Mathematics of Matrix Completion and Numerical Studies

Marwa Musa Babiker Adam (marwa@aims.ac.za)
African Institute for Mathematical Sciences (AIMS)

Supervised by: Prof. Montaz Ali
University of the Witwatersrand, Johannesburg, South Africa

19 May 2020

Submitted in partial fulfillment of a structured masters degree at AIMS South Africa

AIMS | African Institute for Mathematical Sciences SOUTH AFRICA

Abstract

In this research project, we have considered a well-known problem of practical interest: the recovery of a data matrix from a sampling of its entries. The problem is known as the matrix completion problem. The problem arises in a great number of applications, including the famous Netflix Prize, and in collaborative filtering. In general, accurate recovery of a matrix from a small number of entries is impossible but it is well known that the unknown matrix has low rank, which makes this problem solvable. Other requirements are incoherence conditions on the singular vectors of the unknown matrix. We have introduced these and other properties that are needed to obtain an exact recovery for the matrix via rank minimization. The convex formulation is done with the help of nuclear norm formulation. A more tractable formulation of the problem has been carried out via semidefinite programming formulation of the problem. We have also thoroughly presented a number of theoretical results that are needed for unique solution. Results of some numerical experiments have been presented.
Keywords: Matrix completion, convex optimization, sparse matrices, incoherent subspaces, Bernoulli model, nuclear norm minimization.

Declaration

I, the undersigned, hereby declare that the work contained in this research project is my original work, and that any work done by others or by myself previously has been acknowledged and referenced accordingly.

Marwa Musa Babiker Adam, 19 May 2020

Contents

1. Introduction

Matrix completion problem is an important problem with applications on many areas, such as data science, compressed sensing, business, management and biology. It is also called Netflix prize problem, if we consider a data (rating) matrix $\boldsymbol{S}$ with dimension $n_1 \times n_2$, in which the rows represent users or customers n_1 and the columns represents movies or services n_2. We know that the customers are going to rate the movies that they have watched, and the result is a sparse matrix. The question is, are we going to find a linear map $\mathcal{R}$ that maps the matrix $\boldsymbol{S}$ to a complete matrix, i.e. what is the possibility to recover the matrix sufficiently from a small set of it's entries. Moreover, what are the conditions that we need to have on the matrix $\boldsymbol{S}$ given entries. The recovery of any matrix from a small set of it's entries is impossible in general; but knowing that the matrix is a low rank matrix makes a difference.
We will consider that the matrix is a low rank matrix or approximately low rank matrix, since not all of the elements are contributing to the main evaluations. Hence, the size of this matrix is obviously very large and, since each user rates only a few movies, there are many entries of the matrix that are missing, and Netflix is interested in predicting rates. Now, it turns out that only few factors determine a user's preference in movies (e.g. genre, lead actor/actresses, director, year, etc.), that is, there is a relatively small number of "types" of people with respect to movie preferences. Also, customers who agreed on movies ratings in the past will be likely to agree in the future. For these reasons it is a natural assumption to consider that the Netflix matrix is low rank.
Surprisingly, we can use the matrix completion in image, video and signal proccesing. For instance, given an image or a signal which is sparse, then, we can use matrix completion to recover the image or the signal. Also, if we consider the number of words comparing to the number of letters, each word can be represented using some of the 28 letters, so the matrix that represents them will be a low rank matrix, since it will depend in just 28 letters. To have an intuition, if we think about an event, we need to have just a few information, which describe the important contents. We can think of the low rank as clustering problem, since each cluster will have similar features.

1.1 Background

J.Candes and Recht, 2009 have established that it will be sufficient to recover the low rank matrix from a subset of its entries. Moreover, J.Candes et al., 2011 state that, any matrix can be decomposed as a sparse matrix and a low rank matrix. In some applications recovery of low rank matrix is needed while in other applications the recovery of sparse matrix is needed. In this research we have taken the latter case, where we want to recover a low rank matrix given a sparse matrix. The real world problems will contain noise and that makes the problem more challenging. Moreover, Candes and Plan, 2010 have also discussed the recovery of a real data matrix which has a noise. Furthermore, it has been guaranteed in Benjamin Recht, 2010 that we can minimize the rank of the matrix by minimizing the nuclear norm since the problem of minimzing the rank of the matrix is an NP-hard problem. We have first written the ranking minimization in terms of nuclear norm which is convex. We then further convert the problem into semidefinite program which is also convex.
As discussed in J.Candes and Recht, 2009, we cannot hope to guarantee the recovery of any matrix from a subset of its entries. For instance, if the matrix is a 1-rank matrix where it has just 1 non zero row or column, it will be difficult to be recovered. Also, we need to know at least one information about each user and each movie, i.e, we need to have at least one non-zero entry in each row and column. Furthermore, if the original matrix has just a few non-zero entries, then we cannot guarantee the exact recovery.

1.2 Outlines

In chapter 2, we have presented all necessary mathematical preliminaries including convex function, various vector and matrix norms, Bernoulli model, coherence and semidefinite programming (SDP) problem. In Chapter 3 we have addressed some conditions and concepts such as the optimality condition, Golfing scheme which are essential for the existence of dual certificate. In Chapter 4 we have introduced the main result which is a consequence of some lemmas and the proof of some of these lemmas are given.

The notations that we will be using frequently are bold upper case letters for matrices and vector spaces and bold lower case letters for vectors, and capital letters for constants. The transpose of matrix and vector has been denoted by T. Instead of using plenty of letters for constants we may use C but it doesn't mean that it is the same constant everywhere. Moreover, for simplicity we will consider that the matrices are square matrices but the discussion extends to rectangular matrices easily.

2. Fundamentals

In this chapter we introduce some definitions, notations, inequalities and concepts needed in chapter 3.

2.1 Convex Functions

2.1.1 Convex function Osnaga, 2014.

A set $\boldsymbol{A} \subset \mathbb{R}^n$ is a convex set, if for any two points $x, y \in \boldsymbol{A}$ the line that connects them is also in $\boldsymbol{A}$, which means

$$\forall x, y \in \boldsymbol{A}, \lambda \in [0,1], \quad \lambda x + (1-\lambda)y \in \boldsymbol{A}. \tag{2.1.1}$$

Moreover, a real valued function $h : \boldsymbol{A} \to \boldsymbol{A}$ with $\boldsymbol{A} \subset \mathbb{R}^n$, is a convex function if $\forall x, y \in \boldsymbol{A}$ and $\lambda \in [0,1]$

$$h(\lambda x + (1-\lambda)y) \leq \lambda h(x) + (1-\lambda)h(y). \tag{2.1.2}$$

2.1.2 Norm function.

Meyer (2000) A real valued function on a vector space $\boldsymbol{V}$ $h : V \to \mathbb{R}^+$ is called norm function if $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $a \in \mathbb{R}$ the following conditions are satisfied:

1. $h(\boldsymbol{x}) > 0$, and $h(\boldsymbol{x}) = 0$ if and only if $\boldsymbol{x} = 0$ (positive definite).
2. $h(a\boldsymbol{x}) = |a|h(\boldsymbol{x})$ (absolutely homogeneous).
3. $h(\boldsymbol{x} + \boldsymbol{y}) \leq h(\boldsymbol{x}) + h(\boldsymbol{y})$ (triangle inequality).

Therefore, from the function norm conditions 2 and 3 we can conclude that any norm function is convex. Further, we will introduce some norms. The q-norm of a vector $\boldsymbol{x} \in \mathbb{R}^n$ is given by

$$\| \boldsymbol{x} \|_q = \left(\sum_{i=1}^{n} |x_i|^q \right)^{\frac{1}{q}}, \quad q > 0, \tag{2.1.3}$$

whereas, it is called l_1-norm when $q = 1$, hence $\| \boldsymbol{x} \|_1 = \sum_{i=1}^{n} |x_i|$ and while that it is the l_2- norm when $q = 2$, , also known as the Euclidean norm and it is defined by $\| \boldsymbol{x} \|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$. Moreover, the $\infty-$ norm is the spectral norm which is defined by $\| \boldsymbol{x} \|_\infty = \max_{1 \leq i \leq n} |x_i|$.
Furthermore, consider $u, v \in \mathbb{R}^n$, the Cauchy-Schwartz inequality for the inner product of u and v is given by

$$\left(\sum_{i=1}^{n} u_i v_i \right)^2 \leq \sum_{i=1}^{n} |u_i|^2 \sum_{i=1}^{n} |v_i|^2. \tag{2.1.4}$$

2.2 Singular Value Decomposition

Kalman, 1996 Singular value decomposition (SVD) is common appraoch in linear algebra. A real valued matrix $\boldsymbol{M}$ with dimension $m \times n$ can be decomposed as a factorization of three matrices rotation U, scaling Σ and reflection V^T respectively as

$$\mathbf{M} = U \Sigma V^T = \sum_{k=1}^{r} \sigma_k u_k v_k^T, \tag{2.2.1}$$

where $U = [u_1, u_2, \cdots, u_r] \in \mathbb{R}^{m \times r}$, $V = [v_1, v_2, \cdots, v_r] \in \mathbb{R}^{n \times r}$, and $\Sigma = diag(\sigma_1, \sigma_2, \cdots, \sigma_r)$ with $r = rank(\boldsymbol{M})$. Moreover, $u_1, u_2, \cdots, u_r$ and $v_1, v_2, \cdots, v_r$ called the left and right singular vectors, respectively, i.e. the column and row spaces of $\boldsymbol{M}$, which are orthonormal (orthogonal bases). Moreover, U represents the eigenvectors set of $\boldsymbol{M}\boldsymbol{M}^T$, and V is the set of eigenvectors of $\boldsymbol{M}^T\boldsymbol{M}$ which have the property $U^TU = V^TV = I_{r\times r}$. Σ is the matrix of the singular values, i.e. the square roots of the nonzero eigenvalues of $\mathbf{M}\mathbf{M}^T$ and $\boldsymbol{M}^T\boldsymbol{M}$ where the singular values are ordered as

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq \sigma_{r+1} = \cdots \sigma_m = 0, m \leq n.$$

The matrix norm function $\| \,.\, \|: \mathbb{R}^{n\times n} \longrightarrow \mathbb{R}^+$ on the vector space $\mathbb{R}^{n\times n}$ satisfies the same conditions as the vector norm function with an additional condition known as the submultiplicity or consistency condition. Then, $\| \,.\, \|: \mathbb{R}^{n\times n} \longrightarrow \mathbb{R}^+$ is a matrix norm if the following properties are satisfied $\forall \boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{n\times n}$ and $a \in \mathbb{R}$

1. $\| \boldsymbol{X} \| > 0$, and $\| \boldsymbol{X} \| = 0$ if and only if $\boldsymbol{X} = 0$.
2. $\| a\boldsymbol{X} \| = |a| \, \| \boldsymbol{X} \|$.
3. $\| \boldsymbol{X} + \boldsymbol{Y} \| \leq \| \boldsymbol{X} \| + \| \boldsymbol{Y} \|$.
4. $\| \boldsymbol{X}\boldsymbol{Y} \| \leq \| \boldsymbol{X} \| \| \boldsymbol{Y} \|$.

More importantly, the nuclear norm of a matrix $\boldsymbol{M}$ with dimension $n \times n$ is the sum of its non-zero singular values which is

$$\| \mathbf{M} \|_* = \sum_{i=1}^{r} \sigma_i. \tag{2.2.2}$$

While the truncated nuclear norm is the sum of the biggest k singular value. Additionally, the Frobenius norm of a $n \times n$ matrix $\boldsymbol{M}$ is the square root of the sum of absolute squares of its elements

$$\| \boldsymbol{M} \|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} |m_{ij}|^2} = trace\left(\boldsymbol{M}^T\boldsymbol{M}\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{r} (\sigma_i)^2\right)^{\frac{1}{2}}, \tag{2.2.3}$$

where, $trace(\boldsymbol{M})$ denote the trace of the matrix $\boldsymbol{M}$, and σ_i denote the singular values of the matrix $\boldsymbol{M}$. Also, the inner product of two matrices $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{n\times n}$ is given by

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{X}_{ij}\boldsymbol{Y}_{ij} = trace(\boldsymbol{X}^T\boldsymbol{Y}).$$

Moreover, the Schattern q-norm for an $n \times n$ matrix $\boldsymbol{X}$ is given by

$$\| \boldsymbol{X} \|_{S_q} = \left(\sum_{i=1}^{n} \sigma_i^q\right)^{\frac{1}{q}},$$

where, the Schattern 1-norm is the nuclear norm, and the Schattern 2-norm is the Frobenius norm. Moreover, if $\boldsymbol{X}$ is an $n \times n$ matrix with $q \geq \log n$ then there exist a constant e such that

$$\| \boldsymbol{X} \| \leq \| \boldsymbol{X} \|_{S_q} \leq e \, \| \boldsymbol{X} \|$$

thus the Schattern norm is multiplicative by the operator norm, where $\| \boldsymbol{X} \| = \| \boldsymbol{X} \|_2$.
The operator or induced norm of $\boldsymbol{X}$ is defined as follows, if the norm is the l_1-norm then it will be the maximum row sum of the operator which defined by $\| \boldsymbol{X} \|_1 = \max_i \sum_j |\boldsymbol{X}_{ij}|$, and if it is the Euclidean norm then it is called the spectral norm of the operator which is the biggest singular value that defined by $\| \boldsymbol{X} \|_2 = \max\{\sigma_i, \forall i \in \{1, 2, \cdots r\}\}$. Moreover, if the norm is the ∞-norm then it will be the

maximum column sum which is given by $\| \boldsymbol{X} \|_\infty = \max_j \sum_i |\boldsymbol{X}_{ij}|$, we denote $\| \boldsymbol{X} \|_2 = \| \boldsymbol{X} \|$. For any norm $\| \, . \, \|$, there exist a dual norm denoted by $\| \, . \, \|_d$ defined as follows

$$\| \boldsymbol{X} \|_d = \sup_{\boldsymbol{Y}} \{ \langle \boldsymbol{X}, \boldsymbol{Y} \rangle : \boldsymbol{Y} \leq 1 \}.$$

Whereas, the dual of a dual $\| (\| \boldsymbol{X} \|_d) \|_d$ it is just the original norm $\| \boldsymbol{X} \|$. For vectors, the Euclidean norm is self dual, and the dual of l_∞-norm is l_1-norm while the dual of l_1-norm is l_∞-norm. Equivalently for matrix norm, the dual of Frobenius norm is Frobenius norm. Similarly, the dual of matrix l_∞-norm. However, the dual of the spectral norm is the nuclear norm. The induced norms are related with the following inequality

$$\| \boldsymbol{X} \| \leq \| \boldsymbol{X} \|_F \leq \| \boldsymbol{X} \|_* \leq \sqrt{r} \, \| \boldsymbol{X} \|_F \leq r \, \| \boldsymbol{X} \|,$$

where r is the rank of the matrix $\| \boldsymbol{X} \|$.

