On Convergence of the Alternating Projection Method for Matrix Completion and Sparse Recovery Problems

Ming-Jun Lai MJLAI@UGA.EDU
Department of Mathematics,
University of Georgia, Athens, GA, USA 30602.

Abraham Varghese AV@UGA.EDU
Department of Mathematics,
University of Georgia,
Athens, GA, USA, 30602

Editor:

Abstract

This paper study the convergence of the Alternating Projection algorithm for matrix completion and compressed sensing problems and give sufficient condition under which it would converge. Numerical evidence to demonstrate the efficacy and robustness of the algorithm is presented. In the final section, using algebraic-geometric techniques, we prove that, fixing the known positions, if a rank-r matrix can be completed only in *finitely* many ways with one set of known entries, then for almost all set of known entries, the matrix can be completed into a rank-r matrix only in *finitely* many ways.

Keywords: Matrix Completion, , Matrix Recovery, Compressed Sensing, Sparse Recovery, Alternating Projection

1. Introduction

The last two decades have witnessed a resurgence of research in sparse solutions of underdetermined linear systems and matrix completion and recovery. The matrix completion problem was inspired by Netflix problem (cf. Bennett et al. (2007)) and was pioneered by Candès, Recht, 2010 [Candès and Recht (2009)] and Candès and Tao, 2010 [Candès and Tao (2010)]. The problem can be explained as follows. One would like to recovery a matrix $M \in \mathbb{R}^{m,n}$ from a given set of entries $M_{ij}, (i,j) \in \Omega \subset \{1,\cdots,m\} \times \{1,\cdots,n\}$ by filling in the missing entries such that the resulting matrix has the lowest possible rank. In other words, we solve the following rank minimization problem:

$$\min_{X \in \mathbb{R}^{m\times n}} \quad \text{rank } (X): \quad \text{such that} \quad \mathcal{A}_\Omega(X) = \mathcal{A}_\Omega(M), \tag{1}$$

where $\mathcal{A}_\Omega(X) = \mathcal{A}_\Omega(M)$ means the entries of the matrix X are the same entries of matrix M for indices $(i,j) \in \Omega$. Clearly, if we are only given a few entries, say one entry of matrix M of size 2×2, we are not able to recover M even assuming the rank of M is 1. There are necessary conditions on how many entries one must know in order to be able to recover M. Information theoretic lower bound can be found in Candès and Tao (2010).

There are many approaches to recovery such a matrix developed in the last ten years. One popular approach is to find a matrix with minimal summation of its singular values. That is,

$$\min_{X \in \mathbb{R}^{m\times n}} \{\|X\|_*, \qquad \mathcal{A}_\Omega(X) = \mathcal{A}_\Omega(M)\}, \tag{2}$$

where $\|X\|_* = \sum_{i=1}^{k} \sigma_i(X)$ is the nuclear norm of X with $k = \min\{m, n\}$ and $\sigma_i(X)$ are singular values of matrix X. It is known that $f(X) = \|X\|_*$ is a convex function of X, the above problem (2) is a convex minimization problem. By adding $\frac{1}{\lambda}\|X\|_F$ to the minimizing functional in (2), the resulting minimization problem can be solved by using Uzawa type algorithms in [Cai, Candès, Shen, 2010Cai et al. (2010)] or solved by using its dual formulation, e.g. in [Lai and Yin, 2013Lai and Yin (2013). The minimization in (2) can also reformulated as a fixed point iteration and Nestrov's acceleration technique can be used. See [Ma, Goldfarb, Chen, 2011Ma et al. (2011)] and [Toh and Yun, 2010Toh and Yun (2010)]. This constrained minimization (2) is usually converted into an unconstrained minimization using Lagrange multiplier method or augmented Lagrange minimization method. The alternating direction method of multiplier (ADMM) can be used to complete a matrix. See [Tao and Yuan, 2011Tao and Yuan (2011)], and [Yang and Yuan, 2012Yang and Yuan (2013)], Many researchers have studied the matrix completion via variants of the constrained convex minimization approach.

Certainly, the rank completion is also studied by using other approaches. See [Jain, Meka and Dhillon, 2010Jain et al. (2010)] for singular value projection method and [Wen, Yin, Zhang, 2012Wen et al. (2012), [Tanner and Wei, 2016Tanner and Wei (2016)] for alternating least squares, the SOR approaches, steepest descent minimization approaches. See [Lai, Xu, Yin, 2013Lai et al. (2013)] for ℓ_q minimization approach for $q \in (0, 1)$. In addition, a greedy approach, e.g. orthogonal matching pursuit (OMP) and iterative hard thresholding approach can be used as well. See [Wang, Lai, Lu, and Ye, 2015 Wang et al. (2015)] and [Tanner and Wei, 2013Tanner and Wei (2013)]. Iteratively reweighted nuclean norm minimization, Riemannian conjugated gradient method, and alternating projection algorithm in [Mohan and Fazel, 2012Mohan and Fazel (2012)], [Vandereycken, 2013Vandereycken (2013)], [Cai, Wang, and Wei, 2016Cai et al. (2017)], [Wei, Cai, Chan, and Leung, 2016Wei et al. (2016)], [Jiang, Zhong, Liu, and Song, 2017Jiang et al. (2017), and etc.. Among all these algorithms, the computational algorithm proposed in Wang et al. (2015) seems the most efficient one in completing an incomplete matrix. However, the accuracy of the completed matrices is still a question. Usually, the researchers use the relative Frobenius norm errors, i.e., $\|M - M_k\|_F/\|M\|_F$ to measure the accuracy for a given matrix M of size $m \times n$, where M_k is the kth iteration from a matrix completion algorithm. When the size of M is very large and so is $\|M\|_F$ and the missing rate $1 - |\Omega|/(mn) << 1$ small or $|\Omega| \approx mn$, the relative Frobenous norm error will be very small anyway and hence will not be a good measure of errors, where Ω is the set of the indices of known entries. It is better to use a true error such as the maximum norm of all entries of the residual matrix $M - M_k$ to check the accuracy of the completed matrices. Then many of the existing algorithms mentioned above will fail to produce a good recovery. Certainly, the main possible reason may be that the relaxation of rank minimization problem is used. One of the mathematical problems is to find sufficient conditions which ensure the uniqueness of the minimization. However, some sufficient conditions are unrealistic, e.g. only one entry is missing. Another research problem is to design efficient matrix completion algorithms. It is interesting to have an algorithm which is convergent fast, say in a linear fashion.

Recently, the authors just discovered the reference Jiang et al. (2017) which presents a numerical study of a computational algorithm called Alternating Projection(AP) Algorithm which is exactly the algorithm that the authors of this paper have studied for a year. Mainly, we also had the same algorithm and was aware of the good numerical performance. However, we would like to know why the algorithm is convergent, under what kind of conditions the algorithm is convergent, and under what situation the convergence is linear. The study took many months and delayed our announcement of the AP algorithm. Nevetheless, the purpose of this paper is to explain why and when the AP Algorithm will converge and the convergence is linear. In addition, we shall explain the existence of matrix completion and how many matrices can be completed to have the same given entries. In general, for randomly chosen values for a fixed location set Ω to be known entries of a matrix, one will

not be able to complete it by using a rank r matrix. Hence, we shall discuss the convergence of the AP algorithm under the assumption that the given entries are from a matrix with rank r. Also, we will provide an approach to choose a good initial guess such that the AP Algorithm will converge faster than using the simple straightforward initial guess M_Ω as used in Jiang et al. (2017). An application to image process will be shown to demonstrate a nice performance of the AP algorithm. Finally, we shall extend the ideas from the AP Algorithm to find sparse solution of under determined linear systems.

Indeed, the matrix completion problem is closely related to the sparse vector recovery problem. Sparse solutions of underdetermined linear systems have been studied for last twenty years starting from [Chen, Donoho, Saunders, 1998Chen et al. (2001)] and [Tibshirani, 1996Tibshirani (1996)] and then became a major subject of research as a part of compressive sensing study since 2006 due to [Donoho, 2006Donoho (2006)], [Candés, 2006Candès et al. (2006b)], [Candés and Tao, 2005Candes and Tao (2005)], and [Candés, Romberg, and Tao, 2006Candès et al. (2006a)]. Many numerical algorithms have been developed since then. Several algorithms are based on classic convex minimization approach (cf. e.g. Hale et al. (2008), Beck and Teboulle (2009), Lai and Yin (2013), and etc.). Several algorithms are based on iteratively reweighted ℓ_1 minimization or ℓ_2 minimizations (cf. [Candés, Watkin, and Boyd, 2008Candes et al. (2008)], [Daubechies, DeVore, 2010,Daubechies et al. (2010)] and [Lai, Xu, and Yin, 2013Lai et al. (2013)]). Several researchers started the ℓ_q minimization for $q \in (0, 1)$, e.g. in [Foucart and Lai, 2009Foucart and Lai (2009)] and [Lai and Wang, 2011Lai and Wang (2011)]. Various other algorithms are based on greedy or orthogonal matching pursuit (cf. e.g. [DeVore and Temlyakov, 1996DeVore and Temlyakov (1996)], [Tropp, 2004Tropp (2004)], and [Kozlov and Petukhov, 2010Kozlov and Petukhov (2010)]). some algorithms are also based on the hard thresholding technique such as in [Blumensath and Davies, 2009Blumensath and Davies (2009)], [Blumensath and Davies, 2010Blumensath and Davies (2010)], [Foucart, 2011Foucart (2011)] and etc.. Among the various other numerical methods were also proposed. See, e.g. [Dohono, Maleki, and Montanari, 2009Donoho et al. (2009)], [Rangan, 2011Rangan (2011)], [Gong, Zhang, Lu, Huang, and Ye,2013Gong et al. (2013)], [Wang and Ye, 2014Wang et al. (2013)] and etc.. To the best of our knowledge, the method in Kozlov and Petukhov (2010) is the most effective in finding sparse solutions. Thus, we shall extend the alternating projection method to the sparse recovery problem and establish some sufficient conditions that our algorithm is convergent and its convergence is linear.

The paper is simply organized as follows. In the next section, we study the convergence of the AP algorithm. The section is divided into three subsections. We first study the case that the guess rank r_g is the same as the rank of the matrix to be completed. Next we study the remaining case that r_g is not the same as the rank of the matrix whose known entries are given. Finally in this section, we show the excellent performance of the AP algorithm when starting from an initial matrix obtained from the OR1MP algorithm in Wang et al. (2015). In §3, we extend the AP algorithm to the compressive sensing setting. §3 is divided into two subsections. First we study the convergence of the alternating projection algorithm for compressive sensing. Then we present some numerical experiments. Comparing with many known algorithms, the alternating projection method performs exceptionally well. Finally in this paper, we present an algebraic geometry analysis to show the existence of matrix completion and the number of matrices which can be completed from the given known entries of a rank r matrix.

2. The Alternating Projection Algorithm for Matrix Completion

Let $\mathcal{M}_r$ be the manifold in $\mathbb{R}^{n^2}$ consisting of $n \times n$ matrices (without loss of generality) of rank r and denote by $P_{\mathcal{M}_r}$ the projection operator onto the manifold $\mathcal{M}_r$. Next consider the affine space $\mathcal{A}_\Omega$ defined as follows:

$$\mathcal{A}_\Omega := \{X \mid \mathcal{P}_\Omega(X - M) = 0\}.$$

Affine spaces $\mathcal{A}_\Omega$ consists of matrices which has exactly same entries as M with indices in Ω. Although it is a convex set, $\mathcal{A}_\Omega$ is not a bounded set. Starting with an initial guess $X_0 = \mathcal{P}_\Omega(M)$ or a good initial guess (see our numerical experiments near the end of this section), the Alternating Projection (AP) Algorithm can be simply stated as follows:

Algorithm: Alternating Projection Algorithm for Matrix Completion

Data: Rank r of the solution M, the tolerance ϵ whose default value is 1e-6
Result: X_k, a close approximation of M
Initialize $X_0 = \mathcal{P}_\Omega(M)$ or any other good guess;
repeat
  Step 1: $Y_k = P_{\mathcal{M}_r}(X_k)$
  Step 2: $X_{k+1} = P_{\mathcal{A}_\Omega}(Y_k)$
until $\|X_{k+1} - X_k\| < \epsilon$;

In Algorithm 1 above, the computation of the projection $P_{\mathcal{M}_r}$ can be realized easily by using the singular value decomposition. $P_{\mathcal{A}_\Omega}$ is the projection onto $\mathcal{A}_\Omega$. The computation $P_{\mathcal{A}_\Omega}(Y_k)$ is obtained simply by setting the matrix entries of Y_k in positions Ω equal to the corresponding entries in M. Therefore, this algorithm is simple and easy without any minimization. The algorithm is the same as one in [Jiang, Zhong, Liu, and Song, 2017Jiang et al. (2017)]. One of the purposes of our paper is to show the convergence under various conditions.

Before studying the convergence of Algorithm 1, let us comment on the existence of a rank r matrix which has the known entries in position Ω. Let $m = |\Omega|$ be the cardinality of Ω. We shall assume $m > 2nr - r^2$. For convenience, we shall use the complex m dimensional space $\mathbb{C}^m$ to discuss the existence. If one randomly chooses the entries of a matrix M in the positions in Ω from $\mathbb{C}^m$, the probability of completing the matrix M of rank r is zero. See Theorem 28. Thus, we have to assume that the given entries are from a rank r matrix M. In other words, we call a vector $\mathbf{x} \in \mathbb{C}^m$ *r-feasible* if there exist a rank r matrix M such that $M|_\Omega = \mathbf{x}$. If the entries $\mathbf{x} \in \mathbb{C}^m$ over Ω are r-feasible, we would like to know if there is a unique rank-r matrix M satisfying $M|_\Omega = \mathbf{x}$. We can show that, generically, the number of ways to complete a matrix of rank r is less than or equal to $\prod_{i=0}^{n-r-1} \frac{\binom{n+i}{r}}{\binom{r+i}{r}}$. See Theorem 33. To prove these results, we need some background from algebraic geometry. For convenience, the details of the statements and their proofs are thus given in the last section of this paper.

