although this is different than sampling with replacement, the two models are closely related to each other and with high probability the analysis of one can be applied to the other. Given such a random sample set \( \Omega \) and 1-bit measurements of these samples \( Y_\Omega \) we employ a rank-constrained ML estimation to recover the underlying tensor \( T^\sharp \). In particular, we assume \( T^\sharp \) is a \( d \)-dimensional rank-\( r \) tensor with \( \| T^\sharp \|_\infty \leq \alpha \) and we study the following optimization problem:

\[
\hat{T}_r = \arg \min_X \mathcal{L}_{\Omega,Y}(X) \quad \text{subject to} \quad \text{rank}(X) \leq r, \|X\|_\infty \leq \alpha.
\] (4.11)

The main idea of the proof is the same as the ones on [11, 93], i.e., using a second order Taylor expansion on \( \mathcal{L}_{\Omega,Y} \) and some concentration inequalities to get a bound on the error.

**Theorem 51.** Assume that \( T^\sharp \) is a \( d \)-dimensional tensor, where rank \( (T^\sharp) \leq r \) and \( \| T^\sharp \|_\infty \leq \alpha \) and suppose \( \Omega \) is chosen at random (with replacement) according to the probability distribution \( \Pi = \{\pi_\omega\} \) such that \( \pi_\omega = \frac{1}{N^d} \). If \( \hat{T}_r \) is the solution to (4.11), then with probability at least \( 1 - \exp(-\alpha^2 m) \)

\[
\frac{1}{N^d} \| T^\sharp - \hat{T}_r \|_F^2 \leq \max \left( \frac{4\alpha^2 N d}{m}, C_{\alpha,y} \alpha^2 (2r)^{\frac{3d}{2} - \frac{3}{4}} \sqrt{\frac{Nd}{m}} \right).
\] (4.12)

**Remark 52.** Theorem 51, which is a generalization of the results in [11, 93], considers ML estimation using exact rank constraints. The result in [11] is not as good as the one in [93] and therefore, here we compare this theorem with the latter (assuming \( d = 2 \) and taking \( M^\sharp \) to be the original matrix). Theorem 3.3 of [93] proves the upper bound on the recovery error

\[
\frac{\| \hat{M} - M^\sharp \|_F^2}{N^d} \leq C \max(1, \alpha^2) \frac{rN \log(N)}{m}.
\]

Comparing this result with (4.12), we make the following remarks.
- Our result does not have the \( \log(N) \); which is the result of using the strong convexity of the set of low max-norm tensors (refer to Section 4.8.3 for more details).

- Fixing \( p = \frac{m}{N} \) their error is \( O\left(\frac{\log(N)}{p}\right) \) whereas our error is \( O\left(\sqrt{\frac{1}{p}}\right) \). This means that if we fix \( N \) our error bound is weaker in terms of \( m \). The reason behind these differences is that they use the strong convexity of the set of low-rank matrices whereas we consider the set of low M-norm matrices. Notice that information-theoretic lower bounds have been developed for both max-norm and nuclear-norm constrained 1-bit matrix completion in [18, 31]; these bounds are \( O\left(\sqrt{\frac{1}{p}}\right) \) as well. Considering low M-norm tensors is essential for our proof when \( d > 2 \).

**Remark 53.** Comparing exact rank constraints with M-norm and nuclear-norm constraints, notice that regardless of the proof method the upper bound obtained with exact rank constraints can not be worse as the set of low-rank tensors is a subset of the low M-norm and low nuclear-norm tensors considered above. However, we use M-norm and its dual in one of the steps of the proof (see (4.42)) and this is the reason we get similar bounds as using M-norm. One way to solve this issue is proving the restricted strong convexity of the set of low-rank tensors. When \( d = 2 \), this has been studied in [91]; we postpone generalizing this result to \( d > 2 \) to future work.

### 4.6 Information-theoretic lower bound

In this section, we provide a lower bound for the reconstruction error when recovering tensors from 1-bit random measurements. In particular, by using an information-theoretic argument similar to the one in [31], we prove that on a particular set of tensors, any arbitrary algorithm will not be able to recover a close enough approximation of all the tensors in that set. The same approach was taken in both [18, 31] and therefore, we will not go into the details of the proofs and refer to those papers for a more detailed explanation of these type of methods. However, to be more precise, we use a classical information-theoretic technique which shows that with limited amount of information we can distinguish only a limited number of tensors from each other. To that end, we first construct a set of tensors \( \chi \) such that for any two distinct members \( X^i \) and \( X^j \) of \( \chi \) \( \|X^i - X^j\|_F \) is large. Therefore, we should be able to distinguish the tensors in this set if the recovery error of an arbitrary algorithm is small enough. However, Fano’s inequality will imply that the probability of choosing the right tensor among this set is going to be small and therefore force a lower bound on the performance of any arbitrary recovery algorithm.
The lower bound is achieved using Lemma 18 of [45], which constructs a packing set for the set of low M-norm tensors given by

$$K^T_M(\alpha, R_M) := \{ T \in \mathbb{R}^{N_1 \times N_2 \times \cdots \times N_d} : \|T\|_\infty \leq \alpha, \|T\|_M \leq R_M \}.$$ 

We include Lemma 18 of [45] and a reiteration of the proof in [31] in Section 4.8.4. In the following 1.68 < $K_G := c_1 c_2^2 < 1.79$ is Grothendieck’s constant.

**Theorem 54.** Fix $\alpha, r := \left\lceil \left( \frac{R_M}{\alpha K_G} \right)^2 \right\rceil$, $R_M$, and $N$ such that $\alpha \geq 1$, $\alpha^2 r N \geq C_0$, $r \geq 4$ and $r \leq O\left( \frac{N}{\alpha^2} \right)$. Let $\beta_\alpha$ be defined as in (4.2). Suppose $f'(x)$ is decreasing for $x > 0$ and for any tensor $T$ in the set $K^T_M(\alpha, R_M)$, assume we obtained $m$ measurements (denoted by $Y_\Omega$) from a random subset $\Omega$ of the entries following the model (4.1). Considering any arbitrary algorithm that takes these measurements and returns an approximation $\hat{T}(Y_\Omega)$, then there exists a tensor $T \in K^T_M(\alpha, R_M)$ such that

$$\mathbb{P}\left( \frac{1}{N^d} \|T - \hat{T}(Y_\Omega)\|_F^2 \geq \min \left( C_1, C_2 \sqrt{\frac{R_M}{K_G}} \sqrt{\frac{N}{m}} \right) \right) \geq \frac{3}{4}.$$ 

**Remark 55.** Theorem 54 proves that the sample complexity achieved by (4.7) in Theorem 39 is optimal in both $N$ and $R_M$. Evidence in the matrix case suggests that a similar lower bound for the set of low-rank tensors (instead of low M-norm) should be $\frac{N}{m}$ instead of $\sqrt{\frac{N}{m}}$ provided that $m$ is large enough. We postpone an upper bound for exact low-rank tensors to future work.

### 4.7 Numerical results

In this section, we compare some numerical results of using the above algorithms for 1-bit tensor completion. The optimization problem we want to solve is in the form of

$$\arg\min_X \mathcal{L}_{\Omega,Y}(X) \quad \text{subject to} \quad X \in \mathbb{C},$$

where $\mathcal{L}_{\Omega,Y}(X)$ is the negative log-likelihood function defined in (4.5). Similar to the matrix completion problem $\mathcal{L}_{\Omega,Y}$ is a smooth function from $\mathbb{R}^{N^d} \to \mathbb{R}$. The main difference between the three different methods is in the choice of $C$. Before, showing the results, we mention a few practical considerations for each one of the above algorithms.

#### 4.7.1 Max-qnorm constrained 1-bit tensor completion

We are not aware of any algorithm that is able to solve the M-norm constrained problem (4.7). Therefore, instead, we present the results of constraining the max-qnorm of the tensor. To this
end, we employ a similar approach to the one in [45] and use the low-rank factors to estimate the max-qnorm of a tensor. In particular, defining $f(V_1, \cdots, V_d) := \mathcal{L}_{\Omega,Y}(V_1 \circ \cdots \circ V_d)$ the problem becomes

$$\min f(V_1, \cdots, V_d) \text{ subject to } \max_i (\|V_i\|_{2,\infty}) \leq R_{\max}^{\frac{1}{2}}, \|V_1 \circ \cdots \circ V_d\|_{\infty} \leq \alpha. \quad (4.14)$$

In the definition of max-qnorm (2.11), there is no limit on the size of the low-rank factors. Due to computational restrictions, we limit the factor size by twice the dimension size, i.e., $V_i \in \mathbb{R}^{N \times 2N}$. Although we end up approximating the solution of (4.14), our numerical results shows that, for synthetic tensors with rank smaller than $N$, $2N$ is a large enough factor size and the results are not very sensitive to the use of larger factors (this is true for our experiments where $N < 100$). Another practical benefit of (4.14) is that we can use alternating minimization on the small factors instead of the whole tensor which is crucial in applications where saving the whole tensor is too expensive.

The next practical problem that needs to be addressed is in choosing $R_{\max}$. Although theory suggests the upper-bound of $R_{\max} = \|V\|_{F}^2 - \|V\|_F^2 \alpha$ [45], in practice this bound can be much higher than the actual max-qnorm of a rank-$r$ tensor. Moreover, in many practical applications, we do not have an accurate upper bound on the rank of the tensor (or the rank of a good approximation of a tensor). Therefore, similar to [45], we use cross-validation to estimate the optimal value of $R_{\max}$.

In summary, to approximate the solution of (4.14), in our experiments we use 10% of the available data for cross-validating the choice of $R_{\max}$ and optimize over the factors $V_1$ to $V_d$ alternatively while at each iteration we solve the simpler sub-problem

$$\min_{V_i} f(V_1, \cdots, V_d) \text{ subject to } \|V_i\|_{2,\infty} \leq R_{\max}^{\frac{1}{2}}, \quad (4.15)$$

while fixing all the other factors $V_j, j \neq i$. We solve each sub-problem (4.15) by employing a projected gradient algorithm. The algorithm updates all the factors

$$[V_i] \leftarrow \mathbb{P}_C([V_i - \gamma \nabla (f_i) R_i]). \quad (4.16)$$

where $\mathbb{P}_C$ simply projects the factor onto the set of matrices with $\ell_{2,\infty}$ norm less than $R_{\max}^{\frac{1}{2}}$. This projection looks at each row of the matrix and if the norm of a row is bigger than $R_{\max}^{\frac{1}{2}}$, it scales that row back down to $R_{\max}^{\frac{1}{2}}$ and leaves other rows unchanged.
Remark 56. The constraint on the boundedness of the tensor \( \|V_1 \circ \cdots \circ V_d\|_\infty \leq \alpha \) can also be incorporated into (4.15), which introduces new challenges to the algorithm. First, the exact projection onto the set of \( \{V_i\|V_1 \circ \cdots \circ V_d\|_\infty \leq \alpha\} \) is not as straightforward as the projection onto \( \{V_i\|V_i\|_{2,\infty} \leq R_{\text{max}}\} \). An approximate projection by rescaling the factor via

\[
V_i = \frac{\alpha}{\|V_1 \circ \cdots \circ V_d\|_\infty} V_i, \text{ if } \|V_1 \circ \cdots \circ V_d\|_\infty > \alpha
\]

was introduced in [18]. On the other hand, the exact projection can be formulated as a quadratic linear program, which is very computationally expensive. The second complication is that adding this constraint makes the constraint set \( C \) the intersection of two convex sets, which again makes exact projection expensive.

In our synthetic experiments, we noticed that adding this constraint does not change the results significantly, especially for low-rank tensors. Furthermore, in our applications, we concentrate on the performance of the algorithm in recovering the sign of the tensor which reduces the importance of projecting onto the set of bounded tensors and therefore, in this section, we just report the results obtained by (approximately) solving (4.15).

Remark 57. The optimization problem (4.15) can be solved in parallel over the rows of \( V_i \) as both the objective function and the regularizer (\( \|\cdot\|_{2,\infty} \)) are decomposable in the rows.

**Nuclear-norm constrained 1-bit tensor completion**

The nuclear norm of tensors has been thoroughly studied in [35, 57, 130]. Calculating the nuclear norm of a tensor is NP-hard [43] and similarly projections onto the set of low nuclear-norm tensors are NP-hard. Furthermore, we are not aware of any algorithms that can approximate the solution to nuclear-norm constrained (or regularized) tensor completion.