2.3 Bernoulli Model

WolframAlpha The random variable X associated with the Bernoulli distribution has two possible categories $X = 1$, $X = 0$ which are either “success” represented by $X = 1$ with probability p, or “failure” represented by $X = 0$ with probability $1 - p$. Hence, the probabiliy mass function is given by

$$\boldsymbol{P}(X = 1) = p$$
$$\boldsymbol{P}(X = 0) = 1 - p$$

In our case, if $\boldsymbol{S}$ is an $n \times n$ matrix and we have observed just $m << n^2$ of its entries. Let $\Omega = \{(i, j), \forall \boldsymbol{S}_{ij} \neq 0\}$ be the locations of the set of all the non-zero entries of the matrix $\boldsymbol{S}$, assuming that it is distributed uniformally at random. Now suppose that Ω^* is a subset of the matrix $\boldsymbol{S}$ entries sampled uniformally at random using Bernoulli model and let $\{\boldsymbol{\alpha}_{ij}\}_{1 \leq i,j \leq n}$ be a sequence of independent and identically distributed Bernoulli random variables where the probability of the random variable to be equal to 1 is $p = \frac{m}{n^2}$ and to be equal to 0 is $q = 1 - p$. If we consider $\Omega^* = \{(i, j) : \alpha_{ij} = 1\}$ then $\mathbb{E}(|\Omega^*|) = m$, which is the same as for Ω because it represents the non-zero entries of the matrix, i.e. any observed set of the matrix $\boldsymbol{S}$ has to be independent and identically distribued.

2.3.1 Definition. If $f(x)$ is a convex function, and $\boldsymbol{X}$ is a discrete random variable, where $f(\mathbb{E}(\boldsymbol{X}))$ and $\mathbb{E}(f(\boldsymbol{X}))$ are finite, then Jensen's inequality is given by

$$f(\mathbb{E}(\boldsymbol{X})) \leq \mathbb{E}(f(\boldsymbol{X}))$$

2.3.2 Definition. Rademacher sequence is a sign sequence where the probability for the random variable X to be 1 is $\frac{1}{2}$ and the probability to be -1 is also $\frac{1}{2}$. Suppose that $\{\boldsymbol{X}_i\}_{1 \leq i \leq r}$ is a sequence of matrices with the same dimension and let the sequence be Rademacher sequence $\{\epsilon_i\}$. Then

$$\left[\mathbb{E} \, \| \sum_i \epsilon_i \boldsymbol{X}_i \|_{S_q}^q \right]^{\frac{1}{q}} \leq C \sqrt{q} \max \left[\| \left(\boldsymbol{X}_i^T \boldsymbol{X}_i \right)^{\frac{1}{2}} \|_{S_q}, \| \left(\boldsymbol{X}_i \boldsymbol{X}_i^T \right)^{\frac{1}{2}} \|_{S_q} \right],$$

for each $q \geq 2$, where $C = 2^{\frac{-1}{4}} \sqrt{\frac{\pi}{e}}$.

Some of the above definitions can be used to recast the matrix completion problem to an SDP problem. The next section will give an illustrative steps, starting from minimizing the rank and finally approach the SDP programming.

2.4 SDP Formulation

J.Candes and Recht, 2009 If the number of the non-zero entries of the $n \times n$ matrix $\boldsymbol{S}$ is sufficiently large with locations distributed uniformally at random, we can hope there exist only one low rank matrix that fits the data. Then, we can use the following optimization problem to recover the matrix $\boldsymbol{S}$:

$$\begin{array}{ll} \text{minimize} & rank(\boldsymbol{X}) \\ \text{subject to} & \boldsymbol{X}_{ij} = \boldsymbol{S}_{ij} \quad (i,j) \in \Omega \\ & \boldsymbol{X} \in \mathbb{R}^{n \times n} \end{array} \tag{2.4.1}$$

where $\boldsymbol{X}$ is the matrix decision variable. The optimization problem 2.4.1 seeks the simplest solution that fits the data, and it will recover the matrix sufficiently if there exists just one low rank matrix that fits the data. Unfortunately, if the data is sufficiently large then the optimization problem 2.4.1 becomes NP-hard.

Therefore, if the matrix $\boldsymbol{S}$ has rank r then it has r non-zero singular values, because the rank of a matrix is the number of the non-zero singular values. Now, instead of minimizing the rank, we can minimize the nuclear norm using the truncated nuclear norm. Then, the optimization problem 2.4.1 becomes

$$\begin{array}{ll} \text{minimize} & ||\boldsymbol{X}||_* \\ \text{subject to} & \boldsymbol{X}_{ij} = \boldsymbol{S}_{ij} \quad (i,j) \in \Omega \end{array} \tag{2.4.2}$$

which is equivalent to 2.4.1. If the matrix is symmetric and positive semidefinite then the nuclear norm of the matrix $\boldsymbol{X}$ is the sum of the non-negative eigenvalues and also equal to the trace of the matrix $\boldsymbol{X}$. Therefore, for positive semidefinite unknowns of the optimization problem 2.4.2 will minimize the trace of the matrix over the same constraint set, then

$$\begin{array}{ll} \text{minimize} & trace(\boldsymbol{X}) \\ \text{subject to} & \boldsymbol{X}_{ij} = \boldsymbol{S}_{ij} \quad (i,j) \in \Omega \\ & \boldsymbol{X} \geq 0. \end{array} \tag{2.4.3}$$

The above problem can be converted into the following semidefinite programming problem as follows

$$\begin{array}{ll} \text{minimize} & trace(\boldsymbol{W}_1) + trace(\boldsymbol{W}_2) \\ \text{subject to} & \boldsymbol{X}_{ij} = \boldsymbol{S}_{ij} \quad (i,j) \in \Omega \\ & \begin{bmatrix} \boldsymbol{W}_1 & \boldsymbol{X} \\ \boldsymbol{X}^T & \boldsymbol{W}_2 \end{bmatrix} = \begin{bmatrix} U \\ V \end{bmatrix} \Sigma \begin{bmatrix} U \\ V \end{bmatrix}^T = \begin{bmatrix} U\Sigma U^T & U\Sigma V^T \\ V\Sigma U^T & V\Sigma V^T \end{bmatrix} \geq 0. \end{array} \tag{2.4.4}$$

The theoretical results that we will be presenting in Chapter 5 are related to the unique solution for the two problems 2.4.2, 2.4.4.

2.5 Coherence

Given U a subspace of the vector space $\mathbb{R}^n$ of dimension r and the orthogonal projection onto U is $\boldsymbol{P}_U$, then the coherence of U which defines the correlation of its basis with the standard basis e_i is given by

$$\mu(U) = \frac{n}{r} \max_{1 \leq i \leq n} \| \boldsymbol{P}_U e_i \|^2, \tag{2.5.1}$$

where the smallest coherence for a subspace is 1; Suppose that the subspace U basis entries has magnitude $\frac{1}{\sqrt{n}}$ then

$$\| \boldsymbol{P}_U e_i \| = \sqrt{\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}} = \sqrt{\frac{r}{n}},$$

so

$$\mu(U) = \frac{n}{r}\frac{r}{n} = 1,$$

and if we have the standard basis as U eigenvectors which their magnitude is 1, the coherence of U in this case will be

$$\mu(U) = \frac{n}{r}\sqrt{1}^2 = \frac{n}{r},$$

and it is the largest coherence that we can get. We are interested in subspaces with low coherence which means our subspace basis are not sparse. Moreover, if we have a matrix $\boldsymbol{M}$ with dimension $m \times n$ and it has SVD as given in equation 2.2.1, there are two assumptions that U and V should comply:

$S1$ $\max(\mu(U), \mu(V)) \leq \mu_0, \quad \mu_0 \geq 0.$

$S2$ The absolute value of the maximum element of the $m \times n$ matrix $\sum_{k=1}^{r} u_k v_k^T$ is bounded by $\mu_1\sqrt{\frac{r}{mn}}, \quad \mu_1 > 0.$

Moreover, μ_0 and μ_1 might depend on m, n and r, or might not. Additionaly, if $\mu_1 = \mu_0\sqrt{r}$ we can ensure that $\boldsymbol{S2}$ holds. this is because we know that for the matrix $\sum_{1\leq k\leq r} u_k v_k^T$ the (i, j)th entry is $\sum_{1\leq k\leq r} u_{ik} v_{jk}^T$. Using Cauchy-Schwartz inequality we get

$$\left| \sum_{1\leq k\leq r} u_{ik} v_{jk}^T \right| \leq \sqrt{\sum_{1\leq k\leq r} |u_{ik}|^2} \sqrt{\sum_{1\leq k\leq r} |v_{jk}|^2} \leq \frac{\mu_0 r}{\sqrt{mn}}.$$

Now, we will introduce the derandomization condition that is requred for the observed entries of the matrix $\boldsymbol{S}$.

2.6 Derandomization

Given a matrix $\boldsymbol{M}$, it is possible to decompose it as a low rank matrix $\boldsymbol{L}$ and a sparse matrix $\boldsymbol{S}$ as follows

$$\boldsymbol{M} = \boldsymbol{L} + \boldsymbol{S}.$$

Now, consider that the probability of finding a non-zero entry in $\boldsymbol{S}$ is p while the probability of finding a non-zero entry in $\boldsymbol{L}$ is $1 - p = q$, and they are following the Bernoulli model, which means they are distributed uniformly at random. Moreover, the Principal Component Pursuit (PCP) J.Candes et al. (2011) can give us an estimation of $\boldsymbol{L}$ and $\boldsymbol{S}$ by solving

$$\begin{aligned} \text{minimize} \quad & \| \boldsymbol{L} \|_* + \lambda \| \boldsymbol{S} \|_1 \\ \text{subject to} \quad & \boldsymbol{L} + \boldsymbol{S} = \boldsymbol{M} \end{aligned} \tag{2.6.1}$$

which will give an exact recovery of the low rank matrix $\boldsymbol{L}$. Now, consider that the PCP 2.6.1 estimation is $\boldsymbol{L_0}$ and $\boldsymbol{S_0}$ for the low rank matrix and the sparse matrix, respectively.

2.6.1 Theorem. *J.Candes et al. (2011) Assume that* $\boldsymbol{L}_0$ *is an* $n \times n$ *incoherent matrix; Fix an arbitrary sign matrix* $\boldsymbol{N}$ *and assume that the set of entries of the sparse matrix* $\boldsymbol{S}$ *is uniformly distributed at random among all the sets of cardinality* m*, and* $sign([\boldsymbol{S}_0]_{ij}) = \boldsymbol{N}_{ij} \quad \forall(i,j) \in \Omega$*, where* Ω *is the set of non-zero entries of* $\boldsymbol{S}$*. Then, there exist a numerical constant* c *such that with probability* $1 - cn^{-10}$ *at least, the PCP estimation 2.6.1 is exact with* $\lambda = \frac{1}{\sqrt{n}}$ *and*

$$\textit{rank}(\boldsymbol{L_0}) \leq qn\mu^{-1}(\log n)^{-2}, \quad \textit{and} \quad m \leq pn^2.,$$

where $\log n$ *is the natural logarithm.*

In Theorem 2.6.1 we assume that the values of the non-zero entries of $\boldsymbol{S}_0$ are fixed and we state that we can reccover low rank matrices which have scattered singular vectors. Now, we will add the assumption that the signs of the non-zero entries are independent symmetric Bernoulli variables, which means the probability for the sign to be 1 or -1 is $\frac{1}{2}$.

2.6.2 Theorem. *J.Candes et al. (2011) Suppose that the low rank matrix* $\boldsymbol{L}_0$ *satisfies the conditions of Theorem 2.6.1, and the locations of non-zero entries of the matrix* $\boldsymbol{S}_0$ *follow Bernoulli model with probability* $2p$*, and the signs are independent and identically distributed and independent from the locations. If the solution of the* PCP *is exact with high probability, then with at least the same probability, the signs of* $\boldsymbol{S}_0$ *are fixed and the locations are sampled from Bernoulli model with probability* p*.*

The proof of Theorem 2.6.2 can be found in J.Candes et al. (2011).

3. Mathematical Model

Since we have assumed that the matrix $\boldsymbol{S}$ is a low rank matrix, and it has incoherent column and row spaces. We can guarantee the exact recovery for matrices with incoherent column and row spaces that have the singular vectors as given in 2.2.1 model. In addition to matrices with the 2.2.1 model and with singular vectors which have small components, U and V obey the following

$$\max_{ij} |\langle e_i, u_j \rangle|^2 \leq \frac{\mu_B}{n}, \qquad \max_{ij} |\langle e_i, v_j \rangle|^2 \leq \frac{\mu_B}{n}, \tag{3.0.1}$$

for some value $\mu_\beta = O(1)$, and by using the concepts of Section 2.5 the maximum coherence for U and V is less than or equal to μ_β. Moreover, if a matrix obeys Equation 3.0.1 then it is obeying the assumptions $\boldsymbol{S1}$ and $\boldsymbol{S2}$; Also, if the matrix does not obey Equation 3.0.1, i.e. if two rows of U and V are identical with entries having magnitude $\sqrt{\frac{\mu_\beta}{n}}$, in this case, we will have

$$\| \sum_k u_k v_k^* \|_\infty = \frac{\mu_B r}{n},$$

then for $\boldsymbol{S2}$ to hold, μ_1 needs to be bigger than $\mu_0\sqrt{r}$. Now, consider that Ω is the set of locations of the nonzero entries, and T is the space that spanned by U and V J.Candes and Recht, 2009. Also, consider that the orthogonal projection of any matrix $\boldsymbol{A}$ onto T is given by

$$\mathcal{P}_T(\boldsymbol{A}) = \boldsymbol{P}_U\boldsymbol{A} + \boldsymbol{A}\boldsymbol{P}_V - \boldsymbol{P}_U\boldsymbol{A}\boldsymbol{P}_V, \tag{3.0.2}$$

where matrices U and V comes from the SVD of the matrix $\boldsymbol{S}$ Furthermore

$$\mathcal{P}_{T^\perp}(\boldsymbol{A}) = (\mathcal{I} - \mathcal{P}_T)(\boldsymbol{A}) = (\mathcal{I}_n - \boldsymbol{P}_U)\boldsymbol{A}(\mathcal{I}_n - \boldsymbol{P}_V), \tag{3.0.3}$$

where $\mathcal{I}$ is the identity operator, which maps any matrix to itself. A matrix $\boldsymbol{S}$ is sampled from the incoherent basis model if it has the form

$$\boldsymbol{S} = \sum_{i=1}^{r} \epsilon_i \sigma_i u_i v_i^T$$

where $\{\epsilon\}_{1\leq i\leq r}$ is a random sign sequence, and the singular vectors $\{u_i\}_{1\leq i\leq r}$, $\{v_i\}_{1\leq i\leq r}$ have a spectral norm at most $\sqrt{\frac{\mu_\beta}{n}}$.