In the rest of this section, we shall assume that the given entries are from a matrix of rank r. However, in general, we do not know the rank $r > 0$ of M in advance. Thus, we have to make a guess of r. Let r_g be a guessed rank. As we know any reasonable choice of r_g must satisfy $m > 2nr_g - r_g^2$, we still have either $r_g < r$, $r_g = r$ or $r_g > r$. Choose a correct rank $r_g = r$ is a key to have the AP Algorithm, i.e. Algorithm 1 converges with a linear convergence rate. Otherwise, the convergence rate may not be linear. That is, when $r_g = \text{rank}(M)$, we can show that Algorithm 1 converge to M_{r_g} linearly. Otherwise, when $r_g < \text{rank}(M)$, Algorithm 1 converges to a matrix under some conditions and may not be the desired matrix M. Thus, this section is divided into three parts. We shall discuss the two cases in the first two subsections and leave the numerical results in the third subsection.

Another important issue is the distribution of $\Omega \subset \{(i,j), i,j = 1, \cdots, n\}$. Clearly, if a column of M is completely missing, one is not able to recover this column no matter what kind of rank r of M is and how large $m = |\Omega|$ is. If we let $\mathbf{x} \in \mathbb{R}^{n^2-m}$ be the unknown entries of M, the determinant of the sub-matrix of any $r+1$ rows and $r+1$ columns of M will be zero which forms a polynomial equation with coefficients formed from known entries $M|_\Omega$. We have $n^2 - m$ unknowns while $\binom{n}{r+1}^2$ submatrices from

M which will result in $\binom{n}{r+1}^2$ polynomial equations. Since we have $n^2 - m < n^2 - 2nr + r^2 = (n-r)^2$ unknowns and $\binom{n}{r+1}^2$ equations, the system of polynomial equations is overdetermined. We have to assume that the system is consistent, i.e. the system has a solution. Otherwise, the overdetermined system has no solution, i.e. the matrix M can not be completed. Hence, for the rest of the paper, let us assume that the overdetermined system of polynomial equations have a solution, i.e. M can be completed .

2.1 Convergence of Algorithm 1 When $r_g = \mathbf{Rank}(M)$

We start with some preliminary results.

Lemma 1 *Let L be a linear subspace of $\mathbb{R}^n$. Suppose P_L denote the orthogonal projection onto L. Then, for any $x \in \mathbb{R}^n$*

$$\|x\| = \|P_L(x)\| \text{ if and only if } x \in L$$

Equivalently,

$$\|P_L(x)\| < \|x\| \text{ if and only if } x \notin L$$

Proof The 'if' part is clear. So, let us prove the 'only if' part.

Let $l_1, l_2, \cdots l_k$ be a orthonormal basis of L. Extend it to a orthonormal basis $l_1, l_2, \cdots l_n$ of $\mathbb{R}^n$. Then,

$$x = \sum_{i=1}^{n} \langle x, l_i \rangle l_i$$

and

$$\|x\|^2 = \sum_{i=1}^{n} \langle x, l_i \rangle^2 = \|P_L(x)\|^2 + \sum_{i=k+1}^{n} \langle x, l_i \rangle^2$$

Now it follows that if $\|x\| = \|P_L(x)\|$, then $\sum_{i=k+1}^{n} \langle x, l_i \rangle^2 = 0$, which implies $\langle x, l_i \rangle = 0$ for all $i \geq k+1$. Therefore, $x = \sum_{i=1}^{k} \langle x, l_i \rangle l_i \in L$. ■

Lemma 2 *Let L_1 and L_2 be two linear subspaces of $\mathbb{R}^n$. Suppose P_{L_1} and P_{L_2} denote the orthogonal projection onto L_1 and L_2 respectively. Then, $L_1 \cap L_2 = \{0\}$ if and only if*

$$\|P_{L_2} P_{L_1}\| < 1. \tag{3}$$

Proof Assume $L_1 \cap L_2 = \{0\}$. Let $x \neq 0 \in \mathbb{R}^n$. Then if $P_{L_1}(x) = 0$, then $P_{L_2} P_{L_1}(x) = 0 < \|x\|$. Otherwise, $P_{L_1}(x) \neq 0$. Since $L_1 \cap L_2 = \{0\}$, $P_{L_1}(x) \notin L_2$. Therefore, using Lemma 1, we get

$$\|P_{L_2} P_{L_1}(x)\| < \|P_{L_1}(x)\| \leq \|P_{L_1}\| \, \|x\| \leq \|x\| .$$

Hence, we have

$$\|P_{L_2} P_{L_1}(x)\| < \|x\|$$

for all non-zero $x \neq 0 \in \mathbb{R}^n$. So,

$$\|P_{L_2} P_{L_1}\| < 1.$$

To prove the other direction, assume $\|P_{L_2} P_{L_1}\| < 1$. Assume, on the contrary, that $L_1 \cap L_2 \neq \{0\}$. Let $x \neq 0 \in L_1 \cap L_2$ be a nonzero vector in the intersection. Then $P_{L_2} P_{L_1}(x) = P_{L_2}(x) = x$ which implies that $\|P_{L_2} P_{L_1}(x)\| = \|x\|$, contradicting the assumption. ■

Lemma 3 *Let $M \in \mathcal{M}_r$. Then the projection operator $P_{\mathcal{M}_r}$ is well defined (single-valued) in a neighborhood of M and is differentiable with gradient*

$$\nabla P_{\mathcal{M}_r}(M) = \mathrm{P}_{T_{\mathcal{M}_r}(M)}, \tag{4}$$

where $T_{\mathcal{M}}(M)$ is the tangent space of $\mathcal{M}$ at M and $\mathrm{P}_{T_{\mathcal{M}}(M)}$ is the projection operator onto the tangent space.

Proof Since the projection $P_{\mathcal{M}_r}$ of a matrix X is obtained by hard thresholding the least $n-r$ singular values, we see that the projection is unique if $\sigma_r(M) \neq \sigma_{r+1}(M) \geq 0$. Now consider the neighborhood V of M given by

$$V := \left\{ X \in \mathbb{R}^{n\times n} \mid \|X - M\|_F < \frac{\sigma_r(M)}{4} \right\}.$$

Then, by Weyl's Weyl (1912) or more generally Mirsky's Mirsky (1960) perturbation bounds on singular values, we have

$$|\sigma_r(X) - \sigma_r(M)| \leq \|X - M\|_F < \frac{\sigma_r(M)}{4}$$

and

$$|\sigma_{r+1}(X) - \sigma_{r+1}(M)| \leq \|X - M\|_F < \frac{\sigma_r(M)}{4}.$$

Hence, noting $\sigma_{r+1}(M) = 0$, we observe that

$$\sigma_{r+1}(X) < \frac{\sigma_r(M)}{4} < \frac{3\sigma_r(M)}{4} < \sigma_r(X).$$

In particular,

$$\sigma_r(X) \neq \sigma_{r+1}(X).$$

Therefore, $P_{\mathcal{M}_r}$ is single valued in the neighborhood V.

For second part of the result, we refer to Theorem 25 in Feppon and Lermusiaux (2017) which is stated below. We have changed the notations for ease of reading. In particular, note that although the X has rank greater than r in Feppon and Lermusiaux (2017), its easy to see that their proof goes through when X has rank greater than or equal to r. Intuitively, it is easy to see that the gradient vector of the projection $P_{\mathcal{M}_r}$ of smooth manifold $\mathcal{M}_r$ at M will be the projection onto the tangent plane $T_{\mathcal{M}_r}$ at M in general. ■

The following results was used in the proof above.

Theorem 4 (F. Feppon and P.J. Lermusiaux, 2017Feppon and Lermusiaux (2017)) *Consider $X \in \mathbb{R}^{n\times m}$ with rank greater than r and denote $X = \sum_{i=1}^{r+k} \sigma_i u_i v_i^\top$ be its SVD decomposition, where the singular values are ordered decreasingly: $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_{r+k}$. Suppose that the orthogonal projection $P_{\mathcal{M}_r}(X)$ of X onto $\mathcal{M}_r$ is uniquely defined, that is $\sigma_r(X) > \sigma_{r+1}(X)$. Then $P_{\mathcal{M}_r}$, the SVD truncation operator of order r, is differentiable at X and the differential in a direction Y is given by the formula*

$$\begin{aligned}\nabla_Y P_{\mathcal{M}_r}(X) =& \mathrm{P}_{T_{\mathcal{M}_r}(P_{\mathcal{M}_r}(X))}(Y) \\ &+ \sum_{\substack{1\leq i\leq r \\ 1\leq j\leq k}} \left[\frac{\sigma_{r+j}}{\sigma_i - \sigma_{r+j}} \langle Y, \Phi^+_{i,r+j}\rangle \Phi^+_{i,r+j} - \frac{\sigma_{r+j}}{\sigma_i + \sigma_{r+j}} \langle Y, \Phi^-_{i,r+j}\rangle \Phi^-_{i,r+j} \right],\end{aligned} \tag{5}$$

where

$$\Phi^{\pm}_{i,r+j} = \frac{1}{\sqrt{2}}(u_{r+j}v_i^\top \pm u_i v_{r+j}^\top)$$

are the principal directions corresponding to the principal curvature of the manifold of rank-r matrices.

Proof Refer to Theorem 25 in Feppon and Lermusiaux (2017). ■

We are now ready to establish the convergence of Algorithm 1 under a sufficient condition.

Theorem 5 *Assume $T_{\mathcal{A}_\Omega}(M) \cap T_{\mathcal{M}_r}(M) = \{0\}$. Then Algorithm 1 converges to M locally at a linear rate, i.e. there exists a neighborhood V around M such that if $X_0 \in V$, then there exists a positive constant $c < 1$ such that*

$$\|X_k - M\| < c^k \|X_0 - M\|, \tag{6}$$

where X_k is the kth iteration from Algorithm 1.

Proof For notational convenience, let

$$f(X) := P_{\mathcal{A}_\Omega}(P_{\mathcal{M}_r}(X)).$$

Note that $\mathcal{A}_\Omega$ is an affine space, the gradient $\nabla P_{\mathcal{A}_\Omega}$ of the projection $P_{\mathcal{A}_\Omega}$ is the projection onto the tangent space of the affine space $\mathcal{A}_\Omega$. By Lemma 3 and chain rule, we have

$$(\nabla f)(X) = P_{T_{\mathcal{A}_\Omega}(M)}(\mathrm{P}_{T_{\mathcal{M}_r}(M)}(X)).$$

as $T_{\mathcal{A}_\Omega}(M) = T_{\mathcal{A}_\Omega}(X)$ for all X.

Now from the definition of differentiability of f at M, we have

$$\lim_{X \to M} \frac{\|f(X) - f(M) - \nabla f(M) \cdot (X - M)\|}{\|X - M\|} = 0.$$

Hence, there exist an open ball V, say a ball $V = B_{r_0}(M)$ centered at M of radius r_0 around M such that, for all $X \in V$

$$\frac{\|f(X) - f(M) - \nabla f(M) \cdot (X - M)\|}{\|X - M\|} < \epsilon,$$

where $\epsilon = \frac{1 - \|\nabla f\|}{2} > 0$. Using our hypothesis and Lemma 2, we have $\|\nabla f(M)\| = \left\| P_{T_{\mathcal{A}_\Omega}(M)} \mathrm{P}_{T_{\mathcal{M}_r}(M)} \right\| < 1$. Therefore, for all $X \in V$, we use $M = f(M)$ to have

$$\begin{aligned}
\|f(X) - M\| &= \|f(X) - f(M)\| \\
&\leq \|f(X) - f(M) - \nabla f \cdot (X - M)\| + \|\nabla f(M) \cdot (X - M)\| \\
&< \epsilon \|X - M\| + \|\nabla f(M)\| \, \|(X - M)\| \\
&= (\epsilon + \|\nabla f(M)\|) \, \|X - M\| \\
&\leq \frac{1 + \|\nabla f(M)\|}{2} \|X - M\|.
\end{aligned}$$

where $\frac{1 + \|\nabla f(M)\|}{2} < 1$ since $\|\nabla f(M)\| < 1$ as discussed above.

Setting $c = \frac{1 + \|\nabla f(M)\|}{2} < 1$, we can rewrite the above inequality as follows:

$$\|f(X) - M\| < c \|X - M\| \text{ for all } X \in V. \tag{7}$$

Hence, if $X_k \in V = B_{r_0}(M)$, we use $X_{k+1} = f(X_k)$ to have

$$\|X_{k+1} - M\| = \|f(X_k) - M\| < c\,\|X_k - M\| \leq r_0$$

which implies $X_{k+1} \in V = B_{r_0}(M)$. So, if the initial guess $X_0 \in V$, we have, by induction,

$$X_k \in V \text{ for all } k$$

and

$$\|X_k - M\| \leq c^k\,\|X_0 - M\|\,.$$

We have thus completed the proof. ■

We will now derive certain equivalent conditions for hypothesis of the above theorem viz. $T_{\mathcal{A}_\Omega}(M) \cap T_{\mathcal{M}_r}(M) = \{0\}$. Let us recall the following property which is known in the literature. For convenience, we include a proof.

Lemma 6 *The tangent space $T_{\mathcal{M}_r}(M)$ has an explicit description as follows:*

$$T_{\mathcal{M}_r}(M) = \left\{XM + MY \mid X \in \mathbb{R}^{n\times n} \text{ and } Y \in \mathbb{R}^{n\times n}\right\}. \tag{8}$$

Proof First recall that the tangent space $T_{\mathcal{M}_r}(M)$ to a manifold $\mathcal{M}_r$ at a point M is the linear space spanned by all the tangent vectors at 0 to smooth curves $\gamma : \mathbb{R} \to \mathcal{M}_r$ such that $\gamma(0) = M$.

Now let $M \in \mathcal{M}_r$ be a $n \times n$ matrix of rank r. We can write $M = X_0 Y_0^\top$ where $X_0, Y_0 \in \mathbb{R}^{n\times r}$ and both X_0 and Y_0 have full column rank. This is possible because M has exactly rank r.

Let $\gamma(t) = X(t)Y(t)^\top$ be a smooth curve such that $X(0) = X_0$ and $Y(0) = Y_0$. Hence, $\gamma(0) = X_0 Y_0^\top = M$. Since X_0 and Y_0 have full column rank, X_0 and Y_0 have a $r \times r$ minor that does not vanish. Since nonvanishing of a minor is an open condition, there exist an open neighbourhood of M to which if we restrict the curve γ, we can assume $X(t)$ and $Y(t)$ have full column rank. In other words, we can assume, without loss of generality, that $X(t)^\top X(t)$ and $Y(t)^\top Y(t)$ are invertible $r \times r$ matrices for all t.