In the matrix case, i.e., when \( d = 2 \) we know that the nuclear norm of a matrix can be expressed as [117]

\[
\|M\|_* = \min_{\{B,C\}} \frac{1}{2} (\|B\|_F^2 + \|C\|_F^2) \quad \text{s.t.} \quad X = BC^T = B \circ C.
\]  

(4.17)

A generalization of this representation to tensors was used as a regularizer in [9], which has very interesting relations with the CP decomposition of the tensor. Notice that the following representation of nuclear-norm is equivalent to the one in (4.17).

\[
\|M\|_* = \min_{\{B,C\}} (\|B\|_F \|C\|_F) \quad \text{s.t.} \quad X = BC^T = B \circ C.
\]  

(4.18)
When \( d = 2 \), there is an interesting relation between the M-norm \( (2.12) \) and max-norm \( (2.7) \) on one hand, and the two representations of the nuclear norm, i.e., \( (2.5) \) and \( (4.18) \) on the other hand. This motivated us to consider the following norm, which we call factor-frob norm as a surrogate for tensor nuclear-norm.

\[
\| T \|_\diamond := \min_{T = U^{(1)} \circ U^{(2)} \circ \ldots \circ U^{(d)}} \left\{ \prod_{j=1}^{d} \| U^{(j)} \|_F \right\}.
\]  

(4.19)

The relations of this norm with the nuclear norm and the sample complexity of using this norm for tensor completion is an interesting question that we refer to future work. An advantage of \( (4.19) \) is that similar to max-qnorm, we can use an alternating algorithm accompanied with cross-validation to approximate the solution to the following problem

\[
\hat{T} = \arg \min_X \mathcal{L}_{\Omega,Y}(X) \quad \text{subject to} \quad \| X \|_\diamond \leq R, \| X \|_\infty \leq \alpha.
\]  

(4.20)

We postpone analyzing the sample complexity of \( (4.20) \) to future work.

### 4.7.2 1-bit tensor completion with exact rank constraints

The optimization problem \( (4.11) \) is non-convex and very hard to approximate due to local minima. Similar to [93], we use a conditional gradient descent (CGD) algorithm for rank-constrained log-likelihood estimation. In particular, at each iteration \( t \), we first find a descent direction from the gradient of \( \mathcal{L}_{\Omega,Y} \), and then use a tensor rank-1 approximation of the gradient \( \nabla \mathcal{L}_{\Omega,Y}(X^t) \) and update the solution by doing a line search. Notice that at the \( t \)-th iteration of the CGD, the output is a rank-\( t \) tensor. If we know the exact rank, we stop the algorithm when \( t = \text{rank}(T^\#) \). Otherwise, we find the best rank by the estimation error obtained from a validation set of samples.

For the sake of completeness, the details of the algorithm is shown in Algorithm 3. At each iteration of the algorithm, we find the gradient tensor, that has zero values in the non-observed indices and find the best rank-1 approximation to this tensor. For the rank-1 approximations (line 8 in Algorithm 3) we can use the tensor toolbox version 2.5 available in [4]. However, our experiments show that using a tensor completion algorithm with alternating rank-1 factors (e.g., using Algorithm 2) give much better results. This is due to the fact that the values of the gradient tensor on the indices that are not observed is not available and therefore, solving a tensor completion algorithm gives better results than fitting the gradient tensor (with all the zeros of non-observed entries). These approximations are very likely to get stuck in local minima. Therefore, we need to
Algorithm 3 1-bit TC with rank-constraints via a Conditional Gradient Descent (CGD)

1: **Input** $\Omega$, $Y$, $\mathcal{L}_{\Omega,Y}$, $r_{up}$, $\alpha$
2: **Output** $\hat{T}$ using the CGD algorithm
3: Divide $\Omega$ into $\Omega_{train}$ (90%) and $\Omega_{validate}$ (10%)
4: $X^0 = 0$
5: $\hat{T} = 0$
6: $validation_{error} = \mathcal{L}_{\Omega_{validate},Y}$
7: for $t = 0$ to $r_{up}$ do
8: $V^t \leftarrow \arg\min_{\text{rank}(V)=1} \langle V, \nabla \mathcal{L}_{\Omega_{train,Y}}(X^t) \rangle$
9: $(l_t, q_t) \leftarrow \arg\min_{l_t \geq 0, q_t \geq 0} \mathcal{L}_{\Omega_{train,Y}}(l_t X^t + q_t V^t), \text{ s.t. } \|l_t X^t + q_t V^t\|_{\infty} \leq \alpha$
10: $X^{t+1} \leftarrow l_t X^t + q_t V^t$
11: if $\mathcal{L}_{\Omega_{validate,Y}} > validation_{error}$ then
12: break
13: else
14: $validation_{error} = \mathcal{L}_{\Omega_{validate,Y}}(X^{t+1})$
15: $\hat{T} = X^{t+1}$
16: end if
17: end for
18: return $\hat{T}$

run this multiple times with various initial points to make sure that we do not find a local minimum which is far away from the global minimum. This makes the algorithm much slower compared to the other algorithms used in this thesis. Moreover, as we show later, in general, the algorithm performance deteriorates when the order of the tensor increases which is due to higher non-convexity of the problem. We discuss the results in more detail in the next section.

4.7.3 Synthetic experiments

In this section, we present extensive numerical results on synthetic experiments to compare the performance of 1-bit tensor completion with max-qnorm, factor-frob, and exact rank constraints. An important component of the measurements which affects the results is the choice of the function $f$. In particular, we investigate the optimal choice of $\sigma$ when we use the Gaussian noise, i.e., $f(x) = \Phi(\frac{x}{\sigma})$ in the log-likelihood function (4.5). Figure 4.1 shows the results for $d = 2, 3, 4$, where $N = 200, 30, 15$ respectively. In all the figures we average the recovery error over 20 realizations when $r = 5$ and $\sigma$ varies from 0.001 to 10. Moreover, in all cases, $m = |\Omega| = \frac{N^d}{2}$.

Notice that the same experiment can be done with various values of $r$, and $m$. However, finding the optimal value of $\sigma$ for different cases is very time consuming and unnecessary as for most
Figure 4.1: Average relative squared error, $\frac{\|T_{\text{recovered}} - T^\sharp\|^2_F}{\|T^\sharp\|^2_F}$, obtained from matricized 1bit-TC, max-qnorm, factor-frob norm, and CGD over a range of different value of $\sigma$. From left to right the original tensor is 2, 3 and 4-dimensional. Results show that the noise variance should not be too small or too big.

In Figure 4.1, we show the results for matricized nuclear-norm constrained 1-bit matrix completion (MC-Nuclear), Max-qnorm, Factor-frob, and CGD (with exact rank) as explained above. The underlying low-rank tensor is generated from low-rank $N \times r$ factors with entries drawn i.i.d. from a uniform distribution on $[-1, 1]$. The tensor is then scaled to have $\|T^\sharp\|_\infty = 1$. The MC-Nuclear results show the case where we matricize the tensor first and then solve a 1-bit matrix completion problem using the algorithm in [31] but with added cross-validation for the optimal nuclear-norm bound. In Figure 4.1, the left figure shows the result for $T^\sharp \in \mathbb{R}^{100 \times 100}$. In this case, MC-Nuclear and Factor-frob are solving the same problem and as expected we get similar results. Moreover, as expected for the matrix case, the results are close to each other. The CGD algorithm benefits from the extra information of the rank of the tensor. This low-rank structure gives a small advantage to CGD compared to the other algorithms. The figure in the middle shows the results for $T^\sharp \in \mathbb{R}^{30 \times 30 \times 30}$ and as expected from theorems proved above, we see a significant improvement by using 1-bit tensor completion instead of matricizing. Moreover, the performances of max-qnorm and factor-frob are similar. Finally, the figure in the right shows the results for $T^\sharp \in \mathbb{R}^{30 \times 30 \times 30}$. The performance of CGD becomes less and less predictable when we increase the order. One of the reasons behind this is the rank-1 projections of the gradient tensor which is an NP-hard task and becomes less and less accurate when the problem gets harder.

In short, we decided to use $\sigma = 0.1$ as it seems to give the best performance among all the
Figure 4.2: Average relative squared errors, $\|\mathbf{T}_{\text{recovered}} - \mathbf{T}^\#\|_F^2 / \|\mathbf{T}^\#\|_F^2$, for recovering a $200 \times 200$ matrix with various ranks, using $m$ 1-bit measurements sample for a range of different value of $m$.

choices of $\sigma$. Similar to the matrix case [31], using big noise deteriorated the results as we end up measuring the noise which in the 1-bit regime looks like a coin toss. Choosing small noise on the other hand limits the dithering effect which is essential for successful recovery.

Once we fix the choice of $\sigma$ we move on to comparing results for different ranks and tensor sizes. In particular, we present results for various rank $r \in \{3, 5, 10\}$ and various sample sizes $\frac{m}{N^2} \in \{0.1, 0.2, \ldots, 1\}$. Figure 4.2 shows the results of a 2-dimensional case where $\mathbf{T}^\# \in \mathbb{R}^{200 \times 200}$. The results are average relative error obtained from 20 runs of the same sample size $m$ and rank $r$. As expected the results are very similar to each other. In all the cases we can naturally observe that results are better when the underlying tensor has a smaller rank or when the percentage of observed
Figure 4.3: Average relative squared errors, $\frac{\|T_{\text{recovered}} - T^\sharp\|_F}{\|T^\sharp\|_F}$, for recovering a $30 \times 30 \times 30$ tensor with various ranks, using $m$ 1-bit measurements for a range of different value of $m$. Top left is the result of matricization, top right is CGD for exact-rank constraints, bottom left is the results using the factor-frob function as an estimate for nuclear-norm constraint and bottom right is the result of max-qnorm.

entries ($\frac{m}{N^d}$) increases. The CGD algorithm exploits the low-rank structure of the matrix instead of a convex surrogate of the rank and therefore gives a slightly better result when the sample size is large.

Next, we show the results for 3-dimensional tensors $T^\sharp \in \mathbb{R}^{30 \times 30 \times 30}$. A balanced matricization is not possible in this case and therefore, tensor completion has an added advantage when the tensor has an odd order. We can see this advantage very clearly in Figure 4.3. For smaller sampling rates, e.g., when $\frac{m}{N^d} = 0.3$, the average error of 1-bit tensor completion is three to four times better.
Figure 4.4: Average relative squared errors, $\frac{\| T_{\text{recovered}} - T^\sharp \|_F^2}{\| T^\sharp \|_F^2}$, for recovering a $15 \times 15 \times 15 \times 15$ tensor with various ranks, using $m$ 1-bit measurements for a range of different value of $m$. Left is the result of matricization, center is the result of using the factor-frob function as an estimate for nuclear-norm constraint and right is the result of max-qnorm.

The results of max-qnorm and factor-frob constrained completion are very similar. This suggests that similar sample complexity can be obtained by analyzing the factor-frob constrained tensor completion which we postpone to future work. Note that the way we choose the low-rank factors is an important factor that should be considered in any numerical studies. Interestingly, in both these cases, there is not much difference between observing 1-bit samples of the full tensors and observing half of them. This might be due to the fact that the ranks are small for observing the true power of 1-bit tensor completion. The CGD results start deteriorating once we move to higher orders. In particular, when the sample percentage is high, the CGD recovery shows some signs of over-training. In the case of max-qnorm and factor-frob, we used 10% of the samples to prevent over-fitting. However, a similar strategy can’t be used for the CGD unless we change the algorithm significantly, for example using cross-validation for the line-search part. Such modifications are beyond the scope of this chapter.

Figure 4.4 shows the results for $T^\sharp \in \mathbb{R}^{15 \times 15 \times 15 \times 15}$. We removed the CGD results as in all cases they get trapped in local minima and the results are meaningless. It is worth mentioning that although tensor completion results are significantly better than matricizing, the results show some irregular behavior when $r$ is larger.
4.8 Proofs

4.8.1 Proof of Theorem 39

Our proof closely follows the one in [18, Section 7.1]. Therefore, we just briefly explain the steps. In what follows we prove the max-qnorm constrained ML estimation error. The proof for the M-norm constraint case (4.9) follows the exact same steps where the only difference is a constant difference in the Rademacher complexity of the unit balls of these two norms. Define the loss function \( g(x; y) : \mathbb{R} \times \{ \pm 1 \} \rightarrow \mathbb{R} \) as:

\[
g(x; y) = \mathbbm{1}_y \log \left( \frac{1}{f(x)} \right) + \mathbbm{1}_{y = -1} \log \left( \frac{1}{1 - f(x)} \right).
\]

Regarding the tensor completion problem as a prediction problem where we consider the tensor \( T \in \mathbb{R}^{N_1 \times N_2 \times \cdots \times N_d} \) as a function from \([N_1] \times \cdots \times [N_d] \rightarrow \mathbb{R}\) where the value of the function at \((i_1, \ldots, i_d)\) is the corresponding entry, the proof is based on a general excess risk bound developed in [8].