3.0.1 Lemma. J.Candes and Recht, 2009 If a matrix is sampled from the incoherent model, then there exist numerical constants C, c such that the matrix $\boldsymbol{S}$ obeys $\boldsymbol{S2}$ with $\mu_1 \leq C\mu_B\sqrt{(\beta+2)\log n}$ with probability at least $1 - cn^{-\beta}$.

The following Lemma garantees that the random orthogonal model is satisfying $\boldsymbol{S1}$ and $\boldsymbol{S2}$ with high probability.

3.0.2 Lemma. J.Candes and Recht, 2009 Let $\bar{r} = \max(\log n, r)$ then there exist numerical constants such that the random orthogonal model satisfies

1. $\max_i \| \boldsymbol{P}_U e_i \|^2 \leq C\frac{\bar{r}}{n}$.
2. $\| \sum_{1\leq i\leq r} u_i v_i^T \|_\infty \leq C \log n \frac{\sqrt{r}}{n}$

with probability at least $1 - cn^{-3}\log n$.

To prove that $\boldsymbol{S} = \sum_{1\leq i\leq r} \sigma_i u_i v_i^T$ is the unique minimizer of the SDP problem, we will construct a matrix $\boldsymbol{Y}$ which will vanish on Ω^c.
Suppose that we have an operator $O_\Omega : \mathbb{R}^{n\times n} \longrightarrow \mathbb{R}^m$, and for any matrix $\boldsymbol{S}$ the operator consruct a sample set with m element of the matrix $\boldsymbol{S}$.

3.1 Optimality

In the followig section we will introduce the existence of a dual certificate to guarantee the existence of an optimal solution. Moreover, finding a dual is depending on the subgradiant of the nuclear norm of the matrix. The subgradiant of the nuclear norm at an $n \times n$ matrix $\boldsymbol{M}$ with singular value decomposition $\boldsymbol{M} = U\Sigma V^T$ as in Equation 2.2.1 is given by

$$\partial \parallel \boldsymbol{M} \parallel_* = \{UV^T + \boldsymbol{W} : \boldsymbol{W}^T U = 0, \boldsymbol{W}V = 0, \parallel \boldsymbol{W} \parallel \leq 1\}. \tag{3.1.1}$$

3.1.1 Dual certificate.
By abusing the use of notation we also define Ω as follows, Ω is the space of matrices $\{\boldsymbol{X}\}$ that satisfies $\boldsymbol{X}_{ij} = \boldsymbol{S}_{ij}$ for all non-zero entries of the matrix $\boldsymbol{S}$, and $\mathcal{P}_\Omega$ is the sampling operator that maps any matrix $\boldsymbol{M}$ to a sparse matrix $\mathcal{P}_\Omega(M) = \boldsymbol{M}_{ij} \quad \forall (i,j) \in \Omega$, and T is the space that spanned by U and V^T. For the low rank matrix $\boldsymbol{L}_0$ and the sparse matrix $\boldsymbol{S}_0$ to be the unique optimal solution, the existence of a dual will be needed. To achieve that let us introduce the following lemma.

3.1.2 Lemma. J.Candes et al. (2011) Suppose that $\parallel \mathcal{P}_\Omega \mathcal{P}_T \parallel \leq 1$, i.e. $\Omega \cap T = \{0\}$, then $\boldsymbol{L}_0$ and $\boldsymbol{S}_0$ are the unique solution if there exist a pair of matrices $(\boldsymbol{W}, \boldsymbol{F})$, such that $\mathcal{P}_\Omega(UV^T + \boldsymbol{W})$ is a subgradiant of the nuclear norm of the matrix $\boldsymbol{M}$ satisfying

$$UV^T + \boldsymbol{W} = \lambda\left(sgn(\boldsymbol{S}_0) + \boldsymbol{F}\right)$$

with the following conditions $\mathcal{P}_T \boldsymbol{W} = 0$, $\parallel \boldsymbol{W} \parallel < 1$, $\mathcal{P}_\Omega \boldsymbol{F} = 0$ and $\parallel \boldsymbol{F} \parallel_\infty < 1$.

As a result of Lemma 3.1.2 we introduce a dual certificate $\boldsymbol{W}$ satisfying the following

$$\begin{cases} \boldsymbol{W} \in T^\perp \\ \parallel \boldsymbol{W} \parallel < \frac{1}{2} \\ \parallel \mathcal{P}_\Omega \left(UV^T + \boldsymbol{W} - \lambda sgn(\boldsymbol{S}_0)\right) \parallel \leq \frac{\lambda}{4} \\ \parallel \mathcal{P}_\Omega \left(UV^T + \boldsymbol{W}\right) \parallel \leq \frac{\lambda}{2} \end{cases}$$

3.1.3 Golfing scheme for dual certification.
In matrix completion, Golfing scheme is a scheme that we can use to find a dual certificate. We know that $\Omega \sim Ber(p)$, i.e. $\Omega^c \sim Ber(1-p)$ therefore, the distribution of Ω^c is the same as the distribution of $\Omega^c = \Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_k$. Each of the sample sets $\Omega^c = \{\Omega_1, \Omega_2, \cdots, \Omega_k\}$ is following Bernolli model with probability a. Then

$$\boldsymbol{P}((i,j) \in \Omega) = \boldsymbol{P}(Bin(k,a) = 0) = (1-a)^k.$$

And if $p = (1-a)^k$ then the two Bernolli models are equivalent. Now let us introduce the construction of a dual certificate

$$\boldsymbol{W} = \boldsymbol{W}^L + \boldsymbol{W}^S.$$

as follows

- Constructing $\boldsymbol{W}^L$ using Golfing scheme: Fix an integer $k \geq 1$ and let $\Omega^c = \cup_{1\leq i\leq k}\Omega_i$, then

$$\boldsymbol{Y}_i = \boldsymbol{Y}_{i-1} + a^{-1}\mathcal{P}_{\Omega_i}\mathcal{P}_T(UV^T - \boldsymbol{Y}_{i-1}),$$

  with $\boldsymbol{Y}_0 = 0$, then

$$\boldsymbol{W}^L = \mathcal{P}_{T^\perp}\boldsymbol{Y}_k$$

  .

- Furthermore, constructing $\boldsymbol{W}^S$ using least squares methods: Let $\| \mathcal{P}_\Omega\mathcal{P}_T \| < \frac{1}{2}$ so that $\| \mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega \| < \frac{1}{4}$. Therefore, the operator $\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega$ is invertible, and denote its inverse as $(\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega)^{-1}$. Then

$$\boldsymbol{W}^S = \lambda\mathcal{P}_{T^\perp}\left(\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega\right)^{-1} sgn(\boldsymbol{S}_0).$$

  Notice that $\mathcal{P}_\Omega\boldsymbol{W}^S = \lambda\mathcal{P}_\Omega(1-\mathcal{P}_T)\left(\mathcal{P}_\Omega - \mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega\right)^{-1} sgn(\boldsymbol{S}_0) = \lambda sgn(\boldsymbol{S}_0)$. Here we have used $\mathcal{P}_\Omega(1-\mathcal{P}_T) = \mathcal{P}_\Omega\mathcal{P}_T\mathcal{P}_\Omega$, since for any two operators $\mathcal{P}_1, \mathcal{P}_2$, $\mathcal{P}_1\mathcal{P}_2 = \mathcal{P}_2\mathcal{P}_1$ and $\mathcal{P}_\Omega^2 = \mathcal{P}_\Omega$. Now we have for all matrices $\boldsymbol{W} \in T^\perp$ satisfies $\mathcal{P}_\Omega\boldsymbol{W} = \lambda sgn(\boldsymbol{S}_0)$ and that $\boldsymbol{W}^S$ has minimum Frobenius norm. So we can say that $\boldsymbol{W} = \boldsymbol{W}^L + \boldsymbol{W}^S$ is a dual certificate if the following is satisfied

$$\begin{cases} \| \boldsymbol{W}^L + \boldsymbol{W}^S \| < \frac{1}{2}, \\ \| \mathcal{P}_\Omega\left(UV^T + \boldsymbol{W}^L\right) \| \leq \frac{\lambda}{4}, \\ \| \mathcal{P}_\Omega\left(UV^T + \boldsymbol{W}^L + \boldsymbol{W}^S\right) \| \leq \frac{\lambda}{2}. \end{cases}$$

3.1.4 Lemma. J.Candes and Recht, 2009 Suppose $\boldsymbol{S} = \sum_{1\leq i\leq r}\sigma_i u_i v_i^T$ be a matrix with rank r, which is feasible for the SDP problem 2.4.2, and it is the unique solution if

1. There exist a dual certificate such that $\mathcal{P}_T(UV^T + \boldsymbol{W}) = UV^T$ and $\| \mathcal{P}_{T^\perp}(UV^T + \boldsymbol{W}) \| < 1$.
2. The sampling operator O_Ω is injective for matrices in the linear space T.

Now to prove that $\boldsymbol{S} = \sum_{1\leq i\leq r}\sigma_i u_i v_i^T$ is the unique optimal solution to the semi-definite problem 2.4.2, firstly, construct a matrix $\boldsymbol{Y} = UV^T + \boldsymbol{W}$ satisfying the condition of Lemma 3.1.4 and then proving the injectivity as well, and that $\boldsymbol{Y}$ is a solution of

$$\begin{aligned} \text{minimize} \quad & ||\boldsymbol{X}||_F, \\ \text{subject to} \quad & \mathcal{P}_T\mathcal{P}_\Omega(\boldsymbol{X}) = \sum_{i=1}^{r} u_i v_i^T. \end{aligned} \tag{3.1.2}$$

In addition, if the matrix $\boldsymbol{Y}$ doesn't vanish on $\mathcal{P}_{\Omega^c}$ then it is not an optimal solution since $\boldsymbol{Y}$ should obey the constraint of the problem 3.1.2 and $\boldsymbol{Y}$ has smaller Frobenius norm. Using the Pythagoras theorem considering the projection of $\boldsymbol{Y}$ onto T and $T^\perp$ as follows

$$\begin{aligned} \| \boldsymbol{Y} \|_F^2 &= \| \mathcal{P}_T(\boldsymbol{Y}) \|_F^2 + \| \mathcal{P}_{T^\perp}(\boldsymbol{Y}) \|_F^2 \\ &= \| \sum_{i=1}^{n} u_i v_i^T \|_F^2 + \| \boldsymbol{W} \|_F^2 \\ &= r + \| \boldsymbol{W} \|_F^2 \end{aligned}$$

so minimizing the Frobenius norm of $\boldsymbol{Y}$ implies that minimizing the Frobenius norm of $\boldsymbol{W}$ under the constraint $\mathcal{P}_T(\boldsymbol{X}) = \sum_{i=1}^{r} u_i v_i^T$. Furthermore, by forcing $\| \mathcal{P}_{T^\perp}(\boldsymbol{Y}) \|_F^2$ to be small, which could also

garantee that $\| \mathcal{P}_{T^\perp}(\boldsymbol{Y}) \|_\infty^2$ is small and then verifying that $\| \mathcal{P}_{T^\perp}(\boldsymbol{Y}) \|_\infty^2 < 1$ would prove that $\boldsymbol{S}$ is the unique optimal solution of the SDP problem 2.4.2.
Then, to find the solution for 3.1.2 we will propose the following operator $\mathcal{A}_{\Omega T}(\boldsymbol{S}) = \mathcal{P}_\Omega \mathcal{P}_T(\boldsymbol{S})$ and if $\mathcal{A}_{\Omega T}^T \mathcal{A}_{\Omega T} = \mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T$ has a full rank when restricted to T then the minimizer for the problem 3.1.2 is

$$\boldsymbol{Y} = \mathcal{A}_{\Omega T}(\mathcal{A}_{\Omega T}^T \mathcal{A}_{\Omega T})^{-1}(\boldsymbol{E}), \quad \boldsymbol{E} = \sum_{i=1}^{r} u_i v_i^T. \tag{3.1.3}$$

Then, we have to prove two tasks which are as follows

1. Show that $\mathcal{A}_{\Omega T}^T \mathcal{A}_{\Omega T} = \mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T$ is an injective map from T onto itself, which is equivalent to the second condition of Lemma 3.1.4, and the expression of $\boldsymbol{Y}$ which is given in 3.1.3 is well defined.
2. Show that $\| \mathcal{P}_{T^\perp}(\boldsymbol{Y}) \|_\infty = \| \boldsymbol{W} \|_\infty < 1$ which is equivalent to the first condition in Lemma 3.1.4.

3.2 Injectivity

Showing that $\mathcal{A}_{\Omega T}$ is injective will also show that $\boldsymbol{Y}$ is well defined, implies that the Frobenius norm of the operator $p^{-1}\mathcal{P}_T(\mathcal{P}_\Omega - p\mathcal{I})\mathcal{P}_T$ is small as well, so we will verify that the Frobenius norm is small, instead.

3.2.1 Theorem. J.Candes and Recht, 2009 *Assume that the sample set Ω is sampled from Bernoulli model, and the coherence of U and V satisfies $\boldsymbol{S1}$, then there exist a numerical constant K such that*

$$\forall \beta > 1, \quad p^{-1} \| \mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T - p\mathcal{P}_T \| \leq K\sqrt{\frac{\mu_0 n r \beta \log n}{m}}, \tag{3.2.1}$$

with probability at least $1 - 3n^{-\beta}$ provided that $K\sqrt{\frac{\mu_0 n r \beta \log n}{m}} < 1$.

