Proceedings of Machine Learning Research vol 65:1–8, 2017

Matrix Completion from $O(n)$ Samples in Linear Time

David Gamarnik GAMARNIK@MIT.EDU
Quan Li QUANLI@MIT.EDU
Hongyi Zhang HONGYIZ@MIT.EDU
Massachusetts Institute of Technology

Abstract

We consider the problem of reconstructing a rank-k $n \times n$ matrix M from a sampling of its entries. Under a certain incoherence assumption on M and for the case when both the rank and the condition number of M are bounded, it was shown in (Candès and Recht, 2009; Candès and Tao, 2010; Keshavan et al., 2010; Recht, 2011; Jain et al., 2012; Hardt, 2014) that M can be recovered exactly or approximately (depending on some trade-off between accuracy and computational complexity) using $O(n \,\mathrm{poly}(\log n))$ samples in super-linear time $O(n^a \,\mathrm{poly}(\log n))$ for some constant $a \geq 1$.

In this paper, we propose a new matrix completion algorithm using a novel sampling scheme based on a union of independent sparse random regular bipartite graphs. We show that under the same conditions w.h.p. our algorithm recovers an ϵ-approximation of M in terms of the Frobenius norm using $O(n \log^2(1/\epsilon))$ samples and in linear time $O(n \log^2(1/\epsilon))$. This provides the best known bounds both on the sample complexity and computational cost for reconstructing (approximately) an unknown low-rank matrix.

The novelty of our algorithm is two new steps of thresholding singular values and rescaling singular vectors in the application of the "vanilla" alternating minimization algorithm. The structure of sparse random regular graphs is used heavily for controlling the impact of these regularization steps.

Keywords: matrix completion; alternating minimization; singular value thresholding; sparse random graphs.

1. Introduction

We consider the problem of reconstructing a hidden rank-k matrix from a sampling of its entries. Specifically, consider an $n \times n$ matrix M. The goal is to design a sampling index set $\Omega \subseteq [n] \times [n]$ such that M can be reconstructed efficiently from the entries in M associated with Ω, that is, from the entries $M_{ij}, (i,j) \in \Omega$, with the cardinality $|\Omega|$ as small as possible. The problem has wide range of applications in recommendation systems, system identification, global positioning, computer vision , etc. (Candes and Plan, 2010).

For the convenience of discussing various matrix completion results and comparing them to our results, we will assume in the discussion below that the rank, condition number and the incoherence parameter of M (appropriately defined) are bounded in n. The problem of reconstructing M under uniform sampling received considerable attention in recent years. One research direction of matrix completion under this sampling scheme focuses on the exact recovery of M. Recht (2011) and Gross (2011) showed that M can be reconstructed exactly from $O(n \log^2 n)$ samples using trace-norm based optimization. Keshavan et al. (2010) showed that M can be reconstructed exactly from

. Extended abstract. Full version appears as [arXiv:1702.02267, v2].

$O(n \log n)$ samples using singular value decomposition (SVD) followed by gradient descent on Grassmanian manifold. Another research direction of matrix completion under uniform sampling pays more attention to the efficiency of the algorithm, and only requires approximate matrix completion. Jain et al. (2012) showed that an ϵ-approximation of M in the Frobenius norm can be reconstructed from $O(n \log n \log(1/\epsilon))$ samples using alternating minimization in $O(n \log n \log(1/\epsilon))$ time. Then Hardt (2014) refined the analysis of alternating minimization and improved the sample complexity to $O(n \log(n/\epsilon))$. It was shown by Candès and Tao (2010) that $O(n \log n)$ is the information theoretic limit of the number of samples for exact matrix completion of M. With extensive research on this subject, it is tempting to believe that the sample complexity obtained by Jain et al. (2012) or Hardt (2014) are optimal (up to a constant factor) for approximate matrix completion as well. Perhaps surprisingly, we establish that this is not the case and propose a new algorithm, which constructs an ϵ-approximation of M in Frobenius norm using $O(n \log^2(1/\epsilon))$ samples in linear time $O(n \log^2(1/\epsilon))$.

Our proposed algorithm adds two new steps: a thresholding of singular values and a rescaling of singular vectors upon the "vanilla" alternating minimization algorithm. The idea behind the singular value thresholding is regularization of the least square estimation for avoiding the ill-conditioning problem of certain matrices (i.e. the Gramian matrices inverted in (7) and (9)). We call this algorithm Thresholded Alternating Minimization ($\mathcal{TAM}$), referring to the extra singular value thresholding steps added to alternating minimization. A rescaling of the entries of singular vectors is also implemented in the $\mathcal{TAM}$ algorithm in order to maintain the proximity to incoherence. A more specific discussion of the intuition behind these two new steps appears after the introduction of the $\mathcal{TAM}$ algorithm (in Pages 6 and 7).

For the convenience of analysis, $\mathcal{TAM}$ employs a sampling generated from a union of independent random bipartite regular graphs. Although our results of $\mathcal{TAM}$ are established on this special sampling, $\mathcal{TAM}$ can be generalized to uniform sampling in the obvious manner and similar results of $\mathcal{TAM}$ under uniform sampling can be established accordingly. In fact, by considering Poisson cloning model (Kim, 2006) for Erdös-Rényi graphs, (which we intend to research in future), we conjecture that the same sample complexity of $\mathcal{TAM}$ might hold for constructing an ϵ-approximation of M in Frobenius norm under uniform sampling. There is no contradiction between the information theoretic lower bound $O(n \log n)$ for exact matrix completion and this conjecture, due to its approximate nature. Other sampling schemes for matrix completion are also studied in Meka et al. (2009); Király et al. (2015); Pimentel-Alarcón et al. (2016).