For a subset \( S = \{ \omega_1, \omega_2, \ldots, \omega_m \} \) of the set \([N_1] \times \cdots \times [N_d]\) and a tensor \( T, D_S(T; Y) \) is the average empirical loss according to \( g \). Precisely, \( D_S(T; Y) = \frac{1}{m} \sum_{i=1}^{m} g(T(\omega_i); Y(\omega_i)) \). When the sample set \( S \) is drawn i.i.d. according to \( \Pi \) (with replacement), we define:

\[
D_\Pi(T; Y) := \mathbb{E}_{S \sim \Pi}[g(T(\omega_i); Y(\omega_i))] = \sum_{i=1}^{m} \pi_{\omega_i} g(T(\omega_i), Y(\omega_i)).
\]

Notice that since \( \hat{T}_{\text{max}} \) is the optimal solution of optimization problem (4.6) and \( T^\sharp \) is feasible for this problem we have:

\[
D_S(\hat{T}_{\text{max}}; Y) \leq D_S(T^\sharp; Y) = \frac{1}{m} \sum_{i=1}^{m} g(T^\sharp(\omega_i); Y(\omega_i))
\]  

(4.21)

Therefore, we have

\[
\mathbb{E}_Y[D_\Pi(\hat{T}_{\text{max}}; Y) - D_\Pi(T^\sharp; Y)]
\leq \mathbb{E}_Y[D_\Pi(\hat{T}_{\text{max}}; Y) - D_S(\hat{T}_{\text{max}}; Y)] + \mathbb{E}_Y[D_S(T^\sharp; Y) - D_\Pi(T^\sharp; Y)]
\leq \mathbb{E}_Y[D_\Pi(\hat{T}_{\text{max}}; Y) - D_S(\hat{T}_{\text{max}}; Y)] + \mathbb{E}_Y[D_S(T^\sharp; Y) - D_\Pi(T^\sharp; Y)]
\leq \sup_{T \in \mathcal{K}_{\text{max}}^\sharp(\alpha, R_{\text{max}})} \{ \mathbb{E}_Y[D_\Pi(T; Y)] - \mathbb{E}_Y[D_S(T; Y)] \} + \mathbb{E}_Y[D_S(T^\sharp; Y) - D_\Pi(T^\sharp; Y)].
\]  

(4.22)
Notice that the left-hand side of (4.22) is equivalent to weighted Kullback-Leibler divergence between \( f(T^\sharp) \) and \( f(\hat{T}_{\max}) \).

Now we focus on bounding the right-hand side of (4.22). Using Hoeffding’s inequality on the random variable \( Z_\omega := g(T^\sharp(\omega); Y(\omega)) - \mathbb{E}[g(T^\sharp(\omega); Y(\omega))] \) we conclude that with probability \( 1 - \delta \) over choosing the sampling subset \( S \):

\[
D_S(T^\sharp; Y) - D_{\Pi}(T^\sharp; Y) \leq U_\alpha \sqrt{\frac{\log(\frac{1}{\delta})}{2m}}.
\]  

(4.23)

Moreover, a combination of Theorem 8, (4) of Theorem 12 from [8] and the upper bound (10) and noting that \( g(.; y) \) is an \( L_\alpha \)-Lipschitz function yields:

\[
\sup_{T \in K_{\max}(\alpha, R_{\max})} \{ \mathbb{E}_Y [D_{\Pi}(T; Y)] - \mathbb{E}_Y [D_S(T; Y)] \} \leq 12L_\alpha R_{\max} c_1 c_2 d \sqrt{\frac{dN}{m}} + U_\alpha \sqrt{\frac{8\log(\frac{2}{\delta})}{m}}.
\]  

(4.24)

Gathering (4.22), (4.23), and (4.24) we get:

\[
\mathbb{K}_{\Pi}(f(T^\sharp)||f(\hat{T}_{\max})) \leq 12L_\alpha R_{\max} c_1 c_2 d \sqrt{\frac{dN}{m}} + U_\alpha \sqrt{\frac{8\log(\frac{2}{\delta})}{m}} + U_\alpha \sqrt{\frac{\log(\frac{1}{\delta})}{2m}}.
\]  

(4.25)

This, together with (4.4) and [31, Lemma 2] proves (4.8). The upper-bound (4.9) can be proved by following the exact same arguments with the only difference being in the right-hand side of (4.24) where the Rademacher complexity of unit M-norm ball does not have the constant \( c_1 c_2 d \) and the upper bound \( R_M \) is different than \( R_{\max} \).

### 4.8.2 Proof of Theorem 47

To prove this theorem it is more convenient to work with the function

\[
\mathcal{L}_{\Omega,Y}(X) = \mathcal{L}_{\Omega,Y}(X) - \mathcal{L}_{\Omega,Y}(0).
\]

**Lemma 58.** Let \( G \subset \bigotimes_{i=1}^d \mathbb{R}^N \) be

\[
G = \{ X \in \bigotimes_{i=1}^d \mathbb{R}^N : \|X\|_\ast \leq \alpha \sqrt{r^d N^d} \},
\]

85
for some \( r < N \) and \( \alpha \geq 0 \). Then

\[
P \left( \sup_{X \in G} |\mathcal{L}_{\Omega,Y}(X) - \mathbb{E}\mathcal{L}_{\Omega,Y}(X)| > C_d \alpha L \gamma \sqrt{r d} \left( \frac{\log^d(N) \sqrt{N}}{\sqrt{m}} + \frac{\log(N) \sqrt{N} d}{m} \right) \right) \leq \frac{1}{N^{d-1}},
\]

where the probability is both over \( \Omega \) and \( Y \).

Before proving the lemma, note that by assumption, \( T^\sharp \in G \) and from the definition of \( \hat{T}_{\text{nuc}} \), \( \hat{T}_{\text{nuc}} \in G \) and \( \mathcal{L}_{\Omega,Y}(\hat{T}_{\text{nuc}}) \geq \mathcal{L}_{\Omega,Y}(T^\sharp) \). Therefore,

\[
0 \leq \mathcal{L}_{\Omega,Y}(\hat{T}_{\text{nuc}}) - \mathcal{L}_{\Omega,Y}(T^\sharp) = \mathbb{E} \left[ \mathcal{L}_{\Omega,Y}(\hat{T}_{\text{nuc}}) - \mathcal{L}_{\Omega,Y}(T^\sharp) \right] - \mathbb{E} \left[ \mathcal{L}_{\Omega,Y}(\hat{T}_{\text{nuc}}) - \mathcal{L}_{\Omega,Y}(T^\sharp) \right] + 2 \sup_{X \in G} \left| \mathcal{L}_{\Omega,Y}(X) - \mathbb{E} \left[ \mathcal{L}_{\Omega,Y}(X) \right] \right|
\]

\[
= -mD(f(T^\sharp)||f(\hat{T}_{\text{nuc}})) + 2 \sup_{X \in G} \left| \mathcal{L}_{\Omega,Y}(X) - \mathbb{E} \left[ \mathcal{L}_{\Omega,Y}(X) \right] \right|.
\]

And therefore with probability bigger than \( 1 - \frac{1}{N^{d-1}} \)

\[
D(f(T^\sharp)||f(\hat{T}_{\text{nuc}})) \leq 2C_d \alpha L \gamma \sqrt{r d} \left( \frac{\log^d(N) \sqrt{N}}{\sqrt{m}} + \frac{\log(N) \sqrt{N} d}{m} \right) \tag{4.26}
\]

**Proof of Lemma 58:** This proof closely follows the proof of \([31, \text{Lemma 1}]\). Therefore we omit the steps that are exactly similar. Using Markov’s inequality for any \( h > 0 \)

\[
P \left( \sup_{X \in G} |\mathcal{L}_{\Omega,Y}(X) - \mathbb{E}\mathcal{L}_{\Omega,Y}(X)| \geq C_d L \gamma \alpha \sqrt{r} \left( \log^d(N) \sqrt{mN} + \log(N) \sqrt{N} d \right) \right) \leq \mathbb{E} \left[ \sup_{X \in G} \left| \mathcal{L}_{\Omega,Y}(X) - \mathbb{E}\mathcal{L}_{\Omega,Y}(X) \right|^h \right] \leq \left( C_d L \gamma \alpha \sqrt{r} \left( \log^d(N) \sqrt{mN} + \log(N) \sqrt{N} d \right) \right)^h \tag{4.27}
\]

Define \( E \) to be a tensor whose entries are i.i.d. Rademacher random variables and \( \Delta_{\Omega} \) to be the
indicator tensor for $\Omega$. Then we have [31]

$$
\mathbb{E}\left[\sup_{X \in G} |\mathcal{L}_{\Omega,Y}(X) - \mathbb{E}\mathcal{L}_{\Omega,Y}(X)|^h\right] \leq (4L_\gamma)^h \mathbb{E}\left[\sup_{X \in G} |\langle \Delta \ast E \ast Y, X \rangle|^h\right]
= (4L_\gamma)^h \mathbb{E}\left[\sup_{X \in G} |\langle E \ast \Delta, X \rangle|^h\right].
$$

(4.28)

Next, we need to bound the spectral norm of $E \circ \Delta \Omega$ whose entries are i.i.d. zero-mean random variables. For this we define $B := E \circ \Delta \Omega$ and use [92, Theorem 3] to bound $\|B\|$:

**Theorem 59.** Let $B \in \otimes_{i=1}^d \mathbb{R}^{N_i}$ be a random order-$d$ tensor, whose entries are independent zero-mean, random variables. For any $\lambda \leq \frac{1}{64}$, assume $1 \leq h \leq 2d \lambda \ln\left(\frac{5e}{\lambda}\right)$. Then,

$$
\mathbb{E}(\|B\|^h)^{\frac{1}{h}} \leq c_d \sqrt{\ln\left(\frac{1}{\lambda}\right)} \left(\left\lfloor \log_2\left(\frac{1}{\lambda}\right) \right\rfloor - 1\right)^{\frac{d}{2}} \left(\sum_{j=1}^d \mathbb{E}_B \alpha_j^h\right)^{\frac{1}{h}} + \sqrt{\lambda N(\mathbb{E}_B \beta^h)^{\frac{1}{h}}},
$$

(4.29)

where $\alpha_j := \max_{i_1, \ldots, i_{j-1}, i_{j+1}, \ldots, i_d} \left(\sum_{j=1}^N B_{i_1, \ldots, i_{j-1}, i_j, i_{j+1}, \ldots, i_d} \right)$, $\beta := \max_{i_1, \ldots, i_d} |B_{i_1 \ldots i_d}|$ and $c_d \leq c8^{\frac{d}{2}} \sqrt{d}$.

Next, for fixed $i_1, \ldots, i_{j-1}, i_{j+1}, \ldots, i_d$, we define $\delta_j := B_{i_1, \ldots, i_{j-1}, i_j, i_{j+1}, \ldots, i_d}$, where $\mathbb{E} \delta_j = \frac{m}{N^d}$ and $|\delta_j| \leq 1$. Using a Bernstein’s inequality for $t > \frac{6m}{N^d}$,

$$
\mathbb{P}\left(\left|\sum_{j=1}^N \left(\delta_j - \frac{m}{N^d}\right)\right| > t\right) \leq 2 \exp(-t) = 2\mathbb{P}(W_j > t),
$$

(4.30)

where $W_1, \ldots, W_N$ are i.i.d. exponential variables.