Proof. Let $\boldsymbol{G}$ be any matrix, and let us decompose it as

$$\boldsymbol{G} = \sum_{ij} \left\langle \boldsymbol{G}, e_i e_j^T \right\rangle e_i e_j^T,$$

therefore

$$\mathcal{P}_T(\boldsymbol{G}) = \sum_{ij} \left\langle \mathcal{P}_T(\boldsymbol{G}), e_i e_j^T \right\rangle e_i e_j^T = \sum_{ij} \left\langle \boldsymbol{G}, \mathcal{P}_T(e_i e_j^T) \right\rangle e_i e_j^T,$$

since orthogonal projections are self-adjoint. Thus

$$\mathcal{P}_\Omega \mathcal{P}_T(\boldsymbol{G}) = \sum_{ij} \alpha_{ij} \left\langle \boldsymbol{G}, \mathcal{P}_T(e_i e_j^T) \right\rangle e_i e_j^T,$$

where α_{ij} is a random sequence following Bernoulli model with probability p. Then

$$\mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T(\boldsymbol{G}) = \sum_{ij} \alpha_{ij} \left\langle \boldsymbol{G}, \mathcal{P}_T(e_i e_j^T) \right\rangle \mathcal{P}_T(e_i e_j^T),$$

which allow us to formulate the operator $\mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T$ as

$$\mathcal{P}_T \mathcal{P}_\Omega \mathcal{P}_T = \sum_{ij} \alpha_{ij} \mathcal{P}_T(e_i e_j^T) \otimes \mathcal{P}_T(e_i e_j^T).$$

Now, using the formula 3.0.2 we will get

$$\mathcal{P}_T(e_ie_j^T) = (\boldsymbol{P}_U e_i)e_j^T + e_i(\boldsymbol{P}_U e_j)^T - (\boldsymbol{P}_U e_i)(\boldsymbol{P}_U e_j)^T, \tag{3.2.2}$$

then its Frobenius norm will be as follows

$$\| \mathcal{P}_T(e_ie_j^T) \|_F = \left\langle \mathcal{P}_T(e_ie_j^T), e_ie_j^T \right\rangle = \| \boldsymbol{P}_U e_i \|^2 + \| \boldsymbol{P}_V e_j \|^2 - \| \boldsymbol{P}_U e_i \|^2 \| \boldsymbol{P}_V e_j \|^2,$$

and from Section 2.5 we have $\| \boldsymbol{P}_U e_i \|^2 \leq \frac{\mu(U)r}{n}$ and $\| \boldsymbol{P}_V e_j \|^2 \leq \frac{\mu(V)r}{n}$, so we can conclude that

$$\| \mathcal{P}_T(e_ie_j^T) \|_F^2 \leq \frac{2\mu_o}{n}. \tag{3.2.3}$$

The expected value of the operator $\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T$ is

$$\mathbb{E}(\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T) = \mathcal{P}_T\mathbb{E}(\mathcal{P}_\Omega)\mathcal{P}_T = \mathcal{P}_T(p\mathcal{I})\mathcal{P}_T = p(\mathcal{P}_T\mathcal{I}\mathcal{P}_T) = p\mathcal{P}_T,$$

which states that the operator doesn't diverge from its mean in the spectral norm. □

3.2.2 Theorem. *J.Candes and Recht, 2009 Suppose $\{\alpha_{ij}\}$ is an independent Bernoulli variables, with $\mathbb{P}(\alpha_{ij} = 1) = p = \frac{m}{n^2}$, and assume that $\| \mathcal{P}_T(e_ie_j^T) \|_F \leq \frac{2\mu_o}{n}$ as in 3.2.3. Then put*

$$Z = p^{-1} \| \sum_{ij} (\alpha_{ij} - p)\mathcal{P}_T(e_ie_j^T) \otimes \mathcal{P}_T(e_ie_j^T) \| = p^{-1} \| \mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T - p\mathcal{P}_T \|,$$

then

1. *There exist a constant $\bar{K}$ such that*

$$\mathbb{E}(Z) \leq \bar{K}\sqrt{\frac{\mu_0 nr \log n}{m}},$$

   provided that $\bar{K}\sqrt{\frac{\mu_0 nr \log n}{m}} < 1$.

2. *Assume that $\mathbb{E}(Z) \leq 1$; Then for each $\lambda > 0$*

$$\mathbb{P}\left(|Z - \mathbb{E}(Z)| > \lambda\sqrt{\frac{\mu_0 nr \log n}{m}} \right) \leq 3\exp\left(-a \min\left\{ \lambda^2 \log n, \lambda\sqrt{\frac{m \log n}{\mu_0 nr}} \right\} \right), \tag{3.2.4}$$

   for some constant a.

The first task of the theorem 3.2.2 can be constructed from the following theorem.

3.2.3 Theorem. *J.Candes and Romberg 2007 Let U be a matrix of orthogonal vectors in $\mathbb{R}^n$ which are incoherent, and α_i be a Bernolli sequence where $\mathbb{P}(\alpha_i = 1) = p$, so*

$$p^{-1} \| \sum_i (\alpha_i - p)u_i \otimes u_i \| \leq C\sqrt{\frac{\log n}{p}} \max_{1 \leq i \leq r} \| u_i \|, \tag{3.2.5}$$

for some constant $C > 0$, provided that $C\sqrt{\frac{\log n}{p}} \max_{1 \leq i \leq r} \| u_i \| < 1$.
The proof can be found in J.Candes and Romberg 2007.

Now, we can get the first task of Theorem 3.2.2 by setting $u_i = \mathcal{P}_T(e_i e_j^T)$ in 3.2.3 so we will get

$$\begin{aligned} p^{-1} \parallel \sum_{ij} (\alpha_{ij} - p)\mathcal{P}_T(e_i e_j^T) \otimes \mathcal{P}_T(e_i e_j^T) \parallel &\leq C\sqrt{\frac{\log n}{p}} \max_{ij}(\parallel \mathcal{P}_T(e_i e_j^T) \parallel) \\ &\leq C\sqrt{\frac{\log n}{p}} \sqrt{\frac{2\mu_0 r}{n}} \\ &= C\sqrt{2}\sqrt{\frac{\mu_0 r \log n}{np}} = C'\sqrt{\frac{\mu_0 nr \log n}{m}}. \end{aligned}$$

For the second task of Theorem 3.2.2, firstly, set $\lambda = \sqrt{\frac{\beta}{a}}$ in the second part of Theorem 3.2.3, and suppose that $m > \frac{\beta}{a}\mu_0 nr \log n$ then

$$\begin{aligned} \mathbb{P}\left(|Z - \mathbb{E}Z| > \frac{1}{\sqrt{a}}\sqrt{\frac{\mu_0 nr\beta \log n}{m}} \right) &\leq 3\exp\left(-a \min\left\{ \left(\frac{\beta}{a}\right)^2 \log n, \frac{\beta}{a}\sqrt{\frac{m \log n}{\mu_0 nr}} \right\} \right), \\ &= 3exp\{\frac{-a\beta}{a} \log n\} \\ &= 3e^{-\beta \log n} = 3e^{\log n^{-\beta}} = 3n^{-\beta}, \end{aligned}$$

then as a consequence

$$\mathbb{P}\left(|Z - \mathbb{E}Z| \leq \frac{1}{\sqrt{a}}\sqrt{\frac{\mu_0 nr\beta \log n}{m}} \right) = 1 - 3n^{-\beta}.$$

Therefore, if we set $K = \bar{K}\sqrt{\beta} + \frac{1}{\sqrt{a}}$ then

$$Z \leq \frac{1}{\sqrt{a}}\sqrt{\frac{\mu_0 nr\beta \log n}{m}} + \bar{K}\sqrt{\frac{\mu_0 nr \log n}{m}} = K\sqrt{\frac{\mu_0 nr\beta \log n}{m}}.$$

Now, setting m to be sufficiently large, such that $K\sqrt{\frac{\mu_0 nr\beta \log n}{m}} \leq \frac{1}{2}$, then from 3.2.1 the following is valid for all the matrices $\boldsymbol{A}$

$$\frac{p}{2} \parallel \mathcal{P}_T(\boldsymbol{A}) \parallel_F \leq \parallel \mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T(\boldsymbol{A}) \parallel_F \leq \frac{3p}{2} \parallel \mathcal{P}_T(\boldsymbol{A}) \parallel_F, \tag{3.2.6}$$

with high probability.
As a result, we have achieved that the operator $\mathcal{A}_{\Omega T}^T \mathcal{A}_{\Omega T} = \mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T$ which maps T onto itself is bounded, which implies that it is invertible. Further, It follows from Theorem 3.2.1 and Equation 3.2.6 that

$$\parallel \mathcal{P}_\Omega\mathcal{P}_T(\boldsymbol{A}) \parallel_F \leq \sqrt{\frac{3p}{2}} \parallel \mathcal{P}_T(\boldsymbol{A}) \parallel_F .$$

4. Sample Size

In this chapter we will prove the second task of Lemma 3.1.4, equivalently, prove that $\| \mathcal{P}_{T^\perp}\left(UV^T + \boldsymbol{W}\right) \| < 1$. Firstly, note that the operator

$$\mathcal{R} = \mathcal{P}_T - p^{-1}\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T, \tag{4.0.1}$$

satisfying Theorem 3.2.1, then

$$\| \mathcal{R}(\boldsymbol{A}) \|_F \leq K\sqrt{\frac{\mu_0 nr\beta \log n}{m}},$$

with large probability. So we can decompose $(\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T)^{-1}(\boldsymbol{A})$ for any matrix $\boldsymbol{A}$ as

$$(\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T)^{-1}(\boldsymbol{A}) = p^{-1}\left(\boldsymbol{A} + \mathcal{R}(\boldsymbol{A}) + \mathcal{R}^2(\boldsymbol{A}) + \cdots\right), \tag{4.0.2}$$

and $\mathcal{R}$ is decreasing if the sample set m is sufficiently big, and from 3.1.3 we can define $\mathcal{P}_{T^\perp}(\boldsymbol{Y})$ as

$$\mathcal{P}_{T^\perp}(\boldsymbol{Y}) = \mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T(\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T)^{-1}(\boldsymbol{E}), \quad \boldsymbol{E} = \sum_{i=1}^{r} u_i v_i^T, \tag{4.0.3}$$

and using 4.0.2 we can reformulate it as

$$\mathcal{P}_{T^\perp}(\boldsymbol{Y}) = p^{-1}\mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T\left(\boldsymbol{E} + \mathcal{R}(\boldsymbol{E}) + \mathcal{R}^2(\boldsymbol{E}) + \cdots\right), \quad \boldsymbol{E} = \sum_{i=1}^{r} u_i v_i^T, \tag{4.0.4}$$

and we wish to bound the norm of $\mathcal{P}_{T^\perp}(\boldsymbol{Y})$, and verify that it is small, i.e. bounding the new decomposition of $\mathcal{P}_{T^\perp}(\boldsymbol{Y})$ which is the right hand side of 4.0.4.
Firstly let us state the following which will be used in the proofs. For each pair (i, j) and (i', j') it follows from 3.2.2 that

$$\left\langle \mathcal{P}_T(e_{i'}e_{j'}^T), e_i e_j^T \right\rangle = \langle e_i, \boldsymbol{P}_U e_{i'} \rangle \boldsymbol{I}_{j=j'} + \langle e_j, \boldsymbol{P}_V e_{j'} \rangle \boldsymbol{I}_{i=i'} - \langle e_i, \boldsymbol{P}_U e_{i'} \rangle \langle e_j, \boldsymbol{P}_V e_{j'} \rangle, \tag{4.0.5}$$

set μ_0 such that the coherence of U and V obeys $\mu(U) \leq \mu_0$ and $\mu(V) \leq \mu_0$ so that

$$| \langle e_i, \boldsymbol{P}_U e_{i'} \rangle | = | \langle \boldsymbol{P}_U e_i, \boldsymbol{P}_U e_{i'} \rangle | \leq \| \boldsymbol{P}_U e_i \| \| \boldsymbol{P}_U e_{i'} \| \leq \sqrt{\frac{\mu_0 r}{n}}\sqrt{\frac{\mu_0 r}{n}} = \frac{\mu_0 r}{n},$$

and repeating the same for $| \langle e_j, \boldsymbol{P}_V e_{j'} \rangle |$ we will get

$$| \langle e_j, \boldsymbol{P}_V e_{j'} \rangle | \leq \frac{\mu_0 r}{n}.$$

Now if we consider that $i \neq i'$ and $j = j'$ we will get

$$\begin{aligned} \left|\left\langle \mathcal{P}_T(e_{i'}e_j^T), e_i e_j^T \right\rangle\right| &= | \langle e_i, \boldsymbol{P}_U e_{i'} \rangle | - | \langle e_i, \boldsymbol{P}_U e_{i'} \rangle | \, | \langle e_j, \boldsymbol{P}_V e_j \rangle | \\ &= | \langle e_i, \boldsymbol{P}_U e_{i'} \rangle |(1 - | \langle e_j, \boldsymbol{P}_V e_j \rangle |) \\ &= | \langle e_i, \boldsymbol{P}_U e_{i'} \rangle |(1 - \| \boldsymbol{P}_V e_j \|^2) \\ &\leq \frac{\mu_0 r}{n}, \end{aligned}$$

and similarly if $i = i'$ and $j \neq j'$ we will have

$$\left|\left\langle \mathcal{P}_T(e_i e_{j'}^T), e_i e_j^T \right\rangle\right| \leq \frac{\mu_0 r}{n}.$$

Moreover, if $i \neq i'$ and $j \neq j'$ so

$$\left|\left\langle \mathcal{P}_T(e_{i'}e_{j'}^T), e_i e_j^T \right\rangle\right| \leq |\langle e_i, \boldsymbol{P}_U e_{i'}\rangle| \, |\langle e_j, \boldsymbol{P}_V e_{j'}\rangle| \leq \frac{\mu_0 r}{n}\frac{\mu_0 r}{n_2} = \frac{(\mu_0 r)^2}{n^2},$$

and using Equation 3.2.3 we will have the folowing if $i = i'$ and $j = j'$

$$\max_{ij,i'j'} \left|\left\langle \mathcal{P}_T(e_{i'}e_{j'}^T), e_i e_j^T \right\rangle\right| \leq \frac{2\mu_0 r}{n}. \tag{4.0.6}$$

As a result from 3.2.3 we will have the following estimate

$$\sum_{i'j'} \left|\left\langle \mathcal{P}_T(e_{i'}e_{j'}^T), e_i e_j^T \right\rangle\right|^2 = \sum_{i'j'} \left|\left\langle \mathcal{P}_T(e_i e_j^T), e_{i'} e_{j'}^T \right\rangle\right|^2 = \parallel \mathcal{P}_T(e_i e_j^T) \parallel \leq \frac{2\mu_0 r}{n}. \tag{4.0.7}$$

Furthermore, a similar estimate is that

$$\max_i \sum_j |E_{ij}|^2 \leq \frac{2\mu_0 r}{n}, \quad E = \sum_{k=1}^{r} u_k v_k^T. \tag{4.0.8}$$

which is the same if we exchange i and j, now, let

$$\sum_j |E_{ij}|^2 = \parallel e_i^T E \parallel^2 = \parallel \sum_{k \leq r} v_k \langle u_k, e_i \rangle \parallel^2 = \sum_{k \leq r} |\langle u_k, e_i \rangle| = \parallel \boldsymbol{P}_U e_i \parallel^2,$$

and from the coherence property we will have

$$\max_j \parallel \boldsymbol{P}_U e_i \parallel^2 \leq \frac{2\mu_0 r}{n}.$$

Note, we will consider that the matrices are square with dimension $n \times n$, and we will compute the operator norm for some random variables. We will denote $\boldsymbol{Z}$ as the matrix which we want to analyze, and $\boldsymbol{N}$ the matrix that we want to bound its norm.
Now we will introduce the following five lemmas, needed for Theorem 4.0.20.