By product rule, we obtain

$$\begin{aligned}
\dot\gamma(0) &= \dot X(0)Y(0)^\top + X(0)\dot Y(0)^\top \\
&= \dot X(0)Y_0^\top + X_0\dot Y(0)^\top \\
&= \dot X(0)(X_0^\top X_0)^{-1}(X_0^\top X_0)Y_0^\top + X_0(Y_0^\top Y)(Y_0^\top Y_0)^{-1}\dot Y(0)^\top \\
&= \left(\dot X(0)(X_0^\top X_0)^{-1}X_0^\top\right)(X_0Y_0^\top) + (X_0Y_0^\top)\left(Y_0(Y_0^\top Y)^{-1}\dot Y(0)^\top\right) \\
&= \left(\dot X(0)(X_0^\top X_0)^{-1}X_0^\top\right)M + M\left(Y_0(Y_0^\top Y_0)^{-1}\dot Y(0)^\top\right) \\
&\in \left\{XM + MY \mid X \in \mathbb{R}^{n\times n} \text{ and } Y \in \mathbb{R}^{n\times n}\right\}.
\end{aligned}$$

Now to prove the reverse inclusion, let $AM + MB \in \{XM + MY \mid X \in \mathbb{R}^{n\times n} \text{ and } Y \in \mathbb{R}^{n\times n}\}$. Consider the smooth curve $\gamma(t) = X(t)Y(t)^\top$ defined by

$$X(t) = t(AX_0) + X_0$$

and

$$Y(t) = t\left((Y_0B)^\top\right) + Y_0.$$

An easy computation shows that $\gamma(0) = M$ and $\dot\gamma(0) = AM + MB$. Hence we get the equality

$$T_{\mathcal{M}_r}(M) = \left\{XM + MY \mid X \in \mathbb{R}^{n\times n} \text{ and } Y \in \mathbb{R}^{n\times n}\right\}$$

This completes the proof. ■

One can consider $T_{\mathcal{M}_r}(M)$ as a linear space in $\mathbb{R}^{n^2}$ by rewriting it as

$$T_{\mathcal{M}_r}(M) \cong \text{Range}(T_M) = \left\{ T_M \cdot \begin{bmatrix} (X^1)^\top \\ \vdots \\ (X^n)^\top \\ Y_1 \\ \vdots \\ Y_n \end{bmatrix} \mid X \in \mathbb{R}^{n\times n} \text{ and } Y \in \mathbb{R}^{n\times n} \right\}$$

where T_M is a block matrix of size $n^2 \times 2n^2$ consisting of $2n^3$ blocks of size $1 \times n$, X^i and X_j denotes the i^{th} row and j^{th} column of a matrix X respectively.

Explicitly, T_M would take the form

$$T_M = \left[\begin{array}{c|c|c|c|c|c|c|c|c|c|c|c|c|c|c} M_1^\top & 0 & \cdots & 0 & \cdots & \cdots & \cdots & 0 & M^1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline 0 & 0 & \cdots & 0 & M_j^\top & 0 & \cdots & 0 & 0 & \cdots & 0 & M^i & 0 & \cdots & 0 \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array}\right]$$

where the each row corresponds to each index in $\{1, 2, \cdots, n\} \times \{1, 2, \cdots, n\}$.

Let T_M^Ω and $T_M^{\Omega^c}$ denote the matrix obtained from T_M by choosing the rows corresponding to Ω and Ω^c, respectively.

Example 1 *Suppose* $M = \begin{bmatrix} 1 & 4 \\ 2 & 8 \end{bmatrix}$ *and* $\Omega = \{(1,2),(2,1)\}$. *Then*

$$T_M^\Omega = \left[\begin{array}{cc|cc|cc|cc} 4 & 8 & 0 & 0 & 0 & 0 & 1 & 4 \\ \hline 0 & 0 & 1 & 2 & 2 & 8 & 0 & 0 \end{array}\right]$$

$$T_M^{\Omega^c} = \left[\begin{array}{cc|cc|cc|cc} 1 & 2 & 0 & 0 & 1 & 4 & 0 & 0 \\ \hline 0 & 0 & 4 & 8 & 0 & 0 & 2 & 8 \end{array}\right]$$

and

$$T_M = \left[\begin{array}{cc|cc|cc|cc} 4 & 8 & 0 & 0 & 0 & 0 & 1 & 4 \\ \hline 0 & 0 & 1 & 2 & 2 & 8 & 0 & 0 \\ \hline 1 & 2 & 0 & 0 & 1 & 4 & 0 & 0 \\ \hline 0 & 0 & 4 & 8 & 0 & 0 & 2 & 8 \end{array}\right]$$

Example 2 *Suppose* $M = \begin{bmatrix} -3 & -1 & -4 \\ 9 & 3 & 12 \\ 6 & 2 & 8 \end{bmatrix}$ *and* $\Omega = \{(1,1),(1,3),(2,2),(3,1)\}$, *Then*

$$T_M^\Omega = \left[\begin{array}{ccc|ccc|ccc|ccc|ccc|ccc} -3 & 9 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & -1 & -4 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline -4 & 12 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & -1 & -4 \\ \hline 0 & 0 & 0 & -1 & 3 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 9 & 3 & 12 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & -3 & 9 & 6 & 6 & 2 & 8 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$T_M^{\Omega^c} = \left[\begin{array}{ccc|ccc|ccc|ccc|ccc|ccc} -1 & 3 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & -1 & -4 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -3 & 9 & 6 & 0 & 0 & 0 & 9 & 3 & 12 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -4 & 12 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9 & 3 & 12 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & -1 & 3 & 2 & 0 & 0 & 0 & 6 & 2 & 8 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & -4 & 12 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 6 & 2 & 8 \end{array}\right]$$

and

$$T_M = \left[\begin{array}{ccc|ccc|ccc|ccc|ccc|ccc} -3 & 9 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & -1 & -4 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline -4 & 12 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & -1 & -4 \\ \hline 0 & 0 & 0 & -1 & 3 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 9 & 3 & 12 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & -3 & 9 & 6 & 6 & 2 & 8 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline -1 & 3 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & -1 & -4 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -3 & 9 & 6 & 0 & 0 & 0 & 9 & 3 & 12 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -4 & 12 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9 & 3 & 12 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & -1 & 3 & 2 & 0 & 0 & 0 & 6 & 2 & 8 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & -4 & 12 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 6 & 2 & 8 \end{array}\right]$$

Next we need

Lemma 7 *The tangent space $T_{\mathcal{A}_\Omega}(M)$ at M can be given explicitly as follows.*

$$T_{\mathcal{A}_\Omega}(M) = \left\{X \in \mathbb{R}^{n\times n} \mid P_\Omega(X) = 0\right\}. \tag{9}$$

Proof Recall that

$$\mathcal{A}_M := \{X \mid P_\Omega(X - M) = 0\}.$$

Since $P_\Omega(X - M) = P_\Omega(X) - P_\Omega(M) = P_\Omega(X) - P_\Omega(P_\Omega(M)) = P_\Omega(X - P_\Omega(M))$, we get that the set $\mathcal{A}_\Omega$ is a translation of the linear space $\{X \in \mathbb{R}^{n\times n} \mid P_\Omega(X) = 0\}$ by $P_\Omega(M)$, i.e.

$$\mathcal{A}_\Omega = \left\{X \in \mathbb{R}^{n\times n} \mid P_\Omega(X) = 0\right\} + P_\Omega(M)$$

Hence we have that the tangent space of $\mathcal{A}_\Omega$ at M is equal to the tangent space of the vector space $\{X \in \mathbb{R}^{n\times n} \mid P_\Omega(X) = 0\}$ at $M - P_\Omega(M)$. But the tangent space of a vector space at any point is the vector space itself. Hence the result follows. ■

With the above preparation, we have another main result in this section.

Theorem 8 *The following statements are equivalent:*

1. $T_{\mathcal{A}_\Omega}(M) \cap T_{\mathcal{M}_r}(M) = \{0\}$
2. $\text{Rowspace}\left(T_M^{\Omega^c}\right) \subseteq \text{Rowspace}\left(T_M^{\Omega}\right)$
3. $\text{Rank}\left(T_M^{\Omega}\right) = 2nr - r^2$, *where* $r = \text{Rank}(M)$

4. The matrix $V^{\Omega}(M)$ of size $|\Omega| \times |\Omega|$ defined by

$$V^{\Omega}_{(i_1,j_1),(i_2,j_2)}(M) = \begin{cases} 0 & i_1 \neq i_2 \text{ and } j_1 \neq j_2 \\ \langle M_{j_1}, M_{j_2} \rangle & i_1 = i_2 \text{ and } j_1 \neq j_2 \\ \langle M^{i_1}, M^{i_2} \rangle & i_1 \neq i_2 \text{ and } j_1 = j_2 \\ \left\| M^{i_1} \right\|^2 + \| M_{j_1} \|^2 & i_1 = i_2 \text{ and } j_1 = j_2 \end{cases} \tag{10}$$

has rank $2nr - r^2$, where M_j stands for the jth column and M^i for the ith row of M.

Proof (1) $\iff$ (2) Note that the elements of $T_{\mathcal{A}_\Omega}(M) \cap T_{\mathcal{M}_r}(M)$ consists of matrices of the form $XM + MY$ such that the elements in positions Ω is zero by Lemmas 6 and 7. Hence, observing that $T_{\mathcal{M}_r}(M)$ can be considered as the range of T_M and that the rows of T_M correspond to each index in $\{1, 2, \cdots, n\} \times \{1, 2, \cdots, n\}$, we can conclude that $T_{\mathcal{A}_\Omega}(M) \cap T_{\mathcal{M}_r}(M) = \{0\}$ if and only if

$$\text{NullSpace}\left(T_M^{\Omega}\right) \subseteq \text{NullSpace}\left(T_M^{\Omega^c}\right)$$

which is equivalent to

$$\text{NullSpace}\left(T_M^{\Omega}\right)^{\perp} \supseteq \text{NullSpace}\left(T_M^{\Omega^c}\right)^{\perp}$$

The result follows by noting that

$$\text{Rowspace}\left(T_M^{\Omega^c}\right) = \text{NullSpace}\left(T_M^{\Omega^c}\right)^{\perp}$$

and

$$\text{Rowspace}\left(T_M^{\Omega}\right) = \text{NullSpace}\left(T_M^{\Omega}\right)^{\perp}.$$

(2) $\iff$ (3) We begin by recalling that dimension of a tangent space is equal to dimension of the manifold. So, $\dim(T_{\mathcal{M}_r}(M)) = 2nr - r^2$. Now

$$2nr - r^2 = \dim(T_{\mathcal{M}}(M)) = \dim(Range(T_M)) = \text{Rank}(T_M) = \text{Rank}\left(\text{Rowspace}\left(T_M\right)\right).$$

Now the equivalence (2) $\iff$ (3) follows by recalling that T_M^{Ω} and $T_M^{\Omega^c}$ were obtained from T_M by choosing the rows corresponding to Ω and Ω^c, respectively

(3) $\iff$ (4) The equivalence follows from fact that $V^{\Omega}(M) = T_M^{\Omega}\left(T_M^{\Omega}\right)^{\top}$. Hence $\text{Rank}(V^{\Omega}(M)) = \text{Rank}(T_M^{\Omega})$. ■

Remark 9 *Few remarks are in order here. Size of the matrix T_M^{Ω} is $|\Omega| \times 2n^2$ which is considerably larger than the size of the matrix $V^{\Omega}(M)$ which has size $|\Omega| \times |\Omega|$. Therefore, since rank computation is a memory intensive process, it is much efficient to check the statement (4) of above theorem as compared to statement (3).*
In general, the rank of $V^{\Omega}(M)$ is less than or equal to $2nr - r^2$. The equality occurs when the tangent spaces intersect trivially.

The following example is an illustration of the linear convergence of the error when the condition $T_{\mathcal{A}_\Omega}(M) \cap T_{\mathcal{M}_r}(M) = \{0\}$ is satisfied.

Example 3 *We find a 15×15 matrix M of rank 2 which has 28% of entries missing. A straightforward computation shows that $\text{Rank}(V^{\Omega}(M)) = 2nr - r^2$. Hence, M satisfies the condition $T_{\mathcal{A}_\Omega}(M) \cap$*

$T_{\mathcal{M}_r}(M) = \{0\}$. *Hence, by Theorems 8 and 5, we know that Algorithm 1 will converge in a linear fashion.*

$$M =$$

$$\begin{bmatrix}
0.3474 & 0.0897 & 0.3971 & 0.4644 & 0.4168 & 0.7576 & 0.8206 & 0.8161 & 0.3279 & 0.3851 & 0.0825 & 0.4742 & 0.7684 & 0.6113 & 0.3832 \\
0.1502 & 0.0414 & 0.2196 & 0.2450 & 0.2731 & 0.4415 & 0.4293 & 0.4358 & 0.1859 & 0.1574 & 0.0493 & 0.2386 & 0.4502 & 0.3087 & 0.1999 \\
0.3853 & 0.1079 & 0.5912 & 0.6542 & 0.7544 & 1.1986 & 1.1445 & 1.1660 & 0.5024 & 0.3985 & 0.1343 & 0.6315 & 1.2231 & 0.8176 & 0.5325 \\
0.2174 & 0.0577 & 0.2760 & 0.3160 & 0.3141 & 0.5394 & 0.5562 & 0.5582 & 0.2305 & 0.2358 & 0.0594 & 0.3160 & 0.5484 & 0.4080 & 0.2594 \\
0.2124 & 0.0493 & 0.1453 & 0.1940 & 0.0662 & 0.2317 & 0.3503 & 0.3303 & 0.1109 & 0.2539 & 0.0228 & 0.2216 & 0.2302 & 0.2835 & 0.1647 \\
0.1026 & 0.0238 & 0.0701 & 0.0936 & 0.0318 & 0.1117 & 0.1691 & 0.1594 & 0.0535 & 0.1227 & 0.0110 & 0.1070 & 0.1110 & 0.1368 & 0.0795 \\
0.2429 & 0.0600 & 0.2290 & 0.2798 & 0.1972 & 0.4141 & 0.4982 & 0.4864 & 0.1846 & 0.2785 & 0.0439 & 0.2974 & 0.4176 & 0.3823 & 0.2332 \\
0.3848 & 0.0895 & 0.2658 & 0.3538 & 0.1248 & 0.4257 & 0.6385 & 0.6028 & 0.2032 & 0.4595 & 0.0421 & 0.4033 & 0.4232 & 0.5159 & 0.3002 \\
0.3698 & 0.1015 & 0.5311 & 0.5943 & 0.6536 & 1.0640 & 1.0419 & 1.0562 & 0.4488 & 0.3894 & 0.1186 & 0.5806 & 1.0845 & 0.7511 & 0.4852 \\
0.3631 & 0.0880 & 0.3138 & 0.3919 & 0.2395 & 0.5511 & 0.7003 & 0.6776 & 0.2496 & 0.4217 & 0.0575 & 0.4246 & 0.5540 & 0.5451 & 0.3282 \\
0.2081 & 0.0480 & 0.1369 & 0.1850 & 0.0542 & 0.2139 & 0.3347 & 0.3141 & 0.1036 & 0.2498 & 0.0208 & 0.2133 & 0.2120 & 0.2726 & 0.1575 \\
0.5203 & 0.1334 & 0.5792 & 0.6812 & 0.5942 & 1.0977 & 1.2049 & 1.1953 & 0.4769 & 0.5797 & 0.1192 & 0.6992 & 1.1126 & 0.9011 & 0.5627 \\
0.4871 & 0.1231 & 0.5111 & 0.6090 & 0.4961 & 0.9538 & 1.0797 & 1.0652 & 0.4178 & 0.5487 & 0.1028 & 0.6328 & 0.9651 & 0.8148 & 0.5047 \\
0.0287 & 0.0122 & 0.1183 & 0.1173 & 0.2001 & 0.2658 & 0.2007 & 0.2154 & 0.1057 & 0.0156 & 0.0311 & 0.0991 & 0.2738 & 0.1297 & 0.0927 \\
0.2602 & 0.0617 & 0.1997 & 0.2577 & 0.1230 & 0.3353 & 0.4628 & 0.4422 & 0.1558 & 0.3070 & 0.0341 & 0.2867 & 0.3353 & 0.3673 & 0.2173
\end{bmatrix}$$