Following the same line of argument in [31], using $\mathbb{E}q = \int_0^\infty \mathbb{P}(q \geq t)$ for positive random variables
we get

\[
\left( \mathbb{E} \left[ \max_{i_1, \ldots, i_d} \left( \sum_{j=1}^{N} \delta_j \right)^2 \right] \right)^{\frac{1}{2}} \leq \sqrt{\frac{mN}{N^d}} + \left( \mathbb{E} \left[ \max_{i_1, \ldots, i_d} \left( \sum_{j=1}^{N} \left( \delta_j - \frac{m}{N^d} \right)^2 \right) \right] \right)^{\frac{1}{2}}
\]

\[
\leq \sqrt{\frac{m}{N^d-1}} + \left( \frac{6m}{N^d-1} \right)^{\frac{1}{2}} + \frac{1}{2} \int_{\frac{6m}{N^d-1}}^{\infty} \mathbb{P} \left( \max_{i_1, \ldots, i_d} \sum_{j=1}^{N} \left( \delta_j - \frac{m}{N^d} \right)^h \geq t \right) dt
\]

\[
\leq \sqrt{\frac{m}{N^d-1}} + \left( \frac{6m}{N^d-1} \right)^{\frac{1}{2}} + 2 \int_{\frac{6m}{N^d-1}}^{\infty} \mathbb{P} \left( \max_{i_1, \ldots, i_d} W_j^h \geq t \right) dt
\]

\[
\leq \sqrt{\frac{m}{N^d-1}} + \left( \frac{6m}{N^d-1} \right)^{\frac{1}{2}} + 2 \mathbb{E} \left[ \max_{i_1, \ldots, i_d} W_j^h \right]^{\frac{1}{2}}
\]

\[
\leq (1 + \sqrt{6}) \sqrt{\frac{m}{N^d-1}} + 2 \frac{1}{2} \left( \sqrt{\log(N^d-1)} + 2 \frac{1}{2} \sqrt{h} \right).
\]

Above, in the first line we have used triangle inequality, in the second line we have used Jensen’s inequality, in the third line, we have used (4.30), and in the fifth line, we have used the following inequality for the exponential random variables \( W_j \):

\[
\mathbb{E} \left[ \max_{i_1, \ldots, i_d} W_j^h \right] \leq \mathbb{E} \left[ \max_{i_1, \ldots, i_d} W_j - \log(N^d-1) \right]^h + \log^h(N^d-1) \leq 2h! + \log^h(N^d-1)
\]

Now choosing \( h = \left[ \log(N^d-1) \right] > 1 \) we can bound

\[
\mathbb{E}_B \alpha_j^h = \left( \mathbb{E} \left[ \max_{i_1, \ldots, i_d} \left( \sum_{j=1}^{N} \delta_j \right)^2 \right] \right)^{\frac{1}{2}} \leq (1 + \sqrt{6}) \sqrt{\frac{m}{N^d-1}} + (2 + \sqrt{2})(\sqrt{(d-1)\log(N)}).
\]

Plugging above into (4.29) and choosing \( \lambda = \frac{C_\lambda}{N} \) we get:

\[
(\mathbb{E}\|B\|_p^h)^{\frac{1}{2}} \leq c_d \sqrt{\ln(N)} \left( \log_2(N)^{d-1} \right)^p \left( 1 + \sqrt{6} \right) \sqrt{\frac{m}{N^d-1}} + (2 + \sqrt{2})(\sqrt{(d-1)\log(N)}) + \frac{C_\lambda}{N},
\]

where we used the fact that for \( p < 1 \), \((a+b)^p < a^p + b^p\) for \( a, b > 0 \). Moreover, \( C_\lambda \) is chosen such that \( \frac{C_\lambda}{N} < \frac{1}{64} \).
By simplifying the constant we can write
\[(E\|B\|^h)^{\frac{1}{h}} \leq C_d\log^d(N)\sqrt{\frac{m}{N^{d-1}}} + C_d''\log(N), \tag{4.32}\]
where \(\max(C_d', C_d'') \leq c_d 8^d \sqrt{2((2 + \sqrt{2})d^2 + C_d)}.\) Plugging (4.32) into (4.28) we get:
\[E\left[\sup_{X \in \mathcal{G}} |\hat{L}_{\Omega,Y}(X) - \mathbb{E}\hat{L}_{\Omega,Y}(X)|^h\right] \leq \left(4L\alpha\sqrt{r}(C_d\log^d(N)\sqrt{mN} + C_d''\log(N)\sqrt{N^d})\right)^h. \tag{4.33}\]
Plugging this into (4.27), we conclude
\[\mathbb{P}\left(\sup_{X \in \mathcal{G}} |\hat{L}_{\Omega,Y}(X) - \mathbb{E}\hat{L}_{\Omega,Y}(X)| \leq C_dL\alpha\sqrt{r}(\log^d(N)\sqrt{mN} + \log(N)\sqrt{N^d})\right), \tag{4.33}\]
with probability bigger than \(1 - \frac{1}{N^{d-1}}%\) provided that \(C_d > 12 \max(C_d', C_d'').\)
The proof of Theorem \[47\] immediately follows (4.26) and the following lemma [31, Lemma 2].

**Lemma 60.** Let \(\|T\|_\infty, \|\hat{T}\|_\infty \leq \alpha.\) Then
\[D(f(T)\|\hat{f}^2) \geq \inf_{|\xi| \leq \alpha} \frac{(\hat{f}(\xi))^2}{8f(\xi)(1-f(\xi))} \frac{\|T - \hat{T}\|_\infty^2}{N^d}. \tag{4.34}\]

### 4.8.3 Proof of Theorem \[51\]
Define \(F_{\Omega,Y}(X) := -L_{\Omega,Y}(X) = -\sum_{\omega \in \Omega} (\mathbb{I}_{Y_\omega=1}\log(f(X_\omega))) + \mathbb{I}_{Y_\omega=-1}\log(1-f(X_\omega))).\) Further, assume \(\|T\|_\infty \leq \alpha\) and remember that
\[
\gamma_\alpha \leq \min \left\{ \inf_{|x| \leq \alpha} \left\{ \frac{f^2(x)}{f^2(x) - \hat{f}(x)} \right\}, \inf_{|x| \leq \alpha} \left\{ \frac{f^2(x)}{1 - f^2(x)} + \frac{\hat{f}(x)}{f(x)(1-f(x))} \right\} \right\},
\]
\[
L_\alpha \geq \sup_{|x| \leq \alpha} \left\{ \frac{\hat{f}(x)}{f(x)(1-f(x))} \right\}. \tag{4.35}\]
Next define \(C_{\text{rank}}(r, \alpha) := \{ T \in \otimes_{i=1}^d \mathbb{R}_N : \|T\|_\infty \leq \alpha, \text{rank}(T) \leq r \}.

Our proof extends the proof of [11, Theorem 3.1], and [93, Theorem 3.3] to the case of multi-dimensional arrays and the main methodology is similar. In what follows \( \omega = \{i_1, \ldots, i_d\}, \) determines an index of the tensor where \( 1 \leq i_j \leq N. \) For any \( \Omega, Y, F, \) and \( f \) as defined above, we
have
\[ \frac{\partial F_{\Omega,Y}(X)}{\partial T_w} = \left( -\frac{\dot{f}(X_\omega)}{f(X_\omega)} \mathbb{I}_{\{Y_\omega=1\}} + \frac{\dot{f}(X_\omega)}{1-f(X_\omega)} \mathbb{I}_{\{Y_\omega=-1\}} \right) \mathbb{I}_{(\omega \in \Omega)}, \] (4.36)

and
\[ \frac{\partial^2 F_{\Omega,Y}(X)}{\partial X_{\omega}^2} = \left[ \left( \frac{\dot{f}^2(X_\omega) - \dot{f}(X_\omega)}{f^2(X_\omega)} \right) \mathbb{I}_{\{Y_\omega=1\}} + \left( \frac{\dot{f}(X_\omega)}{1-f(X_\omega)} - \frac{\dot{f}^2(X_\omega)}{(1-f(X_\omega))^2} \right) \mathbb{I}_{\{Y_\omega=-1\}} \right] \mathbb{I}_{(\omega \in \Omega)}, \] (4.37)

and \( \frac{\partial^2 F_{\Omega,Y}(X)}{\partial x_{\omega_1} \partial x_{\omega_2}} = 0 \) for \( \omega_1 \neq \omega_2 \).

Let \( \theta^z = \text{vec}(T^z) \in \mathbb{R}^{N^d} \) and \( \tilde{F}_{\Omega,Y}(\theta^z) = F_{\Omega,Y}(T^z) \). The objective function \( F_{\Omega,Y}(T) \) is continuous in \( T \). However, unlike the matrix case the set \( C_{\text{rank}}(r, \alpha) \), defined above, is not necessarily compact. In particular, there exists a sequence of rank-\( r \) tensors that converge to a tensor with rank larger than \( r \) for any \( r > 1 \). This phenomenon raises the definition of the border rank of a tensor which has been extensively studied in the literature. A tensor \( T \) is border-rank \( r \) if it is \( \text{rank}-r\text{-approximable} \) \([13, 34]\). Define \( S_r := \{ X \in \bigotimes_{i=1}^d \mathbb{R}^{N^d} | \text{rank}(X) = r \} \).

**Definition 61.** A tensor \( T \) is border-rank-\( r \) if \( T \in \overline{S}_r \) and \( T \notin \overline{S}_{r-1} \), where \( \overline{S}_r \) is the closure of \( S_r \).

Now consider the set \( \overline{C}_{\text{rank}}(r, \alpha) \). This set is compact and therefore, there exists a tensor in the set \( \overline{C}_{\text{rank}}(r, \alpha) \) that minimizes \( F \). This tensor is either rank-\( r \) or border-rank-\( r \). In either case for any small \( \varepsilon \) there exists a rank-\( r \) tensor \( \hat{T} \), where \( \tilde{F}_{\Omega,Y}(\hat{\theta}) \leq \tilde{F}_{\Omega,Y}(\theta^z) + \varepsilon \), where \( \hat{\theta} = \text{vec}(\hat{T}) \).

By the second-order Taylor’s theorem, for any \( \theta \in \mathbb{R}^{N^d} \), we have:
\[ \tilde{F}_{\Omega,Y}(\theta) = \tilde{F}_{\Omega,Y}(\theta^z) + \langle \nabla_\theta \tilde{F}_{\Omega,Y}(\theta^z), \theta - \theta^z \rangle + \frac{1}{2} \langle \theta - \theta^z, (\nabla^2_{\theta \theta} \tilde{F}_{\Omega,Y}(\tilde{\theta})) (\theta - \theta^z) \rangle, \] (4.38)

where \( \tilde{\theta} = \theta^z + \gamma(\theta - \theta^z) \) for some \( 0 \leq \gamma \leq 1 \). Plugging \( \hat{\theta} \) in (4.38), we get
\[ \varepsilon \geq \langle \nabla_\theta \tilde{F}_{\Omega,Y}(\theta^z), \hat{\theta} - \theta^z \rangle + \frac{1}{2} \langle \hat{\theta} - \theta^z, (\nabla^2_{\theta \theta} \tilde{F}_{\Omega,Y}(\tilde{\theta})) (\hat{\theta} - \theta^z) \rangle. \] (4.39)

Next, we study the two terms in the right-hand side and use them to bound \( \| \theta^z - \hat{\theta} \|_F \). The first one is \( \langle \nabla_\theta \tilde{F}_{\Omega,Y}(\theta^z), \hat{\theta} - \theta^z \rangle \). Define \( X \) to be the tensor where \( \text{vec}(X) = \hat{\theta} \) and let \( \zeta := \text{vec}(X - T^z) = \hat{\theta} - \theta^z \). Then
\[ \langle \nabla_\theta \tilde{F}_{\Omega,Y}(\theta^z), \zeta \rangle = \langle Z, X - T^z \rangle, \] (4.40)
where $Z := \nabla_{\Omega} f(T^\sharp)$. Using (4.36) we have

$$Z_\omega = \left(-\frac{f'(T_\omega^\sharp)}{f(T_\omega^\sharp)} \mathbb{1}_{(Y_\omega = 1)} + \frac{f'(T_\omega^\sharp)}{1 - f(T_\omega^\sharp)} \mathbb{1}_{(Y_\omega = -1)}\right) \mathbb{1}_{(\omega \in \Omega)}.$$

Notice that the entries of $Z$ are independent random variables with $|Z_\omega| \leq L\alpha$. In both [11, 93], the authors have used the inequality $\langle Z, X - T^\sharp \rangle \leq \|Z\| \|X - T^\sharp\|_*$ to bound this inner product. This bound is not optimal in the tensor case for the same reasons we explained in Remark 49. Therefore, here we use the M-norm and its dual to bound this inner product. Notice that here we can use max-qnorm as well with the expense of the constant $c_1 c_2^d$ and larger bounds on the norm of the low-rank tensor. First notice that

$$\langle Z, X - T^\sharp \rangle \leq \|Z\|_M^* \|X - T^\sharp\|_M,$$

where $\|Z\|_M^*$ is the dual norm of $\|Z\|_M$, i.e., $\|Z\|_M^* = \sup_{\|U\|_M \leq 1} |\langle Z, U \rangle| = \sup_{U \in T_\pm} |\langle Z, U \rangle|$. Therefore assuming $\Omega = \{\omega_1, \cdots, \omega_m\}$,

$$\|Z\|_M^* \leq \left[ \sup_{\|U\|_M \leq 1} \left| \sum_{\omega \in \Omega} Z_\omega U(\omega) \right| \right] = \left[ \sup_{U \in T_\pm} \left| \sum_{i=1}^m Z(\omega_i) U(\omega_i) \right| \right],$$

where $Z(\omega_i)$ is a sub-exponential random variable with sub-exponential constant $L\alpha$, i.e., $\max_{i=1, \cdots, m} \mathbb{E}[\exp(\frac{|Z(\omega_i)|}{L\alpha})] \leq e$. Notice that $|T_\pm| \leq 2^{Nd}$. Using Bernstein-type inequality for sub-exponential random variables [124], yields

$$\mathbb{P}\left\{ \left| \frac{1}{m} \sum_{i=1}^m Z(\omega_i) U(\omega_i) \right| \geq t \right\} \leq 2\exp\left\{ -c \cdot \min\left( \frac{mt}{L\alpha^2}, \frac{mt}{L\alpha} \right) \right\}.$$

