

On the Skeleton Decomposition Applied to Matrix and  
Tensor Completion and Data Compression

Kenneth Allen

March 2020

## Abstract

The problem of matrix completion is a significant one. The matrix completion problem is, given a partially known matrix, fill in the missing entries as best as possible. In general we need further assumptions about our matrix in order to define what "as best as possible" means. In this thesis, we will assume that our complete matrix is part of a low rank matrix, leading to the low-rank matrix completion problem.

The problem of low rank matrix completion is, given a partially known matrix, find a completed matrix with low rank. This thesis proposes a novel method of solving the low-rank matrix completion problem which we call the Schur gradient descent method.

This method relies on finding a submatrix of large determinant in modulus called a dominant submatrix which is done with the maxvol algorithm. A greedy version of this algorithm is presented which improves on the original in computational time. Moreover, we present a new upper bound on the number of possible dominant submatrices in terms of the independence number of Johnson graphs. The Schur gradient descent method and the greedy maxvol algorithm are then combined to give us the maxvol Schur gradient descent method.

Similarly to matrix completion, the problem of tensor completion is, given a partially known tensor, find a completed tensor with low rank. There are multiple notions of the rank of a tensor. We give sufficient conditions for a partially known tensor to have a unique low multilinear rank completion.

© 2021

Kenneth Allen

All Rights Reserved

Approval page

# Acknowledgments

This thesis is dedicated to my father Bradford Allen who taught me to love math.

I would like to thank my mother Elaine Allen for all the support and encouragement.

I would like to thank my advisor Ming-Jun Lai without whom this would not have been possible.

I would like to thank Lin Mu for the collaboration and financial support.

I would like to thank David Green and Sebastian De Pascuale at Oak Ridge national lab for being excellent collaborators.

I would like to thank the Department of Energy for the SCGSR award.

I would like to thank James Calhoun for the significant amount of support during the process.

# Contents

<b>Acknowledgments</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Matrix Completion Motivation . . . . .	1
1.2 Mathematical Preliminaries . . . . .	4
<b>2 Low-Rank Matrix Completion Algorithms and Theory</b>	<b>8</b>
2.1 Alternating Projection . . . . .	8
2.2 Alternating Minimization . . . . .	9
2.3 Orthogonal Rank-One Matrix Pursuit . . . . .	10
2.4 Convex Relaxation . . . . .	12
2.5 Singular Value Thresholding . . . . .	14
2.6 Mask Permutations . . . . .	15
2.7 Finite Completability . . . . .	17
2.8 Algebraic Combinatorics of Low-Rank Matrix Completion . . . . .	24
2.9 Zero Sets of Systems of Matrix Minors . . . . .	31
2.10 Matrix Completion Topology . . . . .	33
<b>3 The Maximum Volume Principle and Maximum Volume Algorithms</b>	<b>38</b>
3.1 Schur Complement . . . . .	38
3.2 Then Skeleton Approximation . . . . .	39
3.3 Maximum Volume Algorithms . . . . .	42
3.4 Greedy Maximum Volume Algorithms . . . . .	44
3.5 Greedy Maxvol Numerical Experiments . . . . .	46
3.6 Maxvol Skeleton Approximation on Images . . . . .	48
3.7 Findvol Algorithm . . . . .	50

3.8	A Graph Theoretic Reformulation of Dominant submatrices . . . . .	53
3.9	Upper bounds on the number of dominant submatrices . . . . .	56
3.10	Sharp Inequality Examples . . . . .	59
3.11	Numerical Experiments Approximating the Expected Value of the Number of Dominant submatrices . . . . .	61
3.12	An Upper Bound on the Independence Number of Johnson Graphs . . . . .	64
<b>4</b>	<b>A Schur Complement Based Gradient Descent Method for Matrix Com- pletion</b>	<b>66</b>
4.1	Unique Matrix Completion Example . . . . .	66
4.2	Matrix Completion With a Known Invertible Submatrix . . . . .	66
4.3	General Matrix Completion With Schur Gradient Descent . . . . .	71
4.4	Small Numerical Examples of Schur Gradient Descent . . . . .	72
4.5	Maxvol Schur Gradient Descent . . . . .	73
4.6	Dominant submatrices of partially known matrices . . . . .	73
4.7	Maxvol on partially known matrixes . . . . .	75
4.8	Perturbation Analysis . . . . .	76
4.9	Maxvol Gradient Descent . . . . .	77
4.10	Small Examples with Noise . . . . .	78
4.11	Larger Examples With Noise . . . . .	79
4.12	Comparing Gradient Descent to Maxvol-Gradient Descent . . . . .	81
<b>5</b>	<b>Maximum Volume Based Skeletal Decompositions for Scalable Plasma Physics Applications</b>	<b>84</b>
5.1	Motivation . . . . .	84
5.2	Simulation Data Background Information . . . . .	84

5.3	Maximum Volume Skeleton Decomposition For Plasma Simulation Data Compression . . . . .	86
5.4	Dynamic Mode Decomposition Using the Skeleton Decomposition . . . . .	86
<b>6</b>	<b>Tensor Theory and Background</b>	<b>93</b>
6.1	Types of Tensor Ranks . . . . .	94
6.2	Spaces of Tensors with Rank at Most $r$ . . . . .	100
6.3	Tensor Rank Computation and Low-Rank Approximations . . . . .	102
<b>7</b>	<b>Tensor Completion</b>	<b>105</b>
7.1	Exact Low Multilinear Rank Tensor Completion . . . . .	106
7.2	Generalizing the Schur Gradient Descent Method to Tensors . . . . .	116
7.3	Algebraic Combinatorics of Low-Rank Tensor Completion . . . . .	118
<b>8</b>	<b>Notation and Glossary</b>	<b>121</b>



# 1 Introduction

## 1.1 Matrix Completion Motivation

In 2006, Netflix announced a competition with a grand prize of one million dollars. The problem was, given data on user movie ratings, create a recommendation system which will suggest films users would be likely to watch and enjoy. The grand prize would be given out to anyone who could improve Netflix's existing algorithm by ten percent. This is an example of a data completion problem. We have an incomplete set of data, and we would like to complete that set of data, using the known data, as best as possible.

It is often useful to encode our partially known data in a matrix. In the case of Netflix's problem, on one axis we index the users, and on the other axis we index the movies. In the corresponding entry between a user and a movie, we enter the user's rating of the movie. The result is an incomplete matrix, and our goal is to fill in the matrix as best as possible to predict how users will rate movies that they have not watched. This problem sparked an interest among mathematicians in studying the matrix completion problem.

A first approach to the matrix completion problem is to define a number of features of our data which we can use to calculate intermediate connections between users and movies. In this example, we can use genres such as *adventure* and *comedy* as our features. We assign strengths between users and features, and strengths between features and movies to decide whether or not we should recommend a movie.

To calculate whether or not a user would like a movie, we multiply the strengths between the user-feature score and feature-movie score, and sum over all features. In the example shown in fig. 1, the first user would be given a score of  $3 \cdot 2 + 1 \cdot 5 = 11$  as a prediction for how much they would enjoy the movie *B*. There are multiple ways we can get this data. For example, Netflix had at one point sent users surveys to determine which movie genres they enjoyed the most.

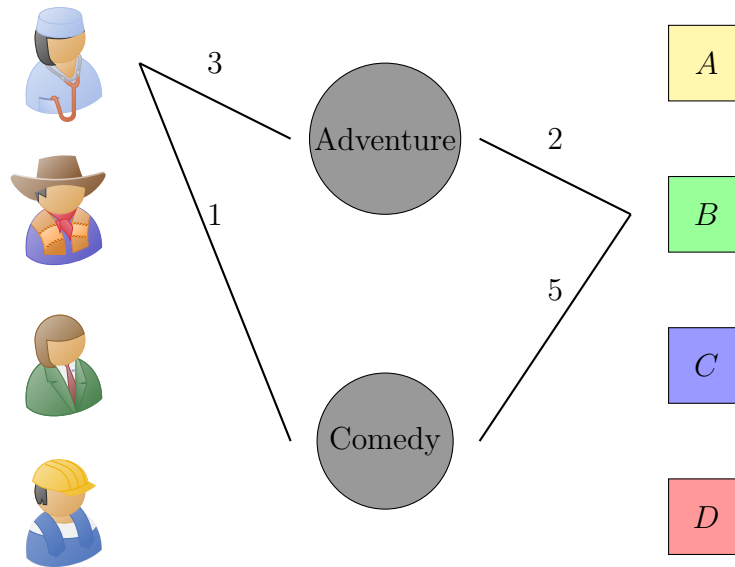


Figure 1: Example of strengths between a users, features, and movies which are represented with colored boxes.

We can represent user-feature and feature-movie data as two matrices as shown in fig. 2. To obtain the corresponding user-movie rating, we simply multiply these two matrices together.

The issues with this method are that it may be unrealistic to gather this data from all users or all movies. Moreover, these features are how we as humans describe movies, and may not be the most general way to represent our data.

Instead, we may reverse the problem. We gather scores on how much users enjoyed a particular movie through data such as ratings as in fig. 3. Then, given some incomplete data on user-movie enjoyment, we produce two factor matrices  $U$  and  $V$  such that the corresponding entries in the multiplication  $UV$  agrees with our known data. If we can find such matrices then we may predict any user-movie enjoyment score by looking at the corresponding entry in  $UV$ .

The key aspect to note here is that the resulting matrix will be low-rank, depending on the size of  $U$  and  $V$ . In particular, if we assume that there are  $r$  features which describe our data, then the resulting matrix will have rank at most  $r$ . Therefore, we don't necessarily

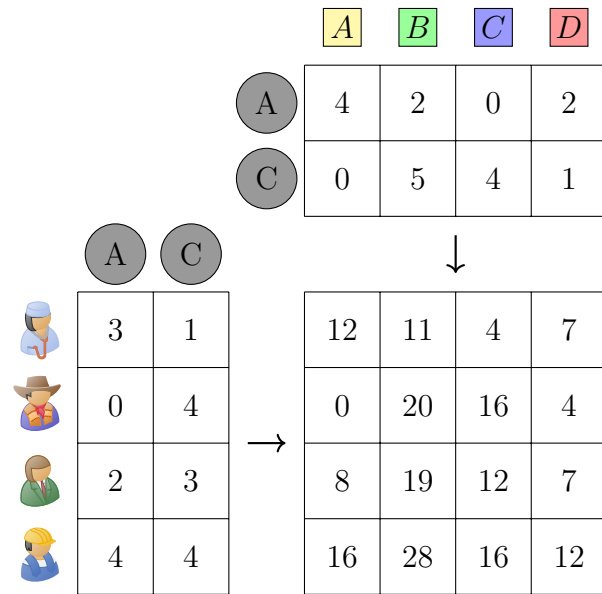


Figure 2: User-Movie rating prediction matrix obtained through matrix multiplication.

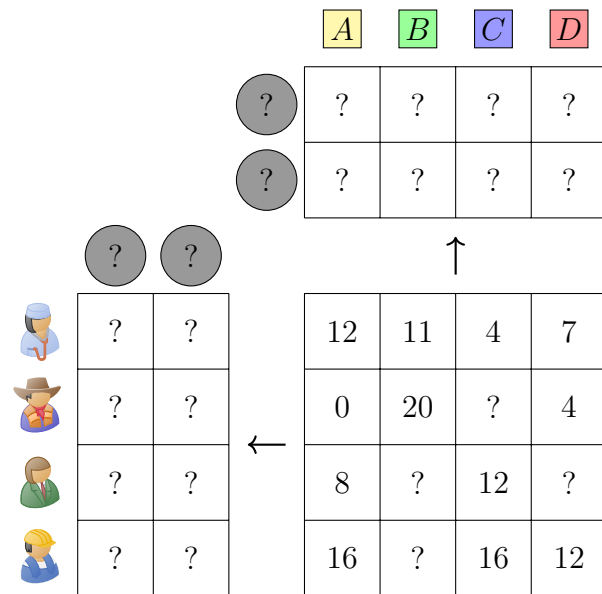


Figure 3: User-Movie prediction matrix obtained through latent factor matrices.

need to find explicit factor matrices  $U$  and  $V$ , all we need to do is to find a rank  $r$  matrix  $M$  with entries equal to the given known entries.

## 1.2 Mathematical Preliminaries

We will now express the matrix completion problem in more formal mathematical terms. Let  $M_{n \times m}$  denote the space of matrices with  $n$  rows and  $m$  columns over the real numbers  $\mathbb{R}$  or the complex numbers  $\mathbb{C}$ . Given a set of observed elements  $\{M_{ij}\}$ , where  $M_{ij}$  is in index  $(i, j)$  of a partially known  $n \times m$  matrix, we let  $\Omega \subset [n] \times [m]$  denote the index set of the known elements  $M_{ij}$ , where  $[n] = \{1, \dots, n\}$ . In other words, if  $M_{ij}$  is a known element, then  $(i, j) \in \Omega$ , and the total number of known elements is  $|\Omega|$ . We may also represent  $\Omega$  as a binary  $n \times m$  matrix where the entry in index  $(i, j)$  is 1 if  $M_{ij}$  is known, and 0 otherwise.

It is desirable to use the fewest number of features possible to represent our data, which corresponds to minimizing the rank of our completed matrix. We may now express the matrix completion problem as finding a solution to the non-convex minimization problem

$$\begin{aligned} \min_{X \in M_{n \times m}} \text{rank}(X) \\ \text{s.t. } X_{ij} = M_{ij} \quad \forall (i, j) \in \Omega \end{aligned}$$

Let  $M_\Omega \in M_{n \times m}$  denote the partially known  $n \times m$  matrix with known entries  $M_{ij}$  in entry  $(i, j) \in \Omega$ , and zeros in unknown entries. We will often denote unknown elements of a matrix with an empty square  $\square$  instead of zero. There may be finitely many, infinitely many, or zero ways to complete  $M_\Omega$  into a rank  $r$  matrix. Consider the following incomplete matrices.

**Example 1.1.** *Let*

$$M_{\Omega} = \begin{bmatrix} \square & 1 & 2 & 3 \\ 1 & \square & 3 & 4 \\ 1 & 3 & \square & 5 \\ 1 & 4 & 5 & \square \end{bmatrix}.$$

*Then  $M_{\Omega}$  has the unique rank 2 completion*

$$M = \begin{bmatrix} 1 & 1 & 2 & 4 \\ 1 & 2 & 3 & 5 \\ 1 & 3 & 4 & 6 \\ 1 & 4 & 5 & 7 \end{bmatrix}.$$

If we change the known entries in the last column of  $M_{\Omega}$  in example 1.1, we have the following example.

**Example 1.2.** *Let*

$$M_{\Omega} = \begin{bmatrix} \square & 1 & 2 & 4 \\ 1 & \square & 3 & 5 \\ 1 & 3 & \square & 6 \\ 1 & 4 & 5 & \square \end{bmatrix}.$$

*Then  $M_{\Omega}$  has exactly two rank 2 completions, which are*

$$M_1 = \begin{bmatrix} 1 & 1 & 2 & 4 \\ 1 & 2 & 3 & 5 \\ 1 & 3 & 4 & 6 \\ 1 & 4 & 5 & 7 \end{bmatrix} \quad M_2 = \begin{bmatrix} -2/3 & 1 & 2 & 4 \\ 1 & 21/8 & 3 & 5 \\ 1 & 3 & 39/11 & 6 \\ 1 & 4 & 5 & 26/3 \end{bmatrix}.$$

To verify that  $M_1$  and  $M_2$  are the only rank 2 completions of  $M_{\Omega}$ , note that a matrix has rank at most  $r$  if and only if all  $(r + 1) \times (r + 1)$  minors vanish. Consider the system of

equations in four variables consisting of all  $3 \times 3$  minors of the matrix

$$M(x, y, z, w) = \begin{bmatrix} x & 1 & 2 & 4 \\ 1 & y & 3 & 5 \\ 1 & 3 & z & 6 \\ 1 & 4 & 5 & w \end{bmatrix}.$$

This gives a system of 16 degree 2 and degree 3 polynomials. One can verify using a computer algebra system that  $M_1$  and  $M_2$  are the only two solutions to this system of equations.

We will now introduce the space of fixed rank  $r$  matrices. Let

$$\mathcal{M}_r = \{M \in M_{n \times m} \mid \text{rank}(M) = r\}$$

denote the space of  $n \times m$  rank  $r$  matrices over  $\mathbb{R}$  or  $\mathbb{C}$ , where  $r \leq \min(n, m)$ . Then  $\mathcal{M}_r$  is a  $(n + m)r - r^2$  dimensional manifold [4]. Note that  $\mathcal{M}_r$  is not a closed set. In particular, we may approximate any low-rank matrix as the limit of a sequence of high-rank matrices, but we may not approximate high-rank matrices as the limit of a sequence of low-rank matrices. So the closure of  $\mathcal{M}_r$  is the space of matrices with rank at most  $r$ , which we will denote

$$\overline{\mathcal{M}}_r = \{X \in M_{n \times m} \mid \text{rank}(X) \leq r\}.$$

The dimension of  $\overline{\mathcal{M}}_r$  is also equal to  $(n + m)r - r^2$ , since the closure operation does not change the dimension of a manifold.

We will now recall some notions from algebraic geometry. A closed set  $V \subset \mathbb{C}^k$  is called an *algebraic variety* if it is the zero set of a set of a system of polynomial equations. The *Zariski closure* of a set  $S \subset \mathbb{C}^k$  is the smallest algebraic variety  $V$  which contains  $S$ . The Zariski closure of  $S$  may be expressed as the intersection of all algebraic varieties which contain  $S$ . Since a matrix  $M$  has rank at most  $r$  if and only if all  $(r + 1) \times (r + 1)$  minors of  $M$  vanish,

$\overline{\mathcal{M}}_r$  is equal to the zero set of all  $(r+1) \times (r+1)$  minors of an  $n \times m$  matrix. Therefore  $\overline{\mathcal{M}}_r$  is an algebraic variety, and it is the Zariski closure of  $\mathcal{M}_r$  since it is the closure of  $\mathcal{M}_r$  and an algebraic variety.  $\overline{\mathcal{M}}_r$  is also sometimes referred to as the determinantal variety.

An algebraic variety  $V$  is called irreducible if it cannot be expressed as the union of two proper sub-varieties. That is, a variety  $V$  is irreducible if it cannot be written as  $V = V_1 \cup V_2$  for proper subvarieties  $V_1 \subset V$  and  $V_2 \subset V$ . Then  $\overline{\mathcal{M}}_r$  is an irreducible variety. It can also be shown that the set of singular points of  $\overline{\mathcal{M}}_r$  is the set of matrices with rank at most  $r-1$ , that is, it is the set  $\overline{\mathcal{M}}_{r-1} \subset \overline{\mathcal{M}}_r$ . To verify this note that the partial derivatives of the all  $(r+1) \times (r+1)$  minors are equal to zero exactly on the set  $\overline{\mathcal{M}}_{r-1}$ .

Define the map  $P_\Omega : M_{n \times m} \rightarrow M_{n \times m}$  such that  $P_\Omega(X)$  fixes entry  $X_{ij}$  if  $(i, j) \in \Omega$ , and sets  $X_{ij}$  equal to zero if  $(i, j) \notin \Omega$ . Here  $P_\Omega$  is the orthogonal projection operator onto the subspace of matrices with entries equal to 0 outside of  $\Omega$ .

**Example 1.3.** Let  $\Omega = \{(1, 1), (2, 2), (2, 3), (3, 2)\}$ . Then  $|\Omega| = 4$ , and

$$P_\Omega \left( \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 6 \\ 0 & 8 & 0 \end{bmatrix}.$$

Given a partially known matrix  $M_\Omega$ , if  $M$  is a completion of  $M_\Omega$ , then  $P_\Omega(M) = M_\Omega$ . Let  $\mathcal{A}_\Omega = P_\Omega^{-1}(M_\Omega)$  be the linear variety of any possible completions of  $M_\Omega$ . In other words, a matrix  $X$  is in  $\mathcal{A}_\Omega$  if  $P_\Omega(X) = M_\Omega$ .  $\mathcal{A}_\Omega$  is irreducible since it is a linear variety.

Note that since  $\overline{\mathcal{M}}_r$  is the space of matrices with rank at most  $r$ , and  $\mathcal{A}_\Omega$  is the space of any possible completion of  $M_\Omega$ , Then finding a rank at most  $r$  completion  $M$  of  $M_\Omega$  is equivalent to finding a point  $M \in \mathcal{A}_\Omega \cap \overline{\mathcal{M}}_r$ .

## 2 Low-Rank Matrix Completion Algorithms and Theory

Given a partially known matrix  $M_\Omega$  along with an index set of known entries  $\Omega$ , the goal of low-rank matrix completion is to find a completed matrix  $M$  with entries equal to the known entries in  $M_\Omega$  such that  $\text{rank}(M)$  is minimized. There are many existing algorithms for finding a low-rank matrix completion. In this section, we will present a few of these algorithms.



### 2.1 Alternating Projection

The notes in this section are taken on [16]. The method of alternating projection has proven to be an effective way of finding intersection points between two manifolds. As the name suggests, starting with an initial guess, we alternate between projecting onto each manifold obtaining successive approximations of a point in the intersection. While most extensively studied in application to find intersections of convex sets, alternating projection methods have also been applied to find intersections non-convex sets. Some issues that may occur are that a projection onto a non-convex set may not be single valued. Moreover, the projection may be difficult to compute.

While  $\mathcal{A}_\Omega$  is convex, as it is a linear variety,  $\mathcal{M}_r$  is not convex for  $r > 0$ . However, given  $X \in M_{n \times m}$ , if  $\text{rank}(X) > r$ , then a closest rank  $r$  projection  $P_{\mathcal{M}_r}(X)$  may be easily calculated using the singular value decomposition (SVD). To calculate  $P_{\mathcal{M}_r}(X)$ , take the SVD of  $X$  obtaining an  $n \times n$  orthogonal matrix of left-singular vectors  $U$ , an  $m \times m$  orthogonal matrix of right-singular vectors  $V$ , and an  $n \times m$  diagonal matrix  $\Sigma$  such that  $X = U\Sigma V^*$ . The diagonal entries of  $\Sigma$  are non-negative real numbers called the singular values of  $X$  denoted  $\sigma_i$  for  $1 \leq i \leq \min(n, m)$ , and are ordered such that  $\sigma_{i+1} > \sigma_i$  for all  $i$ . Given  $1 \leq r \leq \min(n, m)$ , let  $\Sigma_r$  be the  $r \times r$  diagonal matrix with diagonal entries equal



to the  $r$  largest singular values of  $X$ , let  $U_r$  be the  $n \times r$  matrix with columns consisting of the first  $r$  left-singular vectors of  $X$ , and let  $V_r$  be the  $m \times r$  matrix with columns consisting of the first  $r$  right-singular vectors of  $X$ . Then if  $\sigma_i > 0$  for  $1 \leq i \leq r$ , a closest rank  $r$  projection of  $X$  may be calculated as  $P_{\mathcal{M}_r}(X) = U\Sigma_r V^*$ . Moreover, this projection is unique if  $\sigma_r \neq \sigma_{r+1}$ . The projection of  $X$  onto  $\mathcal{A}_\Omega$ ,  $P_{\mathcal{A}_\Omega}(X)$ , is simply calculated by setting all of the entries of  $X$  in the indices of  $\Omega$  to the corresponding known entries of  $M_\Omega$ .

Let us suppose  $M \in \mathcal{A}_\Omega \cap \overline{\mathcal{M}_r}$ . Then given  $X_0$  an initial guess of  $M$  and a tolerance  $\epsilon$ , the alternating projection algorithm runs as follows.

---

**Algorithm 1:** Alternating Projection [15]

---

**Input:** incomplete matrix  $M_\Omega$ , initial guess  $X_0$ , stopping criterion


**Result:**  $X_k$  an approximation of  $M$  such that  $\text{rank}(M) = r$  and  $P_\Omega(M) = M_\Omega$

**for**  $k = 1, \dots$  **do**

$Y_k = P_{\mathcal{M}_r}(X_{k-1});$   
 $X_k = P_{\mathcal{A}_\Omega}(Y_k);$

---

We require some stopping criterion as an input. For example, we could fix a tolerance  $\epsilon$ , and loop until  $\|X_k - X_{k-1}\| < \epsilon$ . Alternatively, if we only want to run the algorithm for a certain number of steps, we could fix the number of steps  $N$  and loop for  $k = 1, \dots, N$ .

Let  $T_{\mathcal{A}_\Omega}(M)$  denote the tangent space of  $\mathcal{A}_\Omega$  at point  $M$ , and let  $T_{\mathcal{M}_r}(M)$  denote the tangent space of  $\mathcal{M}_r$  at point  $M$ . Then if  $M \in \overline{\mathcal{M}_r} \cap \mathcal{A}_\Omega$ , and  $T_{\mathcal{A}_\Omega}(M) \cap T_{\mathcal{M}_r}(M) = \{0\}$ , then algorithm 1 converges to  $M$  linearly. 

## 2.2 Alternating Minimization

The notes in this section are taken on [18]. The alternating minimization method is an empirically successful method for finding a low-rank completion of  $M_\Omega$ . Moreover, it formed a critical component in the winning entry of the Netflix problem. The objective of alternating minimization is to find a completed matrix in bilinear form  $M = LR^\top$  with  $L$  being  $n \times r$  and  $R$  being  $m \times r$  such that the entries of  $M$  correspond with the known entries of  $M_\Omega$ . Such

an  $M$  is found by alternating between optimizing  $L$  and  $R$ . In particular, the non-convex problem to solve is

$$\min_{L,R} \frac{1}{2} \|P_{\Omega}(LR^{\top}) - M_{\Omega}\|^2.$$

The alternating minimization algorithm runs as follows.

---

**Algorithm 2:** Alternating Minimization [18]

---

**Input:** incomplete matrix  $M_{\Omega}$ , initial guess  $R_0 \in M_{m \times r}$ , stopping criterion

**Result:**  $X_k = L_k R_k^{\top}$  an approximation of  $M$  such that  $\text{rank}(M) = r$  and  $P_{\Omega}(M) = M_{\Omega}$

**for**  $k = 1, \dots$  **do**

$$\left[ \begin{array}{l} L_k = \arg \min_{L \in M_{n \times r}} \|P_{\Omega}(L R_{k-1}^{\top}) - M_{\Omega}\|^2; \\ R_k = \arg \min_{R \in M_{m \times r}} \|P_{\Omega}(L_k R^{\top}) - M_{\Omega}\|^2; \end{array} \right.$$


---

We may solve the minimization at each step with the method of least squares.

## 2.3 Orthogonal Rank-One Matrix Pursuit

The notes in this section are on [17]. Recall that we may express a matrix  $X$  as a weighted sum of rank 1 matrices  $M_i$  such that  $\|M_i\| = 1$  for all  $i$ . That is, we may write  $M$  as

$$X = M(\theta) = \sum_{i=1}^r \theta_i M_i.$$

Here  $\theta$  is the vector of weights in the sum. One way to calculate  $M_i$  and  $\theta_i$  is with the singular value decomposition of  $X$ , by setting  $\theta$  equal to the vector of singular values, and  $M_i = u_i v_i^{\top}$  where  $u_i$  and  $v_i$  are the  $i$ th left and right-singular vectors of  $X$  respectively.

Note that the minimum value of  $\|\theta\|_0$  over all possible choices of  $\theta$  is equal to the rank of  $M$ , where  $\|\theta\|_0$  is equal to the number of non-zero elements in the vector  $\theta$ . Therefore, we

may formulate the matrix completion problem as finding a solution to the minimization

$$\begin{aligned} \min_{\theta} & \|P_{\Omega}(M(\theta)) - M_{\Omega}\| \\ \text{s.t.} & \|\theta\|_0 \leq r. \end{aligned}$$

Our goal is to choose proper basis matrices  $M_i$ , and proper weights  $\theta_i$ . We do so by alternating between computing the rank 1 basis matrices  $M_i$  and the weights  $\theta$  accordingly. In particular, on the  $(k-1)$ th step, suppose we have computed  $M_1, \dots, M_{k-1}$  and weights  $\theta^{k-1}$ . To compute  $M_k$ , we first compute the regression residual

$$R_k = M_{\Omega} - \sum_{i=1}^{k-1} \theta_i M_i$$

Here we assume that  $M_{\Omega}$  is a partially known matrix with zeros in unknown indices.

Since it is desired that  $M_k$  is rank one with unit Frobenius norm, we may search for  $M_k$  in the form  $M_k = uv^{\top}$  for some unit vectors  $u$  and  $v$ . We then calculate  $u$  and  $v$  as the solution to

$$\max_{u,v} \{u^{\top} R_k v \mid \|u\| = \|v\| = 1\}.$$

Here the unknown entries of  $R_k$  are replaced with 0. This minimization has optimal solution equal to the first left and right singular vectors of  $R_k$ . After we have computed  $M_k = uv^{\top}$ , we then calculate  $\theta^k$  as the solution to the minimization problem

$$\min_{\theta} \left\| \sum_{i=1}^k \theta_i P_{\Omega}(M_i) - M_{\Omega} \right\|$$

which can be computed with least squares. In particular, let  $m_{\Omega} = \text{vec}(M_{\Omega})$ , and  $m_i = \text{vec}(M_i)$  be the vectorization of  $M_{\Omega}$  and  $P_{\Omega}(M_i)$  respectively. Let  $W_k = [m_1 \cdots m_k]$  be the

vectors  $m_i$  assembled into a matrix for  $i = 1, \dots, k$ . Then

$$\theta^k = (W_k^\top W_k)^{-1} W_k^\top m_\Omega$$

is the solution to the minimization problem.

In summary, the algorithm runs as follows.

---

**Algorithm 3:** Orthogonal Rank-One Matrix Pursuit [17]

---

**Input** : incomplete matrix  $M_\Omega$ , initial guess  $X_0$ , stopping criterion


**Initialize:**  $m_\Omega = \text{vec}(M_\Omega)$

**Result:**  $X_k$  an approximation of  $M$  such that  $\text{rank}(M) = r$  and  $P_\Omega(M) = M_\Omega$

**for**  $k = 1, \dots$  **do**

	$R_k = M_\Omega - X_{k-1};$
	Find the top left- and right-singular vectors $u_k$ and $v_k$ of $R_k$ ;
	$M_k = u_k v_k^\top$ , $m_k = \text{vec}(M_k)$ , and $W_k = [m_1 \cdots m_k]$ ;
	$\theta^k = (W_k^\top W_k)^{-1} W_k^\top m_\Omega;$
	$X_k = \sum_{i=1}^k \theta_i^k M_i;$

---

The orthogonal rank-one matrix pursuit algorithm converges at a linear rate. 

## 2.4 Convex Relaxation

In general, the problem of finding a minimum rank completion of  $M_\Omega$  is difficult because the rank function is non-convex. Moreover, the space of matrices with rank at most  $r$ ,  $\overline{\mathcal{M}}_r$ , is low-dimensional, while real life data often has random noise. If we assume that the random noise is sampled from a continuous distribution, then the probability that the data with noise will belong to  $\overline{\mathcal{M}}_r$  is zero.

Instead of solving  $\min_X \text{rank}(X)$  such that  $X_{ij} = M_{ij}$  for  $(i, j) \in \Omega$ , we may opt to solve a convex relaxation of the problem by replacing the rank function with the nuclear norm. In other words, we opt to solve the convex minimization  $\min_X \|X\|_*$  such that  $X_{ij} = M_{ij}$  for  $(i, j) \in \Omega$ . The *nuclear norm*  $\|\cdot\|_*$  is defined as the sum of the singular values of the matrix.

That is, if  $\sigma(X)$  is the vector of singular values of  $X$  and  $\sigma_i(X)$  are the singular values, we have

$$\|X\|_* = \sum_i \sigma_i(X).$$

This convex relaxation is analogous to the  $l^1$  convex relaxation of the  $l^0$  norm. In fact they are directly related, as we have

$$\begin{aligned} \text{rank}(X) &= \|\sigma(X)\|_0 \\ \|X\|_* &= \|\sigma(X)\|_1. \end{aligned}$$

To understand why we choose this relaxation, we introduce the definition of the convex envelope of a function.

**Definition 1.** *Given a convex domain  $C$ , the convex envelope of a function  $f : C \rightarrow \mathbb{R}$  is the largest convex function  $g$  such that  $g(x) \leq f(x)$ .*

The reason why the  $l^1$  norm is used as a convex relaxation of the  $l_0$  norm is because it is the convex envelope of the  $l_0$  norm on the unit ball. Similarly, the nuclear norm is the convex envelope of the rank function on the unit ball with respect to the spectral norm. That is, on the domain  $\{X \mid \sigma_1(X) \leq 1\}$ .

**Theorem 1.** *On the unit ball  $B = \{X \mid \sigma_1(X) \leq 1\}$ , the convex envelope of the rank function is the nuclear norm function  $\|\cdot\|_*$ .*

*Proof.* First, recall that any norm is a convex function, so the nuclear norm is convex. For  $X \in B$ , we have  $\sigma_1(X) \leq 1$ . Since  $\sigma_1(X)$  is the largest singular value of  $X$ , we have  $\sigma_i(X) \leq 1$  for all  $i$ . Let  $r = \text{rank}(X)$ . Then  $r$  is the number of non-zero singular values of  $X$ . Therefore, we have

$$\|X\|_* = \sum_{i=1}^r \sigma_i(X) \leq \sum_{i=1}^r 1 = r,$$

so  $\|X\|_* \leq \text{rank}(X)$  on  $B$ . Moreover, it is shown in [23] that the nuclear norm is the tightest convex lower bound.  $\square$

The matrix completion problem may then be approximated as the nuclear norm minimization problem

$$\begin{aligned} \min_X \|X\|_* & \tag{1} \\ \text{s.t. } P_\Omega(X) &= M_\Omega. \end{aligned}$$

## 2.5 Singular Value Thresholding

The notes in this section are on [19]. Singular value thresholding is an algorithm for approximately solving the previous convex minimization of the nuclear norm. In particular, instead of minimizing the nuclear norm, we solve the minimization

$$\begin{aligned} \min_X \tau \|X\|_* + \frac{1}{2} \|X\|_F^2 & \tag{2} \\ \text{s.t. } P_\Omega(X) &= M_\Omega \end{aligned}$$

This minimization is easier to solve than minimizing the nuclear norm. Moreover, for large values of  $\tau$ , the term  $\tau \|X\|_*$  dominates the term  $\frac{1}{2} \|X\|_F^2$ , and so the solution is approximately equal to the solution to minimizing the nuclear norm with the same constraints.

We start with defining the *singular value shrinkage operator*  $D_\tau$ . Let  $X = U\Sigma V^*$  be the singular value decomposition of  $X$ . Let  $\Sigma_\tau$  be the diagonal matrix with  $i$ th diagonal entry equal to  $\max(\sigma_i - \tau, 0)$ . Then for  $\tau \geq 0$ , the singular value shrinkage operator  $D_\tau$  is defined as  $D_\tau(X) = U\Sigma_\tau V^*$ .

**Theorem 2.** For  $\tau \geq 0$ , the singular value shrinkage operator satisfies the minimization

$$D_\tau(Y) = \arg \min_X \left\{ \frac{1}{2} \|X - Y\|_F^2 + \tau \|X\|_* \right\}$$

Now in terms of the shrinkage operator, we define the singular value thresholding algorithm.