4.0.1 Lemma. J.Candes and Recht, 2009 Set $\beta \geq 1$ and $\lambda \geq 1$, if $m \geq \lambda \mu_1^2 n r \beta \log n$, then there exist a numerical constant C_0 such that the norm of the first term of Equation 4.0.4 obeys

$$p^{-1} \parallel \mathcal{P}_{T^\perp}\mathcal{P}_\Omega \mathcal{P}_T(\boldsymbol{E}) \parallel \leq \frac{C_0}{\sqrt{\lambda}},$$

with probability at least $1 - n^{-\beta}$.

Proof of Lemma 4.0.1

Proof. In this section, we will find a bound for the first term in Equation 4.0.4. Firstly, since we know that $\mathcal{P}_{T^\perp}\mathcal{P}_T = 0$, $\mathcal{P}_T(\boldsymbol{E}) = \boldsymbol{E}$ and $\parallel \mathcal{P}_{T^\perp}(\boldsymbol{A}) \parallel \leq \parallel \boldsymbol{A} \parallel$ which is true for all matrices, so we can get

$$\begin{aligned} p^{-1} \parallel \mathcal{P}_{T^\perp}\mathcal{P}_\Omega \mathcal{P}_T(\boldsymbol{E}) \parallel &= p^{-1} \parallel \mathcal{P}_{T^\perp}(\mathcal{P}_\Omega - p\mathcal{I})\mathcal{P}_T(\boldsymbol{E}) \parallel \\ &\leq p^{-1} \parallel (\mathcal{P}_\Omega - p\mathcal{I})(\boldsymbol{E}) \parallel . \end{aligned}$$

Now, if we set

$$\boldsymbol{Z} = p^{-1}(\mathcal{P}_\Omega - p\mathcal{I})\boldsymbol{E} = p^{-1}\sum_{ij}(\alpha_{ij} - p)E_{ij}e_ie_j^T \tag{4.0.9}$$

since $\boldsymbol{Z}$ depends on the random variable α_{ij} so it is a random variable and $\mathbb{E}\boldsymbol{Z} = 0$ because $\mathbb{E}\alpha_{ij} = p$. Moreover we will proceed by estimating the size of $(\mathbb{E} \parallel \boldsymbol{Z} \parallel^q)^{\frac{1}{q}}$ for $q \geq 1$. Now, since $f(\boldsymbol{Z}) = \parallel \boldsymbol{Z} \parallel^q$ is convex, and denoting $\boldsymbol{Z}' = p^{-1}\sum_{ij}(\alpha'_{ij} - p)E_{ij}e_ie_j^T$ which is an independent copy of $\boldsymbol{Z}$, then using Definition 2.3.1 inequality we will get

$$\mathbb{E} \parallel \boldsymbol{Z} \parallel^q \leq \mathbb{E} \parallel \boldsymbol{Z} - \boldsymbol{Z}' \parallel^q . \tag{4.0.10}$$

Moreover, while $\alpha_{ij} - \alpha'_{ij}$ is symmetric, then $\boldsymbol{Z} - \boldsymbol{Z}'$ follow the same distribution as

$$p^{-1}\sum_{ij}\epsilon_{ij}(\alpha_{ij} - \alpha'_{ij})E_{ij}e_ie_j^T \equiv \boldsymbol{Z}_\epsilon - \boldsymbol{Z}'_\epsilon$$

denoting that $\boldsymbol{Z}_\epsilon = p^{-1}\sum_{ij}\epsilon_{ij}\alpha_{ij}E_{ij}e_ie_j^T$, where ϵ_{ij} is an independent Rademacher sequence. Now, if we used the triangle inequality we will have

$$(\mathbb{E} \parallel \boldsymbol{Z}_\epsilon - \boldsymbol{Z}'_\epsilon \parallel^q)^{\frac{1}{q}} \leq (\mathbb{E} \parallel \boldsymbol{Z}_\epsilon \parallel^q)^{\frac{1}{q}} + (\mathbb{E} \parallel \boldsymbol{Z}'_\epsilon \parallel^q)^{\frac{1}{q}} = 2\,(\mathbb{E} \parallel \boldsymbol{Z}_\epsilon \parallel^q)^{\frac{1}{q}} ,$$

therefore,

$$(\mathbb{E} \parallel \boldsymbol{Z}_\epsilon \parallel^q)^{\frac{1}{q}} \leq 2p^{-1}\left(\mathbb{E}_\alpha\mathbb{E}_\epsilon \parallel \sum_{ij}\epsilon_{ij}\alpha_{ij}E_{ij}e_ie_j^T \parallel^q\right)^{\frac{1}{q}} .$$

therefore, if $q = q'$ then

$$(\mathbb{E}_\alpha\mathbb{E}_\epsilon \parallel \boldsymbol{Z}_\epsilon \parallel^q)^{\frac{1}{q}} \leq \left(\mathbb{E}_\alpha\mathbb{E}_\epsilon \parallel \boldsymbol{Z}_\epsilon \parallel^q_{S_{q'}}\right)^{\frac{1}{q}} \leq \left(\mathbb{E}_\alpha\mathbb{E}_\epsilon \parallel \boldsymbol{Z}_\epsilon \parallel^{q'}_{S_{q'}}\right)^{\frac{1}{q'}} .$$

Now, using the noncommutayive Khintchine inequality with $q' \geq \log n$ we will get

$$\begin{aligned}
\left(\mathbb{E}_\alpha\mathbb{E}_\epsilon \parallel \boldsymbol{Z}_\epsilon \parallel^{q'}_{S_{q'}}\right)^{\frac{1}{q'}} &= \left(\mathbb{E}_\alpha\mathbb{E}_\epsilon \parallel p^{-1}\sum_{ij}\epsilon_{ij}\alpha_{ij}E_{ij}e_ie_j^T \parallel^{q'}_{S_{q'}}\right)^{\frac{1}{q'}} , \\
&\leq \frac{1}{p}C\sqrt{q'}\left(\mathbb{E}_\alpha \max\left[\parallel \left(\sum_{ij}\alpha_{ij}e_je_j^TE^2_{ij}\right)^{\frac{1}{2}} \parallel^{q'}_{S_{q'}}, \parallel \left(\sum_{ij}\alpha_{ij}E^2_{ij}e_ie_i^T\right)^{\frac{1}{2}} \parallel^{q'}_{S_{q'}}\right]\right)^{\frac{1}{q'}} , \\
&\leq \frac{1}{p}Ce\sqrt{q'}\left(\mathbb{E}_\alpha \max\left[\parallel \sum_{ij}\alpha_{ij}e_je_j^TE^2_{ij} \parallel^{\frac{q'}{2}}, \parallel \sum_{ij}\alpha_{ij}E^2_{ij}e_ie_i^T \parallel^{\frac{q'}{2}}\right]\right)^{\frac{1}{q'}} ,
\end{aligned} \tag{4.0.11}$$

therefore, both of the terms are the same, so we can choose to bound one of them. In the first term, we notice that this $\sum_{ij}\alpha_{ij}e_je_j^TE^2_{ij}$ is a dagonal matrix, so

$$\parallel \sum_{ij}\alpha_{ij}e_je_j^TE^2_{ij} \parallel = \max_j \sum_i \alpha_{ij}E^2_{ij}. \tag{4.0.12}$$

Then, we will use the following lemma to help us bounding Equation 4.0.12.

4.0.2 Lemma. Let q be an ineger obeys $1 \leq q \leq np$ and assume that $np \geq 2\log n$. Then

$$\mathbb{E}\left(\max_j \sum_i \alpha_{ij} E_{ij}^2\right)^q \leq 2\left(2np \parallel \boldsymbol{E} \parallel_\infty^2\right)^q.$$

The proof can be found in J.Candes and Recht, 2009 appendid.

Now, using lemma 4.0.2 and equation 4.0.12, for each $q \geq 1$, then

$$\frac{1}{p} Ce\sqrt{q'}\left(\mathbb{E}_\alpha \max\left[\parallel \sum_{ij} \alpha_{ij} e_j e_j^T E_{ij}^2 \parallel^{\frac{q'}{2}}, \parallel \sum_{ij} \alpha_{ij} E_{ij}^2 e_i e_i^T \parallel^{\frac{q'}{2}}\right]\right)^{\frac{1}{q'}}$$
$$\leq 2(np \parallel \boldsymbol{E} \parallel_\infty^2)^{\frac{q}{2}} = 2\left(\sqrt{2np} \parallel \boldsymbol{E} \parallel_\infty\right)^q.$$

Now, choose $q = \beta \log n$ for some $\beta \geq 1$, and put $q' = q$, then since $\parallel \boldsymbol{E} \parallel_\infty \leq \frac{\mu_1\sqrt{r}}{n}$ so

$$\begin{aligned}(\mathbb{E} \parallel \boldsymbol{Z} \parallel^q)^{\frac{1}{q}} &\leq \frac{C\sqrt{q}e}{p}\sqrt{2}\sqrt{np} \parallel \boldsymbol{E} \parallel_\infty, \\ &\leq Ce\frac{\sqrt{\beta \log n}}{p}\sqrt{np}\frac{\mu_1\sqrt{r}}{n}, \\ &= Ce\mu_1\sqrt{\frac{npr\beta \log n}{p^2n^2}} = Ce\mu_1\sqrt{\frac{nr\beta\log n}{m}} \equiv Q, \end{aligned} \tag{4.0.13}$$

where Q is a representing the constant, and by using Markov's inequality, for each $t \geq 0$

$$\mathbb{P}\left(\parallel \boldsymbol{Z} \parallel > tQ\right) \leq t^{-q}$$

then, if $t = e$ we will obtain

$$\mathbb{P}\left(\parallel \boldsymbol{Z} \parallel > Ce\mu_1\sqrt{\frac{nr\beta\log n}{m}}\right) \leq e^{-\beta \log n} = e^{\log n^{-\beta}} = n^{-\beta}.$$

Therefore, we let $m \geq \max(\beta, 2) n \log n$, for lemma 4.0.2 to hold.

4.0.3 Theorem. *Suppose $\boldsymbol{X}$ is a fixed $n \times n$ matrix. Then there exist a numerical constant C_0 such that for each $\beta > 2$*

$$p^{-1} \parallel (\mathcal{P}_\Omega - p\mathcal{I})\boldsymbol{X} \parallel \leq C_0\left(\frac{\beta n \log n}{p}\right)^{\frac{1}{2}} \parallel \boldsymbol{X} \parallel_\infty, \tag{4.0.14}$$

provided that $np \geq \beta \log n$ with probability at least $1 - n^{-\beta}$.

Now, using theorem 4.0.3 we can obtain

$$\begin{aligned}p^{-1} \parallel (\mathcal{P}_\Omega - p\mathcal{I})\boldsymbol{E} \parallel &\leq C_0\sqrt{\frac{\beta n \log n}{p}} \parallel \boldsymbol{E} \parallel_\infty, \\ &\leq C_0\sqrt{\frac{\beta n \log n}{p}}\frac{\mu_1\sqrt{r}}{n} = C_0\sqrt{\frac{\beta n r \log n}{m}},\end{aligned}$$

which proof lemma 4.0.1. ◻

4.0.4 Lemma. J.Candes and Recht, 2009 Set $\beta \geq 1$ and $\lambda \geq 1$, if $m \geq \lambda \mu_1 \max(\sqrt{\mu_0}, \mu_1) nr\beta \log n$, then there exist numerical constants C_1, c_1 such that the norm of the second term of 4.0.4 obeys

$$p^{-1} \parallel \mathcal{P}_{T^\perp} \mathcal{P}_\Omega \mathcal{P}_T(\mathcal{R}(\boldsymbol{E})) \parallel \leq \frac{C_1}{\lambda},$$

with probability at least $1 - c_1 n^{-\beta}$.

Proof of Lemma 4.0.4

Proof. Here in this section we want to prove Lemma 4.0.4, and our goal will be to bound the second term in Equation 4.0.4. Starting with

$$p^{-1} \parallel \mathcal{P}_{T^\perp} \mathcal{P}_\Omega \mathcal{P}_T \mathcal{R}(\boldsymbol{E}) \parallel \leq p^{-1} \parallel (\mathcal{P}_\Omega - p\mathcal{I})\mathcal{R}(\boldsymbol{E}) \parallel,$$

then we will set

$$\boldsymbol{Z} \equiv p^{-1} \parallel (\mathcal{P}_\Omega - p\mathcal{I})\mathcal{R}(\boldsymbol{E}) \parallel = p^{-2} \sum_{i,j,i',j'} \delta_{ij}\delta_{i'j'} E_{ij} \left\langle \mathcal{P}_T e_{i'} e_{j'}^T, e_i e_j^T \right\rangle e_i e_j^T,$$

where $\delta_{ij} = \alpha_{ij} - p$. Now let us decompose $\boldsymbol{Z}$ as the sum of a diagonal matrix and an off diagonal matrix as follows

$$\boldsymbol{Z} = p^{-2} \sum_{(i,j)=(i',j')} + p^{-2} \sum_{(i,j)\neq(i',j')} = \boldsymbol{Z}_1 + \boldsymbol{Z}_2.$$

As we decompose $\boldsymbol{Z}$ as the sum of two matrices, then instead of bounding $\boldsymbol{Z}$ we will bound the two matrices $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$, seperately.

4.0.5 Bounding the 1st term.
Firstly, we will have

$$\boldsymbol{Z}_1 = p^{-2} \sum_{(i,j)} \delta_{ij}^2 E_{ij} \left\langle \mathcal{P}_T e_i e_j^T, e_i e_j^T \right\rangle e_i e_j^T.$$

Then, we can decompose δ_{ij}^2 as follows

$$\delta_{ij}^2 = (\alpha_{ij} - p)^2 = (1-2p)(\alpha_{ij} - p) + p(1-p) = (1-2p)\delta_{ij} + p(1-p), \tag{4.0.15}$$

then, $\boldsymbol{Z}_1$ will have the form

$$\boldsymbol{Z}_1 = p^{-2}(1-2p) \sum_{(i,j)} \delta_{ij} E_{ij} \left\langle \mathcal{P}_T e_i e_j^T, e_i e_j^T \right\rangle e_i e_j^T + p^{-1}(1-p) \sum_{(i,j)} E_{ij} \left\langle \mathcal{P}_T e_i e_j^T, e_i e_j^T \right\rangle e_i e_j^T.$$

Now, setting $\boldsymbol{N} = p^{-1} E_{ij} \left\langle \mathcal{P}_T e_i e_j^T, e_i e_j^T \right\rangle$, we will have

$$\boldsymbol{Z}_1 = \frac{1-2p}{p} \sum_{(i,j)} \delta_{ij} \boldsymbol{N}_{ij} e_i e_j^T + (1-p) \sum_{(i,j)} \boldsymbol{N}_{ij} e_i e_j^T.$$

Using Theorem 4.0.3 we will obtain the bound for the first term of $\boldsymbol{Z}_1$, so

$$p^{-1} \parallel \sum_{(i,j)} \delta_{ij} \boldsymbol{N}_{ij} e_i e_j^T \parallel \leq C_0 \frac{n\beta \log n}{p} \parallel \boldsymbol{N} \parallel_\infty,$$

with probability at least $1-n^{-\beta}$. Moreover, we can bound $\parallel \boldsymbol{N} \parallel_\infty$ using $\parallel \boldsymbol{E} \parallel_\infty \leq \frac{\mu_1\sqrt{r}}{n}$ and $|\left\langle \mathcal{P}_T e_i e_j^T, e_i e_j^T \right\rangle| \leq \frac{2\mu_0 r}{n}$, then we will have

$$p^{-1} \parallel \sum_{(i,j)} \delta_{ij} \boldsymbol{N}_{ij} e_i e_j^T \parallel \leq C_0 \frac{n^3 \beta \log n}{n^2 p} \frac{1}{p} \frac{\mu_1 \sqrt{r}}{n} \frac{2\mu_0 r}{n}$$
$$= C\mu_0\mu_1 \frac{nr}{m} \sqrt{\frac{nr\beta \log n}{m}}$$

with probability at least $1-n^{-\beta}$. Now to bound the second term, firstly, let us declare that it is deterministic because it doesn't depend on any random variable. We will use the following lemma to help us bound it.