Using $t = \sqrt{\frac{Nd}{m}}$ for $m > Nd$ and a union bound on all $U \in T_\pm$ we get

$$\sup_{U \in T_\pm} \left| \sum_{i=1}^m Z(\omega_i) U(\omega_i) \right| \leq cL\alpha \sqrt{mNd},$$

with probability bigger than $1 - \exp(-Nd)$ and for some absolute constant $c > 0$. Plugging this into (4.41) we get

$$\|Z\|_M^* \leq cL\alpha \sqrt{mNd}.$$

The following lemma summarizes the result above and might be of independent interest.
Lemma 62. Consider a \( d \)-dimensional random tensor \( Z \in \otimes_{i=1}^{d} \mathbb{R}^{N} \) with \( m \) non-zero entries, whose entries are independent sub-exponential random variables taking values in \([-1, 1]\), with sub-exponential constant \( K \). Then
\[
\|Z\|_{M}^{*} \leq c \sqrt{mNd},
\]
with probability bigger than \( 1 - \exp(-Nd) \). Moreover, by the same probability, the dual of the max-qnorm can be bounded in a similar manner but with an order-dependent constant:
\[
\|Z\|_{\max}^{*} \leq C_{d} \sqrt{mNd},
\]
And finally going back to bounding the first term of (4.38)
\[
\langle \nabla_{\theta} \tilde{F}_{\Omega,Y}(\theta^{z}), \zeta \rangle = \langle Z, X - T^{z} \rangle \leq \|Z\|_{M}^{*} \|X - T^{z}\|_{M} \leq cL_{\alpha} \sqrt{mNd}(2r)^{2d/3} 2^{2/3} 2\alpha,
\]
where in the last step we used Theorem 11 along with the fact that \( \text{rank}(X - T^{z}) \leq 2r \) and \( \|X - T^{z}\|_{\infty} \leq 2\alpha. \)

Remark 63. As explained above, both [11] and [93] use nuclear norm and its dual, spectral norm to bound the inner product \( \langle Z, X - T \rangle \). However, in the tensor case, using nuclear norm doesn’t give us optimal bounds. Ignoring the log powers for now, in (4.31) we bounded the spectral norm of a random matrix roughly by \( c_{1} \log(N) + c_{2} \sqrt{\frac{m}{N^{d-1}}} \). In the optimal setting of \( m \sim N \log(N) \), when \( d = 2 \), \( \log(N) \) and \( \sqrt{\frac{m}{N^{d-1}}} \) have the same order, but when we move to \( d > 2 \) the spectral norm is dominated by the log factor which prevents the factor \( \sqrt{\frac{m}{N^{d-1}}} \) from cancelling the \( \sqrt{N^{d}} \) term that comes up in the nuclear bound and this is the main reason that an exact generalization of the method in [93] is not applicable here.

Next, we bound the second term of (4.38), i.e., \( \frac{1}{2} \langle \theta - \theta^{z}, (\nabla_{\theta}^{2} \tilde{F}_{\Omega,Y}(\check{\theta}))(\theta - \theta^{z}) \rangle \). A straightforward generalization of [11, Lemma 6.3] results in the following lemma.

Lemma 64. Let \( \zeta = \text{vec}(X - T^{z}) = \check{\theta} - \theta^{z} \). Then for \( \check{\theta} = \theta^{z} + \gamma(\check{\theta} - \theta^{z}) \), where \( 0 \leq \gamma \leq 1 \), we have
\[
\langle \zeta, [\nabla_{\theta}^{2} \tilde{F}_{\Omega,Y}(\check{\theta})] \zeta \rangle \geq \gamma \alpha \|X - T^{z}\|^{2}_{\tilde{F}}.
\]
Here, \( [(X)_{\Omega}]_{\omega} = X_{\omega} \) if \( \omega \in \Omega \) and zero otherwise.

Now we need to prove that a fixed low-rank tensor is close to its sampled versions. By matricizing \( X - T^{z} \), we end up with a \( N^{d/2} \times N^{d/2} \) matrix whose rank is less than \( 2r \), and then we can use
the result of [91] which proves that the set of low-rank matrices has the strong convexity property with high probability, i.e.,

$$\frac{\| (X - T^\sharp)\Omega \|_F}{\sqrt{\| \Omega \|}} \geq C \frac{\| X - T^\sharp \|_F}{\sqrt{N^d}}. \quad (4.43)$$

However, it also forces the additional condition that $|\Omega| \geq CN^d \log(N)$ which is not optimal. Therefore, we need a similar upper bound on $\| (X - T^\sharp)\Omega \|_F$ that does not force the extra sampling condition. In order to do this, notice that rank($X - T^\sharp$) $\leq 2r$ and $\| X - T^\sharp \|_\infty \leq 2\alpha$, and therefore, $\| X - T^\sharp \|_M \leq (2r)^{3d/2 - 3/2} 2\alpha$. Next, we show that with high probability, tensors with low M-norm have a similar property to the one in (4.43). The following lemma is a direct generalization of the lower bound step used in [19, Section 6.1]; for completeness we include the proof in Section 4.8.5.

**Lemma 65.** Defining

$$C(\beta, \delta) = \{ X \in \bigotimes_{i=1}^d \mathbb{R}^{N^i} : \| X \|_\infty \leq 1, \| X \|_M \leq \beta, \frac{\| X \|^2_F}{N^d} \geq \delta \},$$

for a tensor $X \in C(\beta, \delta)$, with probability bigger than $1 - 2\exp(-c_0 m \delta)$

$$\frac{1}{m} \| X\Omega \|^2_F \geq \frac{1}{2N^d} \| X \|^2_F - C_1 \beta \sqrt{\frac{N^d}{m}} \quad \text{for all } X \in C(\beta, \delta). \quad (4.44)$$

provided $m > \frac{\log^2 c_0 \delta}{c_0 \delta}$.

Applying (4.52) to $X - T^\sharp$ with appropriate rescaling, we get

$$\frac{1}{m} \| (X - T^\sharp)\Omega \|^2_F \geq \frac{1}{2N^d} \| X - T^\sharp \|^2_F - C_1 \frac{3d}{2} - \frac{3}{2} (2\alpha)^2 \sqrt{\frac{N^d}{m}}, \quad \text{if } \frac{\| X - T^\sharp \|^2_F}{N^d} \geq (2\alpha)^2 \delta$$

with probability $1 - 2\exp(-c_0 m \delta)$. Picking $\delta = \frac{\log(\frac{2}{\epsilon})}{c_0 m}$, this means that with probability $1 - t$ either

$$\frac{\| X - T^\sharp \|^2_F}{N^d} \leq (2\alpha)^2 \frac{\log(\frac{2}{\epsilon})}{c_0 m}, \quad (4.45)$$

or

$$\frac{1}{m} \| (X - T^\sharp)\Omega \|^2_F \geq \frac{1}{2N^d} \| X - T^\sharp \|^2_F - C_1 (2r)^{3d - 3} (2\alpha)^2 \sqrt{\frac{N^d}{m}}. \quad (4.46)$$

In the second case, gathering the results of (4.38), (4.42), Lemma 64, and (4.46) we get

$$\epsilon \geq -cL\alpha \sqrt{mN^d (2r)^{3d - 3} 2\alpha} + \frac{1}{2} \gamma a \left( \frac{m}{2N^d} \| X - T^\sharp \|^2_F - C_1 (2r)^{3d - 3} (2\alpha)^2 \sqrt{mN^d} \right).$$
As $\varepsilon$ is arbitrary we can simplify the above result as
\[
\frac{\|X - T^i\|_F^2}{N^d} \leq \frac{c\alpha L_\alpha (2r) \frac{d}{2} - \frac{1}{2}}{\gamma a} \sqrt{Nd} + C_1 (2r) \frac{d}{2} - \frac{1}{2} (2\alpha)^2 \sqrt{Nd} \frac{m}{m}.
\] (4.47)

Combining (4.45) and (4.47) with $t = 2\exp(-Nd)$ finishes the proof.

4.8.4 Proof of Theorem 54

**Lemma 66** ([45, Lemma 18]). Let $r = \lfloor \frac{R}{\alpha K_G}\rfloor$ and let $K_M^T(\alpha, R)$ be defined as in Section 4.6 and let $\gamma \leq 1$ be such that $\frac{r}{\gamma^2}$ is an integer and suppose $\frac{r}{\gamma^2} \leq N$. Then there exists a set $\chi_T^T(\alpha, \gamma) \subset K_M^T(\alpha, R)$ with
\[
|\chi_T^T(\alpha, \gamma)| \geq \exp \frac{rN}{16\gamma^2}
\]
such that

- For $T \in \chi_T^T(\alpha, \gamma)$, $|T(\omega)| = \alpha \gamma$ for $\omega \in \{[N] \times [N] \cdots [N]\}$.
- For any $T(i), T(j) \in \chi_T^T(\alpha, \gamma)$, $T(i) \neq T(j)$
\[
\|T(i) - T(j)\|_F^2 \geq \frac{\alpha^2 \gamma^2 N^d}{2}
\]

Next choosing $\gamma$ in a way that $\frac{r}{\gamma^2}$ is an integer and
\[
\frac{4\sqrt{2}\varepsilon}{\alpha} \leq \gamma \leq \frac{8\varepsilon}{\alpha}
\]
we transform $\chi_T^T(\alpha, \gamma)$ so that the entries of each element of the packing set come from the set $
\{\alpha, \alpha' := (1 - \gamma)\alpha\}$ by defining
\[
\chi := \{X + \alpha(1 - \gamma)\frac{1}{2} : X \in \chi_T^T(\alpha, \gamma)\}
\]
Notice that for any $X \in \chi$, $\|X\|_\infty \leq \alpha$ and $\|X\|_M \leq R_M$ and therefore, we can use $\chi$ as a packing set for $K_M^T$. Once the packing sets are generated, using the exact same proof as [31, Section A.3], we can prove an upper bound. Choose $\varepsilon$ to be
\[
\varepsilon^2 = \min \left( \frac{1}{1024}, C_2 \alpha \sqrt{\frac{B_{3\gamma}}{2}} \sqrt{\frac{rN}{m}} \right).
\] (4.48)
And assume there is an algorithm that, using 1-bit measurements \( Y_\Omega \) of \( X \), returns \( \hat{X} \) such that

\[
\frac{1}{N^d} \|X - \hat{X}\|_F^2 \leq \varepsilon^2,
\]

with probability at least \( \frac{1}{4} \). Notice that this means that with probability at least \( \frac{1}{4} \), we can distinguish the elements of the packing set as for any \( X^{(i)}, X^{(j)} \in \chi \), \( \frac{1}{N^d} \|X^{(i)} - X^{(j)}\|_F^2 \geq 4\varepsilon^2 \). Hence, defining \( X^* \) to be the closest element of \( \chi \) to \( \hat{X} \), if we show that \( P(X \neq X^*) > \frac{3}{4} \) the proof is done. Otherwise, using Fano’s inequality [31] and following the steps in Section A.3.2 of [31] we have

\[
\frac{1}{4} \leq 1 - P(X \neq X^*) \leq 16\gamma^2 \left( \frac{64m\varepsilon^2}{\beta_{\alpha'}} + 1 \right) \leq 1024\varepsilon^2 \left( \frac{64m\varepsilon^2}{\beta_{\alpha'}} + 1 \right),
\]

where \( \alpha' := (1 - \gamma)\alpha \). Comparing \( 64m\varepsilon^2 \) and \( \beta_{\alpha'} \), either

\[
\frac{1}{4} < \frac{2048\varepsilon^2}{\alpha^2 rN},
\]

which is a contradiction when \( C_0 > 8 \) or

\[
\varepsilon^2 > \frac{\alpha\sqrt{\beta_{\alpha'}}}{512\sqrt{2}} \sqrt{\frac{rN}{m}},
\]

which is a contradiction when \( C_2 \leq \frac{1}{512\sqrt{2}} \).