---

**Algorithm 4:** Singular Value Thresholding [19]

---

**Input:** incomplete matrix  $M_\Omega$ , sequence of step sizes  $\{\delta_k\}_{k \geq 1}$ ,  $\tau \geq 0$ , initial guess  $Y_0$ , stopping criterion

**Result:**  $X_k$  an approximation of  $M$  such that  $\text{rank}(M) = r$  and  $P_\Omega(M) = M_\Omega$

**for**  $k = 1, \dots$  **do**

$X_k = D_\tau(Y_{k-1});$   
 $Y_k = Y_{k-1} + \delta_k(M_\Omega - P_\Omega(X_k));$

---

It can be shown that the sequence  $X_k$  converges to the solution to eq. (2), and for large values of  $\tau$ ,  $X_k$  approximates the solution to eq. (1). Moreover, the matrices in the sequence  $\{X_k\}$  empirically have low rank.

## 2.6 Mask Permutations

Given an  $n \times m$  partially known matrix  $M_\Omega$ , it may be useful to permute the rows and columns or take the transpose to simplify the structure of the known and unknown entries. This does not change the number of rank  $r$  completions of  $M_\Omega$  because row permutations, column permutations, and transposes are bijections under which the rank is invariant. More specifically, if  $Q$  is a composition of transposes, row permutations, and column permutations, and  $M$  is a rank  $r$  completion of  $M_\Omega$ , then  $Q(M)$  is a rank  $r$  completion of the partially known matrix  $M_{Q(\Omega)} = Q(M_\Omega)$ .

Two masks  $\Omega_1$  and  $\Omega_2$ , interpreted as binary matrices, are considered equivalent if we may obtain  $\Omega_1$  from  $\Omega_2$  by permuting rows and columns or transposing. How can we tell if two masks are equivalent? An initial attempt may be to count the number of ones in the

rows and columns of  $\Omega_1$  and  $\Omega_2$ . If  $\Omega_1$  and  $\Omega_2$  are equivalent, the number of ones in the rows and columns of  $\Omega_1$  and  $\Omega_2$  must be equal up to order. This is a necessary, but not a sufficient, condition for two masks to be equivalent. For example, consider the masks

$$\Omega_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \qquad \Omega_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

The number of ones in the rows and columns are both  $(1, 2, 2)$  for both  $\Omega_1$  and  $\Omega_2$ , so they may or may not be equivalent. However, note is that that  $\Omega_1$  has a non-trivial stabilizer under the action of  $S_3 \times S_3$  by permutation of rows and columns since swapping rows or columns 2 and 3 does not change  $\Omega_1$ . On the other hand,  $\Omega_2$  is not fixed by swapping any two rows or columns, so it has a trivial stabilizer. Therefore,  $\Omega_1$  and  $\Omega_2$  are not equivalent.



In general, there are many masks up to permutation of rows and columns.

**Theorem 3.** *The number of  $n \times m$  masks up to permutation of rows and columns is at least  $\frac{2^{nm}}{n!m!}$ .*

*Proof.* Let  $B_{n \times m}$  be the set of  $n \times m$  masks, then  $|B_{n \times m}| = 2^{nm}$ . Since row and column permutations commute, there is a group action of  $G = S_n \times S_m$  on  $B_{n \times m}$ , where  $S_n$  is the symmetric group over  $n$  symbols. Then  $|B_{n \times m}/G|$  is the number of masks up to permutation of rows and columns.

By the orbit-counting theorem from [20], we have

$$|B_{n \times m}/G| = \frac{1}{|G|} \sum_{\Omega \in B_{n \times m}} |\text{Stab}_G(\Omega)|,$$



where  $\text{Stab}_G(\Omega)$  is the stabilizer of  $\Omega$ . Since each matrix is fixed by the identity permutation, we have  $|\text{Stab}_G(\Omega)| \geq 1$ , and so

$$|B_{n \times m}/G| \geq \frac{1}{|G|} \sum_{\Omega \in B_{n \times m}} 1 = \frac{|B_{n \times m}|}{|G|} = \frac{2^{nm}}{n!m!}.$$

□



## 2.7 Finite Completability

We call a partially known matrix  $M_\Omega$  finitely completable in rank  $r$  if there are finitely many rank  $r$  completions of  $M_\Omega$ . In this section, we discuss some necessary conditions for an incomplete matrix  $M_\Omega$  to have a unique rank  $r$  completion.

**Theorem 4.** *Having at least  $r$  known entries per row and  $r$  known entries per column is a necessary condition for an  $n \times m$  partially known matrix  $M_\Omega$  to be finitely completable in rank  $r$ .*



*Proof.* Let  $\Omega$  be a set of known indices such that there is a row or column with fewer than  $r$  known entries. By transpose and permutation, suppose the last column of  $M_\Omega$  has  $k$  known entries which is strictly fewer than  $r$ . Without loss of generality, assume the first  $m - 1$  columns are entirely known. Again by permuting the rows, let  $M_\Omega = \begin{bmatrix} A & C \\ B & \square \end{bmatrix}$ , where  $A$  and  $B$  are completely known,  $C$  is a  $k \times 1$  block of known entries with  $k < r$ , and  $\square$  is an  $(n - k) \times 1$  block of unknown entries. It suffices to show that  $M_\Omega$  is not finitely completable in rank  $r$ .

First note that if  $\text{rank}(\begin{bmatrix} A \\ B \end{bmatrix}) > r$ , then any completion of  $M_\Omega$  will have rank greater than  $r$ , so there will be no completions of  $M_\Omega$  in  $\overline{\mathcal{M}}_r$ . If  $\text{rank}(\begin{bmatrix} A \\ B \end{bmatrix}) < r$ , then any completion of  $M_\Omega$  will have rank less than or equal to  $r$ , so  $M_\Omega$  would have infinitely many rank  $r$  completions.

We are left with the case  $\text{rank}(\begin{bmatrix} A \\ B \end{bmatrix}) = r$ . Let  $s = \text{rank}(A)$ . Note that  $s \leq k$  since  $A$  is a  $k \times (m - 1)$  matrix. Suppose  $C$  is not in the column space of  $A$ . In other words, suppose  $\text{rank}(\begin{bmatrix} A & C \end{bmatrix}) = s + 1$ . Then any completion  $M \in \mathcal{A}_\Omega$  will have rank  $r + 1$ , and so there will

be no completion of  $M_\Omega$  in  $\overline{\mathcal{M}}_r$ . So  $C$  must be in the column space of  $A$ . In other words,  $\text{rank}([A \ C]) = s$ .

Since  $\text{rank}(\begin{bmatrix} A \\ B \end{bmatrix}) = r$ , there exists an  $r \times (m-1)$  rank  $r$  submatrix  $\begin{bmatrix} A' \\ B' \end{bmatrix}$ . We may choose  $A'$  such that it consists of  $s$  linearly independent rows of  $A$ . Then, since the remaining rows of  $A$  are in the row space of  $A'$ , we must choose the remaining  $r-s$  linearly independent rows  $B'$  from  $B$ .

Augmenting  $\begin{bmatrix} A' \\ B' \end{bmatrix}$  with the corresponding rows from  $\begin{bmatrix} C \\ \square \end{bmatrix}$ , we get an  $r \times m$  submatrix of  $M_\Omega$  of the form  $M'_\Omega = \begin{bmatrix} A' & C' \\ B' & \square \end{bmatrix}$ , where  $\square$  is an  $(r-s) \times 1$  block of unknown entries. Any completion  $M' = \begin{bmatrix} A' & C' \\ B' & D' \end{bmatrix}$  of  $M'_\Omega$  will be rank  $r$  because the submatrix  $\begin{bmatrix} A' \\ B' \end{bmatrix}$  is rank  $r$ . Moreover, there are  $r-s$  degrees of freedom, which is greater than zero since  $r > k \geq s$ .

Let  $B = \begin{bmatrix} B' \\ B'' \end{bmatrix}$ , where  $B''$  is the  $(n-k-r+s) \times (m-1)$  submatrix of  $B$  consisting of the rows in  $B$  that are not in  $B'$ . Similarly, let  $D = \begin{bmatrix} D' \\ \square \end{bmatrix}$ , where  $\square$  are the remaining  $(n-k-r+s)$  unknown entries of  $M_\Omega$ . Because the rows of  $\begin{bmatrix} A' \\ B' \end{bmatrix}$  form a basis for the row space of  $\begin{bmatrix} A \\ B \end{bmatrix}$ , there exists a unique  $r \times (n-k-r+s)$  matrix  $X$  such that  $\begin{bmatrix} A' \\ B' \end{bmatrix}^\top X = \begin{bmatrix} B'' \end{bmatrix}^\top$ . In particular,  $X = \begin{bmatrix} A'(A')^\top & A'(B')^\top \\ B'(A')^\top & B'(B')^\top \end{bmatrix}^{-1} \begin{bmatrix} A'(B'')^\top \\ B'(B'')^\top \end{bmatrix}$ . So we must have  $\begin{bmatrix} A' & C' \\ B' & D' \end{bmatrix}^\top X = \begin{bmatrix} B'' & \square \end{bmatrix}^\top$ , which implies the remaining unknown entries are equal to  $\begin{bmatrix} C' \\ D' \end{bmatrix}^\top X$ . So there exists a unique  $D$  such that  $\text{rank}(\begin{bmatrix} A' & C' \\ B' & D' \end{bmatrix}) = r$ .

Moreover, since the rank of a matrix is at least as great as the rank of any submatrix, we have

$$s = \text{rank}(A') \leq \text{rank}([A' \ C']) \leq \text{rank}([A \ C]) = s.$$

So  $\text{rank}([A' \ C']) = s$ , which implies that the rows of  $[A' \ C']$  span the row space of  $[A \ C]$ . Therefore  $\text{rank}(\begin{bmatrix} A & C \\ B & D \end{bmatrix}) = \text{rank}(\begin{bmatrix} A' & C' \\ B' & D' \end{bmatrix}) = r$ .

Thus any completion  $M' = \begin{bmatrix} A' & C' \\ B' & D' \end{bmatrix}$  of  $M'_\Omega$  extends to a unique rank  $r$  completion  $M = \begin{bmatrix} A & C \\ B & D \end{bmatrix}$  of  $M_\Omega$ . So  $\dim(\mathcal{A}_\Omega \cap \overline{\mathcal{M}}_r) = r-s > 0$ , and so  $M_\Omega$  has infinitely many rank  $r$  completions.

This exhausts all possible cases of  $M_\Omega$ , so  $M_\Omega$  is not finitely completable in  $r$ .  $\square$

We also observe that a sufficient total number of observed entries are required for  $M_\Omega$  to be finitely completable in  $r$ .

**Theorem 5.** *For  $n \times m$  matrices,  $\Omega$  must contain at least  $(n + m)r - r^2$  known entries for  $M_\Omega$  to be finitely completable.*

*Proof.* From [2], if  $U$  and  $V$  are irreducible affine varieties in  $d$  dimensional space, then we have the inequality  $\dim(U) + \dim(V) \leq \dim(U \cap V) + d$ .

Note that  $\dim(\overline{\mathcal{M}}_r) = (n + m)r - r^2$ ,  $\dim(\mathcal{A}_\Omega) = nm - |\Omega|$ , and  $\dim(M_{n \times m}) = nm$ . Suppose  $M_\Omega$  is finitely completable in  $r$ , then  $\dim(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega) = 0$ . Applying the inequality, we must have  $(n + m)r - r^2 + nm - |\Omega| \leq nm$ , which implies  $|\Omega| \geq (n + m)r - r^2$ .  $\square$

Given  $M_\Omega$ , it may not be known which rank  $r$  we should choose to complete  $M_\Omega$ . If we choose  $r$  too small, then  $\mathcal{A}_\Omega \cap \overline{\mathcal{M}}_r$  will be empty, and if we choose  $r$  too large, then  $M_\Omega$  will have infinitely many completions. We introduce a method of deciding such a rank  $r$ .

Suppose we are given  $\Omega$ , but we do not know  $r$ . Let  $p = |\Omega|$ . Then if  $M_\Omega$  is finitely completable in  $r$ , We must have  $p \geq (n + m)r - r^2$  from theorem 5. This implies



$$r \leq \frac{n + m - \sqrt{(n + m)^2 - 4p}}{2}.$$

So a good guess for a rank  $r$  may be  $r = \lfloor \frac{n+m-\sqrt{(n+m)^2-4p}}{2} \rfloor$ .

It is useful to define the function  $\Phi_\Omega : \overline{\mathcal{M}}_r \rightarrow M_{n \times m}$  as the restriction of  $P_\Omega$  to  $\overline{\mathcal{M}}_r$ . In other words,  $\Phi_\Omega(X)$  is the projection of a matrix  $X \in \overline{\mathcal{M}}_r$  obtained by setting entries with indices not in  $\Omega$  equal to zero. Then given a partially known matrix  $M_\Omega$ , we have  $\Phi_\Omega^{-1}(M_\Omega) = \mathcal{A}_\Omega \cap \overline{\mathcal{M}}_r$ , that is,  $\Phi_\Omega^{-1}(M_\Omega)$  is the space of rank at most  $r$  completions of  $M_\Omega$ .

We now focus on the set of matrices  $X$  such that given  $\Omega$ , the preimage  $\Phi_\Omega^{-1}(\Phi_\Omega(X))$  is a zero dimensional set. In other words, the set of  $X \in \overline{\mathcal{M}}_r$  such that  $\Phi_\Omega(X)$  has finitely many

rank  $r$  completions. We define such a set  $\chi_\Omega \subset \overline{\mathcal{M}}_r$  as

$$\chi_\Omega = \{X \in \overline{\mathcal{M}}_r \mid \Phi_\Omega^{-1}(\Phi_\Omega(X)) \text{ is zero dimensional}\}.$$

**Theorem 6.** *For any  $X \in \chi_\Omega$ ,  $\text{rank}(X) = r$ . That is,  $\chi_\Omega \subset \mathcal{M}_r$ .*

*Proof.* Without loss of generality by permuting rows and columns, suppose the index  $(1, 1) \notin \Omega$ . Consider  $X \in \chi_\Omega$  such that

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots \\ x_{21} & x_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

Suppose  $\text{rank}(X) < r$ . Then

$$Y(t) = \begin{bmatrix} t & x_{12} & \cdots \\ x_{21} & x_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

has rank at most  $r$ , so  $Y(t) \in \overline{\mathcal{M}}_r$  for any  $t$ . Moreover, since  $\Phi_\Omega(X) = \Phi_\Omega(Y(t))$ , then  $Y(t) \in \Phi_\Omega^{-1}(\Phi_\Omega(X))$  for any  $t$ . However, this implies that  $\dim(\Phi_\Omega^{-1}(\Phi_\Omega(X))) > 0$ , contradicting the assumption that  $X \in \chi_\Omega$ . Therefore, we must have  $\text{rank}(X) = r$ .  $\square$

To estimate the number of ways to complete a matrix, we introduce the degree of a variety.

**Definition 2.** *The degree of an affine or projective variety of dimension  $k$  is the number of intersection points of the variety with  $k$  hyperplanes in general position.*

For example, the degree of the algebraic variety  $\overline{\mathcal{M}}_r$  is known [26].

**Theorem 7.** *The degree of the algebraic variety  $\overline{\mathcal{M}}_r$  is*

$$V_{n,m,r} := \prod_{i=0}^{m-r-1} \frac{\binom{n+i}{m-1-i}}{\binom{n-r+i}{m-r-i}}$$

Recall a generalized version of Bézout's Theorem [1].

**Theorem 8.** *Let  $U_1, \dots, U_k$  be irreducible algebraic varieties, and let  $Z_1, \dots, Z_N$  be the irreducible components of  $U_1 \cap \dots \cap U_k$ . Then*

$$\sum_{i=1}^N \deg(Z_i) \leq \prod_{j=1}^k \deg(U_j).$$

Now we are ready to present an upper bound on the number of possible rank  $r$  completions of  $M_\Omega$ .

**Theorem 9.** *Given a partially known  $n \times m$  matrix  $M_\Omega$ , let  $N = |\mathcal{A}_\Omega \cap \overline{\mathcal{M}}_r|$  be the number of rank at most  $r$  completions of  $M_\Omega$ . If  $N < \infty$ , then  $N < V_{n,m,r}$ .*

*Proof.* Given  $M_\Omega$ , recall that  $\mathcal{A}_\Omega = \{X \in M_{n \times m} \mid P_\Omega(X) = M_\Omega\}$  is the algebraic variety of any possible completion of  $M_\Omega$ . Note that  $\mathcal{A}_\Omega$  is a linear variety, and so it is irreducible and  $\deg(\mathcal{A}_\Omega) = 1$ .

Now suppose there are finitely many points  $Z_1, \dots, Z_N \in \mathcal{A}_\Omega \cap \overline{\mathcal{M}}_r$ . That is, there are  $N$  possible rank  $r$  completions of  $M_\Omega$ . Since  $\deg(Z_i) = 1$  for all  $i$ , and both  $\mathcal{A}_\Omega$  and  $\overline{\mathcal{M}}_r$  are irreducible, then by theorem 7 and theorem 8 we have

$$N = \sum_{i=1}^N \deg(Z_i) \leq \deg(\mathcal{A}_\Omega) \deg(\overline{\mathcal{M}}_r) = V_{n,m,r}.$$

□

The number  $V_{n,m,r}$  is very large, and in general is larger than the exact number of rank at most  $r$  completions. One reason for this may be that given  $M_\Omega$ , the hyperplanes  $H_{ij} = \{X \in M_{n \times m} \mid X_{ij} = M_{ij}\}$  such that  $\bigcap_{(i,j) \in \Omega} H_{ij} = \mathcal{A}_\Omega$  may not be in general position. In addition, some intersection points may be at infinity, or the intersection points may have multiplicity.

It is often desirable to have a unique rank  $r$  completion rather than finitely many. If  $\Omega$  has the right structure, then a generic  $X \in \overline{\mathcal{M}}_r$  will be the unique rank  $r$  completion of  $\Phi_\Omega(X)$ . We introduce the following definition.

**Definition 3.** *We say a mask  $\Omega$  is uniquely completable in  $r$  if, for a generic  $X \in \overline{\mathcal{M}}_r$ ,  $X$  is the unique rank  $r$  completion of  $\Phi_\Omega(X)$ .*

We will give a class of such  $\Omega$ . First we introduce the following definition and theorem for motivation.

**Definition 4.** *We say a mask  $\Omega$  is completable entry by entry in  $r$  if we may find an  $(r+1) \times (r+1)$  submatrix of  $\Omega$  with exactly one entry equal to 0, replace that 0 with a 1, and repeat until all entries are equal to 1.*

**Theorem 10.** *If a mask  $\Omega$  is completable entry by entry in  $r$ , then it is uniquely completable in  $r$ .*

To prove theorem 10, will define the entry by entry matrix completion algorithm 5.

---

**Algorithm 5:** Complete Entry by Entry

---

**Input** : mask  $\Omega_0$ , partially known matrix  $M_{\Omega_0}$ , rank  $r$

**Result:** mask  $\Omega_k$  such that  $|\Omega_k| \geq |\Omega_0|$ , partially known matrix  $M_{\Omega_k}$  such that  $\Phi_{\Omega_k}^{-1}(M_{\Omega_k}) = \Phi_{\Omega_0}^{-1}(M_{\Omega_0})$ ;

**for**  $k = 0, 1, \dots$  **do**

search for an  $(r+1) \times (r+1)$  submatrix of  $M_{\Omega_k}$  of the form  $\begin{bmatrix} A_k & b_k \\ c_k & x_k \end{bmatrix}$  where  $x_k$  is the only unknown element,  $A_k$  is  $r \times r$  and nonsingular;

**if** no such submatrix exists **then**

output  $\Omega_k$  and  $M_{\Omega_k}$ ;

**else**

set  $(i_k, j_k)$  equal to the index of  $x_k$  in  $M_{\Omega_k}$ ;

take the cofactor expansion of  $M_k$  along row  $r+1$ , getting an expression of the form  $\det(A_k)x_k - a_k$  for some known constant  $a_k$ ;

set  $\Omega_{k+1} = \Omega_k \cup \{(i_k, j_k)\}$ ;

set  $M_{\Omega_{k+1}}$  equal to  $M_{\Omega_k}$  with the entry in index  $(i_k, j_k)$  equal to  $\frac{a_k}{\det(A_k)}$ ;

We will also introduce the following lemma.

**Lemma 1.** *The set of matrices  $M \in \overline{\mathcal{M}}_r$  that have at least one vanishing  $r \times r$  minor has dimension strictly smaller than  $\dim(\overline{\mathcal{M}}_r)$ .*

*Proof.* Given  $I \subset [n]$ ,  $J \subset [m]$ , such that  $|I| = |J| = r$ , let  $V_{I,J} = \{X \in M_{n \times m} \mid \det(X_{I,J}) = 0\}$ , where  $X_{I,J}$  is the  $r \times r$  submatrix submatrix of  $X$  with indices in  $I \times J$ . Let

$$V = \bigcup_{I,J} V_{I,J}$$

be the set of  $n \times m$  matrices with at least one vanishing  $r \times r$  minor. Then because  $V$  is a finite union of sets of the form  $V_{I,J}$ , it suffices to show that  $\dim(\overline{\mathcal{M}}_r \cap V_{I,J}) < \dim(\overline{\mathcal{M}}_r)$ . Note that  $V_{I,J}$  is an algebraic hypersurface which does not contain  $\overline{\mathcal{M}}_r$ .

**\*\*finish this\*\*** □

We will now prove theorem 10.

*Proof.* Given a mask  $\Omega$ , suppose  $\Omega$  is completable entry by entry in  $r$ . Then given  $M \in \overline{\mathcal{M}}_r$ , input  $M_\Omega = P_\Omega(M)$  into algorithm 5. At each step we may always find a submatrix with exactly one unknown entry because  $\Omega$  is completable entry by entry in  $r$ . The only way algorithm 5 may output a matrix that is not fully completed is if there is at least one  $r \times r$  singular submatrix in  $M$ . However, the set of matrices with a singular  $r \times r$  submatrix has measure zero in  $\overline{\mathcal{M}}_r$ , so algorithm 5 will output a fully completed matrix  $X$  for a generic  $M \in \overline{\mathcal{M}}_r$ . Moreover, since  $M$  satisfies the equations used to complete  $M_\Omega$ , each of which had a unique solution, we must have  $X = M$ . □

More generally, given any mask  $\Omega$  and any  $M \in \overline{\mathcal{M}}_r$ , let  $M_{\Omega'}$  be the output of algorithm 5 with input  $M_\Omega = P_\Omega(M)$ . Then all entries which are completed will be equal to the corresponding entry in  $M$ . Again, this is because each equation used to complete  $M_\Omega$  had a unique solution, and  $M$  satisfies those equations. So the completed entries must be the same for every rank at most  $r$  completion  $X \in \Phi_\Omega^{-1}(M_\Omega)$ , and because algorithm 5 only completed

entries which had unique completions, the space of possible rank at most  $r$  completions of  $M_{\Omega'}$  and  $M_{\Omega}$  must be equal. That is, we have  $\Phi_{\Omega'}^{-1}(M_{\Omega'}) = \Phi_{\Omega}^{-1}(M_{\Omega})$ .

**Example 2.1.** *A special class of partially known matrices which is are completable entry by entry are matrices of the form  $M_{\Omega} = \begin{bmatrix} A & B \\ C & \square \end{bmatrix}$  where  $A$  is  $r \times r$  and nonsingular,  $B$  is  $r \times (m-r)$ ,  $C$  is  $(n-r) \times r$ , and  $A$ ,  $B$ , and  $C$  consist of the known entries of  $M_{\Omega}$ . In fact,  $M_{\Omega}$  may be completed all at once to the unique rank  $r$  completion  $M = \begin{bmatrix} A & B \\ C & CA^{-1}B \end{bmatrix}$ . Note that  $M_{\Omega}$  has  $(n+m)r - r^2$  known entries, which is the minimum number of known entries such that  $M_{\Omega}$  may have a unique completion by theorem 5.*

## 2.8 Algebraic Combinatorics of Low-Rank Matrix Completion

The notes in this section are on [24]. Recall that a mask  $\Omega \subset [n] \times [m]$ , where  $[n] = \{1, \dots, n\}$ . We define  $G(\Omega)$  as the bipartite graph with vertices equal to the disjoint union of  $[n]$  and  $[m]$ , and an edge between  $i \in [n]$  and  $j \in [m]$  if  $(i, j) \in \Omega$ . Moreover, the adjacency matrix of our graph is equal to the binary matrix interpretation of  $\Omega$ .

**Example 2.2.** *The mask*

$$\Omega_1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

*Corresponds to the bipartite graph fig. 4.*

**Example 2.3.** *The mask*

$$\Omega_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

*Corresponds to the bipartite graph fig. 5.*



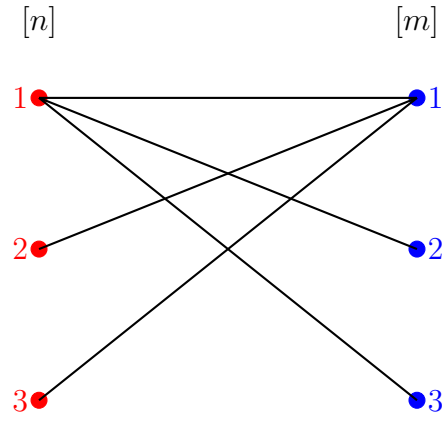


Figure 4: Bipartite graph  $G(\Omega_1)$

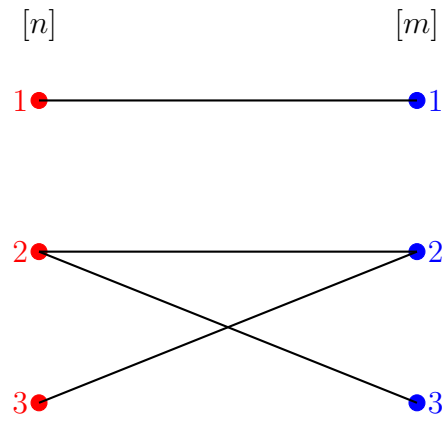


Figure 5: Bipartite graph  $G(\Omega_2)$

We will address the question of whether or not the unknown entry  $(i, j) \in \Omega^c$  is uniquely or finitely completable. In other words, given an unknown entry  $\square$  in  $M_\Omega$  in position  $(i, j)$ , are there only finitely many values that we could fill in for  $\square$  such that there still exists a rank  $r$  completion of the resulting matrix? In general, for any continuous method of sampling, the question of whether or not the unknown entry  $\square$  is uniquely completable depends only on the positions of the known entries in  $M_\Omega$  with probability one.

To answer the question of which entries have finitely many completions as part of a rank  $r$  matrix in further generality, we will introduce the following definition.

**Definition 5.** *Given a set of observed indices  $\Omega$  and a rank  $r$ , we start by defining the rank  $r$  finitely completable closure  $\text{cl}_r(\Omega)$  as the set of positions which are finitely completable in  $\Omega$  for a generic matrix  $M \in \overline{\mathcal{M}}_r$ .*

In order to classify  $\text{cl}_r(\Omega)$ , we introduce the following tools. We define the algebraic matrix multiplication map  $\Upsilon : M_{m,r} \times M_{n,r} \rightarrow M_{m,n}$  by  $\Upsilon : (U, V) \mapsto UV^\top$ . Note that every matrix in the image of this map has rank at most  $r$ . Moreover,  $\overline{\mathcal{M}}_r$  is the image of the map  $\Upsilon$ . Next, we calculate the Jacobian  $J$  of  $\Upsilon$ . At point  $(U, V)$ , the Jacobian has the following representation as an  $mn \times r(m+n)$  matrix in terms of the Kronecker product  $\otimes$  and identity matrices  $I_m$  and  $I_n$ .

$$J(U, V) = \begin{bmatrix} I_m \otimes v_1^\top \\ \vdots \\ I_m \otimes v_n^\top \end{bmatrix} \quad I_n \otimes U$$

Note that each row in  $J(U, V)$  corresponds to some matrix entry  $(i, j)$ . In particular, for an observed position  $(i, j)$ , the row  $J_{(i,j)}$  is defined to be the row of  $J$  corresponding to position  $(i, j)$ . For a collection of known indices  $\Omega$ , the matrix  $J_\Omega$  is defined to be the submatrix of  $J$  with rows corresponding to the known indices in  $\Omega$ . We may use this definition to classify  $\text{cl}_r(\Omega)$ .

**Theorem 11.** *Given a set of known indices  $\Omega$  and a rank  $r$ , we have the following classification of  $\text{cl}_r(\Omega)$ , the rank  $r$  finitely completable closure for generic partially known matrices  $M_\Omega$ .*

$$\text{cl}_r(\Omega) = \{(i, j) \mid J_{(i,j)} \text{ is in the rowspan of } J_\Omega\}$$

Here the linear independence of the rows of  $J_\Omega$  is a generic property, and so it does not depend on the choice of generic  $M_\Omega$ . Using this theorem, we may compute whether or not index  $(i, j)$  is generically finitely completable in  $\Omega$  with the following algorithm.

1. Sample  $U$  and  $V$  from a continuous distribution. This will ensure that they are generic with probability 1.
2. Calculate the Jacobian  $J_\Omega(U, V)$ .
3. Test whether or not  $J_{(i,j)}(U, V)$  is in the row-span of  $J_\Omega(U, V)$ . If it is, then index  $(i, j)$  is finitely completable. If not, then it is not finitely completable.

**Example 2.4.** *For a simple example, consider the case  $U = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$  and  $V = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$ . Then  $UV^\top = \begin{bmatrix} 2 & 10 \\ 3 & 15 \end{bmatrix}$ , and*

$$J(U, V) = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 3 & 0 \\ 5 & 0 & 0 & 2 \\ 0 & 5 & 0 & 3 \end{bmatrix}.$$

In this case we have,

$$\begin{aligned} J_{(1,1)} &= \begin{bmatrix} 1 & 0 & 2 & 0 \end{bmatrix} \\ J_{(2,1)} &= \begin{bmatrix} 0 & 1 & 3 & 0 \end{bmatrix} \\ J_{(1,2)} &= \begin{bmatrix} 5 & 0 & 0 & 2 \end{bmatrix} \\ J_{(2,2)} &= \begin{bmatrix} 0 & 5 & 0 & 3 \end{bmatrix}. \end{aligned}$$

Now let  $\Omega = \{(1, 1), (2, 1), (1, 2)\}$ . Then we have  $M_\Omega = \begin{bmatrix} 2 & 10 \\ 3 & \square \end{bmatrix}$  and

$$J_\Omega = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 3 & 0 \\ 5 & 0 & 0 & 2 \end{bmatrix}$$

If we assume that  $U$  and  $V$  are generic, then in order to check if the entry  $(2, 2)$  is finitely completable, we need to check if  $J_{(2,2)}$  is in the rowspan of  $J_\Omega$ . In this case it is, since  $\frac{3}{2}J_{(1,2)} + 5J_{(2,1)} - \frac{15}{2}J_{(1,1)} = J_{(2,2)}$ . Since  $U$  and  $V$  are generic,  $\text{cl}_1(\Omega) = \Omega \cup \{(2, 2)\}$ . However, if  $U$  and  $V$  are not generic, then the point  $(2, 2)$  may not be uniquely completable. As a counterexample, consider the same  $\Omega$  with  $U = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$  and  $V = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$ . Then  $UV^\top = \begin{bmatrix} 0 & 0 \\ 0 & 15 \end{bmatrix}$ , and  $M_\Omega = \begin{bmatrix} 0 & 0 \\ 0 & \square \end{bmatrix}$ . Clearly we can fill in  $\square$  with any number and the result will be at most a rank one matrix. To check this in terms of the Jacobian, we have

$$J(U, V) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 3 \end{bmatrix}$$

$$J_\Omega = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 5 & 0 & 0 & 0 \end{bmatrix}$$

In this case  $J_{(2,2)}$  is not in the rowspan of  $J_\Omega$ , and so the entry  $(2, 2)$  is not finitely completable in  $M_\Omega$ . However,  $U$  and  $V$  are not generic, and so this does not contradict the statement that  $\text{cl}_r(\Omega)$  is independent of  $M_\Omega$  for generic matrices.

To address the question of whether or not finite completability implies unique completability over complex numbers, we start with the following definition.

**Definition 6.** Similarly to the rank  $r$  finitely completable closure  $\text{cl}_r(\Omega)$ , we define the rank  $r$  uniquely completable closure  $\text{ucl}_r(\Omega)$  as the set of positions which are uniquely completable in  $\Omega$  for a generic matrix  $M \in \overline{\mathcal{M}}_r$ .

In order to characterize the uniquely completable closure  $\text{ucl}_r(\Omega)$ , we will introduce the following.

**Definition 7.** Given  $U \in M_{m \times r}(\mathbb{C})$  and  $V \in M_{n \times r}(\mathbb{C})$ , a rank  $r$  stress of  $M = UV^\top$  is a matrix  $S \in M_{m \times n}(\mathbb{C})$  that, as a vector, is in the kernel of the transpose of the Jacobian of  $U$  and  $V$ . That is, we have

$$J(U, V)^\top \text{vec}(S) = 0$$

Given  $\Omega$  the index set of the known entries, we define the  $\Omega$ -stresses as the stresses  $S$  such that  $S_{ij} = 0$  for all entries in unknown positions, that is, for all  $(i, j) \in \Omega^c$ . Note that if we vectorize  $S$  and remove the zeros corresponding to unknown entries, then the  $\Omega$ -stresses are in the kernel of the submatrix  $J_\Omega(U, V)^\top$ . Also note that the rank  $r$   $\Omega$ -stresses of  $M = UV^\top$  form a complex vector space, which will be denoted  $\Psi_M(\Omega)$ .

Finally, the maximal  $\Omega$ -stress rank of  $M$  in rank  $r$  is defined as

$$\rho_M(\Omega) = \max_{S \in \Psi_M(\Omega)} \text{rank}(S)$$

In general, if  $M$  is generic, then the maximal stress rank  $\rho_M(\Omega)$  depends only on  $\Omega$  and  $r$  and not on  $M$ . In this case, we will denote the generic  $\Omega$ -stress rank as  $\rho(\Omega)$ . These definitions are used to formulate the following theorem that gives conditions for which finite completability implies unique completability.

**Theorem 12.** *Given an index set of known entries  $\Omega$ , if the generic  $\Omega$ -stress rank in  $r$  satisfies the inequality  $\rho(\Omega) \geq \min(m, n) - r$ , then for generic  $M_\Omega$ , the finite completability of an entry in index  $(i, j)$  implies that the entry  $(i, j)$  is uniquely completable. That is  $\text{cl}_r(\Omega) = \text{ucl}_r(\Omega)$ .*

The above theorem gives sufficient conditions for unique completability of all finitely completable entries, but determining whether finite completability implies unique completability relies on calculating the generic  $\Omega$ -stress in rank  $r$ ,  $\rho(\Omega)$ , which can be done by using the following algorithm.

1. Sample  $U \in M_{m \times r}(\mathbb{C})$  and  $V \in M_{n \times r}(\mathbb{C})$  from a continuous distribution. This will ensure that they are generic with probability 1.
2. Calculate  $J_\Omega(U, V)$ .