and

$$M_\Omega =$$

$$\begin{bmatrix}
0 & 0.0897 & 0.3971 & 0 & 0.4168 & 0.7576 & 0.8206 & 0 & 0.3279 & 0.3851 & 0.0825 & 0 & 0.7684 & 0.6113 & 0.3832 \\
0.1502 & 0 & 0 & 0.2450 & 0 & 0.4415 & 0.4293 & 0.4358 & 0 & 0.1574 & 0 & 0.2386 & 0.4502 & 0 & 0.1999 \\
0.3853 & 0.1079 & 0.5912 & 0.6542 & 0.7544 & 1.1986 & 0 & 1.1660 & 0.5024 & 0 & 0.1343 & 0 & 1.2231 & 0.8176 & 0.5325 \\
0.2174 & 0.0577 & 0.2760 & 0 & 0.3141 & 0 & 0 & 0.5582 & 0.2305 & 0 & 0.0594 & 0.3160 & 0.5484 & 0.4080 & 0 \\
0 & 0 & 0.1453 & 0.1940 & 0.0662 & 0.2317 & 0.3503 & 0 & 0.1109 & 0 & 0 & 0.2216 & 0 & 0 & 0.1647 \\
0.1026 & 0.0238 & 0.0701 & 0.0936 & 0.0318 & 0.1117 & 0.1691 & 0.1594 & 0.0535 & 0.1227 & 0.0110 & 0 & 0.1110 & 0.1368 & 0.0795 \\
0.2429 & 0.0600 & 0.2290 & 0.2798 & 0 & 0 & 0.4982 & 0 & 0.1846 & 0 & 0 & 0.2974 & 0.4176 & 0.3823 & 0.2332 \\
0 & 0 & 0 & 0.3538 & 0.1248 & 0 & 0 & 0 & 0 & 0.4595 & 0 & 0 & 0 & 0.5159 & 0.3002 \\
0 & 0.1015 & 0.5311 & 0.5943 & 0.6536 & 1.0640 & 1.0419 & 1.0562 & 0.4488 & 0.3894 & 0 & 0.5806 & 1.0845 & 0 & 0 \\
0.3631 & 0.0880 & 0.3138 & 0 & 0.2395 & 0.5511 & 0.7003 & 0.6776 & 0.2496 & 0 & 0.0575 & 0.4246 & 0.5540 & 0.5451 & 0 \\
0.2081 & 0.0480 & 0.1369 & 0.1850 & 0.0542 & 0.2139 & 0 & 0 & 0.1036 & 0.2498 & 0.0208 & 0.2133 & 0 & 0.2726 & 0.1575 \\
0 & 0.1334 & 0.5792 & 0 & 0.5942 & 1.0977 & 1.2049 & 1.1953 & 0 & 0 & 0.1192 & 0.6992 & 0 & 0.9011 & 0.5627 \\
0.4871 & 0 & 0.5111 & 0.6090 & 0 & 0.9538 & 1.0797 & 0 & 0 & 0.5487 & 0.1028 & 0.6328 & 0.9651 & 0.8148 & 0.5047 \\
0.0287 & 0.0122 & 0.1183 & 0.1173 & 0.2001 & 0.2658 & 0.2007 & 0.2154 & 0 & 0.0156 & 0 & 0.0991 & 0.2738 & 0.1297 & 0.0927 \\
0.2602 & 0.0617 & 0.1997 & 0.2577 & 0.1230 & 0.3353 & 0.4628 & 0 & 0.1558 & 0.3070 & 0.0341 & 0.2867 & 0.3353 & 0.3673 & 0.2173
\end{bmatrix}$$

where 0 *stands for the unknown entries.*

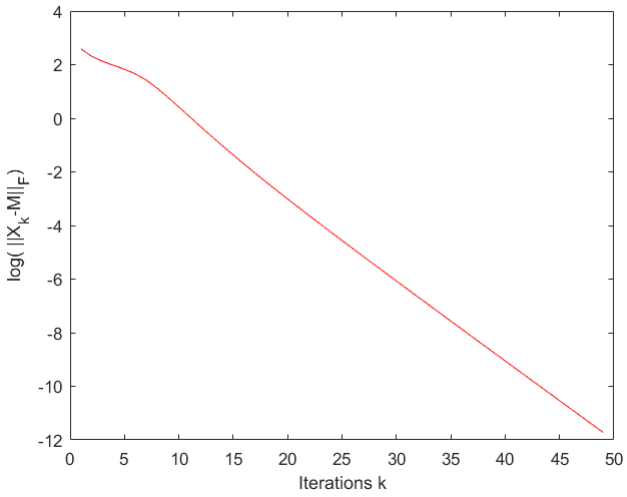

Figure 1: Linear Convergence of the Iterations from Algorithm 1

Notice from the graph in Figure 1 that as the iterations progress, the X_k *would eventually land in a neighborhood of* M *where the convergence become linear.*

The construction of T_M^Ω enables us to choose Ω such that T_M^Ω is of full rank. We end with this subsection with the following

Corollary 10 *Given M with rank r, for any integer m such that $2nr - r^2 \leq m \leq n^2$, there exists a subset Ω with $m = |\Omega|$ such that V^Ω is of full rank, equivalently $T_{\mathcal{A}_\Omega}(M) \cap T_{\mathcal{M}_r}(M) = \{0\}$ and Algorithm 1 can find M in a linear fashion for a good initial guess.*

Proof We mainly choose Ω such that the corresponding rows of T_M which form T_M^Ω of rank $2nr - r^2$. Then Theorems 8 and 5 can be applied. ■

2.2 Convergence of Algorithm 1 When $r_g \neq$ Rank(M)

In this subsection, we show that the algorithm does converge under certain reasonable assumption irrespective of whether our guessed rank r_g is same as the rank r of matrix M or not. We begin with two trivial results.

Lemma 11 *Let Y_k and X_{k+1} be the matrices we obtain in the step 1 and step 2 of the k^{th} iteration of Algorithm 1. Then*

$$X_{k+1} = \begin{cases} (Y_k)_{i,j} & \text{if } (i,j) \notin \Omega \\ M_{i,j} & \text{Otherwise.} \end{cases}$$

That is, X_{k+1} is the orthogonal projection of Y_k onto $\mathcal{A}_\Omega$.

Lemma 12 *Let $X_{k+1} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ be the standard singular value decomposition with $\mathbf{\Sigma} = \text{diag}\{\sigma_1, \cdots, \sigma_n\}$. Then*

$$Y_{k+1} = \mathbf{U}\tilde{\mathbf{\Sigma}}\mathbf{V}^\top,$$

where $\tilde{\mathbf{\Sigma}} = diag\{\sigma_1, \cdots, \sigma_{r_g}, 0, \cdots, 0\}$.

Also Y_{k+1} is the orthogonal projection of X_{k+1} onto $\mathcal{M}_{r_g}$.

$\mathcal{M}_{r_g}$, the collection of $n \times n$ real (complex) matrices of rank r_g, forms a quasi-affine real (complex) variety and is a manifold of real (complex) dimension $r_g(2n - r_g)$.

It is well known that Y_k, obtained from X_k by SVD truncation, is the orthogonal projection of X_k onto $\mathcal{M}_{r_g}$. Hence we $X_k - Y_k$ must be orthogonal to the tangent space of $\mathcal{M}_{r_g}$ at Y_k. Recall from earlier section that tangent space of $\mathcal{M}_{r_g}$ at the point X is given by

$$T_{\mathcal{M}_{r_g}}(X) = \left\{AX + XB, A \in \mathbb{R}^{m\times m}, B \in \mathbb{R}^{n\times n}\right\}$$

Lemma 13 *Y_k satisfies:*

$$\langle AY_k + Y_kB, X_k - Y_k\rangle = 0 \text{ for all } k, A \in \mathbb{R}^{n\times n} \text{ and } B \in \mathbb{R}^{n\times n}.$$

Proof Let $X_k = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ and $Y_k = \mathbf{U}\tilde{\mathbf{\Sigma}}\mathbf{V}^\top$, where $\Sigma = \text{diag}\{\sigma_1, \cdots, \sigma_n\}$ and $\tilde{\Sigma} = \text{diag}\{\sigma_1, \cdots, \sigma_{r_g}, 0, \cdots, 0\}$ be the singular value decompositions of X_k and Y_k respectively.

$$\begin{aligned}
\langle AY_k + Y_kB, X_k - Y_k\rangle &= \text{Trace}\left((X_k - Y_k)^\top(AY_k + Y_kB)\right) \\
&= \text{Trace}\left(\mathbf{V}(\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}})\mathbf{U}^\top AY_k\right) + \text{Trace}\left(\mathbf{V}(\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}})\mathbf{U}^\top Y_kB\right) \\
&= \text{Trace}\left(\mathbf{V}(\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}})\mathbf{U}^\top A\mathbf{U}\tilde{\mathbf{\Sigma}}\mathbf{V}^\top\right) + \text{Trace}\left(\mathbf{V}(\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}})\mathbf{U}^\top\mathbf{U}\tilde{\mathbf{\Sigma}}\mathbf{V}^\top B\right) \\
&= \text{Trace}\left(\mathbf{V}^\top\mathbf{V}(\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}})\mathbf{U}^\top A\mathbf{U}\tilde{\mathbf{\Sigma}}\right) + \text{Trace}\left(\mathbf{V}(\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}})\tilde{\mathbf{\Sigma}}\mathbf{V}^\top B\right) \\
&= \text{Trace}\left(\tilde{\mathbf{\Sigma}}(\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}})\mathbf{U}^\top A\mathbf{U}\right) + \text{Trace}\left(\mathbf{V}(\mathbf{\Sigma} - \tilde{\mathbf{\Sigma}})\tilde{\mathbf{\Sigma}}\mathbf{V}^\top B\right) \\
&= 0.
\end{aligned}$$

The last step uses the fact that $\tilde{\Sigma}(\Sigma - \tilde{\Sigma}) = (\Sigma - \tilde{\Sigma})\tilde{\Sigma} = 0$. ■

From the definitions, it follows that

$$\|X_k - Y_k\| \geq \|X_{k+1} - Y_k\| \geq \|X_{k+1} - Y_{k+1}\| \text{ for all } k. \tag{11}$$

From equation (11), we observe that $\|X_k - Y_k\|$ is a non-increasing sequence bounded below by 0, it thus converges to its infimum. Thus, we have

Lemma 14 *The sequence*

$$\|X_k - Y_k\|$$

converges.

So let

$$L = \lim_k \|X_k - Y_k\|^2 . \tag{12}$$

Next we have

Lemma 15

$$\|X_{k+1} - X_k\|^2 + \|X_{k+1} - Y_k\|^2 = \|X_k - Y_k\|^2 \tag{13}$$

Proof The result (13) follows from Lemmas 11 and 12. In fact we have used the fact $\langle X_{k+1} - X_k, X_{k+1} - Y_k \rangle = 0$ to have (13). ■

Lemma 16 *The series*

$$\sum_{k=1}^{\infty} \|X_{k+1} - X_k\|^2$$

converges. In particular

$$\|X_{k+1} - X_k\| \to 0.$$

Proof We use (13) and (11) to get

$$\|X_k - Y_k\|^2 \geq \|X_{k+1} - X_k\|^2 + \|X_{k+1} - Y_{k+1}\|^2$$

summing both sides from $k = 1$ to n we get

$$\sum_{k=1}^{n} \|X_k - Y_k\|^2 \geq \sum_{k=1}^{n} \|X_{k+1} - X_k\|^2 + \sum_{k=1}^{n} \|X_{k+1} - Y_{k+1}\|^2 .$$

From which it follows that

$$\|X_1 - Y_1\|^2 \geq \|X_n - Y_n\|^2 + \sum_{k=1}^{n} \|X_{k+1} - X_k\|^2 \geq \sum_{k=1}^{n} \|X_{k+1} - X_k\|^2$$

Thus the partial sums of the $\sum_{k=1}^{\infty} \|X_{k+1} - X_k\|^2$ forms an non-decreasing sequence bounded from above. The result follows immediately. ■

Lemma 17 *The series*

$$\sum_{k=1}^{\infty} \|(X_k - Y_k)_{\Omega^c}\|^2$$

converges. In particular

$$\|(X_k - Y_k)_{\Omega^c}\| \to 0.$$

Proof

$$\begin{aligned}
\|X_{k+1} - X_k\|^2 &= \|(X_{k+1} - X_k)_\Omega\|^2 + \|(X_{k+1} - X_k)_{\Omega^c}\|^2 \\
&= \|(X_{k+1})_\Omega - (X_k)_\Omega\|^2 + \|(X_{k+1})_{\Omega^c} - (X_k)_{\Omega^c}\|^2 \\
&= \|M_\Omega - M_\Omega\|^2 + \|(X_{k+1})_{\Omega^c} - (X_k)_{\Omega^c}\|^2 \\
&= \|(X_{k+1})_{\Omega^c} - (X_k)_{\Omega^c}\|^2 .
\end{aligned}$$

Now noting that $(X_{k+1})_{\Omega^c} = (Y_k)_{\Omega^c}$ the above equation simplifies

$$\|X_{k+1} - X_k\|^2 = \|(Y_k)_{\Omega^c} - (X_k)_{\Omega^c}\|^2$$

Summing both sides and using Lemma 16, the result follows. ■

With the above preparation, we are finally ready to establish the main convergence result in this subsection.