### 4.8.5 Proof of Lemma 65

In this section, we prove that with high probability, \( \frac{1}{m} \sum_{i=1}^{m} T(\omega_i)^2 \) doesn’t deviate much from its expectation \( \|T\|_{\Pi}^2 \). For ease of notation, define \( T_S = (T(\omega_1), T(\omega_2), \cdots, T(\omega_m)) \) to be the set of chosen samples drawn from \( \Pi \) where

\[
\|T\|_{\Pi}^2 = \frac{1}{m} \|E_{S \sim \Pi}T_S\|_2^2 = \sum_{\omega} \pi_\omega T(\omega)^2.
\]

Following [19], we prove that with high probability over the samples,

\[
\frac{1}{m} \|T_S\|_2^2 \geq \frac{1}{2} \|T\|_{\Pi}^2 - f_\beta(m, N, d), \tag{4.49}
\]

holds for all all tensors \( T \in C(\beta, \delta) := \{ T \in K^T_M(1, \beta) : \|T\|_{\Pi}^2 \geq \delta \} \) uniformly for some \( \beta \geq 1 \) and...
As suggested in [19], we prove the stronger result that

\[ \frac{1}{m} \| T \|_2^2 - \| T \|_{\Pi}^2 \leq \frac{1}{2} \| T \|_{\Pi}^2 - f_\beta(m, N, d), \]

holds for all \( T \in C(\beta, \delta) \) with high probability. Following the peeling argument of [91], for \( \ell = 1, 2, \cdots \) and \( \theta = \frac{3}{2} \), define a sequence of subsets

\[ C_\ell(\beta, \delta) := \{ T \in C(\beta, \alpha) : \theta^{\ell-1} \delta \leq \| T \|_{\Pi}^2 \leq \theta^\ell \delta \}, \]

and for any radius \( D > 0 \), set

\[ B(D) := \{ T \in C(\beta, \delta) : \| T \|_{\Pi}^2 \leq D \}. \]

If there exist some \( T \in C(\beta, \delta) \) such that

\[ \frac{1}{m} \| T \|_2^2 - \| T \|_{\Pi}^2 > \frac{1}{2} \| T \|_{\Pi}^2 + f_\beta(m, N, d) \]

then there exists a corresponding \( \ell \) such that \( T \in C_\ell(\beta, \delta) \subset B(\theta^\ell \delta) \) and

\[ \frac{1}{m} \| T \|_2^2 - \| T \|_{\Pi}^2 > \frac{1}{2} \theta^\ell \delta + f_\beta(m, N, d). \] (4.50)

And therefore, the main task becomes proving that the probability of the event (4.50) happening is small. To this end, we prove the following lemma that shows that the empirical mean \( \frac{1}{m} \| T \|_2^2 \) doesn’t deviate much from its expectation.

**Lemma 67.** Defining \( \Delta_D(S) := \sup_{T \in B(D)} \frac{1}{m} \| T \|_2^2 - \| T \|_{\Pi}^2 \), there exists a constant \( C_1 \) such that for any \( D > 0 \),

\[ \Pr \{ \Delta_D(S) > \frac{D}{3} + C_1 \beta \sqrt{\frac{Nd}{m}} \} \leq e^{-\frac{md}{16}}. \]

**Proof:** The proof of this lemma follows the proof in [19, Section 6.1.1] which is based on some basic techniques of probability in Banach spaces, including symmetrization, contraction inequality and Bousquet’s version of Talagrand concentration inequality [77].

Regarding the tensor \( T \) as a function in \([N] \times \cdots \times [N] \rightarrow \mathbb{R}\) from its indices to \( \mathbb{R} \), we are interested in bounding the following empirical process:

\[ \sup_{f_T : T \in B(D)} \left| \frac{1}{m} \sum_{i=1}^m f_T(\omega_i) - \mathbb{E}[f_T(\omega_i)] \right|, \quad f_T(\cdot) = \{ T(\cdot) \}^2. \]
Bounding the variance $f_T$ for $T \in B(D)$ and using a standard symmetrization argument

$$
\mathbb{E}_{S \sim \Pi}[\Delta_D(S)] \leq 2\mathbb{E}_{S \sim \Pi}\{\mathbb{E}_{e\sim \Pi}\{\sup_{T \in B(D)} \frac{1}{m} \sum_{t=1}^{m} e_t T(\omega_t)^2 | S}\},
$$

where $\{e_t\}$ is an i.i.d. Rademacher sequence, independent of $S$. Next, noticing that $|T(\omega_t)| \leq 1$ and using Ledoux-Talagrand contraction inequality [77]

$$
\mathbb{E}_e\{\sup_{T \in B(D)} | \frac{1}{m} \sum_{t=1}^{m} e_t T(\omega_t)| | S\} \leq 2\mathbb{E}_e\{\sup_{\|T\|_{\max} \leq \beta} | \frac{1}{m} \sum_{t=1}^{m} e_t T(\omega_t)| | S\} \leq 24\beta \sqrt{\frac{dN}{m}}.
$$

Notice that in the last step we have used Lemma 10 which is a uniform bound on worst-case Rademacher complexity of low M-norm tensors. Hence, $\mathbb{E}_{S \sim \Pi}[\Delta_D(S)] \leq 48\beta \sqrt{\frac{dN}{m}}$ and using Bousquest’s version of Talagrand concentration inequality for empirical processes indexed by bounded functions completes the proof of the Lemma [? ].

Define $f_\beta(m,N,d) := 48\beta \sqrt{\frac{Nd}{m}}$ and consider the sequence of events

$$
\varepsilon_\ell = \{\Delta_{\theta^\ell \delta} > \frac{1}{3} \theta^\ell \delta + f_\beta(m,N,d)\} \text{ for } \ell = 1,2,\ldots .
$$

Since $C(\beta, \delta) = \cup_{\ell \geq 1} C_\ell(\beta, \delta)$, using the union bound we have:

$$
\mathbb{P}\{\exists T \in C(\beta, \delta), \text{ subject to } | \frac{1}{m}\|T S\|_2^2 - \|T\|_{\Pi}^2 | > \frac{1}{2} \|T\|_{\Pi}^2 + f_\beta(m,N,d)\}
\leq \sum_{\ell \geq 1} \mathbb{P}\{\exists T \in C_\ell(\beta, \delta), \text{ subject to } | \frac{1}{m}\|T S\|_2^2 - \|T\|_{\Pi}^2 | > \frac{1}{2} \|T\|_{\Pi}^2 + f_\beta(m,N,d)\}
\leq \sum_{\ell = 1}^{\infty} \mathbb{P}(\varepsilon_\ell) \leq \frac{\exp(-c_0 m \delta)}{1 - \exp(-c_0 m \delta)},
$$

(4.51)

with $c_0 = \log(\frac{3}{10})$. Consequently provided $m > \frac{\log 2}{c_0 \delta}$ we obtain

$$
\frac{1}{m}\|T S\|_2^2 \geq \frac{1}{2} \|T\|_{\Pi}^2 - C_1 \beta \sqrt{\frac{Nd}{m}} \text{ for all } T \in C(\beta, \delta)
$$

(4.52)

with probability greater than $1 - 2\exp(-c_0 m \delta)$ which finishes the proof of lower bound.
Chapter 5

Applications

As explained before, in many applications data are naturally multi-dimensional and the inter-dimensional dependencies are best explained when we use tensors to represent them. In this chapter, we show the results of applying tensor completion and 1-bit tensor completion to various applications ranging from imaging to recommender systems. First, we show applications of tensor completion in Section 5.1 and then we show some applications of 1-bit tensor completion in Section 5.2.

In the results presented in this chapter, we concentrate on recovery errors and do not report their time complexity. Tensor completion algorithms are generally slower than matrix completion algorithms, which is the price we pay for the better sample complexity obtained by not matricizing. Generally our algorithms are about 5 to 10 times slower than their matrix completion counterparts. The algorithms presented in Chapters 3 and 4 are prototypes to show the advantage of using max-qnorm instead of matricizing and they are not optimized to be as fast as they can be. therefore, in this chapter, we just report the recovery errors and do not do a study of the time complexity of the algorithms.

5.1 Applications of tensor completion

5.1.1 Color image inpainting

Color images are usually represented by three-dimensional tensors. The first two dimensions determine the pixel location and the third dimension is the rgb (red, green and blue) values. Although the small length of the third dimension means that the gain over matricization will be small,
Figure 5.1: The Baboon image and the RSE of best rank-$r$ approximations of the $128 \times 128 \times 3$ tensor and its $128 \times 384$ matricization.

our experiments give a nice demonstration of the advantage of using tensors instead of matricizing. To this end, we first consider the Baboon image in Figure 5.1 which is represented by a $128 \times 128 \times 3$ tensor. In Figure 5.1, we also show the relative squared error (RSE) of the best rank-$r$ approximation of the tensor (approximate) and the best rank-$r$ approximation of its matricization along the first dimension (which is exact). For a tensor $T^z$, the RSE of an estimate $\hat{T}$ is defined as

$$RSE := \frac{\|T^z - \hat{T}\|^2_F}{\|T^z\|^2_F}.$$  

Remember that any rank-$r$ tensor approximation is also a rank-$r$ matrix approximation of the matricized tensor and therefore, as expected, the RSE of rank-$r$ matrix approximations are better than the rank-$r$ tensor approximation. It is also worth noting that a tensorized low-rank matrix does not necessarily have a small tensor-rank.

Figure 5.2 shows the results of applying tensor and matrix completion algorithms to the subsampled images in the first column which has 50% sampling rate. We show the Peak signal-to-noise ratio (PSNR) which is a measure of error for image recovery and is defined as

$$PSNR := 10\log\left(\frac{N_1N_2N_3\|T^z\|_\infty^2}{\|\hat{T} - T^z\|_F^2}\right).$$

There are two methods of subsampling which significantly affect the recovery performance. One is missing pixels, i.e., missing all the third dimension information (rgb values) of a pixel of
Figure 5.2: Reconstructed Baboon images with 50% observed entries. The first row shows the results with missing pixels and the second row shows the results with missing random entries. The results of matricization are worse than using tensors and the results of missing pixels are worse than missing random entries of the tensor.

the image. The first row of Figure 5.2 shows this case. The second method is randomly removing indices of the tensor which is illustrated in the second row of 5.2. In the first row, when we miss a pixel, we miss the interdependencies of the values in the third dimension and therefore, the results of using tensors and matricization are very similar. On the other hand, in the second row, the algorithm learns the relation between the rgb valued and used that information for the missing indices and therefore, the results are better. Moreover, the improvement using tensors is more significant in this case.

Comparing the results of Figures 5.1 and 5.2 shows the main reason behind the success of tensor completion compared to matrix completion. Although the matricized version is more low-rank (or is better approximated by low-rank objects), because it has larger dimension we see that the completion results are better when we don’t matricize.

It is worth mentioning that tensor completion has been used for color image inpainting in
Figure 5.3: Results for the Peppers image with only 20% sampling rate.

Our results are comparable (or better) than the most of the algorithms in the literature except for the algorithm in [10] which is a sophisticated image and video completion algorithm. We have not tried implementing those ideas here and postpone it to future work.

In Figure 5.3, we show the results for the Peppers image (which is a 128 by 128 pixels image) with the lower observation sampling rate $p = 0.2$. The results are similar to the ones of the Baboon image.

In Figure 5.4, we show the results for a larger image which is obtained from the Berkeley Segmentation database [86]. In the second row we show the results of some of the state of the art algorithms including LRTC [83], TCTF [134] and t-SVD [132].

---


2[https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/](https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/)
5.1.2 Video completion

Videos can be naturally represented by a tensor where the frames of the video are arranged in the first two (black and white videos) or three (color videos) dimensions and the last dimension determines the frame number. In addition to the fact that each frame is low rank, the video does not change drastically from one frame to the other which makes the video tensor approximately low-rank. This is especially true for videos where objects in the video are more static. An example of such a video is the bridge video\(^3\) where we show one of its frames in Figure 5.5. We consider 100 frames of the video and remove 80% and 95% of the samples randomly. The SNR of the recoveries with \(p = 0.2\) and \(p = 0.05\) is respectively 69 and 57.