4.0.6 Lemma. Suppose that $\boldsymbol{A}$ is a fixed matrix; put $\boldsymbol{X} = \sum_{ij} \boldsymbol{A}_{ij} \left\langle \mathcal{P}_T e_i e_j^T, e_i e_j^T \right\rangle$. Then

$$\parallel \boldsymbol{X} \parallel \leq \frac{2\mu_0 r}{n} \parallel \boldsymbol{A} \parallel .$$

The proof of Lemma 4.0.6 can be found in J.Candes and Recht, 2009.

Using Lemma 4.0.6 on the second term with $\parallel \boldsymbol{E} \parallel = 1$ and $|\left\langle \mathcal{P}_T e_i e_j^T, e_i e_j^T \right\rangle| \leq \frac{2\mu_0 r}{n}$ gives

$$\parallel \boldsymbol{N} \parallel = \parallel p^{-1} E_{ij} \left\langle \mathcal{P}_T e_i e_j^T, e_i e_j^T \right\rangle \parallel \leq \frac{1}{p} \frac{2\mu_0 r}{n} = \frac{2\mu_0 nr}{m}.$$

We can conclude that $\boldsymbol{Z}_1$ is bounded by

$$\parallel \boldsymbol{Z}_1 \parallel \leq C\mu_0\mu_1 \frac{nr}{m} \sqrt{\frac{nr\beta \log n}{m}} + \frac{2\mu_0 nr}{m}$$
$$= C_z \frac{nr}{m} \left(\mu_0\mu_1 \sqrt{\frac{\beta nr \log n}{m}} + \mu_0 \right)$$

for some $C_z > 0$ with probability at least $1-n^{-\beta}$.
Now it remains to bound the matrix $\boldsymbol{Z}_2$. We will start by using the following lemma

4.0.7 Lemma. Assume that $\{a_i\}_{1\leq i\leq n}$ is a sequence of independent random variables, and $\{x_{ij}\}_{i\neq j}$ are elements taken from the Banach space which is a complete norm vector space, then

$$\mathbb{P}\left(\parallel \sum_{i\neq j} a_i a_j x_{ij} \parallel \geq t \right) \leq C_D \mathbb{P}\left(\parallel \sum_{i\neq j} a_i a_j' x_{ij} \parallel > \frac{t}{C_D} \right),$$

where $\{a_j'\}$ is an independent copy of $\{a_j\}$.

4.0.8 Bounding the 2nd term.

Using Lemma 4.0.7, we can bound

$$\boldsymbol{Z}_2' = p^{-2} \sum_{(i,j)\neq(i',j')} \delta_{ij} \delta'_{i'j'} E_{i'j'} \left\langle \mathcal{P}_T e_{i'} e_{j'}^T, e_i e_j^T \right\rangle e_i e_j^T,$$

where $\{\delta'_{ij}\}$ is an independent copy of $\{\delta_{ij}\}$. Now by setting

$$\boldsymbol{N}_{ij} = p^{-1} \sum_{i',j':(i,j)\neq(i',j')} \delta'_{i'j'} E_{i'j'} \left\langle \mathcal{P}_T e_{i'} e_{j'}^T, e_i e_j^T \right\rangle$$

we will obtain

$$\boldsymbol{Z}'_2 = p^{-1} \sum_{ij} \delta_{ij} \boldsymbol{N}_{ij} e_i e_j^T.$$

To bound $\boldsymbol{Z}'_2$ notice that

$$\mathbb{P}\left(\| \boldsymbol{Z}'_2 \| \geq t\right) \leq \mathbb{P}\left(\| \boldsymbol{Z}'_2 \| \geq t | \, \| \boldsymbol{N} \|_\infty \leq K\right) + \mathbb{P}\left(\| \boldsymbol{N} \|_\infty > K\right). \tag{4.0.16}$$

Fortunately, using Theorem 4.0.3 we will be able to bound the first term of the right hand side of Equation 4.0.16 and we will have

$$\mathbb{P} \| \sum_{ij} \delta_{ij} \boldsymbol{N}_{ij} e_i e_j^T \| \leq C_0 \sqrt{\frac{n^3 \beta \log n}{m}} K,$$

with probability at least $1 - n^{-\beta}$. For the second term of the right hand side of Equation 4.0.16 we need to bound $\| \boldsymbol{N} \|_\infty$ and we will use an application of Bernstein's inequality which is defined by the following lemma.

4.0.9 Lemma. Suppose that $\boldsymbol{A}$ is a fixed matrix and define $\mathcal{Y}(\boldsymbol{A})$ as the matrix whose i, j entry is

$$[\mathcal{Y}(\boldsymbol{A})]_{ij} = p^{-1} \sum_{i',j':(i',j')=(i,j)} (\alpha_{ij} - p) \boldsymbol{A}_{i'j'} \left\langle \mathcal{P}_T e_{i'} e_{j'}^T, e_i e_j^T \right\rangle,$$

and $\{\alpha_{ij}\}$ is a Bernolli sequence with $\mathbb{P}(\alpha_{ij} = 1) = p$, then

$$\mathbb{P}\left(\| \mathcal{Y}(\boldsymbol{A}) \|_\infty > \lambda \sqrt{\frac{\mu_0 r}{np}} \| \boldsymbol{A} \|_\infty\right) \leq 2n^2 \exp\left(\frac{-\lambda^2}{2 + \frac{2}{3}\sqrt{\frac{\mu_0 r}{np}}\lambda}\right). \tag{4.0.17}$$

If $\lambda = \sqrt{3\beta \log n}$ then the right hand side of Equation 4.0.17 will be bounded by $2n^{2-\beta}$ and as a consequence $np \geq \frac{4}{3}\beta\mu_0 r \log n$. Moreover, if $\lambda = \sqrt{6\beta \log n}$ and $\beta > 2$ then the right hand side of Equation 4.0.17 will be bounded by $2n^{2-2\beta}$ which is less than $2n^{-\beta}$ giving that $np \geq \frac{8}{3}\beta\mu_0 r \log n$. Furthermore, setting $t = \lambda\sqrt{\frac{\mu_0 r}{np}} \| \boldsymbol{A} \|_\infty$ we will get the inequality 4.0.17.
The proof of Lemma 4.0.9 can be found in J.Candes and Recht, 2009.

Using Lemma 4.0.9, with $\| \boldsymbol{E} \|_\infty \leq \frac{\mu_1 \sqrt{r}}{n}$, considering $\boldsymbol{N} = \mathcal{A}(\boldsymbol{E})$ we will get

$$\| \boldsymbol{N} \|_\infty \leq C \frac{\mu_1 \sqrt{r}}{n} \sqrt{\frac{2\mu_0 r}{n}} \sqrt{\frac{\beta n \log n}{p}} = C_k \frac{\mu_1 \sqrt{r}}{n} \sqrt{\frac{\mu_0 n r \beta \log n}{m}},$$

with probability at least $1 - 2n^{-\beta}$ for each $\beta > 2$. Then

$$\begin{aligned}
\| \boldsymbol{Z}'_2 \| &= p^{-1} \| \sum_{ij} \delta_{ij} \boldsymbol{N}_{ij} e_i e_j^T \|, \\
&\leq C^* \sqrt{\frac{n^3 \beta \log n}{m}} C_k \frac{\mu_1 \sqrt{r}}{n} \sqrt{\frac{\mu_0 n r \beta \log n}{m}}, \\
&= C \sqrt{\mu_0} \mu_1 \frac{n r \beta \log n}{m},
\end{aligned}$$

with probability at least $1 - 3n^{-\beta}$.
In conclusion, now we can obtain the bound for $\boldsymbol{Z}$ as follows

$$\boldsymbol{Z} \equiv p^{-1} \parallel (\mathcal{P}_\Omega - p\mathcal{I})\mathcal{R}(\boldsymbol{E}) \parallel \leq C\frac{nr}{m}\left(\sqrt{\mu_0}\mu_1\left(\sqrt{\frac{\mu_0 nr\beta\log n}{m}} + \beta\log n\right) + \mu_0\right),$$

with probability at least $1 - (1 + 3C_D)n^{-\beta}$. We will expess the proof of Lemma 4.0.4 by establishing the following lemma.

4.0.10 Lemma. Suppose that $\boldsymbol{A}$ is a fixed $n \times n$ matrix. There exist a constant C_0' such that

$$p^{-2} \parallel \sum_{(i,j)\neq(i',j')} \delta_{ij}\delta_{i'j'}\boldsymbol{A}_{ij}\left\langle \mathcal{P}_T e_i e_j^T, e_i e_j^T\right\rangle e_i e_j^T \parallel \leq C_0' \frac{\sqrt{\mu_0 r}\beta\log n}{p} \parallel \boldsymbol{A} \parallel_\infty,$$

with probability at least $1 - O(n^{-\beta}) \quad \forall\beta > 2$ provided that $n \geq 3\mu_0 r\beta\log n$.

□

4.0.11 Lemma. J.Candes and Recht, 2009 Set $\beta \geq 1$ and $\lambda \geq 1$, if $m \geq \lambda\mu_0^{\frac{4}{3}} nr^{\frac{4}{3}}\beta\log n$, then there exist numerical constants C_2, c_2 such that the norm of the third term of 4.0.4 obeys

$$p^{-1} \parallel \mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T(\mathcal{R}^2(\boldsymbol{E})) \parallel \leq \frac{C_2}{(\sqrt{\lambda})^3},$$

with probability at least $1 - c_2 n^{-\beta}$.

Proof of Lemma 4.0.11

Proof. Proving Lemma 4.0.11 means bounding the third term of the Equation 4.0.4, so using $\mathcal{P}_T(\boldsymbol{E}) = \boldsymbol{E}$ and $\parallel \mathcal{P}_{T^\perp}(\boldsymbol{A}) \parallel \leq \parallel \boldsymbol{A} \parallel$ which is true for all matrices, we will get

$$\begin{aligned} p^{-1}(\mathbb{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T)\mathcal{R}^2(\boldsymbol{E}) &= p^{-1}(\mathcal{P}_\Omega - p\mathcal{I})\mathcal{R}^2(\boldsymbol{E}), \\ &= p^3 \sum_{i_1,j_1,i_2,j_2,i_3,j_3} \delta_{i_1j_1}\delta_{i_2j_2}\delta_{i_3j_3}E_{i_3j_3}\left\langle \mathcal{P}_T(e_{i_3}e_{j_3}^T), e_{i_2}e_{j_2}^T\right\rangle\left\langle \mathcal{P}_T(e_{i_2}e_{j_2}^T), e_{i_1}e_{j_1}^T\right\rangle e_{i_1}e_{j_1}^T, \end{aligned} \tag{4.0.18}$$

where $\delta_{ij} = \alpha_{ij} - p$. Just as a notation, we will use $t_k = (i_k, j_k)$, $F_{t_k} = e_{i_k}e_{j_k}^T$ and $P_{t_k t_l} = \left\langle \mathcal{P}_T(e_k e_k^T), e_l e_l^T\right\rangle$, we will be using them as long as we are proceeding in this proof. We can rewrite Equation 4.0.18 as follows

$$p^{-1}(\mathcal{P}_\Omega - p\mathcal{I})\mathcal{R}^2 = p^{-3}\sum_{t_1,t_2,t_3} \delta_{t_1}\delta_{t_2}\delta_{t_3}E_{t_3}P_{t_3t_2}P_{t_2t_1}F_{t_1}. \tag{4.0.19}$$

The next step is to decompose the matrix 4.0.19 as the sum of matrices depending on whether the t_k's are similar or not, as follows

$$p^{-1}((\mathcal{P}_\Omega - p\mathcal{I}))\mathcal{R}^2 = \frac{1}{p^3}\left[\sum_{t_1=t_2=t_3} + \sum_{t_1\neq t_2=t_3} + \sum_{t_1=t_3\neq t_2} + \sum_{t_1=t_2\neq t_3} + \sum_{t_1\neq t_2\neq t_3}\right]. \tag{4.0.20}$$

From this point we need to bound each term of the right hand side of Equation 4.0.20.

4.0.12 Bounding the 1st term.

Starting with the first term where the sum is over $t_1 = t_2 = t_3 = t$ and we will denote it as $\boldsymbol{X}$, so

$$\boldsymbol{X} = \frac{1}{p^3} \sum_t \delta_t^3 E_t P_{tt}^2 F_t,$$

and by using the identity

$$\delta_t^3 = (1 - 3p + 3p^2)\delta_t + p(1 - 3p + 2p^2),$$

we will get

$$\boldsymbol{X} = \frac{(1 - 3p + 3p^2)}{p^3} \sum_t \delta_t^3 E_t P_{tt}^2 F_t + \frac{(1 - 3p + 2p^2)}{p^2} \sum_t E_t P_{tt}^2 F_t, \tag{4.0.21}$$

firstly, let $N_t = p^{-2} E_t P_{tt}^2$ then starting by bounding the first term of Equation 4.0.21, and using Lemma 4.0.6 with $\| \boldsymbol{E} \|_\infty \leq \frac{\mu_1 \sqrt{r}}{n}$ we will have

$$|N_t| \leq \left(\frac{2\mu_0 r}{np}\right)^2 \| \boldsymbol{E} \|_\infty \leq \left(\frac{2\mu_0 r}{np}\right)^2 \frac{\mu_1 \sqrt{r}}{n}.$$

Now, using Theorem 4.0.3, then for each $\beta > 2$ we will get

$$\begin{aligned} \| \boldsymbol{Z}_1 \| &= p^{-1} \| \sum_t \delta_t N_t F_t \|, \\ &\leq C \left(\frac{2\mu_0 r}{np}\right)^2 \frac{\mu_1 \sqrt{r}}{n} \sqrt{\frac{\beta n \log n}{p}}, \\ &= C\mu_0^2 \left(\frac{2nr}{m}\right)^2 \mu_1 \sqrt{\frac{\beta n r \log n}{m}}, \\ &= C\mu_0^2 \mu_1 \sqrt{\beta \log n} \left(\frac{nr}{m}\right)^{\frac{5}{2}}, \end{aligned}$$

with probability at least $1 - n^{-\beta}$.