Theorem 18 *There exist a subsequence of $(Y_k)_\Omega$ that converges, say without loss of generality, $(Y_k)_\Omega \to y^\star$. Assume that there are only finitely many rank-r matrices Y such that $P_\Omega(Y) = y^\star$. Then there exist subsequences X_{k_j} and Y_{k_j} which converge, say $Y^\star$ and $X^\star$ such that*

$$X_{k_j} \to X^\star \text{ and } Y_{k_j} \to Y^\star.$$

Furthermore, we have $X^\star|_{\Omega^c} = Y^\star|_{\Omega^c}$ and

$$X^\star \in \mathcal{A}_\Omega \text{ and } \operatorname{rank}(Y^\star) \leq r_g. \tag{14}$$

Proof By Lemma 14, $\|X_k - Y_k\| \to \sqrt{L}$, we see that the sequence $\|M_\Omega - (Y_k)_\Omega\| = \|(X_k)_\Omega - (Y_k)_\Omega\| \leq 2\sqrt{L}$ for all $k \geq 1$ without loss of generality. It follows that $\|(Y_k)_\Omega\|, k \geq 1$ are a bounded sequence and hence, $\|(Y_k)_\Omega\| \leq C_1 < \infty$ for a positive constant C_1 and $(Y_k)_\Omega \to y^*$ without loss of generality

Under the assumption that there are finitely many $Y \in \overline{\mathcal{M}_{r_g}}$ such that $P_\Omega(Y) = y^*$, we next claim that $Y_k, k \geq 1$ are bounded. Indeed, for any matrix $Y \in \overline{\mathcal{M}_{r_g}}$, the set of matrices with rank $\leq r_g$, if we write the entries in Y_{Ω^c} as variables, say $\mathbf{x} \in \mathbb{R}^{n^2-m}$ while the entries $Y|_\Omega$ are known, the determinant of any $(r+1) \times (r+1)$ minor of Y will be zero and is a polynomial function of variables $\mathbf{x}$ with coefficients based on the known entries $Y|_\Omega$. Thus, vanishing of all $(r+1) \times (r+1)$ minors would form a set of $(\binom{n}{r_g+1})^2$ polynomial equations with variables $\mathbf{x}$ and coefficients from entries in $Y|_\Omega$. By our assumption, this set of polynomial equations have *finitely* many solutions when the coeffficients of the system is derived from the Ω entries of $y^\star$. Since the zeros of these polynomial equations are continuously dependent on the coefficients of polynomial functions, we see that there are finitely many solutions to the polynomial system when coefficients are derived from $(Y_k)_\Omega$ that are sufficiently close to $y^\star$. We can bound the zeros by using the coefficients. More precisely, these

polynomial equations can be reduced to a triangular system (cf. Chen and Moreno Maza (2011)), that is, writing $\mathbf{x} = (x_1, \cdots, x_{n^2-m})$ for a fixed order of these unknown entries,

$$\begin{cases} f_1(x_1) = 0, \\ f_2(x_1, x_2) = 0, \\ \cdots\cdots, \\ f_{n^2-m}(x_1, \cdots, x_{n^2-m}) = 0 \end{cases} \tag{15}$$

for a set of polynomial functions $f_1, \cdots, f_{n^2-m}$ by using one of the computational methods discussed in Aubry and Maza (1999). Certainly, for each $k \geq 1$, these f_i are dependent on k in the sense that the coefficients of f_i are dependent on the values $Y_k|_\Omega$. Then we can use any standard bound of the zeros of univariate polynomials to find a bound of these variables $\mathbf{x}$ iteratively from the reduced system above. Indeed, the bound on x_1 of this system is obtained by $\max\{1, |a_i|, i = 1, \cdots, r+1\}$ with coefficients a_i of the first univariate equation $f_1 = 0$ which are dependent on $Y_k|_\Omega$. Since $Y_k|_\Omega$ is bounded by C_1, we see x_1 is bounded in terms of C_1. Then x_2 can be bounded from the second equation which is now univariate if assuming x_1 is known. x_2 can be bounded in terms of the coefficients of f_2 and the bound on x_1. And so on. In summary, all the entries of Y_k with indices in Ω^c can be bounded in terms of the entries in $Y_k|_\Omega$. In other words, $\|Y_k\| \leq C_2 < \infty$ with a positive constant C_2 for all $k \geq 1$ which is dependent on C_1 above.

It now follows that there exists a subsequence Y_{k_j} which converges to $Y^\star$. Next by (12), X_k are bounded because of Y_k are bounded and hence, $X_k, , k \geq 1$ have a convergent subsequence and $X_{k_j} \to X^\star$ when $k_j \to \infty$ without loss of generality. By Lemma 17, we have $(Y^\star)_{\Omega^c} = (X^\star)_{\Omega^c}$. Finally, it is easy to see (14) which follows from the facts that set $\mathcal{A}_\Omega$ and set $\overline{\mathcal{M}_{r_g}}$ are closed sets. These complete the proof. ■

Although we do not know how to check if there are only finitely many matrices $Y \in \overline{\mathcal{M}_r}$ satisfying $(Y)_\Omega = \mathbf{x}$, we can see if the norms of Y_k are bounded or not from the algorithm. If they are bounded, the conclusions of Theorem 18 hold. In general, $X^\star \neq Y^\star$ as r_g is not equal to $\text{rank}(M)$. For example, when $r_g < \text{rank}(M)$, Y^* will not be equal to M and hence, $Y^\star$ does not satisfy $(Y^\star)_\Omega = M_\Omega$ in general. Of course X^* satisfies the interpolation conditions $(X^*)_\Omega = M_\Omega$, but $\text{rank}(X^\star)$ may be bigger than r_g. That is, informally speaking, when $r_g < \text{rank}(M)$, the chance of $X^\star = M$ is bigger than the chance $Y^* = M$. On the other hand, when $r_g > \text{rank}(M)$, there are more possibilities of matrices with rank $= r_g$ satisfying the interpolatory conditions. Anyway, if $X^\star - Y^\star \neq 0$, the guess r_g is not correct and we need to increase r_g.

Finally, even though $X^* \neq Y^*$ in general, they satisfy the following nice property.

Proposition 19 *Let $X^\star$ and $Y^\star$ be matrices in (14) Then,*

$$Y^\star(X^\star)^\top = Y^\star(Y^\star)^\top \text{ and } (Y^\star)^\top X^\star = (Y^\star)^\top Y^\star.$$

Proof Using Lemma 13, we obtain

$$\langle AY^\star + Y^\star B, X^\star - Y^\star \rangle = 0$$

for all $A, B \in \mathbb{R}^{n\times n}$ which implies

$$\langle A^\top, Y^\star(X^\star - Y^\star)^\top \rangle + \langle B, (Y^\star)^\top (X^\star - Y^\star) \rangle = 0$$

for all $A, B \in \mathbb{R}^{n\times n}$. Hence,

$$Y^\star(X^\star - Y^\star)^\top = 0 \text{ and } (Y^\star)^\top (X^\star - Y^\star) = 0.$$

Rearranging the above equations, we obtain the required result. ■

3. Numerical Results

In this section, we first present some results based on the simple initial guess $X_0 = P_\Omega(M)$. The robustness of Algorithm 1 was demonstrated in Jiang et al. (2017). We shall not repeat the similar numerical experimental results. We mainly present numerical results based on a good strategy to choose quality initial guesses which lead even better performance of Algorithm 1. That is, we recall an efficient computational algorithm called OR1MP for matrix completion in Wang et al. (2015). We use the OR1MP algorithm to get a completed matrix which serves as an initial guess X_0. Our numerical experimental results show that this new initial guess gives more accurate completion. We measure the error matrices by using the maximum norm of all entries of the matrices. One can see that the maximum norm error is very small and hence, the recovered matrix is very accurate. We shall also use Algorithm 1 to recover images from their partial pixel values and demonstrate that Algorithm 1 is able to recover the images better visually. Thus, this section is divided into two subsections.

3.1 Numerical Results: Initial Matrices from the OR1MP Algorithm

In all the experiments in this subsection, we used the initial matrix X_0 from the OR1MP algorithm in Wang et al. (2015) based on the $P_\Omega(M)$ using a few iterations, that is, $X_0 = \text{OR1MP}(P_\Omega(M))$.

Example 4 *In this example, we show the maximum missing rate that Algorithm 1 can recover a matrix when its rank is fixed. Together we show the computational times. Abbreviations used in Tables in this example are as follows:*
M.R. *= Missing Rate, the fraction of missing entries* $= \frac{m}{n^2}$,
O.R. *= oversampling ratio* $= \frac{m}{2nr-r^2}$,
M.C.E. *= Maximum Component Error* $= \max_{i,j}|(X_{recovered})_{i,j} - M_{i,j}|$,
A.R.E. *= Average Relative Error* $= \|P_\Omega(Y_k) - P_\Omega(M)\|_F / \|P_\Omega(M)\|_F$,

Table 1: Numerical results based on 100×100 matrices averaged over 20 runs

Rank	M.R.	O.R.	M.C.E	A.R.E	Time
2	0.80	5	9.5202e-04	3.7217e-05	0.4171
5	0.61	4	6.0350e-04	1.0894e-05	0.2648
10	0.43	3	4.4343e-04	4.4977e-06	0.2778
20	0.28	2	6.0317e-04	2.0486e-06	0.8492
35	0.25	1.3	1.2698e-06	0.0015	2.8798
50	0.025	1.3	0.0013	7.2350e-07	1.2605

Example 5 *Next we provide another tables to show that our algorithm is very effective in recovering the original matrix. We let the missing rate* $= 0.1, 0.2, \cdots, 0.9$ *and find the largest rank our algorithm can complete within maximum norm error* $< 1e-3$*, that is, every entry of the completed matrix is accurate to the first three digits. That is, for a fixed missing rate* δ*, we randomly find the known indices set* Ω *with* $|\Omega|/(n^2) = 1-\delta$ *and then we randomly generate a matrix* M *of size* $n \times n$ *with rank* $r \geq 1$*. We use* M_Ω*,* Ω*, and* r *to recover* M *(the stopping criterion is* $1e-5$ *of the consecutive iterations), check if the completed matrix* $\widehat{M}$ *approximates* M *in the maximum norm within* $\epsilon = 1e-3$*, and repeat*

Table 2: Numerical results based on 250×250 matrices averaged over 20 runs

Rank	M.R.	O.R.	M.C.E	A.R.E	Time
10	0.76	3	6.6930e-04	3.0595e-06	1.2283
20	0.53	3	2.2215e-04	1.0460e-06	1.3495
50	0.28	2	2.0560e-04	3.3083e-07	2.2624
75	0.18	1.6	2.6951e-04	2.0955e-07	4.5208
100	0.168	1.3	3.9345e-04	1.6690e-07	14.3622
125	0.025	1.3	6.2102e-04	1.1374e-07	8.8464

Table 3: Numerical results based on 500×500 matrices averaged over 10 runs

Rank	M.R.	O.R.	M.C.E	A.R.E	Time
25	0.70	3	2.8565e-04	5.5169e-07	4.5253
50	0.62	2	1.6818e-04	2.4458e-07	10.7270
100	0.28	2	8.6199e-05	8.0210e-08	11.3425
150	0.23	1.5	1.2031e-04	5.6053e-08	35.3097
200	0.04	1.5	1.5896e-04	3.2623e-08	24.1821
250	0.0250	1.3	3.3090e-04	2.8449e-08	46.9267

Table 4: Numerical results based on 1000×1000 matrices averaged over 10 runs

Rank	M.R.	O.R.	M.C.E	A.R.E	Time
50	0.70	3	6.8718e-05	1.3722e-07	30.1813
100	0.52	2.5	3.7074e-05	5.2213e-08	50.0631
200	0.10	2.5	2.6120e-05	1.2338e-08	42.7043
300	0.05	1.85	5.1339e-05	1.0448e-08	83.9782
400	0.04	1.5	7.1099e-05	8.0391e-09	186.2271
500	0.0025	1.33	2.4592e-04	6.5708e-09	226.3912

the computation in 10 times. If all 10 computations are able to accurately recover M, we advance r by $r+1$ and repeat the above procedures until the accurate recovery is less than 10 times for a fixed r. In this way, we can find the largest rank for a fixed missing rate. As we use two initial guesses, we summarize the computational results in Table 5.

missing rates	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
largest ranks	30	16	19	14	9	7	5	2	1	OR1MP
largest ranks	13	14	13	10	9	7	3	2	1	M_Ω

Table 5: maximum ranks are based on matrices of size 100×100 with initial values from OR1MP (second row) and from the initial matrix M_Ω (third row)

From Table 5, we can see that using OR1MP algorithm to generate an initial guess for Algorithm 1 is much better when the rates of missing entries are small. When the rate of missing entries are large, the performance is similar. If this table is compared with the ones in Wei et al. (2016), we remind the reader that we use a much tougher criterion $\epsilon = 1e-3$ in the maximum norm to find the maximum rank than the relative Frobenius norm error used in Wei et al. (2016).

If we use the standard relative Frobenius norm error, we have largest ranks that Algorithm 1 can recover 100% times listed in Table 6 with two different initial guesses. We can see that the performance increases greatly when using a completed matrix from OR1MP algorithm.

missing rates	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
largest ranks	132	106	83	68	41	40	25	14	4	OR1MP
largest ranks	33	29	25	24	19	15	11	8	4	M_Ω

Table 6: maximum ranks are based on matrices of size 200×200 with initial values from OR1MC (second row) and from the initial matrix M_Ω (third row)

3.2 Image Recovery from Partial Pixel Values

We shall use Algorithm 1 to recover images from partial pixel values.