3. Sample a random vector  $\text{vec}(S)$  in the kernel of  $J_\Omega(U, V)^\top$  from a continuous distribution, this will again ensure genericness with probability 1. Reshape  $\text{vec}(S)$  as a matrix  $S$  with entries in  $\Omega$  corresponding to entries in  $\text{vec}(S)$ , and entries in  $\Omega^c$  as zeros.
4. Output  $\rho(\Omega) = \text{rank}(S)$

## 2.9 Zero Sets of Systems of Matrix Minors

In this section we will introduce a decomposition of the zero set of a system of minors which will be useful later on. We first start with a lemma.

**Lemma 2.** *Consider the space  $M_{(r+1) \times n}$  of all  $(r+1) \times n$  matrices. Let  $V$  be the zero set of all  $(r+1) \times (r+1)$  minors containing the first  $k$  columns with  $k \leq r+1$ . Then  $V = \overline{\mathcal{M}}_r \cup W$ , where  $W = \{M \in M_{(r+1) \times n} \mid \text{the rank of the first } k \text{ columns of } M \text{ is } < k\}$ , where  $\overline{\mathcal{M}}_r$  is the space of  $(r+1) \times n$  matrices with rank  $\leq r$ .*

*Proof.* Note that  $\overline{\mathcal{M}}_r \subset V$  since the set of equations which define  $\overline{\mathcal{M}}_r$  contains the set of equations which define  $V$ . Also,  $W \subset V$ , since if  $M \in W$ , then the first  $k$  columns of  $M$  are linearly dependent, so every  $(r+1) \times (r+1)$  minor containing the first  $k$  columns vanishes, so  $M \in V$ . Therefore,  $\overline{\mathcal{M}}_r \cup W \subset V$ .

For the opposite inclusion, we will induct on  $n$ , the number of columns in  $M_{r+1 \times n}$  and backwards induction on  $k$ . Consider  $n = r+1$ . Then there is exactly one  $(r+1) \times (r+1)$  minor. Note that in this case for all  $k \leq r$ ,  $W \subset \overline{\mathcal{M}}_r$ , and so  $V = \overline{\mathcal{M}}_r = \overline{\mathcal{M}}_r \cup W$ . Fix  $n$ . For the case  $k = r+1$ , both  $V$  and  $W$  are the zero set of the first  $(r+1) \times (r+1)$  minor. In this case  $V = W$ , and  $\overline{\mathcal{M}}_r \subset W$ . Therefore,  $V = W = \overline{\mathcal{M}}_r \cup W$ .

Now by induction we will assume that  $V = \overline{\mathcal{M}}_r \cup W$  for all  $k$  for  $(r+1) \times (n-1)$  matrices, and that  $V = \overline{\mathcal{M}}_r \cup W$  for the first  $k+1$  columns. Let  $M \in V$ . If  $\text{rank}(M) \leq r$ , then  $M \in \overline{\mathcal{M}}_r$ . Suppose  $\text{rank}(M) = r+1$ . Consider the submatrix  $M'$  obtained by deleting the  $(k+1)$ st column. Then since  $M \in V$ , then by our inductive hypothesis on  $n$  we have

that the first  $k$  columns of  $M'$  are linearly dependent, or  $\text{rank}(M') \leq r$ . In the first case, we have that  $M \in W$ , so suppose  $\text{rank}(M') \leq r$ . Then since  $\text{rank}(M) = r + 1$ , we have that  $\text{rank}(M') = r$ , and the  $k + 1$ st column of  $M$  is linearly independent from the rest of the columns. In particular, it is linearly independent from the first  $k$  columns. Now note that  $M$  is in the zero set of all  $(r + 1) \times (r + 1)$  minors containing the first  $k + 1$  columns. By backwards induction on  $k$ ,  $M \in \overline{\mathcal{M}}_r$ , or the first  $k + 1$  columns are linearly dependent. However,  $M \notin \overline{\mathcal{M}}_r$ , so the first  $k + 1$  columns must be linearly independent. We also have that the  $k + 1$ st column is linearly independent from the first  $k$  columns. Therefore, the first  $k$  columns are linearly dependent, so  $M \in W$ . Therefore,  $M \in \overline{\mathcal{M}}_r \cup W$ , and so  $V \subset \overline{\mathcal{M}}_r \cup W$ . So  $V = \overline{\mathcal{M}}_r \cup W$ .  $\square$

We shall use the above lemma to prove the following:

**Theorem 13.** *Let  $A$  be the top-left  $k \times k$  submatrix with variables in  $M_{m \times n}$  and  $k \leq r$ . Consider the variety  $V$  which is the zero set of all  $(r + 1) \times (r + 1)$  minors which contain  $A$ . Then  $V = \overline{\mathcal{M}}_r \cup W$ , for some  $W$  such that for all  $M \in W$ , the first  $k \times k$  block of  $M$  is not invertible.*

*Proof.* We will induct on  $m$ . For the base case, let  $m = r + 1$ . Consider some  $M \in V$ , and suppose  $\text{rank}(A) = k$ . Then by lemma 2, we have that  $\text{rank}(M) \leq r$  or the rank of the first  $k$  columns is less than  $k$ . However, since  $\text{rank}(A) = k$ , the first  $k$  columns have rank  $k$ , and so we have  $\text{rank}(M) \leq r$ , so  $M \in \overline{\mathcal{M}}_r$ .

Now by induction we will assume that  $V = \overline{\mathcal{M}}_r \cup W$  for  $(m - 1) \times n$  matrices. Consider an  $m \times n$   $M \in V$ , and suppose  $\text{rank}(A) = k$ . Then we will show that  $M \in \overline{\mathcal{M}}_r$ .

Consider the submatrix  $M'$  which consists of the first  $m - 1$  rows of  $M$ . Now by the inductive hypothesis, since  $\text{rank}(A) = k$ , we must have that  $\text{rank}(M') \leq r$ . If  $\text{rank}(M') < r$ , then by adding in the last row we will have  $\text{rank}(M) \leq r$ , which means that  $M \in \overline{\mathcal{M}}_r$ . So suppose  $\text{rank}(M') = r$ . Since  $\text{rank}(A) = k$ , then the first  $k$  rows are linearly independent. So



without loss of generality by permuting the rows, suppose that the first  $r$  rows are linearly independent and span the row space of  $M'$ . Now consider the submatrix  $M''$  of  $M$  which consists of all but the second to last row. Then again by the inductive hypothesis and since  $\text{rank}(A) = k$ , we must have that  $\text{rank}(M'') \leq r$ . However, since the first  $r$  rows of  $M''$  span the row space, we have that the last row of  $M$  is contained in the span of the first  $r$  rows. Therefore, since  $\text{rank}(M') = r$ , and since the last row of  $M$  is contained in the row space of  $M'$  we must have  $\text{rank}(M) = r$ , so  $M \in \overline{\mathcal{M}}_r$ .  $\square$

## 2.10 Matrix Completion Topology



In this section we will discuss the topology of the spaces  $\overline{\mathcal{M}}_r$  and  $\mathcal{A}_\Omega$  and how they intersect. For a topological space  $X$ , let  $H_i(X)$  be the  $i$ th homology group of  $X$ . Also let  $h_i(X)$  be  $i$ th Betti number, which is equal to the dimension of the group  $H_i(X)$ . In particular,  $h_0(X)$  is equal to the number of connected components of  $X$ , so if  $X$  is a finite number of points, then  $h_0(X)$  is equal to that number of points. Moreover,  $h_1(X)$  is equal to the number of cycles in  $X$ . Intuitively speaking, a cycle is a non-trivial loop in the topological space  $X$ .

First, we will introduce the Mayer-Vietoris sequence from algebraic topology.

**Theorem 14.** [3] *Let  $U$  and  $V$  be two topological spaces whose interiors cover  $U \cup V$ . Then there exists a long exact sequence of the form*

$$\begin{array}{ccccccc}
 & & & & \dots & \longrightarrow & H_{n+1}(U \cup V) \\
 & & & & & \nearrow & \\
 H_n(U \cap V) & \longleftarrow & H_n(U) \oplus H_n(V) & \longrightarrow & H_n(U \cup V) & & \\
 & & & & & \nearrow & \\
 H_{n-1}(U \cap V) & \longleftarrow & \dots & \longrightarrow & H_1(U \cup V) & & \\
 & & & & & \nearrow & \\
 H_0(U \cap V) & \longleftarrow & H_0(U) \oplus H_0(V) & \longrightarrow & H_0(U \cup V) & \longrightarrow & 0
 \end{array}$$

which is

called the Mayer-Vietoris sequence.

Using this sequence, we obtain the following theorem on the relationship between  $h_0(\overline{\mathcal{M}}_r \cap$

$\mathcal{A}_\Omega$ ) and  $h_1(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega)$ .

**Theorem 15.**  $h_0(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega) = h_1(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega) + 1$ . In particular,  $h_0(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega)$  is equal to the number of connected components of  $\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega$ , so if  $\dim(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega) = 0$ , that implies that there is some finite number of points  $N$  such that  $N = |\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega| = h_1(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega) + 1$ .

*Proof.* Let us assume  $\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega \neq \emptyset$ . By Mayer-Vietoris, we have the long exact sequence

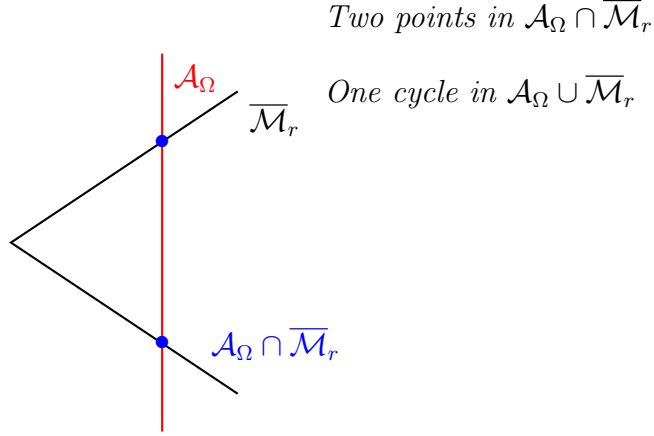
$$\begin{array}{ccccccc} \cdots & \longrightarrow & H_1(\overline{\mathcal{M}}_r) \oplus H_1(\mathcal{A}_\Omega) & \longrightarrow & H_1(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega) & & \\ & & & \searrow & & & \\ & & H_0(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega) & \longrightarrow & H_0(\overline{\mathcal{M}}_r) \oplus H_0(\mathcal{A}_\Omega) & \longrightarrow & H_0(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega) \longrightarrow 0 \end{array}$$

Note that  $\overline{\mathcal{M}}_r$  is homotopy equivalent to a point. Since scaling does not change the rank of a matrix,  $\overline{\mathcal{M}}_r$  deformation retracts to  $\{0\}$  from the map  $f : \overline{\mathcal{M}}_r \times [0, 1] \rightarrow \{0\}$ , where  $f(M, t) = (1 - t)M$ . Similarly, since  $\mathcal{A}_\Omega$  is an affine-plane, it also deformation retracts to a point, so both  $\overline{\mathcal{M}}_r$  and  $\mathcal{A}_\Omega$  have trivial homology. Moreover, since  $\overline{\mathcal{M}}_r$  and  $\mathcal{A}_\Omega$  are both connected and have non-empty intersection, then their union is also connected. Therefore, our long exact sequence simplifies to:

$$0 \longrightarrow H_1(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega) \longrightarrow H_0(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega) \longrightarrow \mathbb{Z}^2 \longrightarrow \mathbb{Z} \longrightarrow 0$$

It is a fact that if we have a long exact sequence with zeros on the ends, then the alternating sum of the dimensions is equal to zero, so we have  $h_1(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega) - h_0(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega) + 2 - 1 = 0$ , which implies  $h_0(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega) = h_1(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega) + 1$ .  $\square$

**Example 2.5.** For visual intuition on theorem 15, consider the case where there are two intersection points between  $\overline{\mathcal{M}}_r$  and  $\mathcal{A}_\Omega$ . That is,  $h_0(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega) = 2$ . Then we have  $h_1(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega) = 1$ , meaning that there is one cycle on  $\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega$ .



Note that if the number of rank  $r$  completions is finite, that is if  $\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega$  is a finite number of points, then  $h_0(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega)$  is equal to that number of points. This means that if it is possible to calculate  $H_1(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega)$ , then we can find the number of points in the intersection  $\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega$ . One strategy is to calculate the fundamental group  $\pi_1(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega)$ , since its abelianization is equal to  $H_1(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega)$ .

We have the following theorems in terms of the Euler characteristic. Let  $\chi(X)$  be the Euler characteristic of a topological space of  $X$ .

**Theorem 16.** *We have  $\chi(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega) = 2 - \chi(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega)$ . In particular, if  $\dim(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega) = 0$ , let  $N$  be the number of points in  $\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega$ . Then we have  $N = 2 - \chi(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega)$ .*

*Proof.* Since both  $\overline{\mathcal{M}}_r$  and  $\mathcal{A}_\Omega$  are homotopy equivalent to a point, and the Euler characteristic is invariant under homotopy,  $\chi(\overline{\mathcal{M}}_r) = 1$  and  $\chi(\mathcal{A}_\Omega) = 1$ . By the inclusion-exclusion property of the Euler characteristic,  $\chi(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega) = \chi(\overline{\mathcal{M}}_r) + \chi(\mathcal{A}_\Omega) - \chi(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega) = 2 - \chi(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega)$  □

In terms of the Euler characteristic of the manifold of rank  $r$  matrices  $\mathcal{M}_r$ , we have the following theorem.

**Theorem 17.**  *$\chi(\mathcal{M}_r \cap \mathcal{A}_\Omega) = 1 + \chi(\mathcal{M}_r) - \chi(\mathcal{M}_r \cup \mathcal{A}_\Omega)$ . Moreover, if  $\dim(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega) = 0$ , let  $N$  be the number of points in  $\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega$ . Then  $N = 1 + \chi(\mathcal{M}_r) - \chi(\mathcal{M}_r \cup \mathcal{A}_\Omega)$ .*

*Proof.* By the inclusion-exclusion principle of the Euler characteristic,

$$\chi(\mathcal{M}_r \cap \mathcal{A}_\Omega) = \chi(\mathcal{A}_\Omega) + \chi(\mathcal{M}_r) - \chi(\mathcal{M}_r \cup \mathcal{A}_\Omega).$$

Since  $\mathcal{A}_\Omega$  is homotopy equivalent to a point,  $\chi(\mathcal{A}_\Omega) = 1$ . Moreover, we have shown that if  $\dim(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega) = 0$ , then every point in the intersection has rank equal to  $r$ . Therefore,  $N = \chi(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega) = \chi(\mathcal{M}_r \cap \mathcal{A}_\Omega)$ .  $\square$

The Euler characteristic of the set of fixed rank  $r$ ,  $n \times m$  matrices,  $\chi(\mathcal{M}_r)$ , is known [37]. In particular, if our base field is  $\mathbb{C}$ , we have  $\chi(\mathcal{M}_r) = 0$  for  $r \geq 1$  and  $\chi(\mathcal{M}_r) = 1$  if  $r = 0$ . If our base field is  $\mathbb{R}$ , we have

$$\chi(\mathcal{M}_r) = \begin{cases} 1 & \text{if } r = 0 \\ 0 & \text{if } r \geq 2 \\ \frac{(1+(-1)^{n-1})(1+(-1)^{m-1})}{2} & \text{if } r = 1 \end{cases}$$

Note that since  $\overline{\mathcal{M}}_r$  and  $\mathcal{A}_\Omega$  are not open subsets of  $\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega$ , we cannot directly use Mayer-Vietoris. One approach is to instead consider thickened versions of  $\overline{\mathcal{M}}_r$  and  $\mathcal{A}_\Omega$ . Let

$$U = \{X \mid \exists M \in \overline{\mathcal{M}}_r, \|X - M\| < \epsilon\}$$

$$V = \{X \mid \exists M \in \mathcal{A}_\Omega, \|P_\Omega(X) - P_\Omega(M)\| < \delta\}.$$

Then each of  $U$  and  $V$  are open sub-sets of  $U \cup V$ , and so we may use Mayer-Vietoris.

Recall that if  $\dim(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega) = 0$ , then  $\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega$  must be a finite number of points by theorem 9. So suppose  $\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega$  is finite with  $N$  points. Then we may choose both  $\epsilon$  and  $\delta$  small enough so that the number of connected components of  $U \cap V$  is equal to  $N$ , and that  $U \cup V$  is homotopy equivalent to  $\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega$ . So by identical arguments, we have that  $h_0(U \cap V) = h_1(U \cup V) + 1$  and  $\chi(U \cap V) = 2 - \chi(U \cup V)$ , which by homotopy equivalence,

implies  $h_0(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega) = h_1(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega) + 1$  and  $\chi(\overline{\mathcal{M}}_r \cap \mathcal{A}_\Omega) = 2 - \chi(\overline{\mathcal{M}}_r \cup \mathcal{A}_\Omega)$ .

### 3 The Maximum Volume Principle and Maximum Volume Algorithms

In several matrix analysis problems, knowledge of a quality submatrix of a large matrix is required. For example, when using the Schur complement  $S_A = D - CA^{-1}B$ , it is desirable to choose a quality submatrix  $A$ . In particular,  $A$  should not be close to singular, so the quality of  $A$  may be measured by the modulus of the determinant, or the volume. Finding the largest volume submatrix is very difficult in general. However, we may instead opt to search for locally maximal volume submatrices, otherwise known as dominant submatrices, which are much easier to find.

#### 3.1 Schur Complement

First, we will introduce the Schur complement of a matrix  $M$  with respect to a submatrix  $A$ . Let  $M \in M_{n \times m}$ . Without loss of generality by permutation of rows and columns, suppose  $M$  has the structure  $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$  for some  $k \times k$  nonsingular submatrix  $A$ , and corresponding  $B$ ,  $C$ , and  $D$ . Then the *Schur complement* of  $M$  with respect to  $A$  is defined as

$$S_A = D - CA^{-1}B.$$

The Schur complement has the useful formula called the Schur determinant identity.

**Lemma 3.** *For any fixed, full rank,  $k \times k$  submatrix  $A$  in  $M$ , we have*

$$\det(M) = \det(A) \det(S_A).$$

*Moreover, by taking the absolute value, we also have*

$$\text{vol}(M) = \text{vol}(A) \text{vol}(S_A).$$

*Proof.* Note that

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & 0 \\ C & I \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & D - CA^{-1}B \end{bmatrix}.$$

Taking the determinant we get the desired result.  $\square$

We also have the following useful property of the Schur complement.

**Lemma 4.** *For any fixed, nonsingular  $k \times k$  submatrix  $A$  in  $M$ , we have that  $S_A = 0$  if and only if  $M$  is rank  $k$ .*

*Proof.* Suppose  $\text{rank}(M) = k$ . Then since  $A$  is a nonsingular  $k \times k$  submatrix,  $\text{rank}(M) = k$  if and only if the columns of  $\begin{bmatrix} A \\ C \end{bmatrix}$  form a basis for the columns space of  $M$ , if and only if there exists a unique matrix  $X$  such that

$$\begin{bmatrix} A \\ C \end{bmatrix} X = \begin{bmatrix} B \\ D \end{bmatrix}.$$

Solving for  $X$  we must have  $X = A^{-1}B$ . If and only if  $CA^{-1}B = D$ , and  $S_A = D - CA^{-1}B = 0$ .  $\square$

### 3.2 Then Skeleton Approximation

Given an  $n \times m$  matrix  $M$ , after permutation of rows and columns, let  $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$  where  $A$  is an  $r \times r$  invertible submatrix. Then

$$M_r = \begin{bmatrix} A \\ C \end{bmatrix} A^{-1} \begin{bmatrix} A & B \end{bmatrix} = \begin{bmatrix} A & B \\ C & CA^{-1}B \end{bmatrix}$$

is a rank  $r$  *skeleton approximation* with respect to  $A$ . Note that  $\text{rank}(M_r) = r$ . Moreover, the row space is spanned by the first  $r$  rows, and the columns space is spanned by the first  $r$  columns.

In general, the error of this approximation to the original matrix is larger than the error to the best rank  $r$  approximation obtained by the singular value decomposition. However, the skeleton approximation has the benefit being parameterized by a rational function of the actual entries of  $M$ . Moreover, we do not need to calculate a singular value decomposition of  $M$  to calculate a skeleton approximation of  $M$ .

The infinity norm error of this approximation is minimized over all choices of  $k \times k$  submatrix  $A$  when  $\text{vol}(A) = |\det(A)|$  is maximized. In particular, when  $A$  is chosen with maximum volume we have the inequality from [21] that

$$\|M - M_r\|_\infty \leq (r + 1)\sigma_{r+1}(M).$$

Note that  $\sigma_{r+1}(M)$  is the error to the best rank  $r$  approximation in the spectral norm.

Now relating the skeleton approximation to the Schur complement, we have

$$M - M_r = \begin{bmatrix} A & B \\ C & D \end{bmatrix} - \begin{bmatrix} A & B \\ C & CA^{-1}B \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & S_A \end{bmatrix}$$

This implies that

$$\|M - M_r\|_\infty = \|S_A\|_\infty$$

In other words, the infinity norm of the error of our original matrix to the skeleton approximation is equal to the infinity norm of the Schur complement of  $M$  with respect to  $A$ . We have the following theorem on how the choice of  $A$  affects the error of the skeleton approximation.

**Theorem 18.** [21] *The infinity norm of the Schur complement,  $\|S_A\|_\infty = \|D - CA^{-1}B\|_\infty$ , is minimized over all possible choices of submatrices  $A$  with corresponding  $B, C$ , and  $D$  when  $\text{vol}(A)$  is maximized.*



We will denote a  $k \times k$  submatrix of maximum volume over all choices of  $k \times k$  submatrix by  $A_{\blacksquare}$ . In general, finding  $A_{\blacksquare}$  is an NP-hard problem [22]. However, we may instead search for a submatrix which is locally maximum in volume as opposed to globally. Such submatrices are called dominant submatrices.

**Definition 8.** For an  $n \times r$  matrix  $M$ , we call a submatrix  $A$  of  $M$  dominant if all entries of  $MA^{-1}$  are no larger than 1 in modulus. Equivalently,  $A$  is dominant if  $\|MA^{-1}\|_{\infty} = 1$ . If  $M$  is  $n \times m$ . For a general  $n \times m$  matrix  $M$ , a submatrix  $A$  is dominant if it is dominant in it's respective rows and columns of  $M$ .

We refer to dominant submatrices of  $M$  by  $A_{\square}$ . We introduce the concept of a dominant submatrix is because they are easier to search for that maximum volume submatrices, and are not too far off in volume from the maximum volume submatrix.

For an equivalent definition of dominant submatrices for an  $n \times r$  matrix  $M$ , a submatrix  $A$  is dominant in  $M$  if we may not increase the volume of  $A$  by swapping a row in  $A$  with a row in  $M$ . More generally for a  $n \times m$  matrix  $M$ , a submatrix  $A$  is dominant in  $M$  in we may not increase the volume of  $M$  by swapping either two rows or two columns of  $M$ .

Analogous to the statement that global maximums are also local maximums, we have the following theorem.

**Theorem 19.** A submatrix of maximum volume  $A_{\blacksquare}$  in  $M$  is always a dominant submatrix.

*Proof.* Let  $A_{\blacksquare}$  be a  $k \times k$  submatrix of maximum volume over all  $k \times k$  submatrices in  $M$ ,  $n \times k$ . Suppose  $x_{ij}$  an element of  $MA^{-1}$  is larger than 1 in modulus. Let  $A'$  be the  $k \times k$  submatrix of  $M$  obtained by swapping the  $i$ th row of  $M$  with the  $j$ th row of  $A$ . Then since multiplication by an invertible matrix does not change the ratio of determinants of submatrices, we have  $\frac{\text{vol}(A)}{1} = \frac{\text{vol}(A')}{x_{ij}}$  which implies that  $\text{vol}(A') = x_{ij} \text{vol}(A)$ . However, this is a contradiction, since  $x_{ij} > 1$  implies  $\text{vol}(A') > \text{vol}(A)$  when  $A$  is maximum volume.  $\square$

One nice property of dominant submatrices is that the volume of a dominant submatrix  $A_{\square}$  is not too far from the volume of a maximum volume submatrix  $A_{\blacksquare}$ .

**Theorem 20.** *For any matrix  $M$ , we have*

$$\text{vol}(A_{\blacksquare}) \leq r^{r/2} \text{vol}(A_{\square})$$

*Proof.* This proof is given in [21]. □

### 3.3 Maximum Volume Algorithms

We start with the standard maxvol algorithm described in [21] for finding a close to dominant  $r \times r$  submatrix of an  $n \times r$  matrix  $M$ .

---

#### Algorithm 6: Maximal Volume Algorithm

---

**Input:**  $n \times r$  matrix  $M$ ,  $r \times r$  nonsingular submatrix  $A_0$ , tolerance  $\epsilon > 0$

**Result:**  $A_l$  a close to dominant submatrix of  $M$ .

Let  $l = 0$ ,  $B_0 = MA_0^{-1}$ ;

Set  $b_{ij}$  equal to the largest in modulus entry of  $B_0$ ;

**while**  $|b_{ij}| > 1 + \epsilon$  **do**

Replace the $j$ th row of $A_l$ with the $i$ th row of $M$ ;
$l := l + 1$ ;
Let $B_l = MA_l^{-1}$ ;
Set $b_{ij}$ equal to the largest in modulus entry of $B_l$ ;

---

The Maxvol algorithm gives up an increasing sequence of volumes of submatrices. In other words, we have the following theorem.

**Theorem 21.** [21] *The sequence  $\{v_l\} = \{\text{vol}(A_l)\}$  is increasing.*

We may generalize this algorithm to find a  $r \times r$  dominant submatrix of an  $n \times m$  matrix  $M$  by searching for the largest in modulus entry of both  $B_l = M(:, J_l)A_l^{-1}$ , and  $C_l = A_l^{-1}M(I_l, :)$  at each step.

---

**Algorithm 7:** 2D Maximal Volume Algorithm

---

**Input:**  $n \times m$  matrix  $M$ ,  $r \times r$  nonsingular submatrix  $A_0$ , tolerance  $\epsilon > 0$ ,  $l = 0$ ,  
 $b_{ij} = \infty$

**Result:**  $A_l$  a close to dominant submatrix with indices  $(I_l, J_l)$  in  $M$ .

**while**  $|b_{ij}| > 1 + \epsilon$  **do**

    Let  $B_l = M(:, J_l)A_l^{-1}$ , and  $C_l = A_l^{-1}M(I_l, :)$ ;

    Set  $b_{ij}$  equal to the largest in modulus entry of both  $B_l$  and  $C_l$ ;

**if**  $b_{ij}$  is from  $B_l$  **then**

        | Replace the  $j$ th row of  $A_l$  with the  $i$ th row of  $M(:, J_l)$

**else**

        | Replace the  $i$ th column of  $A_l$  with the  $j$ th column of  $M(I_l, :)$

$l := l + 1$ ;

---

However, this algorithm requires two backslash operations at each step. To simplify this to one backslash operation at each step, we may consider an alternating maxvol algorithm where we alternate between optimizing swapping rows and columns. Note that this converges to a dominant submatrix because the sequence  $\text{vol}(A_l)$  is again increasing and bounded above.

---

**Algorithm 8:** Alternating Maximal Volume Algorithm

---

**Input:**  $n \times m$  matrix  $M$ ,  $r \times r$  nonsingular submatrix  $A_0$ , tolerance  $\epsilon > 0$ ,  $l = 0$ ,  
 $b_{ij} = \infty$

**Result:**  $A_l$  a close to dominant submatrix of  $M$  with index set  $(I_l, J_l)$  in  $M$ .

**while**  $\max\{|b_{ij}|, |c_{ij}|\} > 1 + \epsilon$  **do**

    Let  $B_l = M(:, J_l)A_l^{-1}$ ;

    Set  $b_{ij}$  equal to the largest in modulus entry of  $B_l$ ;

**if**  $|b_{ij}| > 1 + \epsilon$  **then**

        | Replace the  $j$ th row of  $A_0$  with the  $i$ th row of  $M(:, J_0)$

    Let  $C_l = A_l^{-1}M(I_l, :)$ ;

    Set  $c_{ij}$  equal to the largest in modulus entry of  $C_l$ ;

**if**  $|c_{ij}| > 1 + \epsilon$  **then**

        | Replace the  $i$ th column of  $A_l$  with the  $j$ th column of  $M(I_l, :)$

$l := l + 1$ ;

---

### 3.4 Greedy Maximum Volume Algorithms

We may reduce the number of backslash operations needed to find a dominant submatrix by swapping more rows at each step, which we will call a greedy maxvol algorithm. The greedy maxvol algorithm is similar to the maxvol algorithm. The main difference is that instead of swapping one row every iteration, we may swap two or more rows. First, we will describe the algorithm for swapping at most two rows of an  $n \times r$  matrix, which we will call greedy maxvol, or 2-greedy maxvol.

Given an  $n \times r$  matrix  $M$ , initial  $r \times r$  nonsingular submatrix  $A_0$ , and tolerance  $\epsilon > 0$ , we do the following.

---

**Algorithm 9:** 2-Greedy Maximal Volume Algorithm

---

**Input:**  $n \times r$  matrix  $M$ ,  $r \times r$  nonsingular submatrix  $A_0$ , tolerance  $\epsilon > 0$

**Result:**  $A_l$  a close to dominant submatrix of  $M$ .

Let  $l = 0$ ,  $B_0 = MA_0^{-1}$ ;

Set  $b_{i_1 j_1}$  equal to the largest in modulus entry of  $B_0$ ;

**while**  $|b_{i_1 j_1}| > 1 + \epsilon$  **do**

    Replace the  $j_1$ th row of  $A_l$  with the  $i_1$ th row of  $M$ ;

    Set  $b_{i_2 j_2}$  equal to the largest in modulus entry of  $B_l$  over all columns excluding the  $j_1$ st column;

    Let  $B'_l = \begin{bmatrix} b_{i_1 j_1} & b_{i_1 j_2} \\ b_{i_2 j_1} & b_{i_2 j_2} \end{bmatrix}$ ;

**if**  $\text{vol}(B'_l) > |b_{i_1 j_1}|$  **then**

        Replace the  $j_2$ th row of  $A_l$  with the  $i_2$ th row of  $M$ .

$l := l + 1$ ;

    Let  $B_l = MA_l^{-1}$ ;

    Set  $b_{i_1 j_1}$  equal to the largest in modulus entry of  $B_l$ ;

---

To prove that this algorithm converges, recall Hadamard's inequality. For an  $n \times n$  matrix  $N$ , we have

$$\text{vol}(N) \leq \prod_{i=1}^n \|N(:, i)\|.$$

For an  $n \times m$  matrix  $M$ , let  $A$  be a square submatrix of  $M$ . Then it follows that  $\text{vol}(A) \leq \prod_{i=1}^n \|M(:, i)\|$ .

**Theorem 22.** *The sequence  $\text{vol}(A_l)$  is increasing and is bounded above by  $\prod_{i=1}^n \|M(:, i)\|$ . Therefore, it converges.*

*Proof.* Note that multiplication by an invertible matrix does not change the ratio of determinants of pairs of corresponding submatrices. Suppose we only swap one row. Then  $\frac{\det(A_{l+1})}{\det(A_l)} = \frac{b_{i_1 j_1}}{\det(I)} = b_{i_1 j_1}$ . Therefore, since  $|b_{i_1 j_1}| > 1$ , we have  $\text{vol}(A_{l+1}) > \text{vol}(A_l)$ .

Now suppose we swap two rows. Then after permutation of rows, the submatrix of  $B_n$  corresponding to  $A_{l+1}$  is  $\begin{bmatrix} B'_l & R \\ 0 & I \end{bmatrix}$  in block form, which has determinant  $\det(B'_l)$ . Therefore similarly to before, we have that  $\det(A_{l+1}) = \det(B'_l) \det(A_l)$ , and since  $\text{vol}(B'_l) > |b_{i_1 j_1}| > 1$ , we have that  $\text{vol}(A_{l+1}) > \text{vol}(A_l)$ , so  $\text{vol}(A_l)$  is an increasing sequence.  $\square$

Note that when we are swapping two rows, the ratio between  $\text{vol}(A_{l+1})$  and  $\text{vol}(A_l)$  is maximized when  $\text{vol}(B'_l)$  is maximized.

For a more general algorithm, we search for a largest element in each of the  $r$  rows of  $B_l$ :  $|b_{i_1 j_1}| \geq |b_{i_2 j_2}| \geq \dots \geq |b_{i_r j_r}|$ . Define  $b_k := b_{i_k j_k}$ . Let

$$B_l^{(k)} = \begin{bmatrix} b_{i_1 j_1} & b_{i_1 j_2} & \dots & b_{i_1 j_k} \\ b_{i_2 j_1} & b_{i_2 j_2} & \dots & b_{i_2 j_k} \\ \vdots & \vdots & & \vdots \\ b_{i_k j_1} & b_{i_k j_2} & \dots & b_{i_k j_k} \end{bmatrix}.$$

Then we replace the  $j_k$ th row of  $A_n$  with the  $i_k$ th row of  $M$  if  $\text{vol}\left(\begin{bmatrix} B_l^{(k)} & x_k \\ y_k & b_{k+1} \end{bmatrix}\right) \geq \text{vol}(B_l^{(k)})$ ,

where  $x_k = \begin{bmatrix} b_{i_1 j_k} \\ b_{i_2 j_k} \\ \vdots \\ b_{i_{k-1} j_k} \end{bmatrix}$ , and  $y_k = [b_{i_k j_1} \ b_{i_k j_2} \ \dots \ b_{i_k j_{k-1}}]$ . Using lemma 3, the Schur determinant formula for the determinant of block-matrices, this condition is equivalent to the condition that  $\left| b_{k+1} - d_k \left[ B_l^{(k)} \right]^{-1} c_k \right| \geq 1$ . In particular, if  $|b_{k+1}| > 1$ , and  $\text{sgn}(b_{k+1}) \neq \text{sgn}(d_k \left[ B_l^{(k)} \right]^{-1} c_k)$ , then the condition will be met. The general  $h$ -greedy maxvol algorithm for swapping at most  $h$  rows at each iteration runs as follows.