In Figure 5.6 we have considered another video which is the rotting tomato video used in [83]. We randomly sample 30% of the pixels where the missing pixels contain blocks of pixels as well (see Figure 5.6). This video is a little harder to reconstruct compared to the bridge video in Figure 5.5 because the tomato gets smaller and smaller as it gets rotten and also we are missing blocks of pixels. Therefore, the results are a little inferior in this case (SNR of the recovery is around 30 dB) but still very good.

\(^3\)Available at http://trace.eas.asu.edu/yuv
**Figure 5.5:** Completing a black and white video from 20% and 5% of the entries.

**Figure 5.6:** Frames 50 and 100 of the tomato video [83]. The subsampling mask includes blocks of pixels as well.
5.1.3 Hyperspectral imaging

Hyperspectral data consist of collections of measurements sensed from contiguous spectral bands [112]. Unlike the human eye that generally sees light in three bands (long wavelengths - perceived as red, medium wavelengths - perceived as green, and short wavelengths - perceived as blue), hyperspectral imaging records images of an object from a wide range of bands (wavelengths). Considering the object as a 2-dimensional object the resulting image can be stored into a 3-dimensional tensor whose bands (third dimension slices) are highly correlated. There are multiple reasons that make a hyperspectral image contaminated. For example, images taken by satellites have missing information due to faulty sensors and the environment situation (such as cloudy sky) [63, 79]. In figure 5.7 we consider the hyperspectral image of [83]4, where 90% of its samples are removed. The missing entries include random pixels and also blocks of pixels that model the effect of a cloud. An example of the observed data is shown in Figure 5.7. In the second row, we add random white noise with $\text{SNR} = 20\text{dB}$ to the image and show the reconstruction of the 30th band. Notice that in this case, we can improve the results by denoising as well (around 1dB in this case).

Figure 5.7: An example of a hyperspectral image with 90% missing entries. In the second row, we show the result with added white noise.

---

4Available at http://www.cs.rochester.edu/u/jliu/publications.html
5.1.4 MRI data recovery

The next application we consider here is recovering dynamic MR images. Figure 5.8 shows a few slices of the original 3D brain MRI data which is in $\mathbb{R}^{191 \times 400 \times 190}$. We randomly remover 95%, 80%, and 50% of the entries and show the sampled and recovered slices in each case. The PSNR for $p = 0.05$, $p = 0.2$ and $p = 0.5$ is 29, 37 and 76 respectively. The results with $p = 0.5$, i.e., the last row is practically full recovery of the image (SNR $\sim$ 80). For $p = 0.2$, it also does a good job of recovering the image except for small blurriness in the corner of the brain image. The main difficulty arises in the final slices where the brain images get smaller; because we only sample 20% of the entries our algorithm does not get enough information to recover those slices.
Figure 5.8: The recovery results for four slices of a brain MR image. The columns are images of different slices and the rows show the original image and different sampling rates. First row shows the original images. Each image in the figure shows the sampled image (top) and the recovered image (bottom).
5.2 Applications of 1-bit tensor completion

In this section, we test the performance of 1-bit tensor completion on real-world data, in particular, for predicting exam scores and user ratings. We show the improvements gained by solving 1-bit tensor completion instead of matricizing. We transform the ratings (or scores) to 1-bit data by replacing them with whether they are above or below some approximate mean-rating. Next, we investigate the ability of 1-bit tensor completion to predict if an observed rating is above or below average. In all the applications we know the maximum value of the ratings and use this value to rescale the data to have unit infinity norm. This helps us in choosing the appropriate function $f$ based on our synthetic experiments in Section 4.7, i.e., Gaussian noise with $\sigma = 0.1$. In Remark 68 (below) we explain the consequences of this choice in more details.

To be more precise, we explain the general application setup we use in this section briefly. Assume $T^\sharp$ is the true tensor and we have access to a subset of its entries for training ($T^\sharp_{\Omega_{\text{train}}}$) and another subset of the entries for testing ($T^\sharp_{\Omega_{\text{test}}}$). $\Omega_{\text{train}}$ and $\Omega_{\text{test}}$ are two disjoint subsets of the entries. Below is a brief recipe for the experiments done in this section.

- Scale the observed entries to have unit infinity norm.
- Using an approximate mean-value $\eta$, take $\text{sign}(T^\sharp_{\Omega_{\text{train}}} - \eta)$ to be the 1-bit measurements. Notice that with synthetic low-rank tensors we add a dithering noise using some function $f$ before taking the sign (see (4.1)). However, our experiments showed that assuming the noise to be intrinsic in the data gives slightly better results than dithering. To be precise instead of dithering the observation tensor $T^\sharp_{\Omega_{\text{train}}}$ by a noise tensor $Z$ and taking the 1-bit measurements to be $\text{sign}(T^\sharp_{\Omega_{\text{train}}} + Z_{\Omega_{\text{train}}} - \eta)$, we assume the noise is hidden in the original ratings. We explain the complications of not knowing the exact noise in Remark 68. The same approach was taken in [31].
- Use a 1-bit tensor completion algorithm to get an initial estimate $\hat{T}_{\text{init}}$.
- Add the approximate mean-value $\eta$ and scale the resulting tensor back to get the final $\hat{T}$, which is an approximation of $T^\sharp$.
- Evaluate $\hat{T}$ by comparing $\hat{T}$ and $T^\sharp$ on the test indices.
Remark 68. In the applications, we empirically observe that we get better predictions if we do not dither the original tensor, i.e., if we assume the noise to be implicit in the data. Moreover, as explained in Remark 56, in the applications section we also ignore the infinity-norm constraint. Therefore, in theory, changing the value of $\sigma$ should not change the results of predicting the sign of the tensor. For example, if we consider recovering the original tensor with two noise functions $f_1(x) = \Phi(\frac{x}{\sigma_1})$ and $f_2(x) = \Phi(\frac{x}{\sigma_2})$, we would have

$$\Phi(\frac{T(\omega)}{\sigma_1}) = \Phi(\frac{\sigma_1 T(\omega)}{\sigma_2}),$$

which shows that the recovered tensors achieved by $f_1$ and $f_2$ should be scalar multiples of each other and the sign predictions should be the same. In Figure 5.9 we illustrate this by showing the results of applying noise functions with different values of $\sigma$ for recovering a rank-5 tensor in $\mathbb{R}^{30 \times 30 \times 30}$ whose 1-bit measurement have been obtained with $\sigma = 0.15$. The plot in the left shows the average RSE of 1-bit max-qnorm 1-bit tensor completion and as expected the best recovery of the original tensor is obtained when we use the true noise function that we originally used to get the observations. However, the right plot shows that the percentage of correct sign predictions is very robust to the choice of $\sigma$ and all the results are very close to each other which supports the above discussion.

5.2.1 Score prediction

In this section, we apply our algorithm to the data on pupil attainments in schools in Scotland which contains the information of 3,435 children who attended 148 primary schools and 19 secondary schools in Scotland. We generate a 5-dimensional tensor in $\mathbb{R}^{148 \times 19 \times 2 \times 4 \times 10}$ which includes the information of primary school, secondary school, gender, social class and attainment score of the students and estimate the verbal reasoning score of the students based on varying number of 1-bit information (from 230 samples to 2100 samples). The scores are ranged from -40 to 40 and we take the 1-bit information to be the sign of the score. We use 10\% of the observations for cross-validating the choice of parameters and another 10\% of the total scores as a test set. Figure 5.10.a shows the percentage of correct sign predictions and Figure 5.10.b shows the mean absolute error

$$MAE := \frac{\sum_{\omega \in \Omega_t}|T(\omega) - \hat{T}(\omega)|}{|\Omega_t|}$$

\footnote{Available at http://www.bristol.ac.uk/cmm/learning/mmsoftware/data-rev.html}
Figure 5.9: The left figure shows the average relative error of recovering a rank-5, $30 \times 30 \times 30$ tensor by max-qnorm constrained 1-bit tensor completion with different values of $\sigma$ when the 1-bit measurements were obtained by using $\sigma = 0.15$. The right figure shows the percentage of correct sign predictions.

on the test set. Notice that the scores are in the range of -40 to 40 and the Mean absolute error is in the range of 8 to 10. The matrix completion results refer to matricizing the tensor to a $592 \times 380$ matrix by putting the first and fourth dimension in the rows and the rest in the columns. This matricization results in the most balanced rearrangement which is the recommended way for better results [90]. In both figures we can see that using tensor completion outperforms matrix completion significantly.

5.2.2 In-car music recommender system

In this section, we apply our algorithm to an in-car music recommender system [6] which contains 4000 ratings from 42 users and 140 music tracks. The ratings are acquired via a mobile application and each rating is accompanied with one of the 26 context information which ranges from the landscape to the weather condition or the driver’s mood. The resulting tensor is of size $42 \times 140 \times 26$ and we use 3200 ratings for the minimization, 400 ratings for validating the choice of max-qnorm bound and the other 400 ratings as test data. Table 5.1 shows the results using tensor completion while considering the context information and matrix completion. The table shows the correct 1-bit predictions made by each algorithm considering their original ratings in the first six columns and the total average in the last row. Bringing context into the reconstruction results in
Figure 5.10: Results of applying 1-bit matrix and tensor completion to partially observed verbal scores tensor [87] to determine whether unobserved scores are above or below average. The left figure shows the percentage of correct predictions and the right figure shows the absolute mean error. The scores are in the range [-40 40]

an impressive improvement of at least 17% and moreover, using 1-bit information does not change the results too much compared to using the full information. Note that the results are averaged over 10 different random training and test sets.

Remark 69. Although the best results are obtained by the conditional gradient descent (Algorithm 3) used for exact-rank constraints, it is worth mentioning that the exact rank constraint is highly sensitive to the algorithm parameters and the results change drastically from one sampling pattern to another. As explained before, to reduce these artifacts, we had to run the code with multiple initial points which results in the algorithm being around 10 times slower than its counterparts. All this being said, in theory, the CGD should give the best results if the original tensor is low-rank.

Remark 70. The results of using the factor-frob algorithm (4.20) are also reported in the last two lines of the table which is better than matricization but a little worse than using max-qnorm. The close outcome of using max-qnorm and factor-frob can be a result of the fact that the underlying rank is not too small in this case. However, we do not have a precise explanation of the recovery error of the factor-frob algorithm.

Remark 71. In [31], a similar experiment was done with movie ratings which showed a significant improvement in using 1-bit data instead of original (multi-bit) data. However, here we just see a small improvement. We believe this is due to the very careful way the in-car music data was gathered. To be precise 1-bit matrix (and tensor) completion seems to be working better when
Table 5.1: Results of a comparison between 1-bit and multi-bit matrix and tensor completion algorithms on incar-music data [6] for predicting whether the unobserved ratings were above or below average. The multi-bit methods use the original ratings from 0 to 5 and the context-aware methods include the context information such as time of the day, weather, location of driving and mood.

<table>
<thead>
<tr>
<th>Method</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-bit matrix completion</td>
<td>75%</td>
<td>65%</td>
<td>54%</td>
<td>51%</td>
<td>70%</td>
<td>65%</td>
<td>60%</td>
</tr>
<tr>
<td>multi-bit matrix completion</td>
<td>76%</td>
<td>62%</td>
<td>50%</td>
<td>54%</td>
<td>53%</td>
<td>61%</td>
<td>57%</td>
</tr>
<tr>
<td>1-bit TC, max-qnorm</td>
<td>80%</td>
<td>89%</td>
<td>58%</td>
<td>65%</td>
<td>78%</td>
<td>85%</td>
<td>77%</td>
</tr>
<tr>
<td>multi-bit TC, max-qnorm</td>
<td>80%</td>
<td>86%</td>
<td>60%</td>
<td>64%</td>
<td>77%</td>
<td>90%</td>
<td>76%</td>
</tr>
<tr>
<td>1-bit TC, factor-frob</td>
<td>82%</td>
<td>85%</td>
<td>57%</td>
<td>69%</td>
<td>71%</td>
<td>80%</td>
<td>75%</td>
</tr>
<tr>
<td>multi-bit TC, factor-frob</td>
<td>84%</td>
<td>78%</td>
<td>57%</td>
<td>63%</td>
<td>67%</td>
<td>85%</td>
<td>72%</td>
</tr>
<tr>
<td>1-bit TC, CGD</td>
<td>82%</td>
<td>89%</td>
<td>60%</td>
<td>63%</td>
<td>80%</td>
<td>88%</td>
<td>78%</td>
</tr>
</tbody>
</table>

there is an implicit noise in the ratings given by the users which can’t be accounted for when we fit the data exactly.