Example 6 *Let us use the standard images knee, penny and thank as testing matrices of pixel values. The image knee is of size* 691×691. *The image penny is a matrix of size* 128×128 *and the image thank is of size* 300×300. *For image knee, we use a missing rate* 0.85 *to generate* M_Ω *and use rank=25 to find an approximation of the image knee by using the well-known matrix completion OR1MP algorithm in Wang et al. (2015), then we feed the approximation as an initial guess to Algorithm 1 to get a better approximation. Also we use the same known entries* M_Ω *as an initial guess in our Algorithm 1 to find an approximation of the image directly. All these images are shown in Figure 2. We do the same for the images penny and thank. See Figures 3 and 4. Visually, we can see that starting from an initial guess obtained from the OR1MP algorithm, our Algorithm 1 produces a much better approximation to the image. For image penny, we are able to see the face of Lincoln and the word as well as number 1984 are much cleaner although the root-mean square error (RMSE) may not be better. Many images have been experimented with similar performance.*

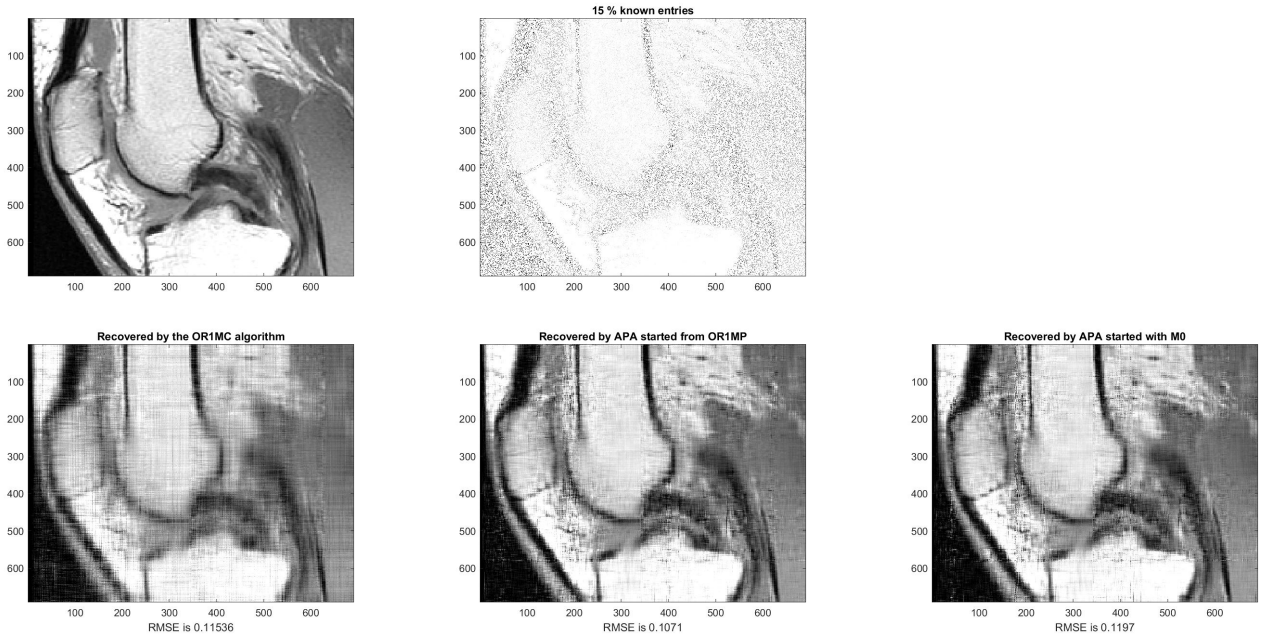

Figure 2: The top row: The original image and the image of 15% known entries; The bottom row: The outputs from Algorithm OR1MP, Algorithm 1 with initial guess from the Algorithm OR1MP and Algorthm 1 from the 15% known entries based on rank 25.

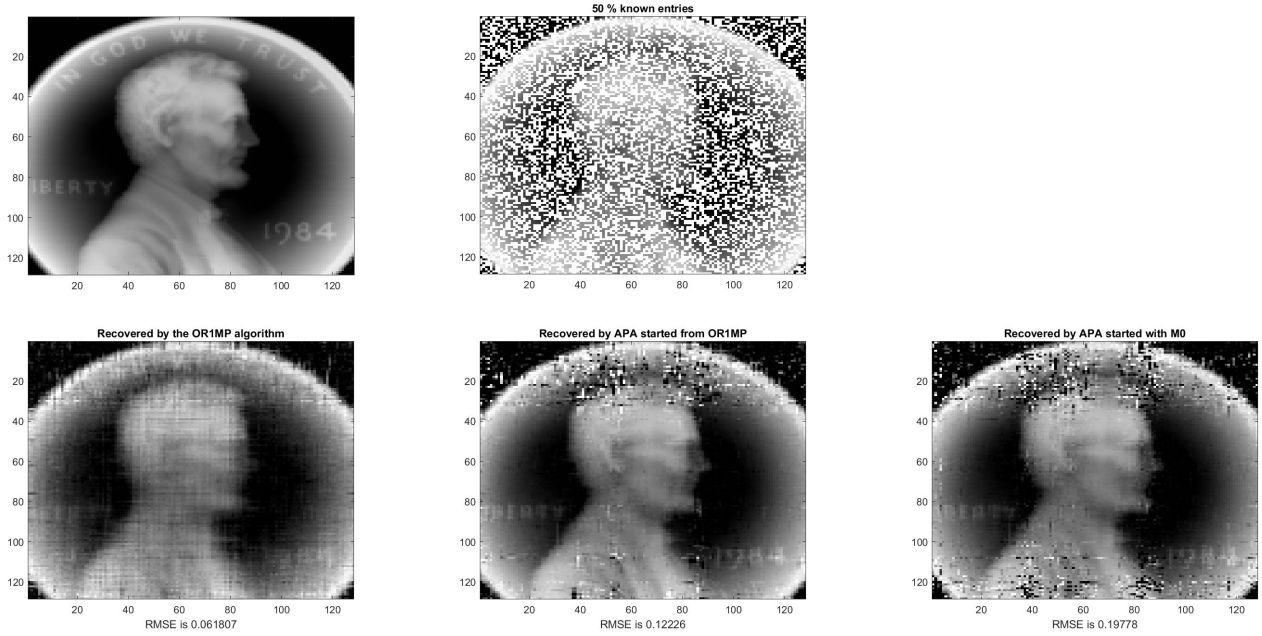

Figure 3: The top row: The original image and the image of 50% known entries; The bottom row: The outputs from Algorithm OR1MP, Algorithm 1 with initial guess from the Algorithm OR1MP and Algorthm 1 from the 50% known entries based on rank 25.

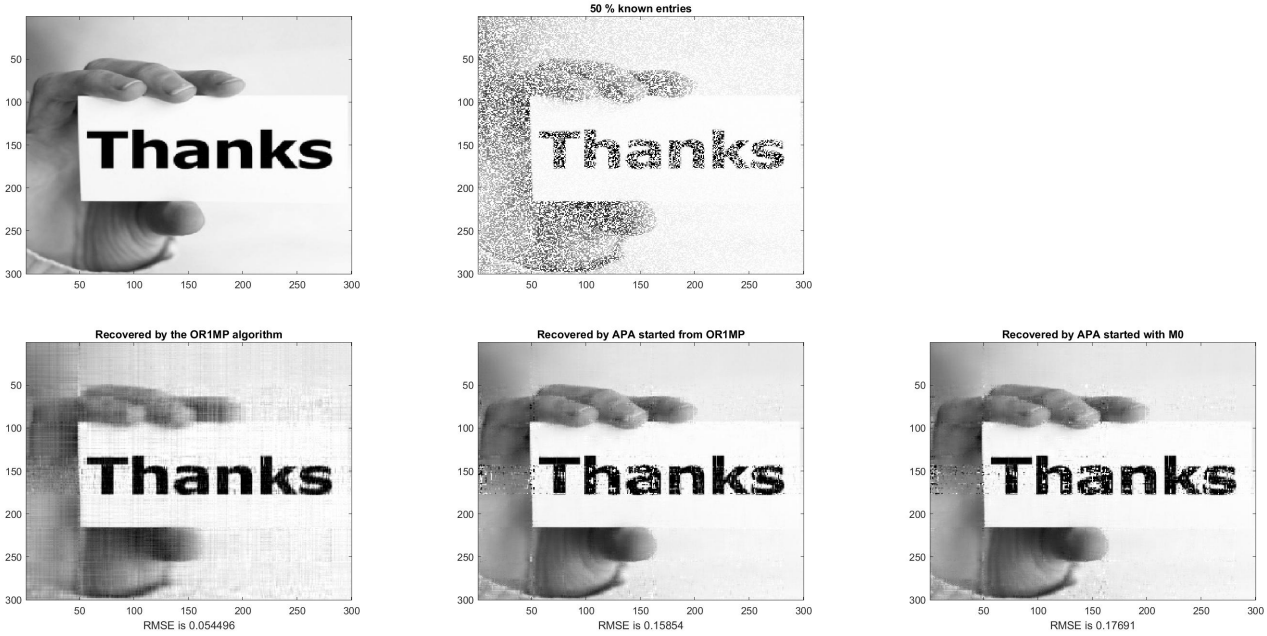

Figure 4: The top row: The original image and the image of 50% known entries; The bottom row: The outputs from Algorithm OR1MP, Algorithm 1 with initial guess from the Algorithm OR1MP and Algorthm 1 from the 50% known entries based on rank 25.

4. Alternating Projection Algorithm for Sparse Solution Recovery Problem

In this section, we will use the same ideas of alternating projection discussed in the previous section to study the following classical problem in the area of compressed sensing:

$$\underset{X}{\text{minimize}} \quad \|X\|_0 \tag{16}$$

$$\text{subject to} \quad A\mathbf{x} = \mathbf{b}, \tag{17}$$

where $A \in \mathbb{R}^{n\times N}, \mathbf{x} \in \mathbb{R}^N, \mathbf{b} \in \mathbb{R}^n, n << N$ and $\|\mathbf{x}\|_0$ is the ℓ_0 quasi-norm of a vector $\mathbf{x}$. Recall that the ℓ_0 quasi-norm of a vector is the number of non-zero components of the vector. Let $\mathcal{L}_s(\mathbb{R}^N)$ denote the collection of all $s-$sparse vectors in $\mathbb{R}^N$,

$$\mathcal{L}_s(\mathbb{R}^N) := \left\{x \in \mathbb{R}^N \mid \quad \|x\|_0 = s\right\}$$

and $\mathcal{P}_{\mathcal{L}_s}$ and $\mathcal{P}_{\mathcal{A}}$ denote the projection onto the set $\mathcal{L}_s(\mathbb{R}^N)$ and the affine space $\mathcal{A} := \{\mathbf{x} : A\mathbf{x} = \mathbf{b}\}$, respectively. It is easy to know $\mathcal{A} = \text{Null}(A) + \mathbf{x}_0$, where $\mathbf{x}_0 \in \mathbb{R}^N$ satisfies $A\mathbf{x}_0 = \mathbf{b}$. Note that the projection $\mathcal{P}_{\mathcal{L}_s}(\mathbf{x}_k)$ can be computed easily by setting the smallest $n - s$ components of the vector $\mathbf{x}_k$ to zero.

Our algorithm can be stated as follows:

Algorithm: Alternating Projection Algorithm for ℓ_0 Minimization

Data: Sparsity s of the solution $\mathbf{x}_\star$, the tolerance ϵ whose default value is 1e-6
Result: $\mathbf{x}_k$ a close approximation of $\mathbf{x}_\star$
Initialize $\mathbf{x}_0$ to a random vector in the affine space $\mathcal{A}$;
repeat
  Step 1: $\mathbf{y}_k = \mathcal{P}_{\mathcal{L}_s}(\mathbf{x}_k)$
  Step 2: $\mathbf{x}_{k+1} = \mathcal{P}_{\mathcal{A}}(\mathbf{y}_k)$;
until *The smallest $n_2 - s$ components of $\mathbf{x}_{k+1}$ have magnitude less than ϵ*;

We first discuss the convergence of Algorithm 2. Then we shall present its numerical performance in the next section. As a good initial guess is very important to have a quick convergence, we shall explain a few approaches to obtain reasonable initial guesses.

4.1 Convergence of Algorithm 2

We begin with some elementary results.

Lemma 20 *Let $\mathcal{L}_s(\mathbb{R}^n)$ be the collection defined as follows.*

$$\mathcal{L}_s(\mathbb{R}^n) = \bigsqcup_{\mathcal{I}} \{x \in \mathbb{R}^n \mid x_j = 0 \quad \forall j \in \mathcal{I}^c\},$$

where the index set $\mathcal{I}$ ranges over all the subsets of $\{1, 2, \cdots, n_1\}$ which has cardinality s. Here, $\bigsqcup_{\mathcal{I}}$ stands for the disjoint union over $\mathcal{I}$. Then $\mathcal{L}_s(\mathbb{R}^n)$ consists of a disjoint union of affine spaces.

Proof It is easy to see that the statement is correct. ■

Lemma 21 *The set of vectors in $\mathbb{R}^N$ for which $\mathcal{P}_{\mathcal{L}_s}(x)$ is single-valued, is given by the open set*

$$V_s = \left\{x \in \mathbb{R}^{n_2} \mid |x_{i_1}| \geq |x_{i_2}| \geq \cdots |x_{i_{n_2}}|, |x_{i_{s+1}}| \neq |x_{i_s}|\right\}$$

consisting of vectors which has the property that if one arrange the components in decreasing order of magnitude, then s^{th} and $(s+1)^{th}$ terms are distinct.