Next, we will bound the second term of Equation 4.0.21. From Lemma 4.0.6, it follows that

$$\| \sum_t E_t P_{tt}^2 F_t \| \leq \left(\frac{2\mu_0 r}{n}\right)^2.$$

So that

$$\| \boldsymbol{N} \| \leq \frac{1}{p^2} \left(\frac{2\mu_0 r}{n}\right)^2 = \left(\frac{2\mu_0 r}{np}\right)^2.$$

Now we can conclude that the first term of Equation 4.0.18 is bounded by

$$\begin{aligned} \| \boldsymbol{X} \| &\leq C\mu_0^2 \mu_1 \sqrt{\beta \log n} \left(\frac{nr}{m}\right)^{\frac{5}{2}} + \left(\frac{2\mu_0 r}{np}\right)^2, \\ &= C \left(\frac{nr}{m}\right)^2 \left[\mu_0^2 \mu_1 \sqrt{\frac{\beta n r \log n}{m}} + \mu_0^2\right], \end{aligned}$$

with probability at least $1 - n^{-\beta}$.

4.0.13 Bounding the 2nd term.
We now consider the second term of Equation 4.0.20 and call it $\boldsymbol{Y}$. By using the identity 4.0.15 we can rewrite it as

$$\begin{aligned}\boldsymbol{Y} &= p^{-3}\sum_{t_1\neq t_2}\delta_{t_1}\delta_{t_2}^2 E_{t_2}P_{t_2t_2}P_{t_2t_1}F_{t_1},\\ &= \frac{1-2p}{p^3}\sum_{t_1\neq t_2}\delta_{t_1}\delta_{t_2}E_{t_2}P_{t_2t_2}P_{t_2t_1}F_{t_1} + \frac{1-p}{p^2}\sum_{t_1\neq t_2}\delta_{t_1}E_{t_2}P_{t_2t_2}P_{t_2t_1}F_{t_1},\end{aligned} \tag{4.0.22}$$

denote the first term by $\boldsymbol{Z}_1$, then to bound it we can use Lemma 4.0.10. Set $\boldsymbol{A}_t = p^{-1}E_tP_{tt}$ and since we know that $\|\ \boldsymbol{E}\ \|\ \infty \leq \frac{\mu_1\sqrt{r}}{n}$ and $|P_{tt}| \leq \frac{2\mu_0 r}{n}$ then we will get

$$\begin{aligned}\|\ \boldsymbol{Z}_1\ \| &\leq C_0'\frac{\sqrt{\mu_0 r}\beta\log n}{p}\ \|\ \boldsymbol{A}\ \|_\infty,\\ &\leq C_0'\frac{\sqrt{\mu_0 r}\beta\log n}{p}\frac{1}{p}\frac{2\mu_0 r}{n}\frac{\mu_1\sqrt{r}}{n},\\ &= C\frac{(\mu_0)^{\frac{3}{n}}r^2\mu_1 n^2}{p^2n^4}\beta\log n,\\ &= C(\mu_0)^{\frac{3}{2}}\mu_1\beta\log n\left(\frac{nr}{m}\right)^2.\end{aligned}$$

Next, we will bound the second term of Equation 4.0.22, denote it by $\boldsymbol{Z}_2$. Set $N_{t_1} = p^{-1}\sum_{t_2:t_2\neq t_1}E_{t_2}P_{t_2t_2}P_{t_2t_1}$ then we can rewrite it as

$$\boldsymbol{Z}_2 = p^{-1}\sum_{t_1}\delta_{t_1}N_{t_1}F_{t_1},$$

observe that N_{t_1} is deterministic, and to bound $\|\ \boldsymbol{N}\ \|_\infty$ we will use the following lemma:

4.0.14 Lemma. The matrix $\boldsymbol{N}$ satisfies that

$$\|\ \boldsymbol{N}\ \|_\infty \leq \frac{\mu_0 r}{np}\left(3\ \|\ \boldsymbol{E}\ \|_\infty + \frac{2\mu_0 r}{n}\right).$$

Using Lemma 4.0.14 and Theorem 4.0.3 we will get

$$\begin{aligned}\|\ \boldsymbol{Z}_2\ \| &\leq C_0\sqrt{\frac{\beta n\log n}{p}}\frac{\mu_0 r}{np}\left(\frac{3\mu_1\sqrt{r}}{n} + \frac{2\mu_0 r}{n}\right),\\ &\leq C_0\sqrt{\beta\log n}\left(\frac{nr}{m}\right)^{\frac{3}{2}}\left(\mu_1\mu_0 + \mu_0^2\sqrt{r}\right),\end{aligned}$$

with probability at least $1 - O(n^{-\beta})$. At this point we can conclude that the second term in equation 4.0.18 is bounded by

$$\boldsymbol{Y} \leq C\sqrt{\beta\log n}\left(\frac{nr}{m}\right)^{\frac{3}{2}}\left(\mu_0\mu_1\sqrt{\frac{\mu_0\beta nr\log n}{m}} + \mu_1\mu_0 + \mu_0^2\sqrt{r}\right).$$

4.0.15 Bounding the 3rd term.
In this step, we will bound the third term of Equation 4.0.18, denote it by $\boldsymbol{L}$. Then, by using the identity 4.0.15 we can rewrite it as

$$\begin{aligned}\boldsymbol{L} &= p^{-3}\sum_{t_1\neq t_2}\delta_{t_1}^2\delta_{t_2}E_{t_1}P_{t_1t_2}P_{t_2t_1}F_{t_1},\\ &= p^{-3}\sum_{t_1\neq t_2}\delta_{t_1}^2\delta_{t_2}E_{t_1}P_{t_2t_1}^2F_{t_1},\\ &= \frac{(1-2p)}{p^3}p^{-3}\sum_{t_1\neq t_2}\delta_{t_1}\delta_{t_2}E_{t_1}P_{t_2t_1}^2F_{t_1}+\frac{(1-2p)}{p^3}p^{-3}\sum_{t_1\neq t_2}\delta_{t_2}E_{t_1}P_{t_2t_1}^2F_{t_1}.\end{aligned} \tag{4.0.23}$$

Then we will start by bounding the first term of Equation 4.0.23, denote it as $\boldsymbol{Z}_1$. Set $N_{t_1} = p^{-2}\sum_{t_2:t_2\neq t_1}\delta_{t_2}^{(2)}P_{t_2t_1}^2$, then we can rewrite it as

$$\boldsymbol{Z}_1 = p^{-1}\sum_{t_1}\delta_{t_1}^{(1)}E_{t_1}N_{t_1}F_{t_1},$$

where $\{\delta_{t_1}^{(1)}\}$ and $\{\delta_{t_2}^{(2)}\}$ are independent copies of $\{\delta_{t_1}\}$ and $\{\delta_{t_2}\}$, respectively.
Using the Bernstein's inequality application given in lemma 4.0.10 with $|P_{tt}| \leq \frac{2\mu_0 r}{n}$ and

$$\sum_{t_2:t_2\neq t_1}|P_{t_2t_1}|^4 \leq \max_{t_2:t_2\neq t_1}|P_{t_2t_1}|^2\sum_{t_2:t_2\neq t_1}|P_{t_2t_1}|^2 \leq \left(\frac{2\mu_0 r}{n}\right)^2\frac{2\mu_0 r}{n} = \left(\frac{2\mu_0 r}{n}\right)^3,$$

therefore, for each $\lambda > 0$

$$\mathbb{P}\left(|N_t| > \lambda\left(\frac{2\mu_0 r}{n}\right)^{\frac{3}{2}}\right) \leq 2n^2\exp\left(\frac{\lambda^2}{2+\frac{2}{3}\lambda\sqrt{\frac{2\mu_0 r}{np}}}\right), \tag{4.0.24}$$

then, it follows from Equation 4.0.24 that

$$\mathbb{P}\left(\|\ \boldsymbol{N}\ \|_\infty > \sqrt{8\beta\log n}\left(\frac{2\mu_0 r}{n}\right)^{\frac{3}{2}}\right) \leq 2n^2\exp\left(\frac{8\beta\log n}{2+\frac{2}{3}\sqrt{8\beta\log n}\sqrt{\frac{2\mu_0 r}{np}}}\right) = 2n^{-2\beta+2},$$

provided that $m \geq \frac{16}{9}\mu_0 nr\beta\log n$. Moreover, using Theorem 4.0.3 we will get for each $\beta > 2$

$$\begin{aligned}\|\ \boldsymbol{Z}_1\ \| &\leq C_0\sqrt{\frac{\beta n\log n}{p}}\ \|\ \boldsymbol{N}\ \|_\infty\|\ \boldsymbol{E}\ \|_\infty,\\ &\leq C_0\sqrt{\frac{\beta n\log n}{p}}\sqrt{8\beta\log n}\left(\frac{2\mu_0 r}{n}\right)^{\frac{3}{2}}\frac{\mu_1\sqrt{r}}{n},\\ &= C_0\mu_0^{\frac{3}{2}}\mu_1\left(\frac{nr}{m}\right)^2\beta\log n,\end{aligned}$$

with proobability at least $1-3n^{-\beta}$. Then, bounding the seconed of Equation 4.0.23, denote it as $\boldsymbol{Z}_2$, firstly, we will write it as

$$\boldsymbol{Z}_2 = (1-p)\sum_{t_1}E_{t_1}N_{t_1}F_{t_1},$$

and then

$$\begin{aligned}\| \sum_{t_1} E_{t_1} N_{t_1} F_{t_1} \| &\leq \| \sum_{t_1} E_{t_1} N_{t_1} F_{t_1} \|_F \\ &\leq \| \boldsymbol{N} \|_\infty \| \boldsymbol{N} \|_F, \\ &\leq C\sqrt{\beta \log n}\left(\frac{\mu_0 nr}{m}\right)^{\frac{3}{2}} \sqrt{r}.\end{aligned}$$

Now, we can conclude that the third term of Equation 4.0.20 is bounded by

$$\| \boldsymbol{L} \| \leq C\mu_0 \sqrt{\beta \log n}\left(\frac{nr}{m}\right)^{\frac{3}{2}} \left(\mu_1 \sqrt{\frac{\mu_0 nr\beta \log n}{m}} + \sqrt{\mu_0 r}\right),$$

with probability at least $1 - O(n^{-\beta})$.

4.0.16 Bounding the 4th term.

Proceeding with the bounding Equation 4.0.20 terms, we will bound the fourth term, denote it by $\boldsymbol{G}$ by firstly using the identity 4.0.15, so we can rewrite it as

$$\begin{aligned}\boldsymbol{G} &= p^{-3} \sum_{t_1 \neq t_3} \delta_{t_1}^2 \delta_{t_3} E_{t_3} P_{t_3 t_1} P_{t_1 t_1} F_{t_1}, \\ &= \frac{1-2p}{p^3} \sum_{t_1 \neq t_3} \delta_{t_1} \delta_{t_3} E_{t_3} P_{t_3 t_1} P_{t_1 t_1} F_{t_1} + \frac{1-p}{p^2} \sum_{t_1 \neq t_3} \delta_{t_3} E_{t_3} P_{t_3 t_1} P_{t_1 t_1} F_{t_1}.\end{aligned} \tag{4.0.25}$$

We denote the first term of Equation 4.0.25 as $\boldsymbol{Z}_1$, then put $N_{t_1} = p^{-2} \sum_{t_1 \neq t_3} \delta_{t_1} \delta_{t_3} E_{t_3} P_{t_3 t_1} F_{t_1}$ and by using Lemma 4.0.9 we have

$$\begin{aligned}\| \boldsymbol{Z}_1 \| &\leq \frac{2\mu_0 r}{np} \| \boldsymbol{N} \|, \\ &\leq \frac{2\mu_0 r}{np} C_0' \sqrt{\mu_0 r} \frac{\beta \log n}{p} \| \boldsymbol{E} \|_\infty, \\ &\leq \frac{2\mu_0 r}{np} C_0' \sqrt{\mu_0 r} \frac{\beta \log n}{p} \frac{\mu_1 \sqrt{r}}{n}, \\ &= C(\mu_0)^{\frac{3}{2}} \mu_1 \beta \log n \frac{nr}{m}.\end{aligned}$$

Now to bound the second term of Equation 4.0.25, we denote it as $\boldsymbol{Z}_2$ and set $N_{t_1} = p^{-1} \sum_{t_1 \neq t_3} \delta_{t_3} E_{t_3} P_{t_3 t_1}$. Using Lemma 4.0.6 we obtain

$$\| \boldsymbol{Z}_2 \| \leq \frac{2\mu_0 r}{np} \| \boldsymbol{N} \|,$$

then, we can rewrite N_{t_1} as

$$N_{t_1} = \sum_{t_3} \delta_{t_3} E_{t_3} P_{t_3 t_1} - \delta_{t_1} E_{t_1} P_{t_1 t_1},$$

then, setting $H_{t_1} = E_{t_1} P_{t_1 t_1}$ we get

$$\boldsymbol{N} = p^{-1} \left[\mathcal{P}_T (\mathcal{P}_\Omega - p\mathcal{I})(\boldsymbol{E}) - (\mathcal{P}_\Omega - p\mathcal{I})(\boldsymbol{H}) \right].$$

Also, for any matrix $\boldsymbol{A}$ we have that

$$\| \mathcal{P}_T(\boldsymbol{A}) \| = \| \boldsymbol{A} - \mathcal{P}_{T^\perp}(\boldsymbol{A}) \| \leq 2 \| \boldsymbol{A} \|,$$

and then,

$$\| \boldsymbol{N} \| \leq 2p^{-1} \| (\mathcal{P}_\Omega - p\mathcal{I})(\boldsymbol{E}) \| + p^{-1} \| (\mathcal{P}_\Omega - p\mathcal{I})(\boldsymbol{H}) \|,$$

and since $\| \boldsymbol{H} \|_\infty \leq \| \boldsymbol{E} \|_\infty$, then using Theorem 4.0.3 we will have for each $\beta > 2$

$$\begin{aligned} \| \boldsymbol{N} \| &\leq 2C_0 \sqrt{\frac{\beta n \log n}{p}} \| \boldsymbol{E} \|_\infty + C_0 \sqrt{\frac{\beta n \log n}{p}} \| \boldsymbol{H} \|_\infty, \\ &\leq 3C_0 \sqrt{\frac{\beta n \log n}{p}} \frac{\mu_1 \sqrt{r}}{n}, \\ &= C\mu_1 \sqrt{\frac{\beta n r \log n}{m}}, \end{aligned}$$

with probability at least $1 - n^{-\beta}$. We can conclude that the fourth term in 4.0.20 is bounded by

$$\| \boldsymbol{G} \| \leq C\mu_0 \mu_1 \sqrt{\beta \log n} \left(\frac{nr}{m} \right)^{\frac{3}{2}} \left(\sqrt{\frac{\mu_0 n r \beta \log n}{m}} + 1 \right),$$

wit hprobability at least $1 - O(n^{-\beta})$.