---

**Algorithm 10:**  $h$ -Greedy Maximal Volume Algorithm

---

**Input:**  $n \times r$  matrix  $M$ ,  $r \times r$  nonsingular submatrix  $A_0$ , tolerance  $\epsilon > 0$ ,  $l = 0$

**Result:**  $A_l$  a close to dominant submatrix of  $M$ .

Let  $B_0 = MA_0^{-1}$ ;

Set  $b_{i_1 j_1}$  equal to the largest in modulus entry of  $B_0$ ;

**while**  $|b_{i_1 j_1}| > 1 + \epsilon$  **do**

    Replace the  $j_1$ th row of  $A_l$  with the  $i_1$ th row of  $M$ ;

**for**  $k = 2:h$  **do**

        Set  $b_{i_k j_k}$  equal to the largest in modulus entry of  $B_l$  over all columns  
        excluding the  $j_1, \dots, j_{k-1}$  columns;

        Let  $B'_k = \begin{bmatrix} b_{i_1 j_1} & b_{i_1 j_2} & \cdots & b_{i_1 j_k} \\ b_{i_2 j_1} & b_{i_2 j_2} & \cdots & b_{i_2 j_k} \\ \vdots & \vdots & & \vdots \\ b_{i_k j_1} & b_{i_k j_2} & \cdots & b_{i_k j_k} \end{bmatrix}$ ;

**if**  $\text{vol}(B'_k) > \text{vol}(B'_{k-1})$  **then**

            replace the  $j_k$ th row of  $A_l$  with the  $i_k$ th row of  $M$ ;

**else**

**break**;

$l := l + 1$ ;

    Let  $B_l = MA_l^{-1}$ ;

    Set  $b_{i_1 j_1}$  equal to the largest in modulus entry of  $B_k$ ;

---

Similarly to before, we may generalize the  $h$ -greedy maxvol algorithm to  $n \times m$  matrices by alternating between the rows and columns.

What is the probability that we actually swap more than one row in our  $h$ -greedy maxvol algorithm? If we assume that  $\text{sgn}(b_{k+1})$  and  $\text{sgn}(d_k [B_n^{(k)}]^{-1} c_k)$  are equal with probability  $\frac{1}{2}$ , then at each iteration we will have a  $\frac{1}{2}$  chance of swapping an additional row. According to these assumptions as  $r$  gets large, the expected number of swaps will be at least 2 in the generalized maxvol algorithm.

### 3.5 Greedy Maxvol Numerical Experiments

The question now is, how many backslash operations, and how much computational time does the  $h$ -greedy maxvol algorithm save when compared to the greedy maxvol algorithm

in practice? We will generate 100 random  $5000 \times r$  matrices, and calculate the average number of backslash operations, and average computational time, needed to find a dominant submatrix within a relative error of  $10^{-8}$  for  $r = 30, 60, 90, 120, 150, 180, 210$  and  $240$ . In table 1 we plot the average number of backslash operations, and in table 2 we plot the average computational time. As we can see the number of backslash operations and the computational time tends to taper off around,  $h = 3$ , and higher  $h$  does not reduce the number of backslash operations calculated or computational time very much.

$r$	Average Number of Backslash Operations				
	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = r$
30	33.92	23.84	20.88	20.33	19.84
60	50.56	34.05	31.78	31.31	29.56
90	62.68	42.91	38.39	37.29	38.06
120	71.64	51.46	44.43	42.57	41.12
150	81.97	54.78	50.08	46.02	46.33
180	89.75	58.93	54.06	52.54	53.10
210	95.68	64.82	57.93	55.25	53.37
240	99.65	69.85	60.77	57.39	55.55

Table 1: Average number of backslash operations needed to find a dominant submatrix of 100 random  $5000 \times r$  matrices using  $h$ -greedy maxvol algorithm within a relative error of  $10^{-8}$ .

$r$	Average Time Taken				
	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = r$
30	0.1310	0.1195	0.1072	0.1048	0.1028
60	0.2653	0.2121	0.1993	0.1973	0.1883
90	0.5128	0.4217	0.3823	0.3716	0.3840
120	0.8644	0.7168	0.6190	0.5968	0.5850
150	1.2885	1.0194	0.9437	0.8723	0.8946
180	1.8529	1.4043	1.2863	1.2581	1.2983
210	2.4726	1.9294	1.7411	1.6637	1.6285
240	3.0215	2.4335	2.1348	2.0247	2.0117

Table 2: Average time in seconds to find a dominant submatrix of 100 random  $5000 \times r$  matrices using  $h$ -greedy maxvol algorithm within a relative error of  $10^{-8}$ .

Next, we will plot the number of backslash operations and the computational time for running an alternating  $h$ -greedy algorithm on  $5000 \times 5000$  matrices.

$r$	Average Number of Backslash Operations				
	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = r$
30	33.63	23.58	21.16	20.49	19.98
60	51.22	34.65	31.81	30.39	30.13
90	65.52	44.31	38.75	38.71	37.06
120	74.13	48.49	45.42	43.79	43.09
150	80.81	55.44	51.24	47.77	48.24
180	89.85	60.85	52.64	50.52	49.04
210	95.74	63.96	57.88	56.89	53.16
240	100.45	67.64	59.38	59.04	55.41

Table 3: Average number of backslash operations needed to find a dominant submatrix of 100 random  $5000 \times 5000$  matrices using  $h$ -greedy maxvol algorithm within a relative error of  $10^{-8}$ .

$r$	Average Time Taken				
	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = r$
30	0.1360	0.0930	0.0795	0.0754	0.0730
60	0.4169	0.2644	0.2457	0.2351	0.2372
90	0.8349	0.5577	0.4921	0.4940	0.4797
120	1.3528	0.8887	0.8391	0.8114	0.8128
150	2.0072	1.3865	1.2946	1.2105	1.2435
180	2.7712	1.8928	1.6516	1.5875	1.5680
210	3.6673	2.4751	2.2535	2.2290	2.1063
240	4.5758	3.1068	2.7553	2.7464	2.6220

Table 4: Average time in seconds to find a dominant submatrix of 100 random  $5000 \times 5000$  matrices using  $h$ -greedy maxvol algorithm within a relative error of  $10^{-8}$ .

As we can see from table 3 and table 4 the computational time and number of backslash operations needed decreases as  $h$  increases up to  $r$ , and choosing  $h = r$  appears to give the best results.

### 3.6 Maxvol Skeleton Approximation on Images

We will use the skeleton approximation to find a low rank approximation of the following  $128 \times 128$  penny picture with entries being integers between 0 and 255.

The first figure, fig. 6, shows the full unmodified penny picture. The next figure, fig. 7, on the left shows 50 rows where their intersecting  $50 \times 50$  submatrix is chosen with respect





Figure 6:  $128 \times 128$  penny picture. Each pixel is an integer from 0 to 255.

to the maxvol algorithm. We then show the skeleton approximation with respect to this submatrix. As can be seen on the right, this is a good approximation and has peak signal to noise ratio equal to 36.4188.

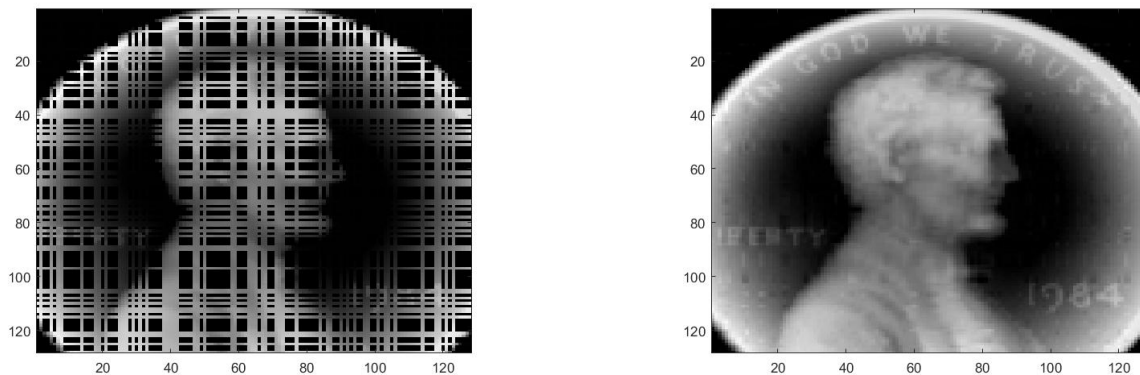


Figure 7: Left: 50 rows and columns chosen from the maxvol algorithm.  $\text{vol}(A) = 1.59 \times 10^{96}$ . Right: Rank 50 skeleton approximation of penny with respect to rows and columns chosen from the max-volume algorithm with peak signal to noise ratio equal to 36.4188.

For comparison, fig. 8 on the left shows 50 random rows and columns. The right shows the skeleton approximation of the penny picture with respect to a random  $50 \times 50$  submatrix. As can be seen, the approximation is not very good, and the picture has a peak signal to

noise ratio equal to 14.2255.

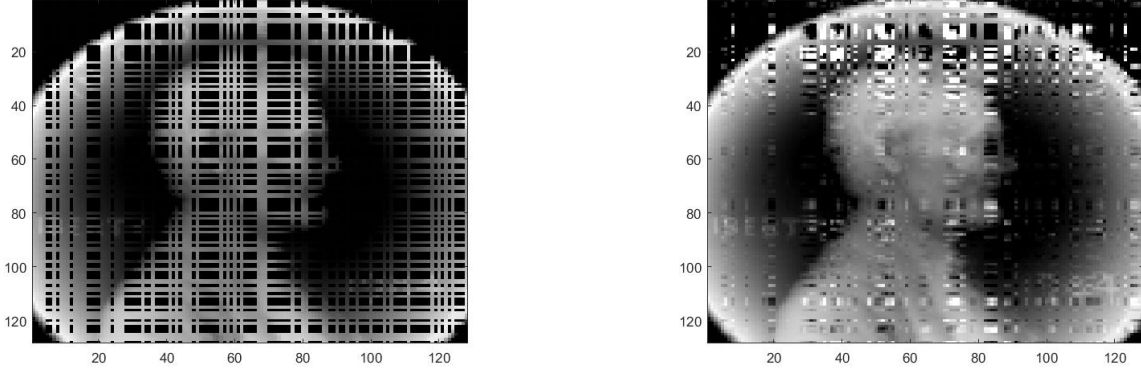


Figure 8: Left: 50 random rows and 50 random columns such that their intersecting submatrix  $A$  has volume  $\text{vol}(A) = 1.46 \times 10^{77}$ . Right: Rank 50 pseudo-skeleton approximation with rows and columns chosen randomly. Peak Signal to noise ratio equal to 14.2255.

In table 5, we study various pictures and find a suitable rank  $r$  such that the skeleton approximation with respect to a submatrix found with the maxvol algorithm has peak signal to noise ratio approximately equal to 32. Note that with the skeleton approximation, we are using the actual entries of the matrix to parameterize the low rank approximation.

### 3.7 Findvol Algorithm

Given an  $n \times m$  matrix  $M$ , we may modify the maxvol algorithm to search for a submatrix of some fixed determinant  $k$ . We will call this modified version of the maxvol algorithm the findvol algorithm. The find volume algorithm is identical to the maxvol algorithm, but instead of searching for  $\max_{(i,j)} \{|(MA^{-1})_{ij}|\}$  at each step, we search for  $\min_{(i,j)} \left\{ \left| (MA^{-1})_{ij} - \frac{k}{\det(A)} \right| \right\}$ . We repeat until there exists an index  $(i, j)$  such that  $(MA^{-1})_{ij} - \frac{k}{\det(A)} = 0$ , which means that  $\det(A) = k$ .

**Theorem 23.** *Let  $A_n$  be the  $n$ th iterate submatrix of the findvol algorithm. Then we have  $|\det(A_n) - k| \geq |\det(A_{n+1}) - k|$ . In other words, the sequence  $\{|\det(A_n) - k|\}$  is decreasing.*

Image	Resolution	r	PSNR
bang	512 × 512	350	32.23
barbara	512 × 512	260	32.31
bike	512 × 512	420	32.18
brain	512 × 512	115	33.53
clock	512 × 512	90	33.20
F16	512 × 512	230	32.22
finger	512 × 512	180	32.67
house	256 × 256	75	31.35
knee	512 × 512	105	32.21
monkey	512 × 512	440	32.38
mri	512 × 512	105	32.73
boat	256 × 256	130	32.28
pepper	512 × 512	375	32.61
saturn	512 × 512	75	33.45

Table 5: Uses a random initial  $r \times r$  matrix, then applies the maxvol algorithm. Uses resulting matrix for skeleton approximation. Shows the peak signal to noise ratio between skeleton approximation and original image.

One application of the findvol algorithm is the following. Given an integer matrix  $M$ , we may want to find a low-rank approximation of  $M$  whose entries remain as integers. Finding the closest low rank integer matrix is in general a very difficult problem. However, it may be possible to find a low-rank integer approximation that is not necessarily the closest.

The problem to solve is given an integer matrix  $M$ , we would like to find a rank  $r$  integer approximation  $M_r$  to  $M$ . One way to do this, is to find an  $r \times r$  submatrix  $A$  of  $M$  such that  $\det(A) = 1$  or  $-1$  using the findvol algorithm. If we can do so, then we may employ a rank  $r$  Skeleton approximation using rows and columns corresponding to  $A$ . Setting  $M_r = CA^{-1}R$ , we get that  $M_r$  is a rank  $r$  integer matrix with rows and columns identical to the corresponding rows and columns in  $M$ . This is true because the inverse  $A^{-1}$  has integer entries by Cramer’s rule, and so  $M_r = CA^{-1}R$  has integer entries as well.

**Theorem 24.** *Without loss of generality, suppose  $\|M(:, 1)\| \geq \|M(:, 2)\| \geq \dots \geq \|M(:, m)\|$ .*

Given  $M_r$  as above, we have

$$\|M - M_r\|_\infty \leq (r + 1)\sigma_{r+1}(M) \prod_{i=1}^r \|M(:, m)\|.$$

Alternatively, if  $|M_{ij}| \leq B$  for some  $B$ , then we have

$$\|M - M_r\|_\infty \leq (r + 1)r^{\frac{r}{2}}B^r\sigma_{r+1}(M)$$

*Proof.* Let  $m$  be largest volume over all  $r \times r$  submatrices of  $M$ . Then we have

$$\|M - M_r\|_\infty \leq m(r + 1)\sigma_{r+1}(M).$$

This follows from Theorem 2.2 in [21]. From Hadamard's inequality, we have

$$m \leq \prod_{i=1}^r \|M(:, m)\|,$$

so we get the first result. Similarly from Hadamard's inequality, we have  $m \leq r^{\frac{r}{2}}B^r$ , giving us the second result.  $\square$

In further generality, given a ring of integers  $R$  of an algebraic number field  $K$ , and a matrix  $M$  with entries in  $R$ , we could attempt to find a low rank approximation  $M_r$  to  $M$  with entries remaining in  $R$ . To do this we could use the findvol algorithm to search for a submatrix with unit determinant  $u$  for all units  $u$  in the unit group  $U(R)$ . Once a submatrix  $A$  with unit determinant is found, we calculate the skeleton approximation  $M_r$  with respect to that submatrix. The resulting matrix has entries remaining in  $R$ , since  $A^{-1}$  has entries in  $R$  by Cramer's rule.

### 3.8 A Graph Theoretic Reformulation of Dominant submatrices

In this section, we will re-formulate the concept of a dominant submatrix in graph theoretic terms. We use submatrices as nodes in the graph, and define an edge between two close submatrices.

First, let  $M \in M_{n,k}$ . We define the graph  $G_k(M)$  as the graph associated to  $M$  where the nodes of  $G_k(M)$  are the  $k \times k$  submatrices  $A_I$  of  $M$  up to permutation of rows. Two nodes  $A_I$  and  $A_{I'}$  are connected if we may obtain  $A_{I'}$  by swapping one row in  $A_I$  with a row in  $M$ .

More generally, for  $M \in M_{n \times m}$ , we let  $G_k(M) = (V_k, E_k)$  be the graph associated to  $M$  where the nodes of  $G_k(M)$  are the  $k \times k$  submatrices  $A_{I,J}$  of  $M$  up to permutations of rows and columns. There is an edge connecting two nodes of  $G_k(M)$  if we can obtain one associated submatrix from another by swapping one row.

We define a function  $f : V_k \rightarrow \mathbb{R}$ , such that  $f(A_{I,J}) = \text{vol}(A_{I,J})$ , and our goal is to find the global maximum of  $f$ , which corresponds to the largest determinant in modulus submatrix of  $M$ . Note that  $f$  is well defined, since swapping any two rows or columns of a matrix does not change its volume.

Dominant submatrices  $A_{\square}$  are the local maximums of  $f$ , since any single row or column swap would decrease the value of  $f$ .

The standard maxvol algorithm from [21] has the following interpretation. Given a nonsingular  $k \times k$  submatrix  $A_{I,J}$ , we can easily calculate which adjacent submatrix in  $G_k$  has the largest increase in  $f$ . We then replace  $A_{I,J}$  with this new submatrix, and repeat until we reach a local maximum of  $f$ , that is, a dominant submatrix.

For  $M \in \mathbb{R}^{m \times n}$ , we define the graph of  $k \times k$  submatrices similarly, where two nodes are connected if we can obtain one submatrix from another by permuting one row or column.

We now recall the definition of the *Johnson graph*  $J(m, k)$ , whose nodes are the subsets with  $k$  elements of a set with  $m$  elements total. There is an edge connecting two nodes if

and only if the intersection of the two subsets contains exactly  $k - 1$  elements.

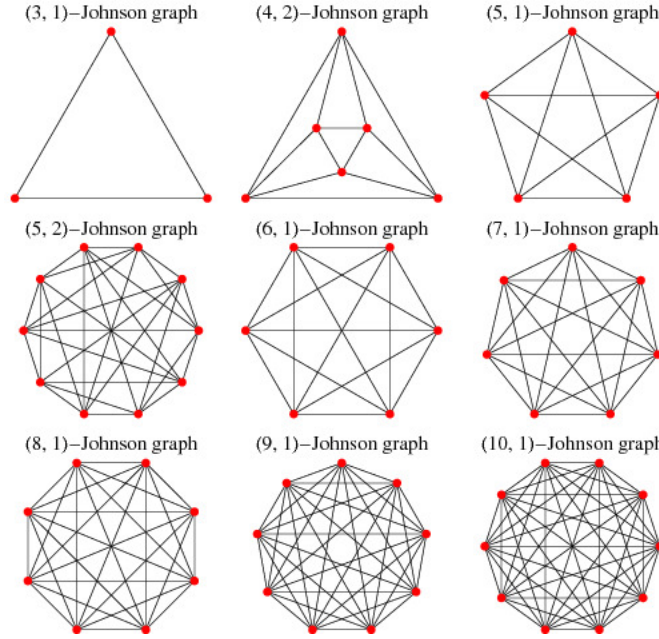


Figure 9: Examples of Johnson graphs from <https://mathworld.wolfram.com/JohnsonGraph.html>.

$J(m, k)$  is a  $k(m - k)$ -regular graph, meaning each node has exactly  $k(m - k)$  connecting edges.  $J(m, k)$  also has  $\binom{m}{k}$  nodes,  $\frac{(m-k)k}{2} \binom{m}{k}$  edges, and has diameter  $\min(k, m - k)$ .

**Theorem 25.** For  $M \in M_{m \times k}$ ,  $G_k(M)$  is isomorphic to the Johnson graph  $J(m, k)$ .

*Proof.* Consider the map  $\phi : G_k(M) \rightarrow J(m, k)$  that takes  $A_I \mapsto I$ . This map is injective since the node  $A_I$  is defined up to permutation of rows, and it is surjective since for every  $k$  element set  $I$ , there exists an  $A_I$ . Moreover, if two nodes  $A_I$  and  $A_{I'}$  are adjacent, then we may obtain  $A_{I'}$  by swapping one row in  $A_I$  with a row in  $M$ . Therefore,  $I$  and  $I'$  differ by exactly one index, and so  $|I \cap I'| = k - 1$ , so  $I$  and  $I'$  are adjacent in  $J(m, k)$ . Similarly, if  $I$  and  $I'$  are adjacent in  $J(m, k)$ , then we may obtain  $A_{I'}$  from  $A_I$  by permuting one row in  $M$ . □

We now introduce the notion of the graph Cartesian product. Given two graphs  $G$  and  $H$ , we define the *graph Cartesian product*  $G \square H$  as the graph with vertices  $V(G) \times V(H)$ ,

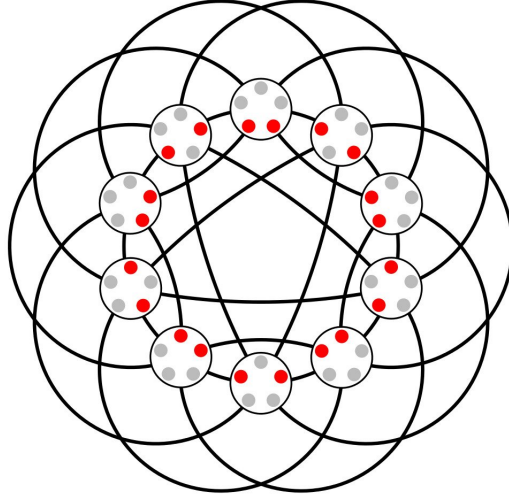


Figure 10: The Johnson graph  $J(5,2)$  from [https://en.wikipedia.org/wiki/Johnson\\_graph](https://en.wikipedia.org/wiki/Johnson_graph).

and an edge between two nodes  $(u, v)$  and  $(u', v')$  if and only if either  $u = u'$  and  $v$  is adjacent to  $v'$  in  $H$  or  $v = v'$  and  $u$  is adjacent to  $u'$  in  $G$ .

**Theorem 26.** *More generally, for  $M \in M_{n \times m}$ ,  $G_k(M)$  is isomorphic to the graph Cartesian product of Johnson graphs  $J(m, k) \square J(n, k)$*

*Proof.* Consider the map  $\phi : G_k(M) \rightarrow J(m, k) \square J(n, k)$  that maps  $A_{I,J} \mapsto (I, J)$ . This map is injective since the node  $A_{I,J}$  is defined up to permutation of rows and columns, and it is surjective since for every  $(I, J)$ , there exists a submatrix  $A_{I,J} \mapsto (I, J)$ .

Moreover, suppose  $A_{I,J}$  and  $A_{I',J'}$  are adjacent in  $G_k(M)$ . Then we may obtain  $A_{I',J'}$  from  $A_{I,J}$  by permuting either two rows or two columns in  $M$ , which means exactly one of the following must be true. Either  $I = I'$  and  $J$  and  $J'$  differ by one element, or  $J = J'$  and  $I$  and  $I'$  differ by one element. Therefore  $(I, J)$  and  $(I', J')$  are adjacent in  $J(m, k) \square J(n, k)$ . Similarly, if  $(I, J)$  and  $(I', J')$  are adjacent in  $J(m, k) \square J(n, k)$ , then we may obtain  $A_{I',J'}$  from  $A_{I,J}$  by permuting two rows or columns in  $M$ .

□

Suppose  $G_1$  has  $v_1$  vertices and  $e_1$  edges, and that  $G_2$  has  $v_2$  vertices and  $e_2$  edges. Then it

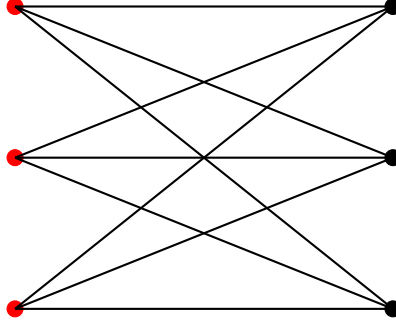


Figure 11: The red vertices are an example of an independent vertex set.

is known that the number of vertices in  $G_1 \square G_2$  is  $v_1 v_2$ , and the number of edges is  $v_1 e_2 + v_2 e_1$ .

Using this, we get that  $J(m, k) \square J(n, k)$  has  $\binom{n}{k} \binom{m}{k}$  nodes, and  $\frac{k(m+n-2k)}{2} \binom{m}{k} \binom{n}{k}$  edges.

### 3.9 Upper bounds on the number of dominant submatrices

In this section we will prove a new upper bound on the number of possible dominant submatrices of a matrix for almost every matrix. We start by introducing a few concepts from graph theory.

**Definition 9.** *An independent vertex set of a graph  $G$  is a subset of the vertices of  $G$  such that no two vertices are adjacent in  $G$ . See for example fig. 11. The independence number of a graph  $G$  is defined to be the maximum size over all independent sets in  $G$ , and is denoted  $\alpha(G)$ .*

We now introduce the following lemma.

**Lemma 5.** *No two dominant  $k \times k$  submatrices in  $M$  of differing volume are adjacent in  $G_k(M)$ .*

*Proof.* Let  $X$  and  $Y$  be  $k \times k$  submatrices of  $M$  which are adjacent in  $G_k(M)$ , and suppose that  $0 < \text{vol}(Y) < \text{vol}(X)$ . Then there exists a constant  $c > 1$  such that  $\text{vol}(X) = c \text{vol}(Y)$ . Since  $X$  and  $Y$  are adjacent in  $G_k(M)$ , then without loss of generality let us assume that



we may obtain  $X$  from  $Y$  by replacing the first  $k$  entries in the  $j$ th row of  $M$  with the first  $k$  entries in the  $i$ th row of  $M$ . Then  $c$  is the  $(i, j)$ th entry of  $M(:, 1 : k)Y^{-1}$ , and so  $Y$  is not dominant because  $c > 1$ .  $\square$

Next, we will provide an upper bound for the number of dominant submatrices for certain classes of matrices.

**Theorem 27.** *For  $M \in M_{m \times k}$ , the number of dominant submatrices of differing volume in  $M$  is at most  $\alpha(J(m, k))$ , the independence number of  $J(m, k)$ . More generally, for  $M \in M_{n \times m}$ , the number of dominant submatrices of differing volume is at most  $\alpha(J(n, k) \square J(m, k))$ .*

*Proof.* Since no two dominant submatrices of differing determinant in modulus are adjacent, the independence number, which is the maximum number of mutually non-adjacent nodes in our graph, provides an upper bound for the number of dominant submatrices of differing minors in modulus.  $\square$

In other words, if the independence number of Johnson graphs is not an upper bound for the number of dominant submatrices, then there must necessarily be at least two dominant submatrices with the same volume. We will now show that another set of matrices which contains this class of matrices has measure zero.

**Theorem 28.** *Let  $T$  be the set of matrices for which each matrix has at least two  $k \times k$  submatrices of equal volume. Then  $T$  has measure zero in  $M_{n \times m}$ .*

*Proof.* We will show that  $T$  is an algebraic variety. Then since  $T$  is an algebraic variety which is not all of  $M_{n \times m}$ , it must have Lebesgue measure zero.

Since there are only finitely many submatrices, without loss of generality we may assume that two specific submatrices  $A_1$  and  $A_2$  have equal volume. Moreover, since two submatrices of equal volume  $v$  may only have determinant  $v$  or  $-v$ , we may assume that the submatrices have equal determinant. That is,  $\det(A_1) - \det(A_2) = 0$ . This is a polynomial equation, and

so the set of matrices which have this property forms a hypersurface in  $M_{n \times m}$ , which has measure zero. More specifically,  $T$  is the finite union of hyper surfaces which are zero sets of polynomials of the form  $\det(A_i) - \det(A_j)$  or  $\det(A_i) + \det(A_j)$ .  $\square$

**Corollary 1.** *The set of matrices  $S$  which have two dominant, adjacent  $k \times k$  submatrices of equal volume has measure zero in  $M_{n \times m}$ .*

*Proof.* From theorem 28, the set of matrices  $T$  which have at least two  $k \times k$  submatrices of equal volume has measure zero. Since  $S \subset T$ , then the measure of  $S$  must also be equal to zero.  $\square$

**Theorem 29.** *For almost every matrix  $M \in M_{n \times k}$ , the number of dominant  $k \times k$  submatrices is at most  $\alpha(J(m, k))$ . Moreover, for almost every  $M \in M_{n \times m}$ , the number of dominant  $k \times k$  submatrices is at most  $\alpha(J(n, k) \square J(m, k))$ .*

*Proof.* This is true for matrices which have no two dominant submatrices of equal determinant from theorem 27. Moreover, the complement of this set has measure zero from corollary 1. Therefore, we have the desired result.  $\square$

Note that it is not necessarily likely that a random  $n \times m$  matrix will have  $\alpha(J(n, k) \square J(m, k))$  number of  $k \times k$  dominant submatrices. For empirical results on the average number of dominant submatrices of a random matrix, see section 3.11.

For an upper bound of the number of dominant  $k \times k$  submatrices of an  $n \times m$  matrix in terms of the independence numbers of Johnson graphs, we have the following result.

**Theorem 30.** *For  $A \in M_{n \times m}$ , the number of dominant submatrices of differing determinant in modulus in  $A$  is at most  $\min\{\alpha(J(m, k)) \binom{n}{k}, \alpha(J(n, k)) \binom{m}{k}\}$ .*

*Proof.* First, note that from [31], we have that for graphs  $G$  and  $H$ , the independence number satisfies  $\alpha(G \square H) \leq \min\{\alpha(G) |V(H)|, \alpha(H) |V(G)|\}$ . Using theorem 27, we have that the number of dominant submatrices is bounded above by  $\alpha(J(m, k) \square J(n, k)) \leq \min\{\alpha(J(m, k)) \binom{n}{k}, \alpha(J(n, k)) \binom{m}{k}\}$ .  $\square$

### 3.10 Sharp Inequality Examples

Given  $M \in M_{n \times k}$  the independence number of the Johnson graph  $\alpha(J(n, k))$  provides an upper bound for number of  $k \times k$  dominant submatrices for almost every  $M$ . More generally, for  $M \in M_{n \times m}$ ,  $\alpha(J(n, k) \square J(m, k))$  is an upper bound for the number of  $k \times k$  dominant submatrices for almost all  $M$ .

In general,  $\alpha(J(n, k))$  is not known. However, it is known in a few cases. For example

1.  $\alpha(J(n, 1)) = 1$
2.  $\alpha(J(n, 1) \square J(m, 1)) = \min\{m, n\}$
3.  $\alpha(J(n, 2)) = \lfloor \frac{n}{2} \rfloor$

Moreover, since  $J(n, k) \cong J(n, n - k)$ , we also have:

4.  $\alpha(J(n, n - 1)) = 1$
5.  $\alpha(J(n, n - 1) \square J(n, n - 1)) = n$

I will provide matrices for which the upper bound on the number of submatrices is reached such that no two dominant submatrices are adjacent.

The first case is trivial, any non-zero vector has at least one element of maximal modulus, and so is a dominant  $1 \times 1$  submatrix.

In the second case, with out loss of generality suppose  $m \leq n$ . Consider the matrix:

$$\begin{bmatrix} a_1 & 0 & 0 & \cdots & 0 \\ 0 & a_2 & 0 & \cdots & 0 \\ 0 & 0 & a_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \cdots & a_m \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

Then assuming  $a_k \neq 0$  for  $k = 1, \dots, m$ , each submatrix  $[a_k]$  is a  $1 \times 1$  dominant submatrix, of which there are  $m$  total, and no two are adjacent.

In the third case, first let us assume that  $n$  is even. Let  $A_k = \begin{bmatrix} \cos(\theta_k) & -\sin(\theta_k) \\ \sin(\theta_k) & \cos(\theta_k) \end{bmatrix}$ , where  $\theta_k = \frac{\pi(k-1)}{n}$  for  $k = 1, \dots, \frac{n}{2}$ . Then we will show that each of  $A_k$  are dominant submatrices of the matrix  $\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_{\frac{n}{2}} \end{bmatrix}$ .

Moreover, we will show that  $A_k$  are not adjacent to any other dominant submatrix. First, note that  $A_k^{-1} = \begin{bmatrix} \cos(\theta_k) & \sin(\theta_k) \\ -\sin(\theta_k) & \cos(\theta_k) \end{bmatrix}$ , and so

$$\begin{aligned} A_j A_k^{-1} &= \begin{bmatrix} \cos(\theta_j) \cos(\theta_k) + \sin(\theta_j) \sin(\theta_k) & \cos(\theta_j) \sin(\theta_k) - \sin(\theta_j) \cos(\theta_k) \\ \sin(\theta_j) \cos(\theta_k) - \cos(\theta_j) \sin(\theta_k) & \cos(\theta_j) \cos(\theta_k) + \sin(\theta_j) \sin(\theta_k) \end{bmatrix} \\ &= \begin{bmatrix} \cos(\theta_j - \theta_k) & -\sin(\theta_j - \theta_k) \\ \sin(\theta_j - \theta_k) & \cos(\theta_j - \theta_k) \end{bmatrix} \\ &= \begin{bmatrix} \cos(\frac{\pi(j-k)}{n}) & -\sin(\frac{\pi(j-k)}{n}) \\ \sin(\frac{\pi(j-k)}{n}) & \cos(\frac{\pi(j-k)}{n}) \end{bmatrix}. \end{aligned}$$

It suffices to show that each entry of  $A_j A_k^{-1}$  is less than 1 in modulus unless  $j = k$ . Note that  $-\frac{\pi}{2} < \theta_j - \theta_k < \frac{\pi}{2}$ , and so  $|\sin(\theta_j - \theta_k)| < 1$ . Moreover,  $\theta_j - \theta_k = 0$  if and only if  $j = k$ , and so  $|\cos(\theta_j - \theta_k)| = 1$  if and only if  $j = k$ .

When  $n$  is odd the matrix  $\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_{\lfloor \frac{n}{2} \rfloor} \\ 0 \end{bmatrix}$  with one row of zeros in the last row similarly has  $\lfloor \frac{n}{2} \rfloor$  dominant submatrices, no two of which are adjacent.

The fourth case is also trivial, any  $n \times (n - 1)$  full rank matrix has a  $(n - 1) \times (n - 1)$  submatrix of maximal volume and is dominant.

For the fifth case, the diagonal matrix

$$\begin{bmatrix} a_1 & 0 & 0 & \cdots & 0 \\ 0 & a_2 & 0 & \cdots & 0 \\ 0 & 0 & a_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_n \end{bmatrix}$$

with all  $a_k \neq 0$  also provides an example where there are exactly  $n$  non-adjacent dominant  $(n - 1) \times (n - 1)$  submatrices. In particular, the cofactor  $C_{ij} = 0$  if  $i \neq j$ , and  $C_{ii} \neq 0$ . Therefore, the minors  $A_{ii}$  are each dominant submatrices which are not adjacent.

### 3.11 Numerical Experiments Approximating the Expected Value of the Number of Dominant submatrices

Although  $\alpha(J(n, k) \square J(m, k))$  provides an upper bound for the total number of possible dominant  $k \times k$  submatrices of an  $n \times m$  matrix for almost every matrix, in general there will be fewer dominant submatrices. For example, consider the  $1 \times 1$  submatrices of  $2 \times 2$  matrices of the form  $M = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ . Assume that no two entries are equal. Also without loss of generality by permuting rows and columns, assume that the largest entry in modulus is

$a_{11}$ . Then we have the following six possibilities for the order of the entries  $a_{ij}$

1.  $|a_{11}| > |a_{12}| > |a_{21}| > |a_{22}|$
2.  $|a_{11}| > |a_{21}| > |a_{12}| > |a_{22}|$
3.  $|a_{11}| > |a_{12}| > |a_{22}| > |a_{21}|$
4.  $|a_{11}| > |a_{21}| > |a_{22}| > |a_{12}|$
5.  $|a_{11}| > |a_{22}| > |a_{12}| > |a_{21}|$
6.  $|a_{11}| > |a_{22}| > |a_{21}| > |a_{12}|$

In each case,  $a_{11}$  is a dominant submatrix since  $\left| \frac{a_{ij}}{a_{11}} \right| \leq 1$  for all  $(i, j)$ . However, in cases 5 and 6,  $a_{22}$  is also a dominant submatrix since  $\left| \frac{a_{12}}{a_{22}} \right| < 1$  and  $\left| \frac{a_{21}}{a_{22}} \right| < 1$ . These are the only possibilities for a random  $2 \times 2$  matrix to have more than one dominant  $1 \times 1$  submatrix. So if we assume that each of cases 1 through 6 happens with equal probability, then the chance of there being one  $1 \times 1$  dominant submatrix in a random  $2 \times 2$  matrix happens with probability  $2/3$ , and the chance of there being two dominant  $1 \times 1$  submatrix of a random  $2 \times 2$  submatrix happens with probability  $1/3$ . Therefore, the expected value of the number of dominant  $1 \times 1$  submatrices of a random  $2 \times 2$  matrix is  $4/3$ .