Proof We first start by noting that the projection $\mathcal{P}_{\mathcal{L}_s}(\mathbf{x})$ is obtained by setting the smallest $N - s$ components in magnitude of the vector x to zero. Hence, the projection is single-valued if the $N - s$ smallest components of $\mathbf{x}$ are in unique positions(indices). Hence we must have that the $(N - s)^{th}$ and $(N - s + 1)^{th}$ components of x must be distinct. Now we will show that the set V_s is an open set. Let $\mathbf{x} = (x_1, x_2, \cdots, x_N) \in \mathbb{R}^N$ with $|x_{i_1}| \geq |x_{i_2}| \geq \cdots |x_{i_N}|, |x_{i_{s+1}}| \neq |x_{i_s}|$. Let

$$\epsilon := \frac{||x_{i_{s+1}}| - |x_{i_s}||}{4}$$

Consider an open ball $B_\epsilon(\mathbf{x})$ centered at $\mathbf{x}$ of radius ϵ. We have, for all $\mathbf{y} \in B_\epsilon(\mathbf{x})$ and $j \in \{1, 2, \cdots, N\}$,

$$||y_j| - |x_j|| \leq |y_j - x_j| \leq \|\mathbf{y} - \mathbf{x}\| < \epsilon$$

Therefore, we have

$$|y_{i_{j+1}}| \leq |x_{i_{j+1}}| + ||y_{i_{j+1}}| - |x_{i_{j+1}}|| < |x_{i_{s+1}}| + \epsilon < \frac{||x_{i_{s+1}}| + |x_{i_s}||}{2}$$

for $j \geq s$. Similarly,

$$|y_{i_j}| \geq |x_{i_j}| - ||x_{i_j}| - |y_{i_j}|| > |x_{i_s}| - \epsilon > \frac{||x_{i_{s+1}}| + |x_{i_s}||}{2}$$

for $j \leq s$. Hence, we deduce that, for all $\mathbf{y} \in B_\epsilon(\mathbf{x})$ and $j \in \{1, 2, \cdots, s\}$, $|y_{i_j}| \geq |y_{i_s}|$, which implies that $\mathbf{y} \in V_s$ and, therefore, $B_\epsilon(\mathbf{x}) \subset V_s$. ■

Next let us recall the following well-known results.

Theorem 22 (Von Neumann, 1950Von Neumann (2016)) *If L_1 and L_2 are two closed subspaces of a Hilbert space X, then the sequence of operators*

$$\mathcal{P}_{L_1}, \mathcal{P}_{L_2}\mathcal{P}_{L_1}, \mathcal{P}_{L_1}\mathcal{P}_{L_2}\mathcal{P}_{L_1}, \mathcal{P}_{L_2}\mathcal{P}_{L_1}\mathcal{P}_{L_2}\mathcal{P}_{L_1}, \cdots$$

converge to $\mathcal{P}_{L_1 \cap L_2}$. In other words,

$$\lim_{k\to\infty} (\mathcal{P}_{L_2}\mathcal{P}_{L_1})^k(x) = \mathcal{P}_{L_1 \cap L_2}(x)$$

for all $x \in X$.

Proof Refer to Von Neumann (2016) Chapter 13, Theorem 13.7 for a proof. ■

Theorem 23 *If $\mathbf{x}_\star$ is an isolated point of $\mathcal{L}_s(\mathbb{R}^N) \cap \mathcal{A}$. Then, Algorithm 2 will locally converge to $\mathbf{x}_\star$ linearly.*

Proof Let $\mathcal{I} = \text{Supp}(\mathbf{x}_\star)$ be the support of $\mathbf{x}_\star$ and $s = \|\mathbf{x}_\star\|_0$. Consider an open set V_s of vectors which has the property that their $n - s$ smallest components are are in unique positions(indices). In fact, V_s can be concretely described as

$$V_s = \left\{\mathbf{x} \in \mathbb{R}^N \mid \quad |x_{i_1}| \geq |x_{i_2}| \geq \cdots |x_{i_{n_2}}|, |x_{i_s}| \neq |x_{i_{s+1}}|\right\}.$$

Clearly $\mathbf{x}_\star \in V_s$. Let $B(r)$ be an open ball centered at $\mathbf{x}_\star$ and of radius r completely contained inside V_s. Since $B(r) \subseteq V_s$, for any $\mathbf{x} \in B(r)$, the projection $\mathcal{P}_{\mathcal{L}_s}(\mathbf{x})$ is uniquely defined. Since affine spaces in a finite dimensional Euclidean space are closed, one can shrink the ball $B(r)$, if necessary, such that the restriction $\mathcal{L}_s(\mathbb{R}^{n_2})|_{B(r)}$ of the set of $s-$sparse vectors to the open set $B(r)$ is an affine space. Then under the assumption the hypothesis in this theorem, the result follows from Theorem 22. ■

Lemma 24 *Assume A has the following property:*

$$\mathcal{L}_s(\mathbb{R}^N) \cap \text{Null}(A) = \{0\}, \tag{18}$$

where Null(A) *is the null space of A. Furthermore, assume that $\mathbf{x}_\star \in \mathcal{L}_s(\mathbb{R}^N) \cap \mathcal{A}$. Then $\mathbf{x}_\star$ is an isolated point of $\mathcal{L}_s(\mathbb{R}^N) \cap \mathcal{A}$.*

Proof Assume, on the contrary, that $\mathbf{x}_\star$ is not an isolated point of the set $\mathcal{L}_s(\mathbb{R}^{n_2}) \cap \mathcal{A}$. Then, since A and $\mathcal{L}_s(\mathbb{R}^N)$ are locally affine spaces, there exist a linear space L of dimension greater than or equal to 1 such that $L + \mathbf{x}_\star \subseteq \mathcal{L}_s(\mathbb{R}^N) \cap \mathcal{A}$. Since each of the intersecting spaces are affine spaces locally, L must lie also in the intersection of their tangent spaces. Hence,

$$L \subseteq T_{\mathcal{L}_s(\mathbb{R}^N)}(\mathbf{x}_\star) \cap \mathrm{Null}(A)$$

where $T_{\mathcal{L}_s(\mathbb{R}^N)}(\mathbf{x}_\star)$ is the tangent space to $\mathcal{L}_s(\mathbb{R}^N)$ at the point $\mathbf{x}_\star$. Now, since $\mathcal{L}_s(\mathbb{R}^N)$ is an union of linear spaces, let us assume $\mathbf{x}_\star \in L_0 \subseteq \mathcal{L}_s(\mathbb{R}^N)$ lies in a linear space L_0 contained in $\mathcal{L}_s(\mathbb{R}^N)$. Therefore, we have

$$L \subseteq T_{L_0}(\mathbf{x}_\star) \cap \mathrm{Null}(A) = L_0 \cap \mathrm{Null}(A) \subseteq \mathcal{L}_s(\mathbb{R}^N) \cap \mathrm{Null}(A) = \{0\}$$

which leads to the contradiction as L is of dimension greater than or equal to 1. Note that, in order to derive the equality in the last equation, we have used the fact that the tangent space of a linear space is the linear space itself. ■

The discussion above leads to our final result in this section.

Theorem 25 *Under the assumption (18) in Lemma 24, Algorithm 2 will converge linearly for any starting initial guess* $\mathbf{x}_0$.

Proof We simply combine Lemma 24 and Theorem 23 together to have this result. ■

4.2 Numerical Results from Algorithm 2 for Sparse Vector Recovery

We have used Algorithm 2 to compute sparse solutions and compare the performance of several existing algorithms. Mainly, we compare with the iteratively reweighted ℓ_1 minimization (CWB for short) in Candes et al. (2008), the L^1 greedy algorithm (KP) proposed in Kozlov and Petukhov (2010), the FISTA in Beck and Teboulle (2009), the hard iterative pursuit (HTP) in Foucart (2011), and generalized approximate message passing algorithm (GAMP) in Donoho et al. (2009), Rangan (2011). LV stands for our Algorithm 2. We present the frequency of recovery of Gaussian random matrices of size 128×256 with sparsity from $10 - -70$ over 500 repeated runs with a tolerance $1e-3$ in maximum norm. In Figure 5(left figure), we show the performance of various algorithms. Next we repeat the same experiments based on uniform random matrices of size 128×256. The performance of frequency of recovery from various algorithm is shown in Figure 5 (right). In this case, it is known that the GAMP is not good.

5. Remarks on Existence of Matrix Completion

Recall $\mathcal{M}_r$ is the set of all matrices of size $n \times n$ with rank r and $\overline{\mathcal{M}_r}$ is the set of all matrices with rank $\leq r$. It is clear that $\overline{\mathcal{M}_r}$ is the closure of $\mathcal{M}_r$ in the Zariski sense (cf. Zariski (1958)). It is easy to see that dimension $\mathcal{M}_r$ is $2nr - r^2$ (cf. Proposition 12.2 in Harris (2013) for a proof). Then the dimension of $\overline{\mathcal{M}_r}$ is also $2nr - r^2$. Also, it is clear that $\overline{\mathcal{M}_r}$ is an algebraic variety. In fact, $\overline{\mathcal{M}_r}$ is an irreducible variety.

Lemma 26 $\overline{\mathcal{M}_r}$ *is an irreducible variety..*

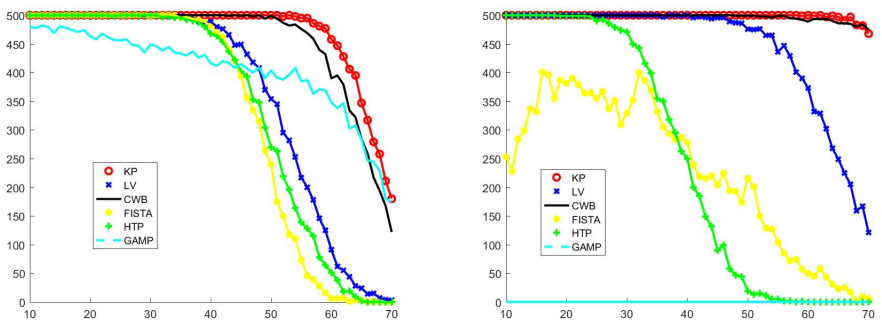

Figure 5: Frequency of Sparse Recovery by Various Algorithms from Gaussian random matrices (left) and from uniform random matrices(right)

Proof Denote by $GL(n)$ the set of invertible $n \times n$ matrices. Consider the action of $GL(n) \times GL(n)$ on $M_n(R)$ given by: $(G_1, G_2) \cdot M \mapsto G_1 M G_2^{-1}$, for all $G_1, G_2 \in GL(n)$. Fix a rank r matrix M. Then the variety $\mathcal{M}_r$ is the orbit of M. Hence, we have a surjective morphism, a regular algebraic map described by polynomials, from $GL(n) \times GL(n)$ onto $\mathcal{M}_r$. Since $GL(n) \times GL(n)$ is an irreducible variety, so is $\mathcal{M}_r$. Hence, the closure $\overline{\mathcal{M}_{r_g}}$ of the irreducible set $\mathcal{M}_{r_g}$ is also irreducible *c.f* (cf. Example I.1.4 in Hartshorne (2013)). ■

Consider the map

$$\Phi_\Omega : \overline{\mathcal{M}_r} \to \mathbb{C}^m$$

given by projecting any matrix $X \in \overline{\mathcal{M}_r}$ to its entries in position Ω which form a vector in $\mathbb{R}^m$. Thus, $\Phi_\Omega(\overline{\mathcal{M}_r})$ are exactly the set of all $r-$feasible vectors in $\mathbb{C}^m$. As the projection Φ_Ω is 'nice' (a polynomial map) unlike a Peano curve mapping $[0,1] \to [0,1]^2$), we expect that $\dim(\Phi_\Omega(\overline{\mathcal{M}_r}))$ is less than or equal to $\dim(\overline{\mathcal{M}_r})$ which is less than the dimension of $\mathbb{C}^m$. Thus, $\Phi_\Omega(\overline{\mathcal{M}_r})$ is not able to occupy the whole space $\mathbb{C}^m$. The Lebesgue measure of $\Phi_\Omega(\overline{\mathcal{M}_r})$ is zero and hence, randomly choosing a vector $\mathbf{x} \in \mathbb{C}^m$ will not be in $\Phi_\Omega(\overline{\mathcal{M}_r})$ most likely. Certainly, these intuitions should be made more precise. Recall the following result from Theorem 1.25 in Sec 6.3 of Shafarevich and Hirsch (1994).

Lemma 27 *Let $f : X \to Y$ be a regular map between irreducible varieties. Suppose that f is surjective: $f(X) = Y$, and that $\dim(X) = n$, $\dim(Y) = m$. Then $m \leq n$, and*

1. *$\dim(F) \geq n - m$ for any $y \in Y$ and for any component F of the fibre $f^{-1}(y)$;*
2. *there exists a nonempty open subset $U \subset Y$ such that $\dim(f^{-1}(y)) = n - m$ for $y \in U$.*

We are now ready to prove

Theorem 28 *If one chooses randomly the entries of a matrix in the positions Ω, probability of completing the matrix to a rank r matrix with given known entries is 0.*

Proof We mainly use Lemma 27. Let $X = \overline{\mathcal{M}_r}$ which is an irreducible variety by Lemma 26. Let $Y = \Phi_\Omega(\overline{\mathcal{M}_r})$ which is also an irreducible variety as it is a continuous image of the irreducible variety $\overline{\mathcal{M}_r}$. Clearly, Φ_Ω is a regular map, we have $\dim \Phi_\Omega(\overline{\mathcal{M}_r}) \leq \dim(\overline{\mathcal{M}_r}) = 2nr - r^2 < m$. Thus, $\Phi_\Omega(\overline{\mathcal{M}_r})$ is a proper lower dimensional closed subset in $\mathbb{C}^m$. For almost all points in $\mathbb{C}^m$, they do not belong to $\Phi_\Omega(\overline{\mathcal{M}_r})$. In other words, for almost all points $\mathbf{x} \in \mathbb{C}^m$, there is no matrix $X \in \overline{\mathcal{M}_r}$ such that

$\Phi_\Omega(X) = \mathbf{x}$. ■

Next define the subset $\chi_\Omega \subset \overline{\mathcal{M}_r}$ by

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

As we are working over Noetherian fields like $\mathbb{R}$ or $\mathbb{C}$, it is worthwhile to keep in mind that all zero dimensional varieties over such fields will have only finitely many points. Next we recall the following result from Proposition 11.12 in Harris (2013).