Remark 72. It is generally true that the prediction should be more accurate when the original rating is further away from the average. This trend can still be seen in Table 5.1 as well except for one case which is due to the very few instances of 0-rating in the test data (generally 3 to 5 ratings).

5.2.3 A Restaurant Context-Aware Recommender System

Next, we apply our algorithm to a restaurant recommender system [102]. The data was gathered by asking 50 users who went to 40 restaurants in the city of Tijuana and rated their experience through a Web-based platform. The contextual information added is the day of the week (midweek and weekend) and the place (school, home and work). The resulting tensor is in $\mathbb{R}^{40 \times 50 \times 6}$ and the ratings are ranged from 1 to 5 with an average of 3.40. Similar to the In-car music recommender system we take the 1-bit information to be whether or not a rating is higher than the average or not. In Table 5.2 we report the percentage of correct above-or-below-average predictions for different ratings in the test set which we will refer to as 1-bit predictions. To be precise we show what percentage of the test ratings have been correctly predicted to be above or below average. In Table 5.3 we show the mean absolute error of the recovered tensor on the test set, i.e., taking $\Omega_t$ to be the test set, we show $MAE := \frac{\sum_{\omega \in \Omega_t} |T(\omega) - \hat{T}(\omega)|}{|\Omega_t|}$. For this restaurant-ratings data set, a naive matricization results in an overall rating of 60% and a more complicated matrix completion version would have repeated user-restaurant ratings (with different context information) and therefore, comparing tensor and matrix completion results is not possible here. Instead, we focus on a few important questions regarding 1-bit tensor completion.

111
• The first question is the importance of using max-qnorm or factor-frob regularizer. In the first two rows of Tables 5.2 and 5.3, we show the results of factorization (without any regularizer) with $r = 10$, i.e., by doing alternating minimization over $N_i \times 10$ factors ($r = 10$ is the best rank found by numerous empirical experiments). Both results of 1-bit prediction and the mean absolute error is generally worse than using max-qnorm and factor-frob results where the difference is significant in the 1-bit predictions. It is worth mentioning though that the tensor factorization is a non-convex problem and a more complex algorithm might get better results. For example the more sophisticated CGD algorithm gives better results but still using a regularizer is better improves the results in this question.

• The second question we investigate is the effect of choosing $f$ (in 4.5) on the results. The 4th and 5th row of Tables 5.2 and 5.3 shows the results for two different values of $\sigma$. In short, when we use max-qnorm, smaller noise (not too small though) result in better 1-bit predictions and larger noise (not too large though) results in better prediction of the actual score. This behavior can be explained by examining the likelihood function 5.3 more closely. When $\sigma$ is small the dithering function $f$ is spiky around $X(\omega) = 0$ and therefore it is more sensitive to whether or not $X(\omega)$ is positive or negative rather not how large it is. Therefore, using larger values of $\sigma$ does a better job of recovering the original rating but has less sensitivity to the sign of $X(\omega)$. This can be seen in Tables 5.2 and 5.3 where we recover the sign of the ratings better when we use $\sigma = 0.1$ but do worse in terms of mean absolute error. This is more evident in Table 5.3 where using smaller $\sigma$ does a better job of recovering the ratings that are close to the average and worse in the ratings that are further from the average. Remark 68 investigates these differences in more details.

As expected the best results for recovering the original ratings (MAE in Table 5.3) is obtained by exact tensor completion. However, notice that 1-bit TC outperform exact tensor completion for ratings that are around the mean significantly and struggles for the correct scale of the ratings that are far away from the mean.
Table 5.2: Results of a comparison between 1-bit and multi-bit matrix tensor completion algorithms on Tijuana restaurant data [102] for predicting whether the unobserved ratings were above or below average. TF refers to using tensor factorization without the max-qnorm regularization. For the 1-bit results we use \( f(x) = \Phi(\frac{x}{\sigma}) \).

<table>
<thead>
<tr>
<th>Original rating</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>multi-bit TC (TF)</td>
<td>73%</td>
<td>58%</td>
<td>63%</td>
<td>43%</td>
<td>73%</td>
<td>74%</td>
</tr>
<tr>
<td>1-bit TC (TF)</td>
<td>70%</td>
<td>67%</td>
<td>74%</td>
<td>57%</td>
<td>79%</td>
<td>77%</td>
</tr>
<tr>
<td>multi-bit TC (max-qnorm)</td>
<td>85%</td>
<td>72%</td>
<td>69%</td>
<td>63%</td>
<td>89%</td>
<td>81%</td>
</tr>
<tr>
<td>1-bit TC (max-qnorm, ( \sigma = 0.1 ))</td>
<td>80%</td>
<td>72%</td>
<td>72%</td>
<td>69%</td>
<td>92%</td>
<td>84%</td>
</tr>
<tr>
<td>1-bit TC (max-qnorm, ( \sigma = 0.5 ))</td>
<td>79%</td>
<td>69%</td>
<td>69%</td>
<td>66%</td>
<td>91%</td>
<td>82%</td>
</tr>
<tr>
<td>multi-bit TC (factor-frob,)</td>
<td>87%</td>
<td>72%</td>
<td>66%</td>
<td>64%</td>
<td>90%</td>
<td>83%</td>
</tr>
<tr>
<td>1-bit TC (factor-frob, ( \sigma = 0.1 ))</td>
<td>86%</td>
<td>72%</td>
<td>74%</td>
<td>63%</td>
<td>89%</td>
<td>82%</td>
</tr>
<tr>
<td>1-bit TC (factor-frob, ( \sigma = 0.5 ))</td>
<td>82%</td>
<td>72%</td>
<td>75%</td>
<td>62%</td>
<td>89%</td>
<td>83%</td>
</tr>
<tr>
<td>1-bit exact-rank (CGD)</td>
<td>83%</td>
<td>75%</td>
<td>74%</td>
<td>66%</td>
<td>86%</td>
<td>80%</td>
</tr>
</tbody>
</table>

Table 5.3: Results of a comparison between 1-bit and multi-bit matrix tensor completion algorithms on Tijuana restaurant data [102] showing the mean absolute error.

<table>
<thead>
<tr>
<th>Original rating</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>multi-bit TC (TF)</td>
<td>1.80</td>
<td>1.22</td>
<td>0.74</td>
<td>0.88</td>
<td>1.10</td>
<td>0.97</td>
</tr>
<tr>
<td>1-bit TC (TF)</td>
<td>1.90</td>
<td>1.02</td>
<td>0.62</td>
<td>0.85</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td>multi-bit TC (max-qnorm)</td>
<td>1.5</td>
<td>1.01</td>
<td>0.54</td>
<td>0.64</td>
<td>0.68</td>
<td>0.76</td>
</tr>
<tr>
<td>1-bit TC (max-qnorm, ( \sigma = 0.1 ))</td>
<td>2.20</td>
<td>1.25</td>
<td>\textbf{0.29}</td>
<td>\textbf{0.46}</td>
<td>1.25</td>
<td>1.11</td>
</tr>
<tr>
<td>1-bit TC (max-qnorm, ( \sigma = 0.5 ))</td>
<td>1.52</td>
<td>1.12</td>
<td>1.08</td>
<td>1.14</td>
<td>0.84</td>
<td>1.02</td>
</tr>
<tr>
<td>multi-bit TC (factor-frob,)</td>
<td>1.41</td>
<td>1.02</td>
<td>0.53</td>
<td>0.70</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>1-bit TC (factor-frob, ( \sigma = 0.1 ))</td>
<td>2.20</td>
<td>1.23</td>
<td>0.27</td>
<td>0.51</td>
<td>1.33</td>
<td>1.16</td>
</tr>
<tr>
<td>1-bit TC (factor-frob, ( \sigma = 0.5 ))</td>
<td>1.49</td>
<td>1.03</td>
<td>0.95</td>
<td>1.00</td>
<td>0.82</td>
<td>0.96</td>
</tr>
<tr>
<td>1-bit exact-rank (CGD)</td>
<td>1.95</td>
<td>1.07</td>
<td>0.41</td>
<td>0.53</td>
<td>1.07</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Chapter 6

Conclusion

In this thesis, we studied the problem of tensor completion from noisy or 1-bit measurements. Below is a list of our contributions in this thesis and some of the open questions related to them.

• We define the M-norm and max-qnorm of tensors as robust proxies for the rank of a tensor. We prove that both M-norm and max-qnorm of a bounded low-rank tensor are upper-bounded by quantities that depend only on the rank and infinity norm of the tensor and are independent of $N$.

• We use a generalization of Grothendieck’s theorem to connect the max-qnorm of a tensor to its nuclear decomposition with unit infinity-norm factors. Using this, we prove that the unit ball of the max-quasi-norm is contained in a constant expansion of the set of rank-1 sign tensors (the unit ball of M-norm) and therefore, the sets of low max-qnorm and low M-norm have small Rademacher complexity. This also establishes a theoretical framework for further investigation of low max-qnorm tensors.

• We prove that with high probability, $m = O(r^{3d}dN)$ (or $m = O(R^2dN)$ if M-norm is bounded by $R$) samples are sufficient to estimate a rank-$r$ bounded tensor using a convex least squares algorithm. Moreover, we derive an information-theoretic lower bound that proves $m = O(R^2N)$ measurements are necessary for recovery of tensors with M-norm less than $R$. This proves that our bound is optimal both in its dependence on $N$ and the M-norm bound $R$. It is worth mentioning though, that the bounds we prove in this thesis are not necessarily optimal in $r$, the rank of the tensor, which is an interesting open problem.

• Through synthetic numerical examples, we illustrate the advantage of using algorithms designed for max-qnorm constrained tensor completion instead of algorithms using matriciza-
These algorithms significantly improve algorithms based on matricization and alternating least squares (ALS). It is worth mentioning that computing the nuclear norm of a general tensor is known to be NP-hard. Although it is not known whether computing the M-norm or max-qnorm of a tensor is NP-hard or not, our numerical results for max-qnorm constrained least squares, using a simple projected quasi-Newton algorithm give promising results. There are two important future directions related to our algorithm. First, is designing methods to approximate the solution of the M-norm constrained tensor completion and second is designing faster algorithms for approximating the max-qnorm constrained algorithms. One straight-forward future direction is solving the max-qnorm constrained tensor completion in parallel (the max-qnorm constrained problems are embarrassingly parallel).

- We formulate and analyze 1-bit tensor completion using M-norm constraints on the underlying tensor. We prove that, with high probability, the mean squared error (MSE) of recovering a rank-$r$ tensor $T^\sharp$ from $m$ 1-bit measurements by solving an M-norm constrained optimization is $O(\sqrt{r^{d-3}} \sqrt{N/m})$. Moreover, we analyze a related non-convex function, max-qnorm, and prove that MSE of optimizing a log-likelihood function constrained by max-qnorm is $O(\sqrt{r^{d-1}} \sqrt{N/m})$.

- We consider using nuclear norm as a convex constraint for the 1-bit tensor completion problem. The theoretical sample complexity we obtain by using nuclear norm is not as tight as using M-norm and we show that with the tools we have used here (or in the matrix-case) there is no hope of getting tighter results. We believe that achieving a tighter upper bound is only possible if we impose some incoherence conditions on the problem. Studying the difference between the tensor nuclear-norm and M-norm is an interesting future direction.

- We also analyze 1-bit tensor completion using exact-rank constraints. The estimation error relies on bounding the M-norm of rank-$r$ tensors and an upper bound on the dual-M-norm of a random tensor which might be of independent interest. However, we think achieving an optimal lower bound for this case depends on proving a restricted strong convexity type inequality for the set of low-rank tensors which is important future work. Notice that [91] studies the matrix case.

- We derive an information-theoretic lower bound on the recovery error of M-norm constrained 1-bit TC that proves the MSE of any arbitrary algorithm is at least $\Omega(r \sqrt{N/m})$ for a rank-$r$ tensor and $\Omega(R \sqrt{N/m})$ for a tensor with M-norm less than $R$. This proves that our upper bound is optimal in $N$ and the M-norm bound $R$ (but not necessarily in $r$).
• We propose a numerical method to approximate the solution of max-qnorm constrained 1-bit tensor completion and show its advantage over 1-bit matrix completion using synthetic and real-world data. Specifically, we show the huge improvement one can get by applying the 1-bit tensor completion to context-aware recommender systems.
Bibliography


121


[80] X. Li, Y. Ye, and X. Xu. Low-rank tensor completion with total variation for visual data inpainting. In Thirty-First AAAI Conference on Artificial Intelligence, 2017. → pages [101]


125


126