The expected value of the number of  $k \times k$  dominant submatrices of a random matrix can be approximated with a Monte Carlo method. One can generate  $a$  random matrices  $A_i$ ,  $1 \leq i \leq a$ . Then, choose  $B_j$ ,  $1 \leq j \leq b$ , a random  $k \times k$  submatrix of  $A_i$ , and determine whether or not  $B_j$  is dominant. We repeated this  $b$  times, and let  $d_i$  be the total number of submatrices tested which were dominant. Then since  $\binom{n}{k} \binom{m}{k}$  is the total number of submatrices of  $A_i$ , we have that  $\binom{n}{k} \binom{m}{k} \frac{d_i}{b}$  is approximately the total number of dominant submatrices of  $M$ . We then average this over all  $i$ , so the  $\binom{n}{k} \binom{m}{k} \frac{1}{ab} \sum_{i=1}^a d_i$  is approximately the expected value of the total number of dominant submatrices of a random  $n \times m$  matrix.

$k$	$n$							
	3	4	5	6	7	8	9	10
2	1.0005	1.0056	1.0177	1.0248	1.0321	1.0379	1.0585	1.0697
3	1.0000	0.9998	1.0232	1.0532	1.0947	1.1224	1.2049	1.2299
4		1.0000	0.9996	1.0373	1.0986	1.1810	1.2581	1.3138
5			1.0000	0.9997	1.0527	1.1127	1.2053	1.3202
6				1.0000	1.0002	0.9992	1.1360	1.2909
7					1.0000	1.0008	1.0669	1.1741
8						1.0000	1.0002	1.0809
9							1.0000	0.9989
10								1.0000

Table 6: Experimental expected number of  $k \times k$  dominant submatrices of a random  $n \times k$  matrix with entries chosen uniformly at random on the interval  $[0, 1]$ . 10000 randomly chosen submatrices of 1000 random matrices.

For  $k \times k$  matrices, there is exactly one  $k \times k$  submatrix, which means there is one dominant submatrix. For  $(k + 1) \times k$  matrices  $M$ , any submatrix may be permuted to any other submatrix by swapping two rows of  $M$ . In other words,  $G_k(M)$  is a complete graph. Therefore, the only way for there to be more than one dominant submatrix would be if there were two submatrices with the same volume, which happens with probability zero. Therefore with probability one, there will be exactly one dominant submatrix.

For a specific example, consider the  $128 \times 128$  penny picture  $P$ . We will approximate the number of  $2 \times 2$  dominant submatrices by sampling  $N$  random matrices  $A$  with some index  $(I, J)$  and determining whether or not those matrices are dominant by calculating  $\|P(:, J)A^{-1}\|_\infty$  and  $\|A^{-1}P(I, :)\|_\infty$ . If we let  $d$  be the total number of dominant submatrices found by this method, then  $\frac{d}{N}$  should approximate the ratio of the total number of dominant submatrices to the total number of  $2 \times 2$  submatrices, which is equal to  $\binom{128}{2}^2$ . Therefore, the total number of dominant submatrices is approximated by  $\binom{128}{2}^2 \frac{d}{N}$ .

We choose  $N = 2000000$ , and get  $d = 17$ . Therefore, there are approximately 562 dominant  $2 \times 2$  submatrices out of 66064384 total  $2 \times 2$  submatrices.

### 3.12 An Upper Bound on the Independence Number of Johnson Graphs

There is lots of literature on the independence number of Johnson graphs, see for example, [27]. In general,  $\alpha(J(n, k))$  is not known. It is known however, that the independence number of Johnson graphs is equal to the size of the largest constant weight code of word length  $n$ , weight  $k$ , and distance at least 4 [28]. In other words, it is equal to the maximum number of binary vectors of length  $n$  having  $k$  ones and  $n - k$  zeros such that any two vectors differ in at least 4 places.

We do have an iterative upper bound on the independence number of Johnson graphs called the Johnson bound. The Johnson bound states that  $\alpha(J(n, k)) \leq \frac{n}{k} \alpha(J(n-1, k-1))$ . Therefore by induction, we have that  $\alpha(J(n, k)) \leq \frac{n!}{(n-k+1)!k!}$ . Since the total number of vertices of  $J(n, k)$  is  $\binom{n}{k}$ , we have that the independence ratio, the ratio of the independence number to the total number of vertices, satisfies the inequality

$$\frac{\alpha(J(n, k))}{\binom{n}{k}} \leq \frac{1}{1 + n - k}.$$

For an alternative proof of this upper bound in terms of the eigenvalues of the adjacency matrix of  $J(n, k)$ , we introduce the following theorem from [29].

**Theorem 31.** *For any connected regular graph  $G$ , with  $v$  vertices, vertex degree  $d$ , and smallest eigenvalue of the adjacency matrix  $s$ , the independence number  $\alpha(G)$  satisfies the inequality*

$$\frac{\alpha(G)}{v} \leq \frac{1}{1 - \frac{d}{s}}$$

Suppose  $G$  is the Johnson graph  $J(n, k)$ . Since  $J(n, k) \cong J(n, n - k)$ , then without loss of generality, suppose  $k \leq \frac{n}{2}$ . Then  $G$  is a regular graph with vertex degree  $d = k(n - k)$ . Moreover, the number of vertices  $v = \binom{n}{k}$ . Also the eigenvalues of the adjacency matrix of  $J(n, k)$  are known, they are  $\lambda_i = (k - i)(n - k - i) - i$  for  $i = 0, \dots, k$ , which has smallest



eigenvalue  $\lambda_k = -k$ .

Therefore by theorem 31, we have the inequality

$$\frac{\alpha(G)}{\binom{n}{k}} \leq \frac{1}{1 - \frac{k(n-k)}{-k}} = \frac{1}{1 + n - k}$$

Now suppose  $G$  is the Cartesian product of Johnson graphs  $J(n, k) \square J(m, l)$ . Again suppose  $k \leq \frac{n}{2}$  and  $l \leq \frac{m}{2}$ . Note that the Cartesian product of connected regular graphs are connected regular, so we may use theorem 31. In this case we have  $d = k(n - k) + l(m - l)$  and  $v = \binom{n}{k} \binom{m}{l}$ . For the eigenvalues of  $G$ , we need to following theorem from [30].

**Theorem 32.** *Let  $G$  and  $H$  be graphs with adjacency eigenvalues  $\{\lambda_i\}$  and  $\{\mu_j\}$  respectively. Then the eigenvalues of  $G \square H$  are of the form  $\lambda_i + \mu_j$  for some  $i$  and  $j$ .*

Then if  $\lambda_i = (k - i)(n - k - i) - i$  and  $\mu_j = (l - j)(m - k - j) - j$  are the eigenvalues of  $J(n, k)$  and  $J(m, l)$  respectively,  $G = J(n, k) \square J(m, l)$  has eigenvalues of the form  $\lambda_i + \mu_j = (k - i)(n - k - i) - i + (l - j)(m - k - j) - j$  for  $i = 0, \dots, k$ , and  $j = 0, \dots, l$ . Therefore the smallest graph eigenvalue is

$$\begin{aligned} s &= \min_{i,j} \{\lambda_i + \mu_j\} \\ &= \min_i \{\lambda_i\} + \min_j \{\mu_j\} \\ &= \lambda_k + \mu_l \\ &= -k - l \end{aligned}$$

So we have the inequality

$$\frac{\alpha(G)}{\binom{n}{k} \binom{m}{l}} \leq \frac{1}{1 + \frac{k(n-k) + l(m-l)}{k+l}}.$$

## 4 A Schur Complement Based Gradient Descent Method for Matrix Completion

In this section we will introduce a new gradient descent method for low-rank matrix completion which we will call the Schur gradient descent. One nice property of the Schur gradient descent is that the gradient is a rational function of the actual entries of the matrix. This means that we do not need to compute a singular value decomposition at each step.

### 4.1 Unique Matrix Completion Example

We may use the Schur complement to find a unique rank  $r$  completion for a certain class of partially known matrices. The following pattern of  $M_\Omega$  is particularly useful. Assume that we may permute a given partially known matrix  $M_\Omega$  into the following  $M_\Omega = \begin{bmatrix} A & B \\ C & \square \end{bmatrix}$ , where  $A$  is  $r \times r$ ,  $B$  is  $r \times (m - r)$ , and  $C$  is  $(n - r) \times k$ , all full of known entries, and  $\square$  is size  $(n - r) \times (m - r)$  and contains only unknowns.

In this case, we will have  $2nr - r^2$  known entries, and we have the following result.

**Theorem 33.** *If  $M_\Omega$  can be represented up to permutation of rows and columns in block form by  $M_\Omega = \begin{bmatrix} A & B \\ C & \square \end{bmatrix}$ , where  $A$  is an  $r \times r$  invertible submatrix, then  $M_\Omega$  has a unique rank  $r$  completion where  $\square = CA^{-1}B$ .*

*Proof.* Let  $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$  be a completion of  $M_\Omega$ . Then from lemma 4, we have that  $\text{rank}(M) = r$  if and only if  $D = CA^{-1}B$ . □

### 4.2 Matrix Completion With a Known Invertible Submatrix

We generalize the previous case where  $A, B$ , and  $C$  are known and  $D$  is unknown to the case where  $A$  is fully known, and  $B, C$ , and  $D$  are partially known. Recall that since  $\overline{\mathcal{M}}_r$  is the zero set of all  $(r + 1) \times (r + 1)$  minors, and  $\mathcal{A}_\Omega$  is the zero set of equations of the form

$x_{ij} - M_{ij}$  for all known  $M_{ij}$ , then their intersection  $\mathcal{A}_\Omega \cap \overline{\mathcal{M}}_r$ , which is the set of all possible rank  $r$  completions of  $M_\Omega$ , is the simultaneous zero set of all  $(r+1) \times (r+1)$  minors and equations of the form  $x_{ij} - M_{ij}$  for all known  $M_{ij}$ . However, these equations are redundant in the sense that we do not need them all to describe the variety  $\mathcal{A}_\Omega \cap \overline{\mathcal{M}}_r$ . In the following theorem, we will provide a smaller set of equations which have zero set equal to  $\mathcal{A}_\Omega \cap \overline{\mathcal{M}}_r$ .

In particular, we will examine the case when  $M_\Omega$  has the form  $M_\Omega = \begin{bmatrix} A & B_\Omega \\ C_\Omega & D_\Omega \end{bmatrix}$ , where  $A$  is some  $k \times k$  fully known submatrix, and  $B_\Omega, C_\Omega$ , and  $D_\Omega$  are partially unknown. Note that  $M_\Omega$  may always be permuted to this form. As a worst case scenario, take  $k = 1$ , and  $A$  to be a  $1 \times 1$  submatrix.

**Theorem 34.** *Suppose that  $M_\Omega$  can be permuted to the form  $M_\Omega = \begin{bmatrix} A & B_\Omega \\ C_\Omega & D_\Omega \end{bmatrix}$ , where  $A$  is a known  $k \times k$  invertible submatrix for some  $k$ , and  $B_\Omega, C_\Omega$ , and  $D_\Omega$  are known entries. Then  $\mathcal{A}_\Omega \cap \overline{\mathcal{M}}_r$  is the zero set of all  $(r+1) \times (r+1)$  minors containing  $A$ , along with the equations  $x_{ij} - M_{ij}$  for all known  $M_{ij}$ . In other words,  $\mathcal{A}_\Omega \cap \overline{\mathcal{M}}_r$  is the variety defined by the equations  $\det(D' - C'A^{-1}B') = 0$  for all  $(r-k+1) \times (r-k+1)$  submatrices  $D'$  of  $D_\Omega$  with corresponding  $C'$  and  $B'$ , along with the equations  $x_{ij} - M_{ij} = 0$  for all known  $M_{ij}$ . In particular, when  $k = r$ , these determinants simplify to the form  $d_{ij} - c_i^\top A^{-1}b_j = 0$ , for all  $d_{ij}$  known and unknown. Where  $b_j$  is the  $j^{\text{th}}$  column of  $B$  and  $c_i^\top$  is the  $i^{\text{th}}$  row of  $C$ .*

*Proof.* Since  $M_\Omega$  contains a known  $k \times k$  rank  $k$  submatrix  $A$ , from theorem 13, we have that  $\mathcal{A}_\Omega \cap \overline{\mathcal{M}}_r = \mathcal{A}_\Omega \cap V$ , since  $\mathcal{A}_\Omega \cap W$  is empty. Therefore,  $\mathcal{A}_\Omega \cap \overline{\mathcal{M}}_r$  is equal to the zero set of all  $(r+1) \times (r+1)$  minors which contain  $A$ . That is, equations of the form  $\begin{vmatrix} A & B' \\ C' & D' \end{vmatrix} = 0$ , where  $D'$  is a  $(r-k+1) \times (r-k+1)$  submatrix of  $D_\Omega$ . By lemma 3, we can express these equations as  $\det(D' - C'A^{-1}B') = 0$ , or  $d_{ij} - c_i^\top A^{-1}b_j = 0$  when  $k = r$ .  $\square$

One strategy is to first recover the unknown elements in  $B_\Omega$  and  $C_\Omega$  from the known elements in  $D_\Omega$ . If we can do these to find all unknowns in  $B_\Omega$  and  $C_\Omega$ , then we can recover the unknown elements of  $D_\Omega$  from theorem 33. For this strategy, the number of known entries in  $D_\Omega$  should be more than the number of unknowns in  $B_\Omega$  and  $C_\Omega$ .

Let us first fix  $k = r$ . To find unknown entries in  $B_\Omega$ ,  $C_\Omega$ , and  $D_\Omega$ , we will cast the matrix completion in a minimization of the norm of the schur complement.

$$\min \frac{1}{2} \|D - CA^{-1}B\|^2 = \min \frac{1}{2} \|S_A\|^2 \quad (3)$$

Where  $A$  is a fixed fully known  $r \times r$  submatrix of  $M_\Omega$ , and  $B, C$ , and  $D$  are corresponding submatrices. We subject this minimization to the constraint that  $P_\Omega(B) = B_\Omega$ ,  $P_\Omega(C) = C_\Omega$ , and  $P_\Omega(D) = D_\Omega$ .

To solve the minimization, we can use the gradient descent method. Let us compute the gradient of the minimizing functional above.

**Theorem 35.** *For some  $d_{ij}$  known, and some  $b_{kj}$  and  $c_{ik}$  unknown, we have the derivatives*

$$\frac{\partial}{\partial b_{kj}} (c_i^\top A^{-1} b_j - d_{ij}) = c_i^\top A^{-1} e_k,$$

and similarly

$$\frac{\partial}{\partial c_{ik}} (c_i^\top A^{-1} b_j - d_{ij}) = e_k^\top A^{-1} b_j,$$

where  $e_k$  is the  $k$ th standard basis vector.

*Proof.* Consider some variable  $x$ . Then by the product rule, we have

$$\frac{\partial}{\partial x} (c_i^\top A^{-1} b_j - d_{ij}) = c_i^\top A^{-1} \frac{\partial b_j}{\partial x} + \frac{\partial c_i^\top}{\partial x} A^{-1} b_j.$$

If  $x = b_{kj}$ , then  $\frac{\partial b_j}{\partial x} = e_k$  and  $\frac{\partial c_i^\top}{\partial x} = 0$ . Similarly, if  $x = c_{ik}$ , then  $\frac{\partial b_j}{\partial x} = 0$  and  $\frac{\partial c_i^\top}{\partial x} = e_k^\top$ .  $\square$

This leads to a matrix completion algorithm for  $M_\Omega$ . First, we permute  $M_\Omega$  to have the structure in  $M_\Omega = \begin{bmatrix} A & B_\Omega \\ C_\Omega & D_\Omega \end{bmatrix}$  where  $A$  is of size  $r \times r$  and is fully known, and  $B_\Omega, C_\Omega$ , and  $D_\Omega$  are partial known and unknown. For convenience, Then the gradient descent method is as follows. Assume the  $k$ th iterate  $B_\Omega^{(k)}$ ,  $C_\Omega^{(k)}$ , and  $D_\Omega^{(k)}$  have been computed. We compute the

$(k + 1)$  iterate

$$\begin{bmatrix} B_{\Omega}^{(k+1)} \\ C_{\Omega}^{(k+1)} \\ D_{\Omega}^{(k+1)} \end{bmatrix} = \begin{bmatrix} B_{\Omega}^{(k)} \\ C_{\Omega}^{(k)} \\ D_{\Omega}^{(k)} \end{bmatrix} - h \nabla f(B_{\Omega}^{(k)}, C_{\Omega}^{(k)}, D_{\Omega}^{(k)}), \quad (4)$$

where our loss function

$$f = \frac{1}{2} \|S_A\|^2 = \sum_{i,j} (c_i^{\top} A^{-1} b_j - d_{ij})^2 / 2$$

with variables being the unknown entries of  $B_{\Omega}$ ,  $C_{\Omega}$ , and  $D_{\Omega}$  and  $h > 0$  is a step size. We run gradient descent until the  $(k + 1)$ th iterative matrix  $M_{\Omega}^{(k+1)} = \begin{bmatrix} A & B_{\Omega}^{(k+1)} \\ C_{\Omega}^{(k+1)} & D_{\Omega}^{(k+1)} \end{bmatrix}$  is of rank  $r$ , or until the  $r + 1$ st singular value of the  $(k + 1)$ st iterate is sufficiently small. The formula for  $\nabla F$  is gives

$$\begin{aligned} \frac{\partial f}{\partial B} &= A^{-\top} C^{\top} (C A^{-1} B - D) \\ \frac{\partial f}{\partial C} &= (C A^{-1} B - D) B^{\top} A^{-\top} \\ \frac{\partial f}{\partial D} &= D - C A^{-1} B \end{aligned}$$

where  $\frac{\partial f}{\partial X}$  is a matrix such that  $(\frac{\partial f}{\partial X})_{ij} = \frac{\partial f}{\partial x_{ji}}$ .

For a more general analysis when  $A$  is not necessarily  $r \times r$ , we will analyze the case where,  $M_{\Omega}$  is of the form  $M_{\Omega} = \begin{bmatrix} A & B_{\Omega} \\ C_{\Omega} & D_{\Omega} \end{bmatrix}$ , where  $A$  is a known  $k \times k$ ,  $k \leq r$ , invertible known submatrix, and  $B_{\Omega}$ ,  $C_{\Omega}$ , and  $D_{\Omega}$  are partially known with  $k < r$ . By theorem 13,  $\mathcal{A}_{\Omega} \cap \overline{\mathcal{M}}_r = \mathcal{A}_{\Omega} \cap V$ , where  $V$  is the vanishing set of all  $(r + 1) \times (r + 1)$  minors containing  $A$ . By lemma 3, these equations may be expressed as  $\det(D' - C' A^{-1} B') = 0$  for all  $(r + 1 - k) \times (r + 1 - k)$  submatrices  $D'$  of  $D_{\Omega}$  with corresponding  $B'$  and  $C'$ . Similarly to before, we cast the matrix completion problem as a minimization problem of the form

$$\begin{aligned} \min \frac{1}{2} \sum_{D'} \det(D' - C' A^{-1} B')^2 \\ \text{s.t. } P_\Omega(B) = B_\Omega, P_\Omega(C) = C_\Omega, \text{ and } P_\Omega(D) = D_\Omega. \end{aligned}$$

To compute the derivative of this function, recall *Jacobi's formula*:

$$\frac{d}{dt} \det(A(t)) = \text{tr}(\text{adj}(A(t)) \frac{dA(t)}{dt}).$$

Where  $\text{adj}(A)$  is the adjugate of the matrix  $A$ . We will use this formula to calculate the derivative of the above equation.

**Theorem 36.** *We have the following derivative information:*

$$\begin{aligned} \frac{\partial}{\partial b_{ij}} \det(D' - C' A^{-1} B') &= \text{tr}(\text{adj}(D' - C' A^{-1} B') (-C' A^{-1} \frac{\partial B'}{\partial b_{ij}})) \\ \frac{\partial}{\partial c_{ij}} \det(D' - C' A^{-1} B') &= \text{tr}(\text{adj}(D' - C' A^{-1} B') (\frac{\partial C'}{\partial c_{ij}} A^{-1} B')) \\ \frac{\partial}{\partial d_{ij}} \det(D' - C' A^{-1} B') &= \text{adj}^\top(D' - C' A^{-1} B')_{ij} \end{aligned}$$

*Proof.* Consider some variable  $x$ . Then by *Jacobi's formula* and the product rule, we have

$$\frac{\partial}{\partial x} \det(D' - C' A^{-1} B') = \text{tr}(\text{adj}(D' - C' A^{-1} B') (\frac{\partial D'}{\partial x} - \frac{\partial C'}{\partial x} A^{-1} B' - C' A^{-1} \frac{\partial B'}{\partial x})).$$

□

Once again, we may employ a gradient descent method to solve this minimization using the above derivative information.

In summary, the minimization in eq. (4) is not convex. A simple gradient descent method starting with any initial matrix will not converge to a minimum in general. A good initial guess is important for convergence.

### 4.3 General Matrix Completion With Schur Gradient Descent

We now consider the most general case of matrix completion where the unknown entries are distributed arbitrarily and do not necessarily contain any specific structure such as a known invertible submatrix.

Let  $X = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ , Where the rows and columns of  $X$  are permuted such that a fixed submatrix  $A$  is in the top left. We now analyze the case when  $A_\Omega$ ,  $r \times r$  is not completely known. For the loss function

$$f(X) = \frac{1}{2} \|D - CA^{-1}B\|^2 = \frac{1}{2} \|S_A\|^2$$

we once again minimize  $f$  using gradient descent. We have the following gradient information

$$\begin{aligned} \frac{\partial f}{\partial A} &= -A^{-\top} C^\top (CA^{-1}B - D) B^\top A^{-\top} = A^{-\top} C^\top S_A B^\top A^{-\top} \\ \frac{\partial f}{\partial B} &= A^{-\top} C^\top (CA^{-1}B - D) = A^{-\top} C^\top S_A \\ \frac{\partial f}{\partial C} &= (CA^{-1}B - D) B^\top A^{-\top} = S_A B^\top A^{-\top} \\ \frac{\partial f}{\partial D} &= D - CA^{-1}B = S_A \end{aligned}$$

where  $\frac{\partial f}{\partial X}$  is the matrix such that  $(\frac{\partial f}{\partial X})_{ij} = \frac{\partial f}{\partial x_{ji}}$ .

*Proof.* To calculate  $\frac{\partial f}{\partial A}$ , we let  $g(A) = \|D - CAB\|_F^2$ , and let  $h(A) = A^{-1}$ . Then  $f = g \circ h$ . Define the dot-product  $(X, Y) = \text{tr}(X^\top Y)$ . Then we have the identities  $df = (C^\top(CAB -$

$D)B^\top, dA)$  and  $dh = A^{-1}(dA)A^{-1}$ . Using the chain rule, we have

$$df = (C^\top(CA^{-1}B - D)B^\top, A^{-1}(dA)A^{-1}) = (A^{-\top}C^\top(CA^{-1}B - D)B^\top A^{-\top}, dA)$$

which implies that  $\frac{df}{dA} = A^{-\top}C^\top(CA^{-1}B - D)B^\top A^{-\top}$ . [36] □

Putting these together, we have we have the following formula for the gradient.

$$\begin{aligned} \nabla f &= P_{\Omega^c} \left( \begin{bmatrix} \left(\frac{\partial f}{\partial A}\right)^\top & \left(\frac{\partial f}{\partial B}\right)^\top \\ \left(\frac{\partial f}{\partial C}\right)^\top & \left(\frac{\partial f}{\partial D}\right)^\top \end{bmatrix} \right) \\ &= P_{\Omega^c} \left( \begin{bmatrix} A^{-1}BS_A^\top CA^{-1} & S_A^\top CA^{-1} \\ A^{-1}BS_A^\top & S_A^\top \end{bmatrix} \right) \end{aligned}$$

Where  $P_{\Omega^c}$  sets elements in the known indices in  $\Omega$  equal to zero. The reason why we include this term is so that we do not change the known entries of our matrix. Note that entries in the gradient is a rational function of our actual matrix elements.

By setting the gradient equal to zero, the formula for the gradient implies that for any critical point, we must have  $(S_A)_{ij} = 0$  for all  $(i, j) \in \Omega_D^c$ .

In general it is difficult to prove the convergence of gradient descent since our function  $f$  is non-convex.

## 4.4 Small Numerical Examples of Schur Gradient Descent

Consider the rank 1 matrix  $M = \begin{bmatrix} 6 & 3 \\ 2 & 1 \end{bmatrix}$ . We will test the gradient descent method on the cases when

$$\Omega = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \text{ and } \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix},$$



with initial guess  $M_0$  containing a zero in the unknown entry. The  $y$ -axis is a log-scale of  $\sigma_2$ , the second singular value of  $M_n$ , and the  $x$ -axis is  $n$ .

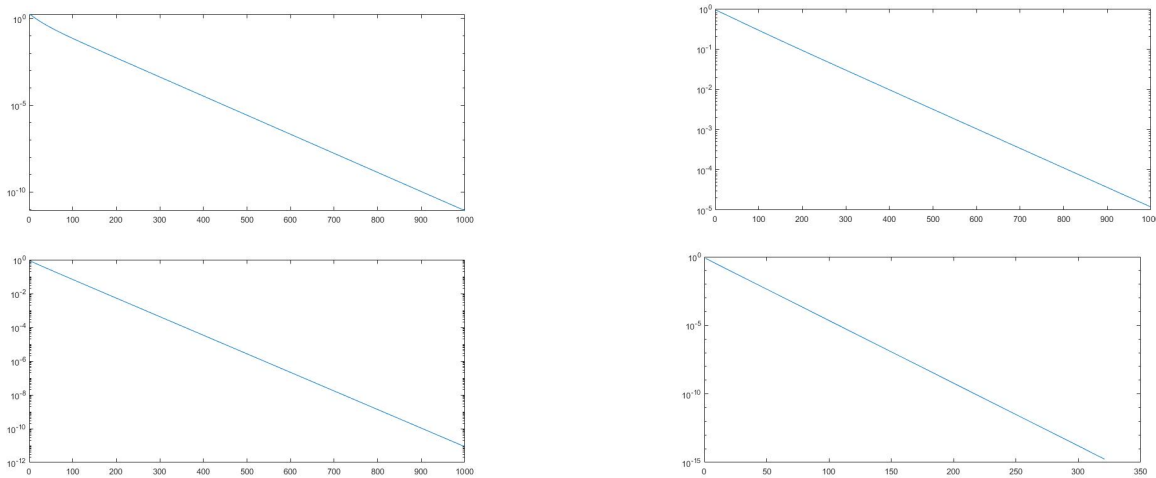


Figure 12: Top left:  $\sigma_2$  of  $M_n$  with  $A$  unknown. Top right:  $\sigma_2$  of  $M_n$  with  $B$  unknown. Bottom left:  $\sigma_2$  of  $M_n$  with  $C$  unknown. Bottom right:  $\sigma_2$  of  $M_n$  with  $D$  unknown.

As we can see, convergence is slowest when  $B$  is unknown, and fastest when  $D$  is unknown.

## 4.5 Maxvol Schur Gradient Descent

We are now ready to combine the greedy maxvol algorithm and the Schur gradient descent algorithm to give us the maxvol Schur gradient descent algorithm. There are multiple ways to combine the greedy maxvol algorithm and the Schur gradient descent. One method is to alternate between gradient steps and maxvol algorithms. Another method is to employ the maxvol method once at the beginning of the algorithm on our initial guess  $M_0$ , or employ the maxvol algorithm once every fixed number of gradient steps.

## 4.6 Dominant submatrices of partially known matrices

In order to make the Schur gradient descent matrix completion method faster and more robust to noise, we should choose our known  $k \times k$  submatrix for our completion with a maxvol

algorithm.

There are multiple methods for doing this. If we allow our submatrix  $A$  to contain unknowns, then we may simply run a maxvol algorithm on our initial guess  $M_0$  to choose  $A$ . Since the entries of  $A$  change,  $A$  may not be dominant after several iterations, so we may repeat our maxvol algorithm to choose a different  $A$  after every fixed number of steps.

On the other hand, if we insist that  $A$  should be fully known, then we may run a modified version of the maxvol algorithm on the known entries of  $M_\Omega$ . One way to do this is to start by generalizing the definition of a dominant submatrix to submatrices of partially known matrices.

**Definition 10.** *A fully known  $k \times k$  submatrix  $A$  of a partially known matrix  $M_\Omega$  is called dominant if we may not increase the volume of  $A$  by swapping either a pair of rows or a pair of columns in  $M_\Omega$ .*

Note that the volume of a submatrix is only defined for fully known submatrices.

Equivalently, Let  $C$  be the largest fully known  $n_\Omega \times k$  submatrix of  $M_\Omega$  containing  $A$ , and let  $R$  be the largest fully known  $k \times m_\Omega$  submatrix containing  $A$ . Then  $A$  is dominant in  $M_\Omega$  if  $\|CA^{-1}\|_\infty = 1$ , and  $\|A^{-1}R\|_\infty = 1$ . Here  $n_\Omega$  is used to denote the height of the largest possible known  $n_\Omega \times k$  submatrix of  $M_\Omega$  containing  $A$ , and similarly  $m_\Omega$  denotes the width of the largest possible  $k \times m_\Omega$  submatrix of  $M_\Omega$  containing  $A$ .

For a graph theoretical formulation, recall that for fully known matrices  $M$ , we define  $G_k(M)$  as the graph where the nodes are the  $k \times k$  submatrices of  $M$ , and we have an edge connecting two nodes if we may obtain one corresponding submatrix from another by swapping either one row or one column. Note that if  $M$  is an  $n \times m$  matrix, then  $G_k(M) \cong J(n, k) \square J(m, k)$ , where  $J(n, k)$  is the  $n, k$  Johnson graph, and  $\square$  is the graph Cartesian product.

For a partially known matrix  $M_\Omega$ , we define  $G_k(M_\Omega)$  to be the graph with nodes corresponding to fully known  $k \times k$  submatrices of  $M_\Omega$ , with an edge connecting two fully known

submatrices if we may obtain one submatrix from another by swapping either one pair of rows or one pair of columns in  $M_\Omega$ . Then the dominant submatrices of  $M_\Omega$  correspond to the nodes of  $G_k(M_\Omega)$  which have locally maximal volume. In other words, the submatrices for which the volume of all connected submatrices is non-increasing. Note that  $G_k(M_\Omega)$  is a sub-graph of  $G_k(M)$  for any completion  $M$  of  $M_\Omega$ , which is obtained by deleting every node and connecting edge in  $G_k(M_\Omega)$  whose corresponding submatrix contains an unknown entry. Since  $G_k(M)$  is isomorphic to  $J(n, k) \square J(m, k)$ , then  $G_k(M_\Omega)$  is isomorphic to a subgraph of  $J(n, k) \square J(m, k)$ .

In the special case when  $M_\Omega$  is an  $n \times k$  partially known matrix, let  $n_\Omega$  be the number of rows of  $M_\Omega$  where every entry is known in that row. Then  $G_k(M_\Omega) \cong J(n_\Omega, k)$ . However, when  $M_\Omega$  is a general  $n \times m$  matrix,  $G_k(M_\Omega)$  may not even be connected, since we may not always obtain one fully known submatrix as a sequence of fully known submatrices by swapping pairs of rows or pairs of columns.

**Theorem 37.** *Similarly to the case of fully known matrices, the independence number  $\alpha(G_k(M_\Omega))$  provides an upper bound to the number of dominant submatrices in  $G_k(M_\Omega)$  for almost all  $M_\Omega$ .*

*Proof.* Similarly to the case of fully known matrices, note that for almost all  $M_\Omega$ , no two dominant submatrices may be adjacent in  $G_k(M_\Omega)$ , since if there were two adjacent dominant submatrices they would necessarily need to have the same volume. Since  $\alpha(G_k(M_\Omega))$  is the largest number of possible non-adjacent nodes in  $G_k(M_\Omega)$ , it provides an upper bound for the number of possible dominant submatrices for almost all partially known matrices  $M_\Omega$ .  $\square$

## 4.7 Maxvol on partially known matrixes

The maxvol algorithm on partially known matrices works similarly to the maxvol algorithm. when  $M_\Omega$  is an  $n \times k$  partially known matrix, we simply find the  $n_\Omega \times k$  submatrix where every row and column is known, and calculate maxvol on that submatrix.

When  $M$  is a general  $n \times m$  matrix, a corresponding alternating or greedy maxvol algorithm according to previous case can be used.

This algorithm converges, since we get a sequence of submatrices with increasing volume which are bounded above.

## 4.8 Perturbation Analysis

Our goal is, given  $M_\Omega = \begin{bmatrix} A & B_\Omega \\ C_\Omega & D_\Omega \end{bmatrix}$ , with some fully known  $k \times k$  submatrix  $A$ , we calculate a rank  $r$  completion  $M$  with gradient descent on all  $r \times r$  minors containing  $A$ . We would like to choose  $A$  such that it is of maximum volume over all fully known  $k \times k$  submatrices.

Suppose our observed entries are noisy. That is, we have  $M_\Omega = Y_\Omega + Z_\Omega$ , where  $Z_\Omega$  is the noise term with  $\|Z_\Omega\|_F < \delta$  and  $Y_\Omega$  are the true entries which have a rank  $r$  completion  $Y$ .

Let us first analyze the case when  $M = Y + Z$ , where  $Y$  is a rank  $r$  matrix, and  $\|Z\|_F < \delta$  is the noise term. Let  $Y = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ , and let  $Z = \begin{bmatrix} Z_A & Z_B \\ Z_C & Z_D \end{bmatrix}$ . Consider the skeleton approximation:  $D + Z_D \sim (C + Z_C)(A + Z_A)^{-1}(B + Z_B)$ . Note that since  $Z_A$  is a small perturbation of  $A$ , we have the approximation  $(A + Z_A)^{-1} = A^{-1} - A^{-1}Z_AA^{-1} + O(\|Z_A\|^2)$ . So we have

$$\begin{aligned}
& (C + Z_C)(A + Z_A)^{-1}(B + Z_B) \\
&= C(A + Z_A)^{-1}B + C(A + Z_A)^{-1}Z_B + Z_C(A + Z_A)^{-1}(B + Z_B) \\
&= CA^{-1}B + CA^{-1}Z_AA^{-1}B + C(A + Z_A)^{-1}Z_B + Z_C(A + Z_A)^{-1}(B + Z_B) + O(\|Z_A\|^2) \\
&= D + CA^{-1}Z_AA^{-1}B + C(A + Z_A)^{-1}Z_B + Z_C(A + Z_A)^{-1}(B + Z_B) + O(\|Z_A\|^2)
\end{aligned}$$

Therefore we have

$$\begin{aligned}
& (C + Z_C)(A + Z_A)^{-1}(B + Z_B) - (D + Z_D) \\
&= CA^{-1}Z_AA^{-1}B + C(A + Z_A)^{-1}Z_B + Z_C(A + Z_A)^{-1}(B + Z_B) - Z_D + O(\|Z_A\|^2)
\end{aligned}$$

where the infinity norm is minimized when  $A + Z_A$  is chosen with maximal volume over all possible choices of  $r \times r$  submatrices of  $M$ .