Lemma 29 *Let X be a quasi-projective variety and $\pi : X \to \mathbb{P}^m$ a regular map; let Y be closure of the image. For any $p \in X$, let $X_p = \pi^{-1}\pi(p)) \subseteq X$ be the fiber of π through p, and let $\mu(p) = \dim_p(X_p)$ be the local dimension of X_p at p. Then $\mu(p)$ is an upper-semicontinuous function of p, in the Zariski topology on X - that is, for any m the locus of points $p \in X$ such that $\dim_p(X_p) > m$ is closed in X. Moreover, if $X_0 \subseteq X$ is any irreducible component, $Y_0 \subseteq Y$ the closure of its image and μ the minimum value of $\mu(p)$ on X_0, then*

$$dim(X_0) = dim(Y_0) + \mu. \tag{19}$$

As we saw that $\dim(\Phi_\Omega(\overline{\mathcal{M}_r}) \leq \dim(\overline{\mathcal{M}_r})$, we can be more precise about these dimensions as shown in the following

Lemma 30 *Assume $m > \dim(\overline{\mathcal{M}_r})$. Then χ_Ω is open subset of $\overline{\mathcal{M}_r}$ and $\dim(\overline{\mathcal{M}_r}) = \dim(\overline{\Phi_\Omega(\overline{\mathcal{M}_r})}) = \dim(\Phi_\Omega(\overline{\mathcal{M}_r}))$ if and only if $\chi_\Omega \neq \emptyset$.*

Proof Assume $\dim(\overline{\mathcal{M}_r}) = \dim(\overline{\Phi_\Omega(\overline{\mathcal{M}_r})}) = \dim(\Phi_\Omega(\overline{\mathcal{M}_r}))$. Then using Lemma 27, there exists a nonempty open subset $U \subset \Phi_\Omega(\overline{\mathcal{M}_r})$ such that $\dim(\Phi_\Omega^{-1}(y)) = 0$ for all $y \in U$. This implies that $\Phi_\Omega^{-1}(y) \in \chi_\Omega$. Hence $\chi_\Omega \neq \emptyset$.

We now prove the converse. Assume $\chi_\Omega \neq \emptyset$. We will apply Lemma 29 above by setting $X = \overline{\mathcal{M}_{r_g}}$, $Y = \Phi_\Omega(\overline{\mathcal{M}_{r_g}})$ and $\pi = \Phi_\Omega$. Couple of things to note here are that it does not matter whether we take the closure in $\mathbb{P}^m$ or in $\mathbb{C}^m$ since $\mathbb{C}^m$ is an open set in $\mathbb{P}^m$ and the Zariski topology of the affine space $\mathbb{C}^m$ is induced from the Zariski topology of $\mathbb{P}^m$. $\overline{\mathcal{M}_{r_g}}$ is an affine variety. Therefore, it is a quasi-projective variety.

By our assumption, χ_Ω is not empty. It follows that there is a point $p \in Y$ such that $\pi^{-1}(p)$ is zero dimensional. Since zero is the least dimension possible, we have $\mu = 0$. Hence, using (19) above, we have $\dim(\overline{\mathcal{M}_r}) = \dim(\overline{\Phi_\Omega(\overline{\mathcal{M}_r})})$. But dimension does not change upon taking closure. So, $\dim(\Phi_\Omega(\overline{\mathcal{M}_r})) = \dim(\overline{\Phi_\Omega(\overline{\mathcal{M}_r})})$. Also, using Lemma 31, $\chi_\Omega = \{x \in X : \dim(\phi^{-1}\phi(x)) < 1\}$ is an open subset of $\overline{\mathcal{M}_r}$. ■

In the proof above, the following result was used. See I.8. Corollary 3 in Mumford (1999).

Lemma 31 *Let $\phi : X \to Y$ be a morphism of affine varieties. Let $\phi^{-1}\phi(x) = Z_1 \cup \cdots \cup Z_j$ be the irreducible components of $\phi^{-1}\phi(x)$. Let $e(x)$ be the maximum of the dimensions of the $Z_i, i = 1, \cdots, j$. Let $S_n(\phi) := \{x \in X : e(x) \geq n\}$. Then, for any $n \geq 1$, $S_n(\phi)$ is a Zariski closed subset of X. Equivalently $\{x \in X : \dim(\phi^{-1}\phi(x)) < n\}$ is an open subset of X.*

Finally, we need the following

Definition 32 *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. See Example 14.4.11 in Fulton (2013), i.e.

Example 7 *Degree of the algebraic variety* $\overline{\mathcal{M}_r}$ *is*

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

We are now ready to prove another main result in this section.

Theorem 33 *Assume that there exist a finite r-feasible vector* $\mathbf{x} \in \mathbb{C}^m$ *over the given* Ω*. Then, with probability 1, the vector* $\mathbf{x}$ *is finite r-feasible. In other words, if one randomly chooses a feasible vector* $\mathbf{x}$ *in the positions* Ω*, the matrix can be completed into a rank-r matrix only in* finitely many ways *with probability 1. In additional, the number of ways to complete will be less than or equal to* $\prod_{i=0}^{n-r-1} \frac{\binom{n+i}{r}}{\binom{r+i}{r}}$.

Proof We begin by noting that, both $\overline{\mathcal{M}_r}$ and $\Phi_\Omega(\overline{\mathcal{M}_r})$ are irreducible varieties. So, the closure $\overline{\Phi_\Omega(\overline{\mathcal{M}_r})}$ is also an irreducible variety. By the assumption and using Lemma 30, $\dim(\overline{\mathcal{M}_r}) = \dim(\overline{\Phi_\Omega(\overline{\mathcal{M}_r})})$. Hence, applying Lemma 27, there exist a nonempty open subset $U \subset \overline{\Phi_\Omega(\overline{\mathcal{M}_r})}$ such that $\Phi_\Omega^{-1}(y)$ is zero-dimensional for all $y \in U$. In other words, If we choose the m entries in positions Ω of a matrix from the open set U, then there are finitely many ways to complete the matrix. The result now follows by recalling that a Zariski open set in an irreducible variety is a dense set whose complement has Lebesque measure zero.

When we fix m entries of a matrix M, the set of matrices of rank r which has those entries in the positions Ω are exactly the intersection points of the variety $\overline{\mathcal{M}_r}$ with m hyperplanes, namely the hyperplanes defined by equations of form $M_{ij} = constant$. Since $m > \dim(\overline{\mathcal{M}_r}) = 2nr - r^2$, the number of intersection points would be lesser than degree of $\overline{\mathcal{M}_r}$ generically. Now using the exact formula for the degree from Example 7, the result follows. ■

Regarding Theorem 18, we have the following open problem: given $\mathbf{x} \in \mathcal{C}^m$, how to check if there are only finitely many matrices $Y \in \overline{\mathcal{M}_r}$ satisfying $(Y)_\Omega = \mathbf{x}$.

Acknowledgments

The first author would like to thank Professors Xiaofei He and Jieping Ye for their hospitality during his visits at Zheijiang University and University of Michigan in 2014. The first author is partially supported by the National Science Foundation under the grant #DMS 1521537. The second author would like to thank Anand Deopurkar for his helpful comments and suggestions in regard to the algebraic-geometric aspects of the paper.

References

Philippe Aubry and Marc Moreno Maza. Triangular sets for solving polynomial systems: a comparative implementation of four methods. *Journal of Symbolic Computation*, 28(1):125 – 154, 1999. ISSN 0747-7171. doi: https://doi.org/10.1006/jsco.1999.0270. URL `http://www.sciencedirect.com/science/article/pii/S0747717199902705`.

Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. doi: 10.1137/080716542. URL https://doi.org/10.1137/080716542.

James Bennett, Stan Lanning, et al. The netflix prize. In *Proceedings of KDD cup and workshop*, volume 2007, page 35. New York, NY, USA, 2007.

T. Blumensath and M. E. Davies. Normalized iterative hard thresholding: Guaranteed stability and performance. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):298–309, April 2010. ISSN 1932-4553. doi: 10.1109/JSTSP.2010.2042411.

Thomas Blumensath and Mike E. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265 – 274, 2009. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2009.04.002. URL http://www.sciencedirect.com/science/article/pii/S1063520309000384.

Jian-Feng Cai, Emmanuel J. Cands, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010. doi: 10.1137/080738970. URL https://doi.org/10.1137/080738970.

Jian-Feng Cai, Tianming Wang, and Ke Wei. Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank hankel matrix completion. *Applied and Computational Harmonic Analysis*, 2017. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2017.04.004. URL http://www.sciencedirect.com/science/article/pii/S1063520317300295.

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

Emmanuel J Candes and Terence Tao. Decoding by linear programming. *IEEE transactions on information theory*, 51(12):4203–4215, 2005.

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

Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006a.

Emmanuel J Candes, Michael B Wakin, and Stephen P Boyd. Enhancing sparsity by reweighted 1 minimization. *Journal of Fourier analysis and applications*, 14(5):877–905, 2008.

Emmanuel J Candès et al. Compressive sampling. In *Proceedings of the international congress of mathematicians*, volume 3, pages 1433–1452. Madrid, Spain, 2006b.

Changbo Chen and Marc Moreno Maza. Algorithms for computing triangular decompositions of polynomial systems. In *Proceedings of the 36th international symposium on Symbolic and algebraic computation*, pages 83–90. ACM, 2011.

Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, and C Sinan Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 63(1):1–38, 2010.

Ronald A DeVore and Vladimir N Temlyakov. Some remarks on greedy algorithms. *Advances in computational Mathematics*, 5(1):173–187, 1996.

David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

Florian Feppon and Pierre FJ Lermusiaux. A geometric approach to dynamical model-order reduction. *arXiv preprint arXiv:1705.08521*, 2017.

Simon Foucart. Hard thresholding pursuit: an algorithm for compressive sensing. *SIAM Journal on Numerical Analysis*, 49(6):2543–2563, 2011.

Simon Foucart and Ming-Jun Lai. Sparsest solutions of underdetermined linear systems via q-minimization for 0¡ q 1. *Applied and Computational Harmonic Analysis*, 26(3):395–407, 2009.

William Fulton. *Intersection theory*, volume 2. Springer Science & Business Media, 2013.

Pinghua Gong, Changshui Zhang, Zhaosong Lu, Jianhua Huang, and Jieping Ye. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In *International Conference on Machine Learning*, pages 37–45, 2013.

Elaine T Hale, Wotao Yin, and Yin Zhang. Fixed-point continuation for \ell_1-minimization: Methodology and convergence. *SIAM Journal on Optimization*, 19(3):1107–1130, 2008.

Joe Harris. *Algebraic geometry: a first course*, volume 133. Springer Science & Business Media, 2013.

Robin Hartshorne. *Algebraic geometry*, volume 52. Springer Science & Business Media, 2013.

Prateek Jain, Raghu Meka, and Inderjit S Dhillon. Guaranteed rank minimization via singular value projection. In *Advances in Neural Information Processing Systems*, pages 937–945, 2010.

Xue Jiang, Zhimeng Zhong, Xingzhao Liu, and Hing Cheung So. Robust matrix completion via alternating projection. *IEEE Signal Processing Letters*, 24(5):579–583, 2017.

Inna Kozlov and Alexander Petukhov. Sparse solutions of underdetermined linear systems. In *Handbook of Geomathematics*, pages 1243–1259. Springer, 2010.

Ming-Jun Lai and Jingyue Wang. An unconstrained \ell_q minimization with 0q\leq1 for sparse solution of underdetermined linear systems. *SIAM Journal on Optimization*, 21(1):82–101, 2011.

Ming-Jun Lai and Wotao Yin. Augmented \ell_1 and nuclear-norm models with a globally linearly convergent algorithm. *SIAM Journal on Imaging Sciences*, 6(2):1059–1091, 2013.

Ming-Jun Lai, Yangyang Xu, and Wotao Yin. Improved iteratively reweighted least squares for unconstrained smoothed \ell_q minimization. *SIAM Journal on Numerical Analysis*, 51(2):927–957, 2013.

Shiqian Ma, Donald Goldfarb, and Lifeng Chen. Fixed point and bregman iterative methods for matrix rank minimization. *Mathematical Programming*, 128(1):321–353, 2011.

Leon Mirsky. Symmetric gauge functions and unitarily invariant norms. *The quarterly journal of mathematics*, 11(1):50–59, 1960.

Karthik Mohan and Maryam Fazel. Iterative reweighted algorithms for matrix rank minimization. *Journal of Machine Learning Research*, 13(Nov):3441–3473, 2012.

David Mumford. *The red book of varieties and schemes: includes the Michigan lectures (1974) on curves and their Jacobians*, volume 1358. Springer Science & Business Media, 1999.

Sundeep Rangan. Generalized approximate message passing for estimation with random linear mixing. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 2168–2172. IEEE, 2011.

Igor Rostislavovich Shafarevich and Kurt Augustus Hirsch. *Basic algebraic geometry*, volume 2. Springer, 1994.

Jared Tanner and Ke Wei. Normalized iterative hard thresholding for matrix completion. *SIAM Journal on Scientific Computing*, 35(5):S104–S125, 2013.

Jared Tanner and Ke Wei. Low rank matrix completion by alternating steepest descent methods. *Applied and Computational Harmonic Analysis*, 40(2):417–429, 2016.

Min Tao and Xiaoming Yuan. Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM Journal on Optimization*, 21(1):57–81, 2011.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

Kim-Chuan Toh and Sangwoon Yun. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific Journal of Optimization*, 6(615-640):15, 2010.

Joel A Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information theory*, 50(10):2231–2242, 2004.

Bart Vandereycken. Low-rank matrix completion by riemannian optimization. *SIAM Journal on Optimization*, 23(2):1214–1236, 2013.

John Von Neumann. *Functional Operators (AM-22), Volume 2: The Geometry of Orthogonal Spaces.(AM-22)*, volume 2. Princeton University Press, 2016.

Jie Wang, Jiayu Zhou, Peter Wonka, and Jieping Ye. Lasso screening rules via dual polytope projection. In *Advances in Neural Information Processing Systems*, pages 1070–1078, 2013.

Zheng Wang, Ming-Jun Lai, Zhaosong Lu, Wei Fan, Hasan Davulcu, and Jieping Ye. Orthogonal rank-one matrix pursuit for low rank matrix completion. *SIAM Journal on Scientific Computing*, 37(1):A488–A514, 2015.

Ke Wei, Jian-Feng Cai, Tony F Chan, and Shingyu Leung. Guarantees of riemannian optimization for low rank matrix completion. *arXiv preprint arXiv:1603.06610*, 2016.

Zaiwen Wen, Wotao Yin, and Yin Zhang. Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. *Mathematical Programming Computation*, pages 1–29, 2012.

Hermann Weyl. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen*, 71(4):441–479, 1912.

Junfeng Yang and Xiaoming Yuan. Linearized augmented lagrangian and alternating direction methods for nuclear norm minimization. *Mathematics of computation*, 82(281):301–329, 2013.

Oscar Zariski. On the purity of the branch locus of algebraic functions. *Proceedings of the National Academy of Sciences*, 44(8):791–796, 1958.