4.0.17 Bounding the 5th term.

Finally, we will estimate the last term of the Equation 4.0.20 which is

$$\sum_{t_1 \neq t_2 \neq t_3} \delta_{t_1} \delta_{t_2} \delta_{t_3} E_{t_3} P_{t_3 t_2} P_{t_2 t_1} F_{t_1}.$$

Now using the decopling inequality for triple we will have

$$\boldsymbol{J} = \sum_{t_1 \neq t_2 \neq t_3} \delta_{t_1}^{(1)} \delta_{t_2}^{(2)} \delta_{t_3}^{(3)} E_{t_3} P_{t_3 t_2} P_{t_2 t_1} F_{t_1}.$$

where the sequences $\{\delta_t^{(1)}\}, \{\delta_t^{(2)}\}$ and $\{\delta_t^{(3)}\}$ are an independent copies of $\{\delta_t\}$. Then setting $H_{t_2} = p^{-1} \sum_{t_3 : t_3 \neq t_2} \delta_{t_3}^{(3)} E_{t_3} P_{t_3 t_2}$, $N_{t_1} = p^{-1} \sum_{t_2 : t_2 \neq t_1} \delta_{t_2}^{(2)} H_{t_2} P_{t_2 t_1}$ then $\boldsymbol{J} = p^{-1} \sum_{t_1} \delta_{t_1}^{(1)} N_{t_1} F_{t_1}$. By using Lemma 4.0.9 we will get for each $\beta > 2$

$$\| \boldsymbol{H} \|_\infty \leq C \sqrt{\frac{\mu_0 n r \beta \log n}{m}} \| \boldsymbol{E} \|_\infty,$$

with large probability, and by using the same approach for the second time we will obtain

$$\| \boldsymbol{N} \|_\infty \leq C \sqrt{\frac{\mu_0 n r \beta \log n}{m}} \| \boldsymbol{H} \|_\infty \leq C \frac{\mu_0 n r \beta \log n}{m} \| \boldsymbol{E} \|_\infty,$$

with priobability at least $1 - 4n^{-\beta}$. Which means

$$\begin{aligned} \| \boldsymbol{J} \| &\leq C_0 \sqrt{\frac{\beta n \log n}{p}} \| \boldsymbol{N} \|_\infty, \\ &\leq C_0 \sqrt{\frac{\beta n^3 \log n}{m}} C \frac{\mu_0 n r \beta \log n}{m} \frac{\mu_1 \sqrt{r}}{n}, \\ &= C\mu_0 \mu_1 \left(\frac{n r \beta \log n}{m} \right)^{\frac{3}{2}}, \end{aligned}$$

with probability at least $1 - O(n^{-\beta})$.
At the end of this section, after bounding all the terms of Equation 4.0.20, with the given estimates $\mu_0 \geq 1$ and $\mu_1 \leq \mu_0\sqrt{r}$ then, if $m \geq \mu_0 nr\beta \log n$ we have

$$\begin{aligned}\| p^{-1}((\mathcal{P}_\Omega - p\mathcal{I}))\mathcal{R}^2(\boldsymbol{E}) \| \leq C\left(\frac{nr}{m}\right)^2 &\left[\mu_0^2\mu_1\sqrt{\frac{\beta nr\log n}{m}} + \mu_0^2\right] \\ &+ C\sqrt{\beta\log n}\left(\frac{nr}{m}\right)^{\frac{3}{2}}\left(\mu_0\mu_1\sqrt{\frac{\mu_0\beta nr\log n}{m}} + \mu_1\mu_0 + \mu_0^2\sqrt{r}\right) \\ &+ C\mu_0\sqrt{\beta\log n}\left(\frac{nr}{m}\right)^{\frac{3}{2}}\left(\mu_1\sqrt{\frac{\mu_0 nr\beta\log n}{m}} + \sqrt{\mu_0 r}\right) \\ &+ C\mu_0\mu_1\left(\frac{nr\beta\log n}{m}\right)^{\frac{3}{2}}, \qquad (4.0.26)\end{aligned}$$

with probability at least $1 - O(n^{-\beta})$. Furthermore, if $m = \lambda\mu_0^{\frac{4}{3}}nr^{\frac{4}{3}}\beta\log n$ for a fixed $\beta > 2$ and $\lambda > 1$, then using 4.0.26 we will get that

$$\| p^{-1}((\mathcal{P}_\Omega - p\mathcal{I}))\mathcal{R}^2(\boldsymbol{E}) \| \leq C_1\lambda^{\frac{3}{2}},$$

with probability at least $1 - O(n^{-\beta})$. □

The proof of Lemma 4.0.11 is now complete.

4.0.18 Lemma. Set $\beta \geq 1$ and $\lambda \geq 1$, if $m \geq \lambda\mu_0^2 nr^2\beta\log n$, then there exist numerical constants C_3, c_3 such that the norm of the fourth term of 4.0.4 obeys

$$p^{-1} \| \mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T(\mathcal{R}^3(\boldsymbol{E})) \| \leq \frac{C_3}{\sqrt{\lambda}},$$

with probability at least $1 - c_3 n^{-\beta}$.
The proof of Lemma 4.0.18 can be found in J.Candes and Recht, 2009.

4.0.19 Lemma. J.Candes and Recht, 2009 If theorem 3.2.1 assumptions are satisfied and if $m \geq (2K)^2\mu_0 n^{\frac{5}{4}}r\beta\log n$, then there exist a numerical constant C_a such that

$$p^{-1} \| \mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T(\sum_{k\geq a}\mathcal{R}^k(\boldsymbol{E})) \| \leq C_a\left(\frac{n^2 r}{m}\right)^{\frac{1}{2}}\left(\frac{\mu_0 nr\beta\log n}{m}\right)^{\frac{a}{2}},$$

with probability at least $1 - n^{-\beta}$.
The proof of Lemma 4.0.19 can be found in J.Candes and Recht, 2009.

Now, combining the four Lammas 4.0.1, 4.0.4, 4.0.11 and 4.0.19 and considering $a = 3$ in Lemma 4.0.19, then generally they garantee that the recovery will be exact, and there exist numerical constants C and c such that

$$m \geq C\max\left(\mu_1^2, \mu_0^{\frac{1}{2}}\mu_1, \mu_0^{\frac{4}{3}}r^{\frac{1}{3}}, \mu_0 n^{\frac{1}{4}}\right)nr\beta\log n, \qquad (4.0.27)$$

and $\| \mathcal{P}_{T^\perp}(\boldsymbol{W}) \| < 1$ with probability at least $1 - cn^{\beta}$. In the case where $m \geq \mu_0 n^{\frac{1}{4}}nr\beta\log n$ if $\mu_0 n^{\frac{1}{4}}r\beta\log n \approx n$ then it would be problematic since it means we should observe almost all the entries

of the matrix, but if $\mu_0 n^{\frac{1}{4}} r\beta \log n \geq pn$ where $p < 1$, then it would be possible to recover the matrix efficiently. To solve our problem we need p to be in a range which allow us to complete a sparse matrix. Moreover, if $\mu_0 r \leq n^{\frac{1}{5}}$ then from Lemma 4.0.19 we can garantee that the recovery is exact with $m \geq \mu_0 n^{\frac{6}{5}} r\beta \log n$ with high probability.
Now we can introduce the main theorems that garantee the exact recovery of a matrix from a small set of its entries.

4.0.20 Theorem. *J.Candes and Recht, 2009 Suppose that $\boldsymbol{S}$ has dimension $n_{\times} n$ and $rank(\boldsymbol{S}) = r$ with incoherent column and row spaces satisfying $\boldsymbol{S}_1$, $\boldsymbol{S}_2$. Assume that we observe m entries of the matrix $\boldsymbol{S}$ with locations sampled uniformally at random. Then there exist numerical constants c, C, such that the minimizer for the problem 2.4.2 is unique and can recover M efficiently if*

$$m \geq C \max\left(\mu_1^2, \mu_0^{\frac{1}{2}} \mu_1, \mu_0 n^{\frac{1}{4}}\right) nr(\beta \log n), \tag{4.0.28}$$

with probability at least $1 - cn^{-\beta}$. Moreover, if $r \leq n^{\frac{1}{5}}$ so the recovery for M is exact with the same probability if

$$m \geq C\mu_0 n^{\frac{6}{5}} r(\beta \log n). \tag{4.0.29}$$

4.0.21 Theorem. *J.Candes and Recht, 2009 Let $\boldsymbol{S}$ be an $n \times n$ matrix with rank r sampled from the random orthogonal model. Assume we observe m entries of $\boldsymbol{S}$ with locations sampled uniformaly at random. Then there exist numerical constants c, C such that if*

$$m \geq Cn^{\frac{5}{4}} r \log n, \tag{4.0.30}$$

then the minimizer for the problem 2.4.2 is unique and can recover $\boldsymbol{S}$ perfectly with probability at least $1 - cn^{-3}$. Furthermore, if $r \leq n^{\frac{1}{5}}$ then the recovery for $\boldsymbol{S}$ is exact with probability at least $1 - cn^{-3}$ provided that

$$m \geq Cn^{\frac{6}{5}} r \log n. \tag{4.0.31}$$

As a conclusion, if the number of the observed entries is small, then the recovery of the matrix $\boldsymbol{S}$ will be sufficient since it is a low rank matrix. And if the rank is too small, we consider the number of the entries of order $n^{\frac{6}{5}}$, which is quite smaller than the total number of the matrix entries n^2. This guarantees the existance of a unique low rank matrix, and also solving the optimization problem 2.4.2 is equivalent to solving the problem 2.4.1.

5. Results and Conclusion

In this chapter, we will discuss the numerical results and the conclusion.

5.1 Numerical Results

We have used the convex optimization solver (cvxpy) which is a Python-embedded solver to solve the convex optimization problem 2.4.2. The first remark is the limited dimension of this software, also, when the dimension grows, the CPU time grows as well, and, it was difficult to run matrices with dimension 1000 and even 500.
Firstly, to have a generic matrix, we generate random matrices following the Gaussian distribution, $\boldsymbol{L}, \boldsymbol{R} \in \mathbb{R}^{n \times r}$ and setting the matrix $\boldsymbol{X} = \boldsymbol{L}\boldsymbol{R}^T \in \mathbb{R}^{n \times n}$. Hence, $rank(\boldsymbol{X}) = r$ and $\boldsymbol{X}$ meant to be a low-rank matrix, so we have chosen r to be small. Secondly, we delete some of the $\boldsymbol{X}$ entries randomly, and make it sparse as much as possible. Finally, solve the optimization problem 2.4.2 and then calculate the Frobenius norm $\frac{\|\boldsymbol{X}_{predicted} - \boldsymbol{X}\|_F}{\|\boldsymbol{X}\|_F}$ to check the error between the predicted values and the original values. We have used the following pseudocode

Algorithm 1 Matrix completion pseudocode

Choose $r << n$
Generate $L, R \in \mathbb{R}^{n*r}$
Compute $X = LR^T \in R^{n*n}$
Sparse matrix $P_\Omega(\boldsymbol{X}) = Y$
Check incoherence
Solve: minimize $||Z||_*$
Subject to $P_\Omega(\boldsymbol{Z}) = \boldsymbol{Y}$
Calculate Frobenius norm

We were able to recover the matrices of our experiments with a small error. The biggest error that we had was 8×10^{-5}. The experiments have been done on matrices with dimension 200 at most because of the long CPU time and also because the solver failed to solve higher dimension.
We will state some of our results here. We have taken a matrix of size 100×100 with rank 50; we have used 6000 non-zero entries. The recovery was possible surprisingly with a low error 3.7×10^{-4}. We have also used the matrix of size 200×200 with rank 150 and with 30000 non-zero entries. In this case, the recovery was possible with error 1.5×10^{-4}.

5.2 Conclusion

We have studied the matrix completion problem. We have thoroughly studied the theoretical concepts that are needed to solve the problem. We have presented the incoherence condition and the bounds on the missing entries of the matrix and the related theorems. The convex formulation of the optimization problem has been presented. We have used a known software and tested the completion problem where we have taken the matrix with low rank and required missing numbers. Other assumptions such as incoherence stated in Theorem 4.0.20 were maintained. The results obtained clearly justified the theoretical results stated in the literature. This finding has encouraged us to performed further research on the topic. We would like to concentrate on designing efficient algorithm for the problem so that large scale problems can be solved. We will also theoretically study the problem using concepts like truncated nuclear norm in the recovery of the matrix.

Acknowledgements

First and formost I would like to thank almighty Allah for giving me the strength and protection to finish this project. Secondly, I would like to express my sincere gratitude to my supervisor Professor Monaz Ali for his guidance, continuous support and his patience, it is my priviledge to work with him. I would like to thank my tutor Emilia Magnani for her help, support and motivation.
My humble effort I dedicate to my sweet and loving ones, my family and friends whose affection, encouragement and prayers of day and night have made me succeed.
Last but not the least, I would like to acknowledge AIMS director Barry Green, the academic director Simukai Utete, IT manager Jan Groenewald, tutors, especially Alice, Reem and Samar. I am grateful to all of those whom I have had the pleasure to work with in this year.

References

Benjamin Recht, P. A. P., Maryam Fazel. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *Society for industrial and applied mathemaics*, 52:471–501, 2010.

Candes, E. J. and Plan, Y. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, 2010.

J.Candes, E. and Recht, B. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9:717–772, 2009.

J.Candes, E. and Romberg, J. Sparsity and incoherence in compressive sampling. *IOP science*, 23: 969–985, 2007.

J.Candes, E., Li, X., Ma, Y., and Wright, J. Robust principal component analysis. *Jornal of the ACM*, 58, 2011.

Kalman, D. A singularly valuable decomposition: The SVD of a matrix. *The College Mathematics Journal*, 27, 1996.

Meyer, C. D. *Matrix analysis and applied linear algebra*, volume 71. Siam, 2000.

Osnaga, S. M. *Low Rank Representations of Matrices Using Nuclear Norm Heuristics*. Phd, Colorado State University, 2014.

Wikipedia. Norm function. Wikipedia, the Free Encyclopedia, https://en.wikipedia.org/wiki/Norm_(mathematics), Accessed March 2020.

WolframAlpha. Bernoulli distribution. WolframAlpha, Computational Knowlage Ingine, https://mathworld.wolfram.com/BernoulliDistribution.html, Accessed March 2020.