## 4.9 Maxvol Gradient Descent

Given  $M_\Omega$  with initial guess  $M_0$ , and an initial invertible  $r \times r$  submatrix  $A_0$  of  $M_0$ , we can employ maxvol on the initial guess, followed by gradient descent. Or we could alternate between a maxvol algorithm and gradient descent. More specifically, we have the following algorithm.

The alternating maxvol-gradient descent algorithm runs as follows. Given  $M_\Omega$  a partially known matrix, and an initial guess  $M_0$  such that  $P_\Omega(M_0) = M_\Omega$ , a known invertible  $r \times r$  submatrix  $A_0$  of  $M_0$ , step size  $h$ , and a tolerance  $\epsilon$  we do the following:

While  $\|C_n A_n^{-1} B_n - D_n\| > \epsilon$

1. Let  $A_{n+1}$  be the resulting submatrix from the maxvol algorithm on  $M_n$  with initial submatrix in index set of  $A_n$ , with corresponding  $B_n$ ,  $C_n$ , and  $D_n$ .
2. let  $f_n = \frac{1}{2} \|C_n A_n^{-1} B_n - D_n\|_F^2$
3. Let  $M_{n+1} = M_n - h \nabla f_n(M_n)$

Suppose  $\Omega$  is a binary matrix with a 1 in position where that entry is known, and a 0 in positions where entries are unknown. Then let  $\Omega^c$  be the binary matrix with a 0 in positions where entries are known, and a 1 in position where entries are unknown. Let  $P_{\Omega^c}$  be the operator which sets all entries in known positions to 0. Then in particular we have

$$M_{n+1} = M_n - h \nabla f_n(M_n) = \begin{bmatrix} A_n & B_n \\ C_n & D_n \end{bmatrix} - h P_{\Omega^c} \left( \begin{bmatrix} \frac{\partial f_n}{\partial A_n} & \frac{\partial f_n}{\partial B_n} \\ \frac{\partial f_n}{\partial C_n} & \frac{\partial f_n}{\partial D_n} \end{bmatrix} \right)$$

Where  $\frac{\partial f_n}{\partial A_n}$ ,  $\frac{\partial f_n}{\partial B_n}$ ,  $\frac{\partial f_n}{\partial C_n}$ , and  $\frac{\partial f_n}{\partial D_n}$  were calculated previously.

## 4.10 Small Examples with Noise

Let  $M_\Omega$  be a partially known matrix with the unique rank  $r$  completion  $M$ . Let  $Z_\Omega$  be a noise term with  $\|Z_\Omega\| < \epsilon$ . Let  $Y_\Omega = M_\Omega + Z_\Omega$  be the known entries. Then we would like to find a completion  $Y$  of  $Y_\Omega$  such that  $\|Y - M\|$  is small.

Another formulation is that we would like to find a matrix  $Y$  with  $\sigma_{r+1}(Y)$  minimized and  $P_\Omega(Y) = Y_\Omega$ . Note that  $\sigma_{r+1}(Y)$  is equal to the error of  $Y$  to a closest rank  $r$  approximation.

**Example 4.1.** Let  $r = 1$ ,  $M_\Omega = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & \square \\ 2 & \square & \square \end{bmatrix}$  and  $Z_\Omega = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & \square \\ 0 & \square & \square \end{bmatrix}$ . Then  $M_\Omega$  has the unique rank 1 completion  $M = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix}$ . However,  $Y_\Omega = M_\Omega + Z_\Omega = \begin{bmatrix} 2.1 & 2 & 2 \\ 2 & 2.1 & \square \\ 2 & \square & \square \end{bmatrix}$  has no rank 1 completion since  $Y_\Omega$  contains a rank 2 submatrix. However, we can complete  $Y_\Omega$  to a matrix  $Y$  such that  $\sigma_2(Y)$  is small, and that  $Y$  is close to  $M$ .

We will employ maxvol followed by gradient descent for the initial guess

$$Y_0 = \begin{bmatrix} 2.1 & 2 & 2 \\ 2 & 2.1 & 0 \\ 2 & 0 & 0 \end{bmatrix}.$$

After 200 steps with step size 0.1, we recover the matrix

$$Y_{200} = \begin{bmatrix} 2.1 & 2 & 2 \\ 2 & 2.1 & 1.9042 \\ 2 & 1.9976 & 1.9045 \end{bmatrix},$$

where  $\sigma_2(Y_{200}) = 0.1147$ , and  $\|M - Y_{200}\|_F = 0.1353$ . Moreover, for each  $n$  we have a plot of  $\sigma_2(Y_n)$  in fig. 13.

We try again with the same  $\Omega$ , and  $M_\Omega$ , but we let  $Z_\Omega$  consist of uniformly distributed random numbers between 0 and 0.1. Then again,  $Y_\Omega$  will contain a rank two submatrix with

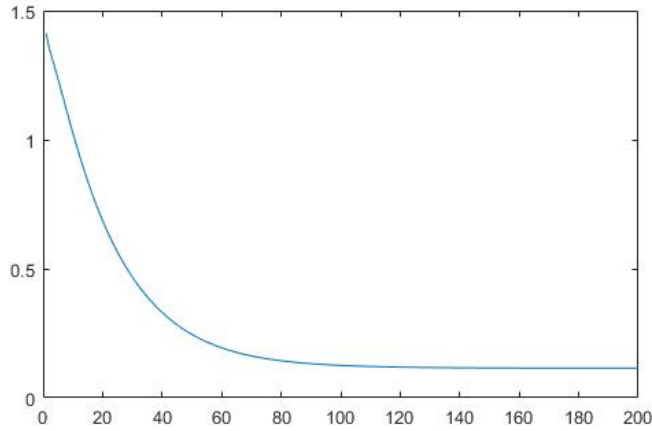


Figure 13:  $\sigma_2$  of the  $3 \times 3$  matrix  $Y_n$ .

probability 1. In particular, we start with

$$Y_\Omega = \begin{bmatrix} 2.0611 & 2.0107 & 2.0399 \\ 2.0812 & 2.0885 & \square \\ 2.0201 & \square & \square \end{bmatrix},$$

and after 200 steps, we recover the matrix

$$Y_{200} = \begin{bmatrix} 2.0611 & 2.0107 & 2.0399 \\ 2.0812 & 2.0885 & 2.0593 \\ 2.0201 & 2.0272 & 1.9985 \end{bmatrix}$$

where  $\sigma_2(Y_{200}) = 3.8546e - 02$ , and  $\|M - Y_{200}\|_F = 1.5662e - 01$ . Moreover, for each  $n$  we have the plot of  $\sigma_2(Y_n)$  in fig. 14.

## 4.11 Larger Examples With Noise

Once again, we will let  $Y$  be the  $128 \times 128$  penny picture. We may consider  $Y$  as a rank  $r$  matrix with noise. In particular we let  $M = P_{\mathcal{M}_r}(Y)$ , and let  $Z = Y - M$ . Then given some

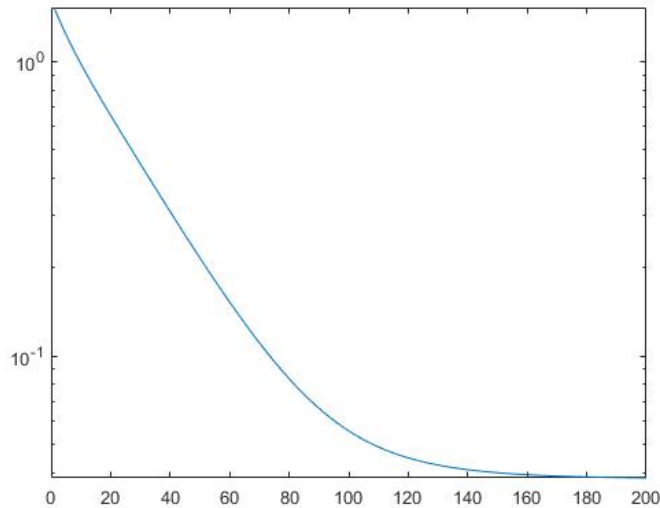


Figure 14:  $\sigma_2$  of the  $3 \times 3$  matrix with random noise  $Y_n$ .

index of unknown entries  $\Omega$  and initial guess  $Y_0$ , we can run maxvol followed by gradient descent to approximately recover  $M$ .

$M$  will be a good approximation of  $Y$  if  $r$  is chosen large enough. In other words, if  $r$  is chosen large enough, then  $\|Z\|$  will be small. Table 7 shows  $\|Z\|$  for some fixed  $r$ .

$r$	$\ Z\ $
10	6.0523e+02
20	1.9172e+02
30	1.1082e+02
40	6.6198e+01
50	4.3853e+01
60	3.1554e+01
70	2.2414e+01
80	1.5001e+01
90	1.0220e+01
100	6.5856e+00
110	3.7256e+00
120	1.4032e+00

Table 7: Error to closest rank  $r$  approximation of  $128 \times 128$  penny picture.

Let us choose  $r = 100$ , in which case  $\|Z\| = 6.59$ . We will pick  $\Omega$  by randomly keep  $2nr - r^2 = 15600$  entries, and we will set the rest equal to zero for  $Y_0$ . We will set the step



size  $h = 0.005$ , and run for 10000 steps.

Once again, we plot  $\sigma_{r+1}(Y_n)$  vs  $n$  and get fig. 15.

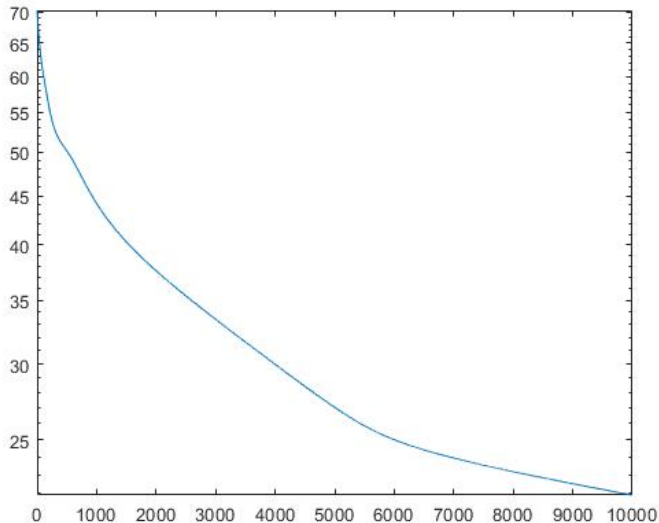


Figure 15:  $\sigma_{101}$  of  $Y_n$ .

As we can see,  $\sigma_{101}(Y_n)$  decreases from 69.48 to 21.96.

## 4.12 Comparing Gradient Descent to Maxvol-Gradient Descent

The reason why we use maxvol in our gradient descent method is to improve numerical stability, speed up convergence, and allow us to take larger step sizes. As a simple example, consider the  $2 \times 2$  rank 1 matrix  $M = \begin{bmatrix} 5 & 5 \\ 1 & 1 \end{bmatrix}$ , with  $\Omega = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ ,  $M_0 = \begin{bmatrix} 5 & 5 \\ 1 & 0.5 \end{bmatrix}$ , and step size  $h = 0.1$ . If we choose our matrix  $A$  to be in either index (1,1) or index (1,2), then the gradient descent method converges to  $M$ . However, if  $A$  is chosen to be the submatrix 1 in index (2,1), then the gradient descent method diverges. In order to obtain convergence, we must reduce our step size to  $h = 0.07$ .

For a larger example, we once again consider the  $n \times n$  penny picture with  $n = 128$  and integer entries between 0 and 255. We fix  $r = 30, 60$ , and 90, and we keep  $2nr - r^2$  random

entries and replace the rest with uniformly sampled independent random numbers between 0 and 255.

We first choose  $A$  by randomly picking  $r$  rows and  $r$  columns and run gradient descent. We compare this to choosing  $A$  by using the maxvol algorithm and running gradient descent. We create a semi-log plot of  $\sigma_{r+1}(M_n)$  for each method. We fix our step size  $h = 0.005$  and run for 10000 steps for each method.

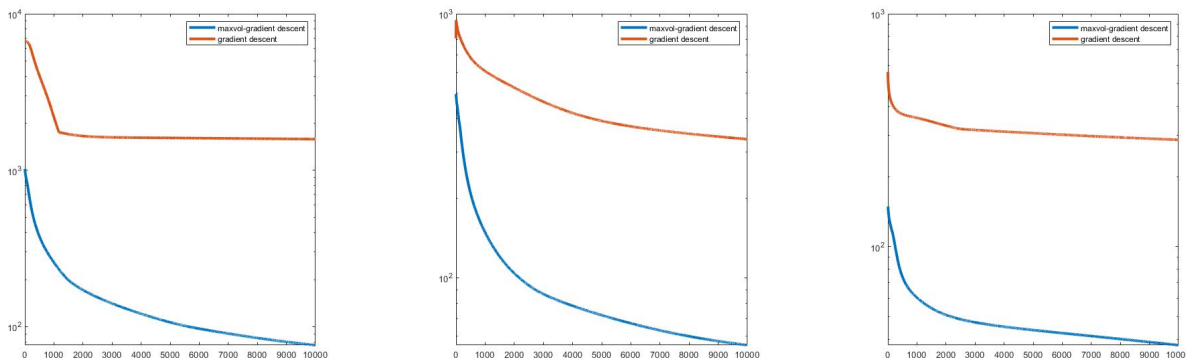


Figure 16: Left: Using the penny picture, we fix  $r = 30$ . Plot of  $\sigma_{31}(M_n)$  vs  $n$  for maxvol-gradient descent and gradient descent. Middle: Using the penny picture, we fix  $r = 60$ . Plot of  $\sigma_{61}(M_n)$  vs  $n$  for maxvol-gradient descent and gradient descent. Right: Using the penny picture we fix  $r = 90$ . Plot of  $\sigma_{91}(M_n)$  vs  $n$  for maxvol-gradient descent and gradient descent.

As we can see from fig. 16 choosing  $A$  through maxvol provides a significant improvement compared to choosing  $A$  randomly.

We now consider the  $256 \times 256$  house image with entries as integers between 0 and 255. In the house picture, the 51st singular value is equal to 214.6764, and the 101st singular value is equal to 75.2592.

We set  $r = 50$ , and 100 and randomly keep  $2nr - r^2 + 10000$  entries, and delete the rest. We run maxvol gradient descent by setting our step size  $h = 0.005$  and running for 1000 steps.

First, for  $r = 10$ , we have fig. 18.

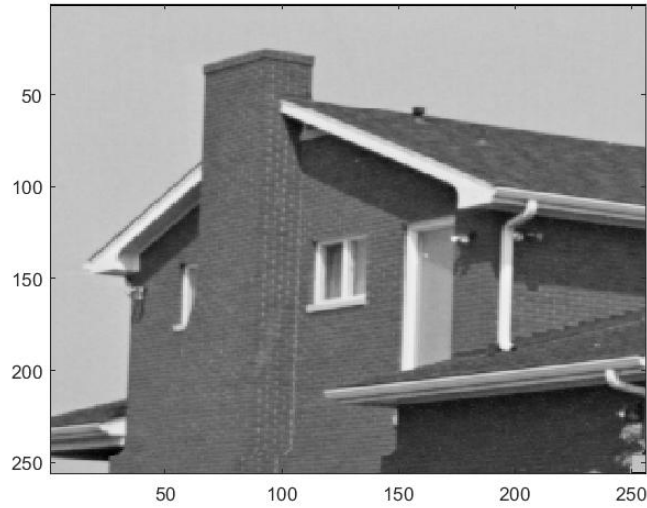


Figure 17:  $256 \times 256$  house image used for numerical tests.

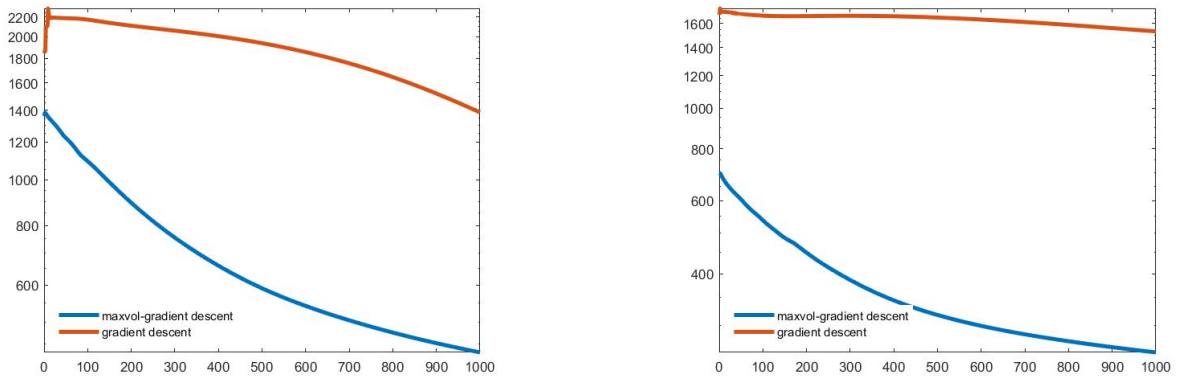


Figure 18: Left: Using the house picture, we fix  $r = 50$ . Plot of  $\sigma_{51}(Y_n)$  vs  $n$  for maxvol-gradient descent and gradient descent. Right: Using the house picture, we fix  $r = 100$ . Plot of  $\sigma_{101}(Y_n)$  vs  $n$  for maxvol-gradient descent and gradient descent.

# 5 Maximum Volume Based Skeletal Decompositions for Scalable Plasma Physics Applications

## 5.1 Motivation

Researchers within the Fusion Energy Division at ORNL aim to reduce the computational complexity of fusion plasma simulation via construction of linear time advance operators for multi-timescale problems. That work is based on the data-driven dynamic mode decomposition (DMD) method, which has been proven to be a useful tool for the analysis of fluid dynamics [32]. However, for high dimensional data the SVD scales poorly, on the order of  $O(m^2n)$  for  $n \times m$  matrices with  $m \leq n$ . For the ORNL application, the matrices to be approximated are generated by the multi-fluid SOLPS code for modeling the flow of plasma at the edge region of a fusion device [34]. A typical  $n \times m$  SOLPS produced data matrix can have  $n$  greater than 10,000 and  $m$  on the order of 100. For kinetic plasma simulations these sizes may be much greater again.

Application of the SVD to the high dimensional data required for fusion simulation is computationally prohibitive and alternative methods are desirable. We aim to leverage the maxvol skeleton decomposition to allow for the extension of the ORNL projective integration algorithm to the scales of simulation data inherent in the physics of fusion plasmas.

## 5.2 Simulation Data Background Information

1D data, high noise corresponds to temperature (outboard diverter target) and low noise corresponds to density (midplane density) (outboard midplane corresponds to thickest cut of the midsection)

Simulations at the boundary of the magnetically confined plasma in a tokamak are carried out to compute either the exhaust (heat and particle flux) due to magnetically driven plasma

dynamics from the core, such as deposition of escaping heat flux onto the diverter target. The boundary also permits an accessible region for controlling this plasma through actuators such as pellet injection or neutral gas puff. SOLPS simulates these dynamics through the coupling of the fluid plasma equations and the kinetic monte carlo neutral equation.

kinetic monte carlo neutral equation compute the trajectories of the kinetic particles and supply the source terms

$s_n$  continuity equation source term,  $s_m$  momentum source term,  $s_E$  energy source term.

$$\begin{aligned} \frac{d\eta_n}{dt} + \eta_s \nabla \cdot V_s &= s_n \\ m_s \eta_s \frac{dV_s}{dt} + \nabla p_s + \nabla \pi_s - e_s \eta_s (E + V_s \times B) &= s_m \\ \frac{3}{2} \frac{dp_s}{dt} + \frac{5}{2} p_s \nabla \cdot V_s + \pi_s \cdot \nabla V_s + \nabla \cdot q_s &= s_E \end{aligned}$$

Interaction between observable quantities such as temperature, density, pressure, and the separate population of neutral particles (do not interact with large scale electric or magnetic fields).

The system of coupled ODEs are solve implicitly and utilize a Picard iteration to converge onto the solution. This calculation is also subject to the independent convergence of the Monte Carlo simulation.

Each time step requires the iteration over the internal calculation of the Monte Carlo equations that converges to a solution to the source terms of the plasma equations. With these source terms you can then attempt to solve the odes which can be vastly different timescales, both implicit time stepping and Picard iteration is used to obtain a solution at sensible time steps.

If a plasma is to reach thermal equilibrium with the ions and electrons their momenta will be orders of magnitude apart. For the fluid equations the difference in mass between

the electrons and protons is the cause for the exhibited difference in timescales.

(Look at diego's and david hatches paper for explanations on why compression is useful.)

### 5.3 Maximum Volume Skeleton Decomposition For Plasma Simulation Data Compression

Simulations are carried out for several input parameters that can produce a high quantity of data that prohibits efficient analysis. Not efficient to load.

We present numerical results on applying the skeleton decomposition to plasma simulation data.

### 5.4 Dynamic Mode Decomposition Using the Skeleton Decomposition

First, we will present the standard definition for the *dynamic mode decomposition*, or DMD, from [32]. Consider a sequence of data vectors  $\{z_0, \dots, z_m\}$  where  $z_k \in \mathbb{R}^n$  for all  $n$ . We assume that the data satisfies the linear relationship  $z_{k+1} = Az_k$  for some matrix  $A$ . We define the  $n \times m$  matrices  $X = [z_0 \ \dots \ z_{m-1}]$  and  $Y = [z_1 \ \dots \ z_m]$ , which satisfies  $Y = AX$ .

We will now define the *psuedoinverse*  $X^+$  of  $X$ . To compute the psuedoinverse, assume  $X$  is rank  $r$ . Let  $X = U\Sigma V^*$  be the SVD of  $X$ , where  $\Sigma$  is the  $r \times r$  diagonal matrix of non-zero singular values. Then the psuedoinverse of  $X$  is defined as  $X^+ = V\Sigma^{-1}U^*$ . In general, we compute the SVD of  $X$  obtaining  $X = U\Sigma V^*$ , where  $\Sigma$  is the diagonal matrix consisting of the singular values of  $X$ . We compute a rank  $r$  approximation  $X_r$  using the first  $r$  singular values of  $X$  and the first  $r$  columns of  $U$  and  $V$ , obtaining

$$X_r = U_r \Sigma_r V_r^*$$

Where  $U_r$  is  $n \times r$ ,  $V_r$  is  $m \times r$ , and  $\Sigma_r$  is the  $r \times r$  diagonal matrix consisting of the first  $r$

singular values of  $X$ .

We then define the matrices

$$A_r = YX_r^+ = YV_r\Sigma_r^{-1}U_r^* \quad (5)$$

$$\tilde{A}_r = U_r^* A_r U_r = U_r^* Y V_r \Sigma_r^{-1} \quad (6)$$

We would like to replace the low-rank approximation using the singular value decomposition with a low rank approximation using the skeleton decomposition.

When computing a low-rank approximation of  $X$ , we can instead run the maxvol algorithm on  $X$  obtaining an  $r \times r$  dominant submatrix  $X_\square$  with corresponding columns  $C$  and rows  $R$ . We then define a low-rank approximation of  $X$  as

$$X_r = CX_\square^{-1}R$$

.

We will show that many of the theorems from [32] translate to this context. Analogously to eq. (5) and eq. (6), we define the matrices

$$A_r = YX_r^+ = YR^+X_\square C^+ \quad (7)$$

$$\tilde{A}_r = C^+ Y R^+ X_\square = C^+ A_r C \quad (8)$$

Note that since  $C$ ,  $X_\square$ , and  $R$  are full rank, we have that  $X_r^+ = (CX_\square^{-1}R)^+ = R^+X_\square C^+$ . Moreover, since the columns of  $C$  are linearly independent, we have that  $C^+C = I_r$ , where  $I_r$  is the  $r \times r$  identity matrix.

We now calculate the eigenvalues  $\lambda$  of  $\tilde{A}_r$  and corresponding eigenvectors  $w$ , giving  $\tilde{A}_r w = \lambda w$ .

Then the projected skeletal DMD mode with respect to the eigenvalue  $\lambda$  is given by

$$\hat{\varphi} = Cw \quad (9)$$

and the exact skeletal DMD mode with respect to  $\lambda$  is given by

$$\varphi = \frac{1}{\lambda} Y R^+ X_{\square} w. \quad (10)$$

Define  $B = Y R^+ X_{\square}$ . Then in terms of  $B$ , we have  $A_r = BC^+$ ,  $\tilde{A}_r = C^+B$ , and  $\varphi = \frac{1}{\lambda} Bw$ .

**Theorem 38.**  *$\varphi$  is an eigenvalue of  $A_r$  with eigenvalue  $\lambda$ . Moreover, every non-zero eigenvalue of  $A_r$  is also an eigenvalue of  $\tilde{A}_r$ .*

*Proof.* We have

$$\begin{aligned} A_r \varphi &= (BC^+) \left( \frac{1}{\lambda} Bw \right) \\ &= B \left( \frac{1}{\lambda} \tilde{A}_r w \right) \\ &= Bw \\ &= \lambda \left( \frac{1}{\lambda} Bw \right) \\ &= \lambda \varphi \end{aligned}$$

Moreover,  $\varphi \neq 0$ , since if  $\varphi = \frac{1}{\lambda} Bw = 0$ , then  $C^+Bw = \tilde{A}_r w = \lambda w = 0$ , so  $\lambda = 0$ .

To show that every eigenvalue of  $A_r$  is an eigenvalue of  $\tilde{A}_r$ , suppose  $\lambda$  is a non-zero



eigenvalue of  $A_r$  with eigenvector  $\varphi$ . Let  $w = C^+\varphi$ . Then we have

$$\begin{aligned}
\tilde{A}_r w &= (C^+B)(C^+\varphi) \\
&= C^+A_r\varphi \\
&= \lambda C^+\varphi \\
&= \lambda w.
\end{aligned}$$

Moreover,  $w \neq 0$ , since if  $w = C^+\varphi = 0$ , then  $BC^+\varphi = A_r\varphi = \lambda\varphi = 0$ . So  $\lambda = 0$ , which is a contradiction. So  $\lambda$  is an eigenvalue of  $\tilde{A}_r$  with eigenvector  $w$ .  $\square$

Let  $\mathbb{P}_{X_r}$  be the orthogonal projection operator onto the column space of  $X_r$ . Then since the column space of  $X_r$  is equal to the column space of  $C$ , we have that  $\mathbb{P}_{X_r} = CC^+$ .

**Theorem 39.** *Let  $\hat{\varphi}$  be defined as in eq. (9), and let  $\varphi$  be defined as in eq. (10). Then  $\hat{\varphi}$  is an eigenvector of  $\mathbb{P}_{X_r}A_r$  with eigenvalue  $\lambda$ . Moreover,  $\hat{\varphi} = \mathbb{P}_{X_r}\varphi$ .*

*Proof.* First, we have

$$\begin{aligned}
\mathbb{P}_{X_r}A_r\hat{\varphi} &= (CC^+)(BC^+)(Cw) \\
&= C(C^+B)I_r w \\
&= C\tilde{A}_r w \\
&= \lambda Cw \\
&= \lambda\hat{\varphi}.
\end{aligned}$$

So  $\lambda$  is an eigenvalue of  $\hat{\varphi}$ . Moreover, we have

$$\begin{aligned}
\mathbb{P}_{X_r}\varphi &= (CC^+)(\frac{1}{\lambda}Bw) \\
&= \frac{1}{\lambda}C\tilde{A}_r w \\
&= Cw \\
&= \hat{\varphi}
\end{aligned}$$

□

Note that we have made no reference to how good of an approximation  $X_r^+$  is to  $X^+$ . In particular, the above argument works for any rank  $r$  approximation  $X_r = LR^*$  of  $X$ , regardless of the error between  $X$  and  $X_r$ , where  $L$  is an  $n \times r$  full rank matrix and  $R$  is an  $m \times r$  full rank matrix. What is the error between  $X^+$  and  $X_r^+$  when we choose  $X_r$  using the SVD? Suppose  $\text{rank}(X) = s$ . Let  $X = U\Sigma V^*$ , and let  $X_r = U\Sigma_r V^*$ , where  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_s, 0, \dots, 0)$  is the  $n \times m$  diagonal matrix with diagonal entries equal to the singular values of  $X$ . Similarly, let  $\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0)$  where  $r \leq s$ . Then  $X^+ = V\Sigma^{-1}U^*$ , and  $X_r^+ = V\Sigma_r^{-1}U$ . Then we have

$$\begin{aligned}
\|X^+ - X_r^+\| &= \|X^+ - X_r^+\| \\
&= \|V\Sigma^{-1}U^* - V\Sigma_r^{-1}U^*\| \\
&= \|U(\Sigma^{-1} - \Sigma_r^{-1})V^*\| \\
&= \|\Sigma^{-1} - \Sigma_r^{-1}\| \\
&= \left\| \text{diag}\left(0, \dots, 0, \frac{1}{\sigma_{r+1}}, \dots, \frac{1}{\sigma_s}, 0, \dots, 0\right) \right\| \\
&= \frac{1}{\sigma_s}
\end{aligned}$$

where  $\|\cdot\|$  is the spectral norm. This means that the error between  $A$  and  $A_r$  can get arbitrar-

ily large as the smallest non-zero singular value  $\sigma_k \rightarrow 0$ . This is because the pseudoinverse function is discontinuous at matrices which are not full rank [35]. Moreover, this implies that

$$\begin{aligned} \|A - A_r\| &= \|YX^+ - YX_r^+\| \\ &\leq \|Y\| \|X^+ - X_r^+\| \\ &= \|Y\| \frac{1}{\sigma_k} \end{aligned}$$

The error between  $X^+$  and  $X_r^+$  is more difficult to analyze when  $X_r$  is chosen with respect to the skeleton approximation. In particular, if  $X = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ , then the skeleton approximation of  $X$  with respect to a  $r \times r$  nonsingular submatrix  $A$  is defined as

$$X_r = \begin{bmatrix} A \\ C \end{bmatrix} A^{-1} \begin{bmatrix} A & B \end{bmatrix} = \begin{bmatrix} A & B \\ C & CA^{-1}B \end{bmatrix}.$$

Because  $A$  is full rank, we have

$$\begin{aligned} X_r^+ &= \left( \begin{bmatrix} A \\ C \end{bmatrix} A^{-1} \begin{bmatrix} A & B \end{bmatrix} \right)^+ \\ &= \begin{bmatrix} A & B \end{bmatrix}^+ A \begin{bmatrix} A \\ C \end{bmatrix}^+ \\ &= \begin{bmatrix} A^* \\ B^* \end{bmatrix} (AA^* + BB^*)^{-1} A (A^*A + C^*C)^{-1} \begin{bmatrix} A^* & C^* \end{bmatrix} \end{aligned}$$

So letting  $H = (AA^* + BB^*)^{-1}A(A^*A + C^*C)^{-1}$ , we have

$$X_r^+ = \begin{bmatrix} A^*HA^* & A^*HC^* \\ B^*HA^* & B^*HC^* \end{bmatrix}.$$

Now let  $S_A = D - CA^{-1}B$  denote the Schur complement of  $X$  with respect to  $A$ . Then from [33], if the column space of  $C$  is contained in the column space of  $S_A$ , and the row space of  $B$  is contained in the row space of  $S_A$ , then we have

$$X^+ = \begin{bmatrix} A^{-1} + A^{-1}BS_A^+CA^{-1} & -A^{-1}BS_A^+ \\ -S_A^+CA^{-1} & S_A^+ \end{bmatrix}.$$

It is not so clear what the error  $\|X^+ - X_r^+\|$  would be in this case.

## 6 Tensor Theory and Background

We may generalize some of our results on matrix completion to the case of tensors, which may be considered as high dimensional matrices or arrays. Let  $U$  be a vector space of dimension  $n$  with basis  $u_1, \dots, u_n$  and let  $V$  be a vector space of dimension  $m$  with basis  $v_1, \dots, v_m$ . Then the tensor product  $U \otimes V$  is the vector space of dimension  $nm$  spanned by elements of the form  $u_i \otimes v_j$  over all  $i$  and  $j$  with the following relations. For elements  $u$  and  $v \in V$ ,  $w \in W$ , and scalar  $\alpha$  we have

1.  $\alpha(v \otimes w) = (\alpha v) \otimes w = v \otimes (\alpha w)$
2.  $(v + u) \otimes w = v \otimes w + u \otimes w$
3.  $w \otimes (v + u) = w \otimes v + w \otimes u.$

If  $W$  is a vector space of dimension  $p$  with basis  $w_1, \dots, w_p$ , then  $U \otimes V \otimes W$  is a vector space of dimension  $nmp$  spanned by elements of the form  $u_i \otimes v_j \otimes w_k$ . If  $T \in U \otimes V \otimes W$  is expanded with respect to this basis,  $T = \sum_{ijk} a_{ijk} u_i \otimes v_j \otimes w_k$ , then  $T$  may be represented as the three dimensional array  $[a_{ijk}]$ .

For a tensor product of vector spaces  $V_1 \otimes \dots \otimes V_d$ , an element  $T \in V_1 \otimes \dots \otimes V_d$  is called a tensor of degree, or order,  $d$ . We express theorems and definitions in terms of degree 3 tensors, but many may be generalized to higher degree tensors.

One of the main issues with generalizing the methods of matrix completion to tensor completion is generalizing the notion of rank to that of tensors. There are multiple ways to do so, but calculating the rank of a tensor is not always easy. In this section we will present some known results on the geometry of low-rank tensors.

## 6.1 Types of Tensor Ranks

The definitions and theorems from this section are from [5]. We start by reformulating the rank of a linear map in terms of tensors.

Let  $U, V$  be vector spaces, and  $U^*$  the dual of  $U$ . Let  $\alpha \in U^*$ , and  $b \in V$ . Then given  $\alpha \otimes b \in U^* \otimes V$  one can define a rank 1 linear map  $\alpha \otimes b : U \rightarrow V$  by  $(\alpha \otimes b)(a) = \alpha(a)b$  where  $a \in U$ , and  $\alpha \otimes b \in U^* \otimes V$ . In more familiar terms the linear map  $\alpha \otimes b$ , is essentially the rank 1 matrix  $b\alpha$ , where  $\alpha$  is a row vector,  $b$  is a column vector.

In general, for  $T \in U^* \otimes V$ , the rank of the linear map  $T : U \rightarrow V$  is the smallest  $r$  such that there exists  $\alpha_1, \dots, \alpha_r \in U^*$  and  $b_1, \dots, b_r \in V$  such that  $T = \sum_{i=1}^r \alpha_i \otimes b_i$ . This definition of rank agrees with the standard definitions for the rank of a matrix  $M$ , since the rank of  $M$  is equal to the fewest possible number of terms when writing  $M$  as a sum of rank 1 matrices.

We may generalize the definition of rank to tensors of any degree. To start, we will extend the definition to bilinear operators. Let  $\alpha \in U^*, \beta \in V^*, c \in W$  with  $a \in U$ , and  $b \in V$ . Then the map  $\alpha \otimes \beta \otimes c : U \times V \rightarrow W$ , defined by  $(a, b) \mapsto \alpha(a)\beta(b)c$  is a rank 1 bilinear map.

In general, a bilinear map of the form  $T : U \times V \rightarrow W$  can be represented as a sum

$$T(a, b) = \sum_{i=1}^r \alpha^i(a)\beta^i(b)c_i$$

for some  $r$ , where  $\alpha^i \in U^*, \beta^i \in V^*$ , and  $c_i \in W$ . The smallest such  $r$  is the rank of  $T$ , denoted  $\mathbf{R}(T) = r$ .

We will generalize this definition of rank to degree 3 tensors.

**Definition 11.** *An element  $T \in U \otimes V \otimes W$  is said to have rank one if  $T = u \otimes v \otimes w$ , for some  $u \in U$ ,  $v \in V$ , and  $w \in W$ . The rank of a tensor, is the smallest  $r$  such that  $T = \sum_{j=1}^r Z_j$ , with each  $Z_j$  rank one, and is denoted  $\mathbf{R}(T) = r$ .*

A tensor  $T \in U \otimes V \otimes W$  may be considered as the linear maps

$$T^{(1)} : U^* \rightarrow V \otimes W$$

$$T^{(2)} : V^* \rightarrow U \otimes W$$

$$T^{(3)} : W^* \rightarrow U \otimes V.$$

To explain these linear maps, consider the first example  $T^{(1)} : U^* \rightarrow V \otimes W$ . Recall that we may represent an element in  $X^* \otimes Y$  as a linear map from  $X$  to  $Y$ . In this context, we set  $Y = V \otimes W$ , and  $X = U^*$ , since  $U^{**}$  is canonically isomorphic to  $U$ .

Each of these linear maps has a rank for which is equal to the dimension of the image of the map. That is, the ranks are equal to  $\dim(T^{(1)}(U^*))$ ,  $\dim(T^{(2)}(V^*))$ , and  $\dim(T^{(3)}(W^*))$  respectively.

**Definition 12.** *The multilinear rank, also known as the Tucker rank, of  $T \in U \otimes V \otimes W$  is denoted as  $\mathbf{R}_m(T)$ , and is defined to be the 3-tuple of natural numbers*

$$\mathbf{R}_m(T) = (\text{rank}(T^{(1)}), \text{rank}(T^{(2)}), \text{rank}(T^{(3)}))$$

The multilinear rank of higher degree tensors is defined similarly. That is, the  $i$ th component of the multi-linear rank is equal to the rank of the mode- $i$  unfolding of  $T$ . Note that for degree two tensors  $T \in U \otimes V$ ,  $\dim(T(U^*)) = \dim(T(V^*))$  since the row rank of a matrix is equal to its column rank.

In general, if  $T$  is an  $n \times m \times p$  tensor, with  $\mathbf{R}_m(T) = (r_1, r_2, r_3)$ , then we must have

$$r_1 \leq \min(mp, n)$$

$$r_2 \leq \min(np, m)$$

$$r_3 \leq \min(nm, p).$$

This is because  $T^{(1)}$  is an  $mp \times n$  matrix,  $T^{(2)}$  is an  $np \times m$  matrix, and  $T^{(3)}$  is an  $nm \times p$  matrix.

Another way to view the multilinear rank of degree three tensors is that it is the 3-tuple of integers which are the maximum number of linearly independent mode-1 (column) fibers, the maximum number of linearly independent mode-2 (row) fibers, and the maximum number of linearly independent mode-3 (tube) fibers.

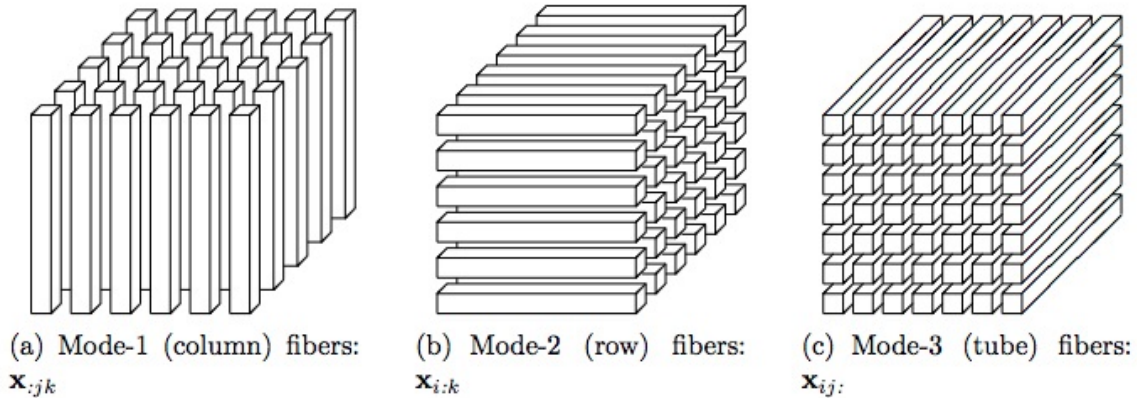


Figure 19: Fibers of a 3rd order tensor [13].

The matrix  $T^{(i)}$  is also known as the mode- $i$  unfolding matrix of the tensor  $T$  [6]. If  $T$  is expanded with respect to a basis,  $T = \sum_{jkl} a_{jkl} u_j \otimes v_k \otimes w_l$ , and is represented as a the array  $[a_{jkl}]$ , then  $T^{(i)}$  can be represented as a matrix by unfolding the slices of  $[a_{jkl}]$  along the  $i$ th coordinate.

These two definitions of the rank and the multilinear rank are related in the following



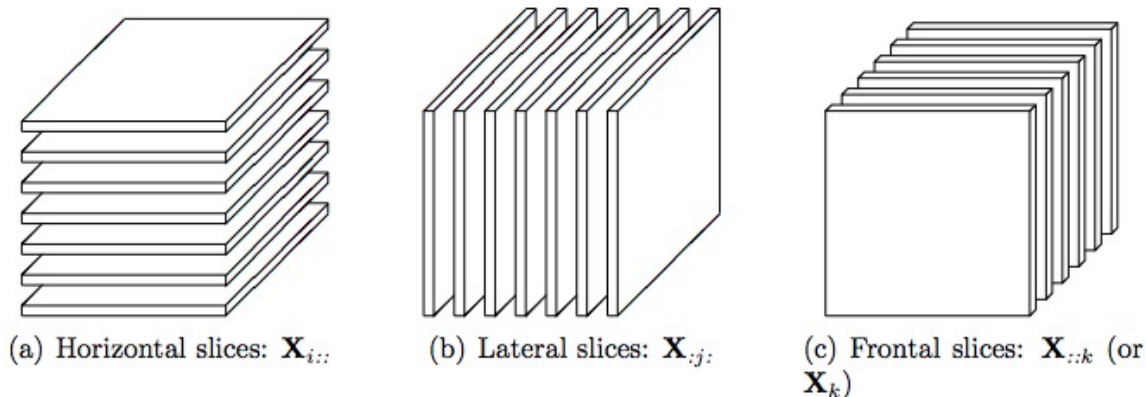


Figure 20: Slices of a 3rd order tensor [13].

way.

**Lemma 6.** *For any tensor  $T$ ,  $\max(\mathbf{R}_m(T)) \leq \mathbf{R}(T)$ .*

*Proof.* Let  $T \in U \otimes V \otimes W$ , and suppose  $\mathbf{R}(T) = r$ . Then  $T$  may be expressed as the sum of  $r$  rank one tensors  $T = \sum_{j=1}^r Z_j$  for all  $Z_j$  rank one. Note that  $\dim(Z_j(U^*)) \leq 1$  for all  $j$ , which implies that  $\dim(T(U^*)) \leq r$ , and similarly for  $V$  and  $W$ . So each component of the multilinear rank is at most  $r$ .  $\square$

One issue with the definition of the rank of a tensor is that the space of tensors with rank at most  $r$  may not be a closed set. In other words, it may be possible to express a high rank tensor as the limit of low rank tensors. This is in contrast to matrices, where  $\overline{\mathcal{M}}_r$ , the space of matrices of rank at most  $r$ , is closed. To study the closure of the space of matrices of rank at most  $r$ , we introduce the notion of the border rank. A tensor has border rank at most  $r$  if it can be approximated by rank  $r$  tensors.

**Definition 13.** *The border rank of a tensor  $T$  is the smallest  $r$  such that there exists a sequence of tensors  $\{T_i\}_i$  of rank  $r$  such that  $\lim_{i \rightarrow \infty} T_i = T$ , and is denoted  $\underline{\mathbf{R}}(T)$ . Equivalently, the border rank of a tensor  $T$  is the smallest  $r$  such that there exists a tensor of rank  $r$  in the ball  $B_\epsilon(T) = \{T_\epsilon \mid \|T_\epsilon - T\| < \epsilon\}$  for all  $\epsilon > 0$ , where  $\|\cdot\|$  is the Euclidean norm.*

Note that the border rank of a tensor is no more than the rank of a tensor. That is,  $\underline{\mathbf{R}}(T) \leq \mathbf{R}(T)$ . We will give an example of a tensor whose rank and border rank differ.

**Example 6.1.** Let  $T = u \otimes u \otimes v + u \otimes v \otimes u + v \otimes u \otimes u$ , where  $u$  and  $v$  are linearly independent. Then  $\mathbf{R}(T) = 3$ . However,  $\underline{\mathbf{R}}(T) = 2$ . To see this, let

$$T_n = n(u + \frac{1}{n}v) \otimes (u + \frac{1}{n}v) \otimes (u + \frac{1}{n}v) - n(u \otimes u \otimes u).$$

Then  $T_n$  is rank 2, and  $\lim_{n \rightarrow \infty} T_n = T$ .

We will prove by contradiction that  $\mathbf{R}(T) = 3$  in the case where  $u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$  and  $v = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ , we will suppose  $\text{rank}(T) \leq 2$ . Then we can write

$$T = \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix} \otimes \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} \otimes \begin{bmatrix} w_{11} \\ w_{12} \end{bmatrix} + \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix} \otimes \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} \otimes \begin{bmatrix} w_{21} \\ w_{22} \end{bmatrix}.$$

for some variables  $u_{11}, u_{12}, u_{21}, u_{22}, v_{11}, v_{12}, v_{21}, v_{22}, w_{11}, w_{12}, w_{21}, w_{22}$ . Comparing this with the definition of  $T$ , we get the following system of equations:

$$\begin{array}{ll} u_{11}v_{11}w_{11} + u_{21}v_{21}w_{21} = 0 & u_{12}v_{11}w_{11} + u_{22}v_{21}w_{21} = 1 \\ u_{11}v_{11}w_{12} + u_{21}v_{21}w_{22} = 1 & u_{12}v_{11}w_{12} + u_{22}v_{21}w_{22} = 0 \\ u_{11}v_{12}w_{11} + u_{21}v_{22}w_{21} = 1 & u_{12}v_{12}w_{11} + u_{22}v_{22}w_{21} = 0 \\ u_{11}v_{12}w_{12} + u_{21}v_{22}w_{22} = 0 & u_{12}v_{12}w_{12} + u_{22}v_{22}w_{22} = 0 \end{array}$$

Using the software Macaulay2, we can verify that this system of equations has no solutions.

An open question is, given a tensor  $T$  of border rank  $r$ , what is the largest rank the tensor can have? For example, it is known that a border rank 4 tensor in  $\mathbb{C}^4 \otimes \mathbb{C}^4 \otimes \mathbb{C}^4$  can have rank at most 7, but in general the question is open [14].

Let  $\sigma_r$  denote the set of tensors in  $U \otimes V \otimes W$  with rank at most  $r$ .

**Definition 14.** Let the  $B_\epsilon(T) = \{T_\epsilon \mid \|T_\epsilon - T\| < \epsilon\}$  denote the ball centered at  $T$  with radius  $\epsilon$ . Then the generic rank of a tensor  $T$ , denoted  $\mathbf{R}_g(T)$ , is the least  $r$  such that the intersection  $B_\epsilon(T) \cap \sigma_r$  is of positive measure for all  $\epsilon > 0$ .

Note that if  $r$  is the generic rank of  $T$ , then there exists a sequence  $\{T_i\}_i$  where  $\mathbf{R}(T_i) = r$  such that  $\lim_{i \rightarrow \infty} T_i = T$ . Therefore the border rank of  $T$  is no larger than the generic rank of  $T$ . That is,  $\underline{\mathbf{R}}(T) \leq \mathbf{R}_g(T)$ . In general, the rank of a tensor  $T$  could be either larger or smaller than the generic rank of  $T$ .

A rank  $r$  is a typical rank if the set of tensors of rank  $r$  in  $U \otimes V \otimes W$  has non-zero measure. Equivalently,  $r$  is a typical rank if it is the generic rank of some tensor  $T$ . Over  $\mathbb{C}$ , the typical rank is unique. Moreover, it is equal to the maximum border rank [10]. However, over  $\mathbb{R}$  the typical rank may not be unique.

For  $n \times m$  matrices, the typical rank is  $\min\{n, m\}$ , and is also the maximum possible rank of any  $n$  by  $m$  matrix. For general tensors, unlike in the case of matrices, there could be measure zero sets of tensors with rank larger than a typical rank.

It is known that the typical rank in  $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$  is 2. However, in  $\mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R}^2$ , both 2 and 3 are typical ranks. More generally, in  $\mathbb{R}^n \otimes \mathbb{R}^m \otimes \mathbb{R}^p$ , every rank between the largest border rank and the largest rank is a typical rank. [9]

Let  $n = \dim(U)$ ,  $m = \dim(V)$ , and  $p = \dim(W)$ . Note that the set of rank one tensors in  $U \otimes V \otimes W$  has dimension  $n + m + p - 2$ , since it is parameterized by elements  $u, v, w$  up to scale, plus one scalar. Rank  $r$  tensors are sums of  $r$  rank one tensors, so from [10] we have

$$\dim(\sigma_r) \leq r(n + m + p - 2). \quad (11)$$

To approximate the typical rank over  $\mathbb{C}$ , since the dimension of  $\dim(U \otimes V \otimes W) = nmp$ , by substituting  $nmp$  into the left hand side of eq. (11) we conclude that if  $r$  is the typical rank, we have

$$\lceil \frac{nmp}{n+m+p-2} \rceil \leq r.$$

We define the *expected rank* of a tensor in  $\mathbb{C}^n \otimes \mathbb{C}^m \otimes \mathbb{C}^p$  to be  $\lceil \frac{nmp}{n+m+p-2} \rceil$ . In general, there is no easy way to calculate the true typical rank explicitly. However, it is known in some cases. For example, it is known that for all  $n \neq 3$ , the typical rank of an element of  $\mathbb{C}^n \otimes \mathbb{C}^n \otimes \mathbb{C}^n$  is the expected  $\lceil \frac{n^3}{3n-2} \rceil$ . When  $n = 3$ , the typical rank is five [8].

Typical ranks are a useful notion, because if we have an incomplete tensor  $T_\Omega$  with entries chosen from a continuous distribution, then  $T_\Omega$  will have a rank  $r$  completion with non-zero probability only if  $r$  is a typical rank.

## 6.2 Spaces of Tensors with Rank at Most $r$

The definitions and theorems from this section are from [5]. In contrast to the space of matrices with rank at most  $r$ ,  $\overline{\mathcal{M}}_r$ , because we have several notions of the rank of a tensor, we have several ways to define spaces of tensors of rank at most  $r$ . First, we let  $\sigma_r$  denote the space of tensors in  $U \otimes V \otimes W$  of rank at most  $r$ . That is,

$$\sigma_r = \{T \in U \otimes V \otimes W \mid \mathbf{R}(T) \leq r\}.$$

Let  $\hat{\sigma}_r$  denote the set of tensors of border rank at most  $r$  in  $U \otimes V \otimes W$ . That is,

$$\hat{\sigma}_r = \{T \in U \otimes V \otimes W \mid \underline{\mathbf{R}}(T) \leq r\}.$$

By definition 13,  $\hat{\sigma}_r$  is the closure of  $\sigma_r$ . More strongly, it is the Zariski closure of  $\sigma_r$ . In other words,  $\hat{\sigma}_r$  is an algebraic variety. That is, it is the zero set of a collection of polynomials. In general, the defining equations for  $\hat{\sigma}_r$  are unknown. We will introduce some known results. Other known results can be found in Chapter 7 of [5].

**Definition 15.** The subspace variety  $\hat{S}ub_r = \hat{S}ub_r(U \otimes V \otimes W)$  is the space of tensors such that each entry of the multilinear rank is at most  $r$ . In other words, a tensor  $T$  is in  $\hat{S}ub_r$  if and only if  $\text{rank}(T^{(i)}) \leq r$  for all  $i$ . So we have

$$\begin{aligned}\hat{S}ub_r &= \{T \in U \otimes V \otimes W \mid \max(\mathbf{R}_m(T)) \leq r\} \\ &= \{T \in U \otimes V \otimes W \mid \text{rank}(T^{(i)}) \leq r \forall i\}.\end{aligned}$$

Recall that a linear map  $M$  has rank less than or equal to  $r$  if and only if all  $(r+1) \times (r+1)$  minors of  $M$  vanish. Since we require  $\text{rank}(T^{(i)}) \leq r$  for all  $i$ , we have that a tensor  $T$  is in  $\hat{S}ub_r$  if and only if all  $(r+1) \times (r+1)$  minors of  $T^{(i)}$  vanish for each  $i$ . Therefore,  $\hat{S}ub_r$  is a zero set of a system of the system of polynomials which are the  $(r+1) \times (r+1)$  minors of all mode- $i$  unfoldings. The minors of a tensor  $T$  can be explicitly calculated by calculating the minors of  $T$  unfolded along each dimension.

What is the relationship between  $\sigma_r$ ,  $\hat{\sigma}_r$ , and  $\hat{S}ub_r$ ?

**Theorem 40.** We have the following inclusions describing the relationship between  $\sigma_r$ ,  $\hat{\sigma}_r$ , and  $\hat{S}ub_r$ :

$$\sigma_r \subset \hat{\sigma}_r \subset \hat{S}ub_r$$

which implies

$$\max(\mathbf{R}_m(T)) \leq \underline{\mathbf{R}}(T) \leq \mathbf{R}(T).$$

*Proof.* Recall that if the rank of a tensor  $T$  is at most  $r$ , this implies that the maximum multilinear rank is at most  $r$ . That is,  $\mathbf{R}(T) \leq r$  implies  $\max(\mathbf{R}_m(T)) \leq r$ . Therefore if  $\mathbf{R}(T) \leq r$ , then the  $(r+1) \times (r+1)$  minors of every mode- $i$  unfolding of  $T$  vanish. In other words, we have  $\sigma_r \subset \hat{S}ub_r$ . Moreover, since  $\hat{S}ub_r$  is the zero set of a system of polynomials,

it is a closed set, so the closure of  $\sigma_r$ , that is  $\hat{\sigma}_r$ , is also a subset of  $\hat{Sub}_r$ . So we have

$$\sigma_r \subset \hat{\sigma}_r \subset \hat{Sub}_r.$$

□

Since polynomials are continuous, then the limit of tensors in the zero set of a system of polynomials must also be a zero of that system of polynomials. It follows that the  $(r + 1) \times (r + 1)$  minors of the mode- $i$  unfolding of any tensor  $T$  with border rank at most  $r$  must vanish.  $T \in \hat{Sub}_r$  is a necessary, but not always a sufficient condition to imply  $T \in \hat{\sigma}_r$ . However, it is sufficient when  $r = 1$ . Moreover, it is sufficient to imply  $T \in \sigma_1$  [5]. That is, we have

$$\sigma_1 = \hat{\sigma}_1 = \hat{Sub}_1(U \otimes V \otimes W). \quad (12)$$

### 6.3 Tensor Rank Computation and Low-Rank Approximations

We will now discuss how to explicitly compute different tensor ranks.

The easiest type of rank to compute is the multilinear rank. We simply compute the rank of the mode- $i$  unfolding matrix  $T^{(i)}$  for all  $i$  by unfolding the tensor in all three directions.

The border rank is difficult to compute in practice, since  $\mathbf{R}(T) = \min\{\mathbf{R}(t), t \in B(T, \epsilon)\}$  for arbitrary  $\epsilon$ , where  $B(T, \epsilon)$  is the ball with center  $T$  and radius  $\epsilon$ . In theory, since the set of tensors of border rank at most  $r$  is an algebraic variety, then there exists a finite number of polynomial equations such that if  $T$  is in the zero set of all such equations, then  $T$  has border rank at most  $r$ . We could use these equations to find the border rank, however, these equations are not all known in general. For necessary conditions, we may check that the maximum component of the multilinear rank is at most  $r$ .

To compute the rank of a tensor  $T \in U \otimes V \otimes W$ , we need to find the smallest  $r$  such

that  $T$  can be written as a sum of  $r$  rank 1 tensors. For example, to check if an  $n \times n \times n$  tensor  $T$  can be written as a sum of  $r$  rank one tensors for some  $r$ , assume we can express  $T = \sum_{i=1}^r Z_i$ , where each  $Z_i = u_i \otimes v_i \otimes w_i$ , where the  $n \times 1$  vectors  $u_i$ ,  $v_i$ , and  $w_i$  have variable components. We expand  $\sum_{i=1}^r Z_i$  into an  $n \times n \times n$  tensor, which should have  $3nr$  variables, and set that equal to the known entries in  $T$ . This will give us  $n^3$  equations, which can be solved for example using Gröbner bases for example. If there is a solution, then  $T$  has rank at most  $r$ . If there is no solution then  $T$  has rank greater than  $r$ .

We may attempt to calculate a closest rank  $r$  approximation problem with respect to some tensor norm. That is, given  $T \in U \otimes V \otimes W$ , we would like to find a solution to

$$\min_{u_i, v_i, w_i} \left\| T - \sum_{i=1}^r u_i \otimes v_i \otimes w_i \right\|$$

*s.t.*  $u_i \in U, v_i \in V, w_i \in W$

i In other words, we would like to solve

$$\min_{\mathbf{R}(T_r) \leq r} \|T - T_r\|$$

with  $T_r \in \sigma_r$ . Note that in the case of matrices, this problem is solved with respect to any unitary invariant norm by zeroing out small singular values by the Eckart–Young theorem [25]. However, for the case of tensors with order at least three, this problem is ill-posed [11]. This is because  $\sigma_r$ , the set of tensors of rank at most  $r$ , is not a closed set. Recall from example 6.1 that we have the sequence of rank 2 tensors  $\{T_n\}_{n=1}^\infty$  which converges to a rank 3 tensor, where

$$T_n = n(u + \frac{1}{n}v) \otimes (u + \frac{1}{n}v) \otimes (u + \frac{1}{n}v) - n(u \otimes u \otimes u).$$

and  $\lim_{n \rightarrow \infty} T_n = T = u \otimes u \otimes v + u \otimes v \otimes u + v \otimes u \otimes u$ . Therefore,  $T$  does not have a closest

rank 2 approximation, since there exist rank 2 tensors arbitrarily close to  $T$ . Even worse, the norms of the individual rank one terms are unbounded. That is,

$$\lim_{n \rightarrow \infty} \left\| n \left( u + \frac{1}{n} v \right) \otimes \left( u + \frac{1}{n} v \right) \otimes \left( u + \frac{1}{n} v \right) \right\| = \infty$$

and

$$\lim_{n \rightarrow \infty} \|n(u \otimes u \otimes u)\| = \infty.$$

This is sometimes called the problem of diverging components [12]. One workaround when trying to compute a low-rank approximation is to impose additional constraints that bound the norms of the rank one components.



## 7 Tensor Completion

Similar to matrix completion, the problem of tensor completion is, given a partially known tensor  $T_\Omega$ , complete the unknown entries of  $T_\Omega$  subject to the constraint that the resulting tensor is low rank. There are multiple notions of low rank for tensors, for example the tensor could have low rank, low border rank, or low max multilinear rank.

Let  $U, V, W$  be vector spaces of dimension  $n, m, p$  respectively, and let  $\Omega$  denote the known index set of known entries of our tensor. Let  $T_\Omega$  denote a partially known tensor, and let  $P_\Omega : U \otimes V \otimes W \rightarrow U \otimes V \otimes W$  denote the projection of a tensor  $T$  such that  $P_\Omega(T)$  fixes known entries with indices in  $\Omega$ , and zeros out other entries. Also let  $\mathcal{A}_\Omega = P_\Omega^{-1}(T_\Omega)$  be the linear variety of any possible completion of  $T_\Omega$ . Then the goal of tensor completion is to find a solution to the minimization problem

$$\begin{aligned} \min_{T \in U \otimes V \otimes W} \text{rank}(T) \\ \text{s.t. } P_\Omega(T) = T_\Omega \end{aligned}$$

where the rank function is either the rank, border rank, or max multilinear rank of  $T$ . Similar to matrix completion, the problem of tensor completion is, given a partially known tensor  $T_\Omega$ , we would like to fill in the missing entries of  $T_\Omega$  such that the resulting tensor is low rank. We have introduced various notions of low rank for tensors. In particular, a tensor could have low rank, low border rank, or low maximum multilinear rank. The issue with the rank of a tensor is that the set of tensors of rank at most  $r$  is not closed, and the rank of a tensor is difficult to compute. The issue with the border rank is that the equations which define the space of tensors of border rank at most  $r$  are not completely known, and the border rank is difficult to compute. In this section, we will complete tensors with respect to the condition that the maximum component of their multilinear rank is small.

Recall that there is a relationship between rank, border rank, and multilinear rank. In particular,  $\sigma_r$  the set of tensors of rank at most  $r$ , is contained in  $\hat{\sigma}_r$ , the set of tensors of border rank at most  $r$ , which is contained in  $\hat{Sub}_r$ , the set of tensors of maximum multilinear rank at most  $r$ . Therefore, given  $T_\Omega$ , if there exists a rank at most  $r$  completion, and there exists a unique maximum multilinear rank  $r$  completion, then the rank at most  $r$  completion is unique.

More formally, we have the following theorem.

**Theorem 41.** *Given  $T_\Omega$ , suppose  $\mathcal{A}_\Omega \cap \sigma_r$  is non-empty. That is, there exists at least one rank at most  $r$  completion of  $T_\Omega$ . Suppose there exists a unique multilinear rank at most  $r$  completion  $T$  of  $T_\Omega$ . That is, suppose  $\mathcal{A}_\Omega \cap \hat{Sub}_r = \{T\}$ . Then  $\mathcal{A}_\Omega \cap \sigma_r = \{T\}$ . Similarly, if  $\mathcal{A}_\Omega \cap \hat{Sub}_r = \{T\}$  and  $\mathcal{A}_\Omega \cap \hat{\sigma}_r$  is non-empty, then  $\mathcal{A}_\Omega \cap \hat{\sigma}_r = \{T\}$ .*

*Proof.* Since  $\sigma_r \subset \hat{\sigma}_r \subset \hat{Sub}_r$ , we have

$$\mathcal{A}_\Omega \cap \sigma_r \subset \mathcal{A}_\Omega \cap \hat{\sigma}_r \subset \mathcal{A}_\Omega \cap \hat{Sub}_r = \{T\}.$$

Then by assumption since there is at least one border rank or multilinear rank at most  $r$  completion, we must have  $\mathcal{A}_\Omega \cap \sigma_r = \{T\}$  and  $\mathcal{A}_\Omega \cap \hat{\sigma}_r = \{T\}$ .  $\square$

In this section we will provide sufficient conditions for there to be a unique multilinear rank  $(r, r, r)$  completion of  $T_\Omega$ .

## 7.1 Exact Low Multilinear Rank Tensor Completion

We will introduce sufficient conditions for an incomplete degree 3 tensor  $T_\Omega$  to have a unique multilinear rank  $(r, r, r)$  completion.

First, we will introduce sufficient conditions for an incomplete tensor to have a unique multilinear rank  $(1, 1, 1)$  completion. In this case when  $r = 1$ , this is equivalent to an incomplete tensor having a unique rank 1 completion.

**Theorem 42.** *Given  $T_\Omega$ , suppose  $T_{i,1,1}$  is known for all  $i$ ,  $T_{1,j,1}$  is known for all  $j$ , and  $T_{1,1,k}$  is known for all  $k$ , and suppose  $T_{1,1,1} \neq 0$ . If  $\mathcal{A}_\Omega \cap \hat{S}ub_1(U \otimes V \otimes W)$  is non-empty, then  $T_\Omega$  has a unique rank 1 completion.*

*Proof.* Since  $\hat{S}ub_1(U \otimes V \otimes W) = \{T \in U \otimes V \otimes W \mid \text{rank}(T^{(j)}) \leq 1 \forall j\}$ , then the equations that furnish  $\hat{S}ub_1(U \otimes V \otimes W)$  are the zero sets of all  $2 \times 2$  minors of the mode-1, mode-2, and mode-3 unfoldings.

$$T^{(1)} : U^* \rightarrow V \otimes W$$

$$T^{(2)} : V^* \rightarrow U \otimes W$$

$$T^{(3)} : W^* \rightarrow U \otimes V.$$

In other words, the equations are the zero set of all  $2 \times 2$  minors of the unfoldings of  $T$  into a matrix along each mode. By assumption, the entry  $T_{ijk}$  in  $T_\Omega$  is known if exactly two or more indices in  $(i, j, k)$  are to 1.

First, we will show that we can recover all entries with one index in  $(i, j, k)$  equal to 1. That is, we can recover all entries of the form  $T_{1jk}$ ,  $T_{i1k}$ , or  $T_{ij1}$ .

Note that the equations of flattening give us equations of the form

$$\begin{aligned} \begin{vmatrix} T_{111} & T_{11k} \\ T_{1j1} & T_{1jk} \end{vmatrix} &= 0 \\ \begin{vmatrix} T_{111} & T_{11k} \\ T_{i11} & T_{i1k} \end{vmatrix} &= 0 \\ \begin{vmatrix} T_{111} & T_{1j1} \\ T_{i11} & T_{ij1} \end{vmatrix} &= 0 \end{aligned}$$

for all  $i, j$ , and  $k$ . Here the bottom-right entry of each matrix is unknown and all other entries are known. Moreover since  $T_{111} \neq 0$ , we may solve for each unknown entry which is equal to

$$\begin{aligned} T_{1jk} &= \frac{T_{11k}T_{1j1}}{T_{111}} \\ T_{i1k} &= \frac{T_{11k}T_{i11}}{T_{111}} \\ T_{ij1} &= \frac{T_{i11}T_{1j1}}{T_{111}}. \end{aligned}$$

Next, we may complete an arbitrary entry  $T_{ijk}$  by considering the equation of flattening

$$\begin{vmatrix} T_{111} & T_{1jk} \\ T_{i11} & T_{ijk} \end{vmatrix} = 0$$

and solving for  $T_{ijk} = \frac{T_{1jk}T_{i11}}{T_{111}}$ . □

Again, from eq. (12) we have  $\sigma_1 = \hat{\sigma}_1 = \hat{S}ub_1(U \otimes V \otimes W)$ . So  $T_\Omega$  also has a unique rank one completion which is equal to the constructed completion  $T$ .

**Example 7.1.** Consider  $\Omega$  as above, and suppose  $T_{ijk} = 1$  for all  $(i, j, k) \in \Omega$ . Then  $T_\Omega$  has the unique rank 1 completion, border rank 1 completion, and multilinear rank 1 completion  $T$ , where  $T_{ijk} = 1$  for all  $(i, j, k)$ .  $T$  may also be written in the rank one form

$$T = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

We now generalize this theorem for  $r \geq 1$ .

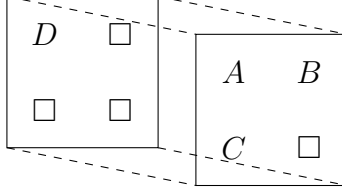


Figure 21: An incomplete tensor  $T_\Omega$ , with known subtensors  $A$ ,  $B$ ,  $C$ , and  $D$ .

**Theorem 43.** *Given  $T_\Omega$ , suppose that for an index  $(i, j, k)$ , if there are two or more of  $i$ ,  $j$ , or  $k$  less than or equal to  $r$ , then  $(i, j, k) \in \Omega$ . Let  $A$  denote the  $r \times r \times r$  known subtensor of  $T_\Omega$  consisting of entries with indices  $(i, j, k)$  where  $i \leq r$ ,  $j \leq r$ , and  $k \leq r$ . Suppose  $A$  has multilinear rank equal to  $(r, r, r)$ . Also, suppose  $\mathcal{A}_\Omega \cap \hat{S}ub_r$  is non-empty. Then  $T_\Omega$  has a unique multilinear rank  $(r, r, r)$  completion  $T \in \mathcal{A}_\Omega \cap \hat{S}ub_r$ . Moreover, if  $\mathcal{A}_\Omega \cap \sigma_r$  or  $\mathcal{A}_\Omega \cap \hat{\sigma}_r$  are non-empty, then  $T$  is a rank  $r$  or border rank  $r$  completion of  $T_\Omega$  respectively.*

*Proof.* Since  $\hat{S}ub_r = \{T \in U \otimes V \otimes W \mid \text{rank}(T^{(j)}) \leq r \forall j\}$ , the equations that furnish  $\hat{S}ub_r$  are the zero sets of all  $(r+1) \times (r+1)$  minors of each mode- $i$  unfolding

$$T^{(1)} : U^* \rightarrow V \otimes W$$

$$T^{(2)} : V^* \rightarrow U \otimes W$$

$$T^{(3)} : W^* \rightarrow U \otimes V.$$

In other words, the equations that furnish  $\hat{S}ub_r$  are the  $(r+1) \times (r+1)$  minors of the mode- $i$  unfoldings of  $T$  for all  $i$ .

Let  $T_\Omega$  be given as in theorem 43. Let  $B$  denote the known subtensor of  $T_\Omega$  consisting of entries with indices  $(i, j, k)$  such that  $i \leq r$ ,  $j > r$ , and  $k \leq r$ . Let  $C$  denote the known subtensor of  $T_\Omega$  consisting of entries with indices  $(i, j, k)$  such that  $i > r$ ,  $j \leq r$ , and  $k \leq r$ . Let  $D$  denote the known subtensor of  $T_\Omega$  consisting of entries with indices  $(i, j, k)$  such that

$i \leq r$ ,  $j \leq r$ , and  $k > r$ .

First, we will show that we may complete entries  $T_{ijk}$  where exactly one of  $j$  or  $k$  is less than or equal to  $r$  by using the mode-1 unfolding. Then, we will complete the rest of the entries by using the mode-2 unfolding. Since the known subtensor  $A$  has multilinear rank  $(r, r, r)$ , the mode-1 unfolding  $A^{(1)}$  has at least one rank  $r$  submatrix. Choose a rank  $r$  submatrix of  $A^{(1)}$ , and denote it  $A_J$ . Define  $J = \{(j_\alpha, k_\alpha)\}_{1 \leq \alpha \leq r}$  as the set of  $r$  pairs of indices such that the entry in position  $(i, \alpha)$  of  $A_J$  is equal to  $T_{ij_\alpha k_\alpha}$ .

Let  $G$  denote the subtensor of  $T$  with entries  $T_{ijk}$  such that  $k \leq r$ ,  $i > r$ , and  $j > r$ . Then each entry  $T_{ijk}$  of  $G$  is unknown, and there is a  $(r+1) \times (r+1)$  submatrix of the mode-1 flattening of  $T_\Omega$  of the form

$$\begin{bmatrix} A_J & b_{jk} \\ c_{iJ} & T_{ijk} \end{bmatrix}$$

where  $b_{jk} = [T_{ljk}]_{1 \leq l \leq r}$  is the  $r \times 1$  submatrix of  $B^{(1)}$  consisting of entries of the form  $T_{ljk}$ , with  $j$  and  $k$  fixed, and  $l$  ranging from 1 to  $r$ . Also  $c_{iJ} = [T_{ij_\alpha k_\alpha}]_{1 \leq \alpha \leq r}$  is the  $1 \times r$  submatrix of  $C^{(1)}$  where  $i$  is fixed, and  $(j_\alpha, k_\alpha) \in J$  with  $\alpha$  ranging from 1 to  $r$ . Since every  $(r+1) \times (r+1)$  minor of  $T^{(1)}$  must vanish, and since  $A_J$  is invertible, we may set the determinant of this submatrix equal to zero and solve for  $T_{ijk}$ , getting  $T_{ijk} = c_{iJ} A_J^{-1} b_{jk}$ , which completes  $G$ .

Let  $E$  denote the subtensor of  $T$  with entries  $T_{ijk}$  such that  $j \leq r$ ,  $i > r$ , and  $k > r$ . Then  $T_{ijk}$  is unknown, and there is a  $(r+1) \times (r+1)$  submatrix of the mode-1 flattening of  $T_\Omega$  of the form

$$\begin{bmatrix} A_J & d_{jk} \\ c_{iJ} & T_{ijk} \end{bmatrix}$$

where  $d_{jk} = [T_{ljk}]_{1 \leq l \leq r}$  is the  $r \times 1$  submatrix of  $D^{(1)}$  consisting of entries of the form  $T_{ljk}$ , with  $j$  and  $k$  fixed, and  $l$  ranging from 1 to  $r$ . Again, setting the determinant of this submatrix equal to zero we may solve for  $T_{ijk}$ , getting  $T_{ijk} = c_{iJ} A_J^{-1} d_{jk}$ , which completes  $E$ .

Now we consider the mode-2 unfolding of  $T_\Omega$ . Again, since  $A$  has multilinear rank  $(r, r, r)$ , the mode-2 unfolding  $A^{(2)}$  has at least one rank  $r$  submatrix. Choose a rank  $r$  submatrix of  $A^{(2)}$ , and denote it  $A_I$ . Define  $I = \{(i_\beta, k_\beta)\}_{1 \leq \beta \leq r}$  as the set of  $r$  pairs of indices such that the entry in position  $(j, \beta)$  of  $A_I$  is equal to  $T_{i_\beta j k_\beta}$ .

Let  $F$  denote the subtensor of  $T$  with entries  $T_{ijk}$  such that  $i \leq r$ ,  $j > r$ , and  $k > r$ . Then each entry  $T_{ijk}$  of  $F$  is unknown, and there is a  $(r+1) \times (r+1)$  submatrix of the mode-2 flattening of  $T_\Omega$  of the form

$$\begin{bmatrix} A_I & d_{ik} \\ b_{jI} & T_{ijk} \end{bmatrix}$$

where  $d_{ik} = [T_{ilk}]_{1 \leq l \leq r}$  is the  $r \times 1$  submatrix of  $D^{(2)}$  consisting of entries of the form  $T_{ilk}$ , with  $i$  and  $k$  fixed, and  $l$  ranging from 1 to  $r$ . Also  $b_{jI} = [T_{i_\beta j k_\beta}]_{1 \leq \beta \leq r}$  is the  $1 \times r$  submatrix of  $B^{(2)}$  where  $j$  is fixed, and  $(i_\beta, k_\beta) \in I$  with  $\beta$  ranging from 1 to  $r$ . Since every  $(r+1) \times (r+1)$  minor of  $T^{(2)}$  must vanish, and since  $A_I$  is invertible, we may set the determinant of this submatrix equal to zero and solve for  $T_{ijk}$ , getting  $T_{ijk} = b_{jI} A_I^{-1} d_{ik}$  which completes  $F$ .

Finally, let  $H$  denote the subtensor of  $T$  with entries  $T_{ijk}$  such that  $i > r$ ,  $j > r$ , and  $k > r$ . Then each entry  $T_{ijk}$  of  $H$  is unknown, and there is a  $(r+1) \times (r+1)$  submatrix of the mode-2 flattening of  $T_\Omega$  of the form

$$\begin{bmatrix} A_I & e_{ik} \\ b_{jI} & T_{ijk} \end{bmatrix}$$

where  $e_{ik} = [T_{ilk}]_{1 \leq l \leq r}$  is the  $r \times 1$  submatrix of  $E^{(1)}$  consisting of entries of the form  $T_{ilk}$ , with  $i$  and  $k$  fixed, and  $l$  ranging from 1 to  $r$ . Again setting the determinant of this submatrix equal to zero we may solve for  $T_{ijk}$ , getting  $T_{ijk} = b_{jI} A_I^{-1} e_{ik}$ , which completes  $H$ , and finishes completing  $T_\Omega$ .

Note that every tensor in  $\mathcal{A}_\Omega \cap \hat{S}ub_r$  must be in the zero set of the system of equations used to solve for our completion  $T$ . Since each equation had a unique solution, each unknown entry is uniquely determined, and since  $\mathcal{A}_\Omega \cap \hat{S}ub_r$  is non-empty, our completion  $T$  exists and is unique. Moreover, since  $T \in \hat{S}ub_r$ , each component of the multilinear rank is at most  $r$ . Also, since  $A$  is a subtensor of  $T$ , and  $\mathbf{R}_m(A) = (r, r, r)$ , then each component of the multilinear rank of  $T$  is at least  $r$ , so  $\mathbf{R}_m(T) = (r, r, r)$

Also note that by theorem 41, if  $\mathcal{A}_\Omega \cap \sigma_r$  is non-empty, then  $T$  is a rank at most  $r$  completion of  $T_\Omega$ , and since  $\max(\mathbf{R}_m(T)) = r \leq \mathbf{R}(T) \leq r$ , the rank of  $T$  is equal to  $r$ . Similarly, if  $\mathcal{A}_\Omega \cap \hat{\sigma}_r$  is non-empty, then  $T$  is a border rank  $r$  completion of  $T_\Omega$ .  $\square$

In total  $r^2(n + m + p) - r^3$  of  $nmp$  entries are known. This is an improvement from [7] in which  $O(nr^2 + r^4)$  known entries are required. This gives us the following corollary.

**Corollary 2.** *Given an  $n \times m \times p$  partially known tensor  $T_\Omega$ , if the positions of the entries in  $\Omega$  are distributed correctly, and  $\mathcal{A}_\Omega \cap \hat{S}ub_r$  is non-empty, then knowing  $r^2(n + m + p) - r^3$  entries in  $T_\Omega$  is a sufficient condition for  $T_\Omega$  to have a unique completion in  $\hat{S}ub_r$ .*

Note that there were entries in  $A$  that we did not use to compute unknown entries, we only used entries in  $A_J$  and  $A_I$ . Moreover, we did not use the fact that the mode-3 unfolding of  $A$  is rank  $r$ , so it should be possible to express a similar unique tensor completion theorem under weaker assumptions.

We will now construct an explicit multilinear rank  $(2, 2, 2)$  completion as an example of theorem 43.

**Example 7.2.** *Consider the partially known  $3 \times 3 \times 3$  tensor  $T_\Omega$  from fig. 22. Note that any way to unfold  $T_\Omega$  results in an incomplete row or column. Therefore unfolding  $T_\Omega$  in one way is not sufficient to complete  $T_\Omega$ , we must unfold  $T_\Omega$  in multiple ways.*

*We will construct a multilinear rank  $(2, 2, 2)$  completion  $T$  of  $T_\Omega$ . Note that the subtensor,  $A = [T_{ijk}]$  with  $i, j, k \leq 2$ , is fully known and has multilinear rank  $(2, 2, 2)$ . Moreover,  $T_{ijk}$*



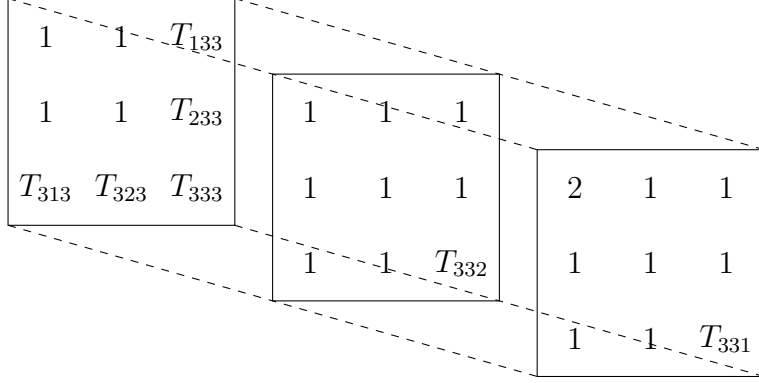


Figure 22:  $3 \times 3 \times 3$  incomplete tensor  $T_\Omega$  from example 7.2 with unknown entries  $T_{ijk}$ .

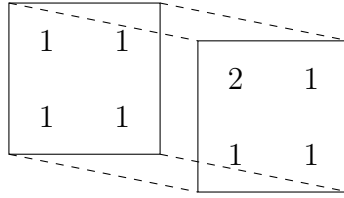


Figure 23: The  $2 \times 2 \times 2$  known subtensor  $A$  of  $T_\Omega$  with multilinear rank  $(2, 2, 2)$ .

is known if two or more of  $i, j, \text{ or } k$  are at most 2. Therefore, we may use theorem 43 to attempt to construct a unique multilinear rank  $(2, 2, 2)$  completion of  $T_\Omega$ .

Consider the mode-1 unfoldings of  $T_\Omega$  and  $A$

$$T_\Omega^{(1)} = \left[ \begin{array}{ccc|ccc|ccc} 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & T_{133} \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & T_{233} \\ 1 & 1 & T_{331} & 1 & 1 & T_{332} & T_{313} & T_{323} & T_{333} \end{array} \right]$$

$$A^{(1)} = \left[ \begin{array}{cc|cc} 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{array} \right]$$

First, we recover  $T_{331}, T_{332}, T_{313}$ , and  $T_{323}$  by solving the equations of the form

$$\begin{vmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & T_{ijk} \end{vmatrix} = 0 \quad (13)$$

getting  $T_{ijk} = 1$ . Next, we consider the mode-2 unfolding of  $T_\Omega$ ,

$$T_\Omega^{(2)} = \left[ \begin{array}{ccc|ccc|ccc} 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & T_{313} \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & T_{323} \\ 1 & 1 & T_{133} & 1 & 1 & T_{233} & T_{331} & T_{332} & T_{333} \end{array} \right].$$

Filling in the completed entries from the mode-1 unfolding, we get

$$T^{(2)} = \left[ \begin{array}{ccc|ccc|ccc} 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & T_{133} & 1 & 1 & T_{233} & 1 & 1 & T_{333} \end{array} \right].$$

We may now recover entries  $T_{133}, T_{233}$ , and  $T_{333}$  similarly to eq. (13). There also exists a rank two completion of  $T_\Omega$ , and so our completed tensor  $T$  is also rank two, and border rank two. That is, we have

$$T = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

If  $T_\Omega$  may be permuted to the appropriate form, theorem 43 gives a constructive way to complete  $T_\Omega$  to a multilinear rank  $(r, r, r)$  completion, assuming one exists. Here we will more explicitly express this tensor completion algorithm. Recall that  $T$  is partitioned into the following subtensors.

Subtensor of $T$	Indices $T_{ijk}$ with
$A$	$i \leq r, j \leq r, k \leq r$
$B$	$i \leq r, j > r, k \leq r$
$C$	$i > r, j \leq r, k \leq r$
$D$	$i \leq r, j \leq r, k > r$
$E$	$i > r, j \leq r, k > r$
$F$	$i \leq r, j > r, k > r$
$G$	$i > r, j > r, k \leq r$
$H$	$i > r, j > r, k > r$

Then if  $T_\Omega$  satisfies the assumptions under theorem 43, we have the following algorithm to complete  $T_\Omega$  into a multilinear rank  $(r, r, r)$  tensor. Suppose  $T_\Omega$  is a  $n \times m \times p$  partially known tensor.

1. Find a full rank  $r \times r$  submatrix  $A_J$  of  $A^{(1)}$ . Define  $J = \{(j_\alpha, k_\alpha)\}_{1 \leq \alpha \leq r}$  as the set of  $r$  pairs of indices such that the entry in position  $(i, \alpha)$  of  $A_J$  is equal to  $T_{ij_\alpha k_\alpha}$ .
2. Let  $C_J$  denote the  $(n - r) \times r$  submatrix of  $C^{(1)}$  such that the entry in position  $(i, \alpha)$  of  $C_J$  is equal to the entry of  $T$  in position  $(i + r, j_\alpha, k_\alpha)$ .
3. Set  $G^{(1)} = C_J A_J^{-1} B^{(1)}$ .
4. Set  $E^{(1)} = C_J A_J^{-1} D^{(1)}$ .
5. Refold  $E^{(1)}$  and  $G^{(1)}$  to obtain  $E$  and  $G$ .
6. Find a full rank  $r \times r$  submatrix  $A_I$  of  $A^{(2)}$ . Define  $I = \{(i_\beta, k_\beta)\}_{1 \leq \beta \leq r}$  as the set of  $r$  pairs of indices such that the entry in position  $(j, \beta)$  of  $A_I$  is equal to  $T_{i_\beta j k_\beta}$ .
7. Let  $B_I$  denote the  $(m - r) \times r$  submatrix of  $B^{(2)}$  where the entry in position  $(j, \beta)$  of  $B_I$  is equal to the entry of  $T$  in position  $(i_\beta, j + r, k_\beta)$ .

8. Set  $F^{(2)} = B_I A_I^{-1} D^{(2)}$ .

9. Set  $H^{(2)} = B_I A_I^{-1} E^{(2)}$ .

10. Refold  $F^{(2)}$  and  $H^{(2)}$  to obtain  $F$  and  $H$ , and assemble  $E, F, G$ , and  $H$  to complete  $T$ .

Do analogous algorithms work to complete higher order tensors? We conjecture yes.

**Conjecture 1.** *Given an order  $d > 3$  partially known  $n_1 \times \cdots \times n_d$  tensor  $T_\Omega$  and given an index  $(i_1, \dots, i_d)$ , if there are  $d - 1$  or more entries less than or equal to  $r$ , then  $(i_1, \dots, i_d)$  is in  $\Omega$ . Suppose the known subtensor  $A = [T_{i_1, \dots, i_d}]$ ,  $i_j \leq r$  for all  $j$  has multilinear rank equal to  $(r, \dots, r)$ , and suppose also that there exists at least one multilinear rank  $(r, \dots, r)$  completion of  $T_\Omega$ . Then  $T_\Omega$  has a unique multilinear rank  $(r, \dots, r)$  completion.*

This means that in total  $(\sum_{i=1}^d n_i)r^{d-1} - (d-1)r^d$  entries are known out of  $\prod_{i=1}^d n_i$  entries total. In this case, the expected way to recover  $T$  from  $T_\Omega$  is to first complete entries with index of the form  $(i_1, \dots, i_d)$  where at least  $d - 2$  entries are at most  $r$ , then recover entries with indices where at least  $d - 3$  entries are at most  $r$ , and so on.

## 7.2 Generalizing the Schur Gradient Descent Method to Tensors

We now generalize the Schur gradient descent method from section 4 to the case of degree three tensors. Let  $T_\Omega$  be a partially known tensor. Suppose for an index  $(i, j, k)$ , if each of  $i, j$ , and  $k$  are at most  $r$ , then  $(i, j, k) \in \Omega$ . Call  $A$  the known subtensor  $A = [T_{ijk}]$  where each of  $i, j$ , and  $k$  are at most  $r$ . Our goal is to find a tensor  $T$  of multilinear rank  $(r, r, r)$  such that  $P_\Omega(T) = T_\Omega$ . We will cast this as a minimization problem.

Recall that  $T \in U \otimes V \otimes W$  can be unfolded in the following three ways as linear maps

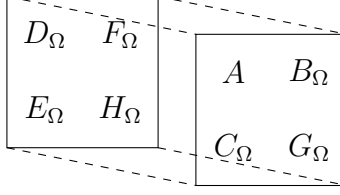


Figure 24: Incomplete tensor  $T_\Omega$ , with known  $r \times r \times r$  known subtensor  $A$  with multilinear rank  $(r, r, r)$ , and partially known subtensors  $B_\Omega, C_\Omega, D_\Omega, E_\Omega, F_\Omega, G_\Omega, H_\Omega$ .

$$T^{(1)} : U^* \rightarrow V \otimes W$$

$$T^{(2)} : V^* \rightarrow U \otimes W$$

$$T^{(3)} : W^* \rightarrow U \otimes V$$

If  $T$  has multilinear rank  $(r, r, r)$ , this means that each of these linear maps has rank  $r$ . Similarly, we have the mode-1, mode-2, and mode-3 unfoldings of  $A$ . By assumption,  $A$  has multilinear rank  $(r, r, r)$ , which means each of these linear maps has full rank  $r$ , and thus each one contains an  $r \times r$  invertible submatrix. We run maxvol on each of these three flattenings of  $A$  and record the resulting dominant submatrices with non-zero determinant denoted  $A_{\square}^{(1)}$ ,  $A_{\square}^{(2)}$ , and  $A_{\square}^{(3)}$  respectively.

In terms of the Schur complement, we want to find a solution to the minimization problem

$$\begin{aligned} \min_T & \left( \frac{1}{2} \left\| S_{A_{\square}^{(1)}}^{(1)} \right\|^2 + \frac{1}{2} \left\| S_{A_{\square}^{(2)}}^{(2)} \right\|^2 + \frac{1}{2} \left\| S_{A_{\square}^{(3)}}^{(3)} \right\|^2 \right) \\ & s.t. \quad P_\Omega(T) = T_\Omega \end{aligned} \tag{14}$$

where  $S_{A_{\square}^{(i)}}^{(i)}$  denotes the Schur complement of  $T^{(i)}$  with respect to  $A_{\square}^{(i)}$ . Note that the objective function is the sum of squares of norms, so it is non-negative. Therefore, if the objective

function is equal to zero,  $T$  is a minimizer.

**Theorem 44.** *Given a tensor  $T$  such that  $P_\Omega(T) = T_\Omega$ , the objective function in minimization 14 is equal to zero if and only if the multilinear rank of  $T$  is equal to  $(r, r, r)$ .*

*Proof.*  $T$  has multilinear rank  $(r, r, r)$  if and only if  $\text{rank}(T^{(i)}) = r$  for all  $i$ , if and only if the Schur complement  $T^{(i)}/A_{\square}^{(i)} = 0$  for all  $i$ , if and only if the objective function is equal to zero. □

To calculate a minimizer  $T$ , we may employ a gradient descent method similarly to the Schur gradient descent method in section 4.

### 7.3 Algebraic Combinatorics of Low-Rank Tensor Completion

A hypergraph is a pair  $(V, E)$  where  $V$  are the vertices and  $E \subset \mathcal{P}(V)$  are the hyperedges which consist of any number of vertices. Here  $\mathcal{P}(V)$  denotes the power set of  $V$ . A 3-partite hypergraph is a hypergraph where the vertices are partitioned into 3 sets, and each hyperedge contains one vertex from each set. We may model  $\Omega$  as a 3-partite hypergraph  $H_\Omega = (V_\Omega, E_\Omega)$ . Suppose  $\Omega$  is an  $n \times m \times p$  binary tensor, where a 1 in entry  $(i, j, k)$  means that corresponding entry is known, and a 0 in entry  $(i, j, k)$  means corresponding that entry is unknown. Let  $V_\Omega$  consist of three groups of vertices

$$V_\Omega = \{x_1, \dots, x_n\} \cup \{y_1, \dots, y_m\} \cup \{z_1, \dots, z_p\}$$

Suppose there is an edge  $\{x_i, y_j, z_k\} \in E_\Omega$  if and only if entry  $(i, j, k)$  in  $\Omega$  is equal to one. Then there is a one to one correspondence between 3-partite hypergraphs and 3rd order masks  $\Omega$  by mapping  $\Omega$  to  $H_\Omega$ . Moreover, the adjacency tensor of  $H_\Omega$  is equal to  $\Omega$ .

We may generalize some of the notions from the combinatorics of matrix completion in section 2.8 to the combinatorics of tensor completion.

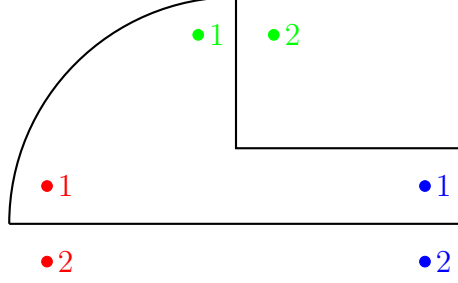


Figure 25: Hypergraph of indices of a  $2 \times 2 \times 2$  tensor with one hyperedge corresponding to the known entry  $(1, 1, 1)$  in  $\Omega$ .

**Definition 16.** Given a mask  $\Omega$ , we say an index  $(i, j, k)$  in  $\Omega^c$  is finitely completable in  $r$  if entry  $(i, j, k)$  of the partially known tensor  $P_\Omega(T)$  has finitely many completions for generic  $T \in \hat{Sub}_r$ . We define the rank  $r$  finitely completable closure  $\text{cl}_r(\Omega)$  as the set of indices which are finitely completable in  $\Omega$ .

In terms of the rank  $r$  finitely completable closure  $\text{cl}_r(\Omega)$ , we may reformulate theorem 43 as follows.

**Theorem 45.** Given  $\Omega$ , suppose if there are two or more entries of  $(i, j, k)$  less than or equal to  $r$ , then  $(i, j, k) \in \Omega$ . Then  $\text{cl}(\Omega)$  is equal to  $\Omega \cup \Omega^c$ .

*Proof.* Given  $\Omega$ , suppose if there are two or more entries of  $(i, j, k)$  less than or equal to  $r$ , then  $(i, j, k) \in \Omega$ . We have shown in theorem 43 that for  $T \in \hat{Sub}_r$ , if the subtensor  $A = [T_{ijk}]$ ,  $1 \leq i, j, k \leq r$  has multilinear rank equal to  $(r, r, r)$ , then  $P_\Omega(T)$  can be uniquely completed to  $T$ . Note that for a generic  $T$ ,  $A$  is also generic, which implies any mode- $i$  unfolding  $A^{(i)}$  of  $A$  is full rank  $r$ . So  $A^{(i)}$  will contain a rank  $r$  submatrix almost surely for all  $i$ . Therefore, the subtensor  $A$  will have multilinear rank  $(r, r, r)$  almost surely. So a generic tensor  $T$  will have subtensor  $A$  with multilinear rank equal to  $(r, r, r)$ , which means  $P_\Omega(T)$  can be uniquely completed to  $T$ , and so every entry is finitely completable.  $\square$

We may also generalize the unique completability closure to the case of tensors.

**Definition 17.** *Similarly to the rank  $r$  finitely completable closure  $\text{cl}_r(\Omega)$ , we define the rank  $r$  uniquely completable closure  $\text{ucl}_r(\Omega)$  as the set of positions which are uniquely completable in  $\Omega$  with respect to  $\hat{S}ub_r$  for a generic tensor  $T \in \hat{S}ub_r$ .*

Moreover, from theorem 43 we have that the uniquely completable closure equals the finitely completable closure.

**Theorem 46.** *Given  $\Omega$  as in theorem 43, then we have  $\text{ucl}_r(\Omega) = \text{cl}_r(\Omega) = \Omega \cup \Omega^c$ .*

Since we have already shown that for a generic tensor  $T \in \hat{S}ub_r$ ,  $P_\Omega(T)$  will have a unique completion in  $\Omega$ , the result that  $\text{ucl}_r(\Omega) = \text{cl}_r(\Omega) = \Omega \cup \Omega^c$  follows.



## 8 Notation and Glossary

- $M_{n \times m}$  is the set of matrices with  $n$  rows and  $m$  columns over  $\mathbb{R}$  or  $\mathbb{C}$ .
- $M^\top$  is the transpose of the matrix  $M$ .
- $M^*$  is the conjugate transpose of the matrix  $M$ .
- $M_{ij}$  is the entry of the matrix  $M$  with index  $(i, j)$ .
- $T_{ijk}$  is the element in index  $(i, j, k)$  of the 3rd order tensor  $T$ .
- $M(i, :)$  is the  $i$ th row of the matrix  $M$ .
- $M(:, j)$  is the  $j$ th column of the matrix  $M$ .
- $[n] = \{1, \dots, n\}$  is the set of integers 1 through  $n$ .
- $M_{I,J}$  is the submatrix of  $M$  specified by the sets of indices  $I \subset [n]$ ,  $J \subset [m]$ .
- $\mathcal{M}_r$  is the set of  $n \times m$  matrices with rank equal to  $r$ .
- $\overline{\mathcal{M}}_r$  is the set of  $n \times m$  matrices with rank at most  $r$ .
- $\Omega$  is a matrix or tensor mask which consists of the index set of known entries. It may also be considered as a binary matrix or tensor with a 1 corresponding to a known entry, and a 0 corresponding to an unknown entry.
- $P_\Omega(X)$  is the projection of any matrix or tensor  $X$  obtained by setting entries with indices not in  $\Omega$  equal to zero.
- $\Phi_\Omega(M) : \overline{\mathcal{M}}_r \rightarrow M_{n \times m}$  is the restriction of  $P_\Omega$  to the set  $\overline{\mathcal{M}}_r$ . In other words,  $\Phi_\Omega(M)$  is the projection of a matrix  $M \in \overline{\mathcal{M}}_r$  obtained by setting entries with indices not in  $\Omega$  equal to zero.

- $M_\Omega$  is a partially known matrix with known entries  $M_{ij}$  in index  $(i, j) \in \Omega$  and zeros elsewhere.
- $T_\Omega$  is a partially known 3rd order tensor with known entries  $T_{ijk}$  in index  $(i, j, k) \in \Omega$  and zeroes elsewhere.
- $\mathcal{A}_\Omega = \{X \in M_{n \times m} \mid P_\Omega(X) = M_\Omega\}$  is the linear affine space of all possible completions of  $M_\Omega$  or  $T_\Omega$ .
- $\det(M)$  is the determinant of the matrix  $M$ .
- $\text{vol}(M)$  is the volume of the matrix  $M$ , which is equal to the absolute value of the determinant of  $M$ .
- $\text{rank}(M)$  is the rank of the matrix  $M$ .
- $\sigma_i$  is the  $i$ th singular value of a matrix  $M$ .
- $\|X\|$  is the Euclidean norm of the matrix or tensor  $X$  unless otherwise specified.
- $|S|$  is the cardinality of the set  $S$ .
- $\text{diag}(x_1, \dots, x_n)$  is the diagonal matrix with diagonal entries equal to  $x_1, \dots, x_n$ .
- $\square$  is a missing element of a matrix. In the context of graphs,  $\square$  is used to denote the graph Cartesian product.
- $A_\square$  is a dominant submatrix of a matrix  $M$ .
- $A_\blacksquare$  is a maximum volume submatrix of a matrix  $M$ .
- $S_A$  is the Schur complement of the matrix  $M$  with respect to a submatrix  $A$ .
- $M^+$  is the psuedoinverse of the matrix  $M$ .

- $\text{vec}(X)$  is the vectorization of the matrix or tensor  $X$ .
- $\mathbf{R}(T)$  is the rank of the tensor  $T$ .
- $\underline{\mathbf{R}}(T)$  is the border rank of the tensor  $T$ .
- $\mathbf{R}_m(T)$  is the multilinear rank of the tensor  $T$ .
- $\mathbf{R}_g(T)$  is the generic rank of the tensor  $T$ .
- $\sigma_r$  is the set of tensors in  $U \otimes V \otimes W$  with rank at most  $r$ .
- $\hat{\sigma}_r$  is the set of tensors in  $U \otimes V \otimes W$  with border rank at most  $r$ .
- $\hat{S}ub_r$  is the set of tensors in  $U \otimes V \otimes W$  such that each component of the multilinear rank is at most  $r$ .

## References

- [1] W. Fulton, Intersection Theory, Springer-Verlag, 1984.
- [2] Hartshorne, Robin. Algebraic geometry. Vol. 52. Springer Science & Business Media, 2013.
- [3] Hatcher, Allen. Algebraic topology. Cambridge University Press, Cambridge, 2005.
- [4] Guillemin, Victor, and Alan Pollack. Differential topology. Vol. 370. American Mathematical Soc., 2010.
- [5] J. Landsberg, Tensors: Geometry and Applications, AMS, 2012.
- [6] Song, Qingquan, et al. "Tensor completion algorithms in big data analytics." ACM Transactions on Knowledge Discovery from Data (TKDD) 13.1 (2019): 1-48.

- [7] Cai, Jian-Feng, et al. "Provable Near-Optimal Low-Multilinear-Rank Tensor Recovery." arXiv preprint arXiv:2007.08904 (2020).
- [8] Lickteig, Thomas. "Typical tensorial rank." *Linear algebra and its applications* 69 (1985): 95-120.
- [9] Friedland, Shmuel. "On the generic and typical ranks of 3-tensors." *Linear algebra and its applications* 436.3 (2012): 478-497.
- [10] Strassen, Volker. "Rank and optimal computation of generic tensors." *Linear algebra and its applications* 52 (1983): 645-685.
- [11] De Silva, Vin, and Lek-Heng Lim. "Tensor rank and the ill-posedness of the best low-rank approximation problem." *SIAM Journal on Matrix Analysis and Applications* 30.3 (2008): 1084-1127.
- [12] Stegeman, Alwin. "Low-rank approximation of generic  $p \times q \times 2$  arrays and diverging components in the candecomp/parafac model." *SIAM Journal on Matrix Analysis and Applications* 30.3 (2008): 988-1007.
- [13] Kolda, Tamara G., and Brett W. Bader. "Tensor decompositions and applications." *SIAM review* 51.3 (2009): 455-500.
- [14] Landsberg, Joseph M., and Mateusz Michałek. "Abelian tensors." *Journal de Mathématiques Pures et Appliquées* 108.3 (2017): 333-371.
- [15] A.S. Lewis and J. Malick. Alternating projections on manifolds. *Mathematics of Operations Research*, 33(1):216–234, 2008.
- [16] M. J. Lai, A. Varghese, "On convergence of the alternating projection method for matrix completion and sparse recovery problems", 2017, [online] Available: arXiv:1711.02151.

- [17] Wang, Z., Lai, M., Lu, Z., Fan, W., Davulcu, H., Ye, J.: Orthogonal rank-one matrix pursuit for low rank matrix completion. *SIAM J. Scientific Computing* 37(1) (2015).
- [18] Prateek Jain and Praneeth Netrapalli and Sujay Sanghavi Low-rank Matrix Completion using Alternating Minimization, 2012
- [19] Cai, Jian-Feng, Emmanuel J. Candès, and Zuowei Shen. "A singular value thresholding algorithm for matrix completion." *SIAM Journal on optimization* 20.4 (2010): 1956-1982.
- [20] W. Burnside, *Theory of Groups of Finite Order*, Cambridge University Press, 1911.
- [21] S. A. Goreinov, I. V. Oseledets, D. V. Savostyanov, E. E. Tyrtyshnikov, and N. L. Zamarashkin, How to find a good submatrix, in: *Matrix Methods: Theory, Algorithms, Applications*, (World Scientific, Hackensack, NY, 2010), pp. 247–256.
- [22] Çivril, Ali, and Malik Magdon-Ismael. "On selecting a maximum volume submatrix of a matrix and related problems." *Theoretical Computer Science* 410.47-49 (2009): 4801-4811.
- [23] Fazel, Sarjoui M. "Matrix rank minimization with applications." (2003): 1981-1981.
- [24] Király, Franz J., Louis Theran, and Ryota Tomioka. "The algebraic combinatorial approach for low-rank matrix completion." *The Journal of Machine Learning Research* 16.1 (2015): 1391-1436.
- [25] Eckart, Carl, and Gale Young. "The approximation of one matrix by another of lower rank." *Psychometrika* 1.3 (1936): 211-218.
- [26] Harris, Joe, and Loring W. Tu. "On symmetric and skew-symmetric determinantal varieties." (1984).

- [27] Brouwer, Andries E., et al. "A new table of constant weight codes." *IEEE Transactions on Information Theory* 36.6 (2006): 1334-1380.
- [28] Brouwer, Andries E., and Tzvi Etzion. "Some new distance-4 constant weight codes." *Advances in Mathematics of Communications* 5.3 (2011): 417.
- [29] Brouwer, Andries E., and Willem H. Haemers. *Spectra of graphs*. Springer Science & Business Media, 2011.
- [30] Barik, Sasmita, Ravindra B. Bapat, and Sukanta Pati. "On the Laplacian spectra of product graphs." *Applicable Analysis and Discrete Mathematics* (2015): 39-58.
- [31] Vizing, V. G. The cartesian product of graphs. (Russian) *Vychisl. Sistemy* No. 9 1963 30-43.
- [32] Tu, Jonathan H. *Dynamic mode decomposition: Theory and applications*. Diss. Princeton University, 2013.
- [33] Castro-González, N., M. F. Martínez-Serrano, and J. Robles. "Expressions for the Moore–Penrose inverse of block matrices involving the Schur complement." *Linear Algebra and its Applications* 471 (2015): 353-368.
- [34] Bonnin, Xavier, et al. "Presentation of the new SOLPS-ITER code package for tokamak plasma edge modelling." *Plasma and Fusion Research* 11 (2016): 1403102-1403102.
- [35] Ding, Jiu, and L. J. Huang. "On the continuity of generalized inverses of linear operators in Hilbert spaces." *Linear algebra and its applications* 262 (1997): 229-242.
- [36] @MISC 3829924, TITLE = Calculating the matrix derivative  $\frac{\partial}{\partial X} \|CX^{-1}B\|_F^2$ ,  
AUTHOR = greg (<https://math.stackexchange.com/users/357854/greg>),  
HOWPUBLISHED = Mathematics Stack Exchange, NOTE =

URL:<https://math.stackexchange.com/q/3829924> (version: 2020-09-17), EPRINT = <https://math.stackexchange.com/q/3829924>, URL = <https://math.stackexchange.com/q/3829924>

[37] @MISC 375851, TITLE = Looking for a reference on the Euler characteristic of the manifold of fixed rank matrices, AUTHOR = Powerspawn (<https://mathoverflow.net/users/152336/powerspawn>), HOWPUBLISHED = MathOverflow, NOTE = URL:<https://mathoverflow.net/q/375851> (version: 2020-11-07), EPRINT = <https://mathoverflow.net/q/375851>, URL = <https://mathoverflow.net/q/375851>