$\mathcal{TAM}$ maintains the computational complexity of alternating minimization, which is $O(|\Omega|)$ for bounded k. $\mathcal{TAM}$ only requires $O(n \log^2(1/\epsilon))$, or $O(n \log(1/\epsilon))$ samples, under the standard incoherence Assumption 1 or under both Assumptions 1 and 2, given in Section 2. Hence, $\mathcal{TAM}$ is a linear algorithm of computational complexity $O(n \log^2(1/\epsilon))$ or $O(n \log(1/\epsilon))$. Like alternating minimization, $\mathcal{TAM}$ has computational efficiency advantage over trace-norm based optimization, which requires time $O(n^2 \log n/\sqrt{\epsilon})$ using the singular value thresholding algorithm (Cai et al., 2010) or $O(n^5 \log(1/\epsilon))$ using interior point methods. More specific computational complexity comparison between trace-norm based optimization and alternating minimization is given in (Jain et al., 2012).

2. Problem Formulation and Assumptions

Let $M \in \mathbb{R}^{n\times m}$ be a rank-k matrix and $M = U^*\Sigma^*(V^*)^T$ be its SVD where the singular values are $\sigma_1^* \geq \sigma_2^* \ldots \geq \sigma_k^*$ in decreasing order. The entries in M associated with the index set $\Omega \subseteq [n]\times[m]$ are observed, that is, the entries M_{ij}, $\forall (i,j) \in \Omega$, are known. Define the sampling operator $P_\Omega : \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times m}$ by

$$P_\Omega(M) = \begin{cases} M_{ij} & \text{if } (i,j) \in \Omega, \\ 0 & \text{if } (i,j) \notin \Omega. \end{cases}$$

Let $\mathcal{V}_R$ and $\mathcal{V}_C$ be the sets of rows and columns of matrix M, respectively, indexed by the sets $\{1,2,\ldots,n\}$ and $\{1,2,\ldots,m\}$. Also, let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a bipartite undirected graph on the vertex set $\mathcal{V} = \mathcal{V}_R \cup \mathcal{V}_C$ with edge set $\mathcal{E} \ni (i,j)$ if and only if $(i,j) \in \Omega$. Our goal is to obtain an ϵ-approximation of the matrix M from the observed $P_\Omega(M)$.

For the rest of the paper, we will assume for simplicity that $m = n$. Our results can be easily extended to the more general case $m = \Theta(n)$, using the generalization as in the appendix D of (Hardt, 2014). We say a graph is a random bipartite d-regular graph $\mathbb{G}_d(n,n)$ if it is chosen uniformly at random from all bipartite d-regular graphs with n vertices $\{1,2,\ldots,n\}$ on the left and another n vertices $\{1,2,\ldots,n\}$ on the right. For our proposed algorithm, we choose $\mathcal{G}$ to be a union of several independent random bipartite d-regular graphs $\mathbb{G}_d(n,n)$. Bayati et al. (2010) proposed an algorithm for generating a random bipartite d-regular graph $\mathbb{G}_d(n,n)$ in expected running time $O(nd^2)$.

Let $u_i^{*,T}$, $i \in [n]$, be the i-th row of U^* and $v_j^{*,T}$, $j \in [n]$, be the jth row of V^*. Now we present the incoherence assumptions on M.

- **Assumption** 1**.** There exists a constant $\mu_0 \geq 1$ such that

$$\|u_i^*\|_2^2 \leq \frac{\mu_0 k}{n}, \forall i \in [n] \text{ and } \|v_j^*\|_2^2 \leq \frac{\mu_0 k}{n}, \forall j \in [n]. \tag{1}$$

- **Assumption** 2**.** Given the degree d of $\mathbb{G}_d(n,n)$, let S_n be a subset of $[n]$ chosen uniformly at random from all the subsets of $[n]$ with cardinality d. There exists a constant $\delta \in (0,1)$ such that

$$\mathbb{P}(\|\sum_{i\in S_n} \frac{n}{d} u_i^* u_i^{*,T} - I\|_2 \leq \delta) = 1 - o(1) \text{ and } \mathbb{P}(\|\sum_{j\in S_n} \frac{n}{d} v_j^* v_j^{*,T} - I\|_2 \leq \delta) = 1 - o(1). \tag{2}$$

where Assumption 1 is the standard incoherence condition assumed by most of existing low-rank matrix completion results (Candès and Recht, 2009; Keshavan et al., 2010; Jain et al., 2012; Hardt, 2014) etc. We call Assumption 2 the probabilistic generalized restricted isometry condition, which is strictly weaker, for example, than the incoherence assumption $A2$ in (Bhojanapalli and Jain, 2014). The latter requires

$$\left\|\sum_{i\in S_n^1} \frac{n}{d} u_i^* u_i^{*,T} - I\right\|_2 \leq \delta \quad \text{and} \quad \left\|\sum_{j\in S_n^2} \frac{n}{d} v_j^* v_j^{*,T} - I\right\|_2 \leq \delta, \tag{3}$$

for $\delta \leq 1/6$ and all $S_n^1, S_n^2 \subset [n]$ of cardinality $|S_n^1| = |S_n^2| = d$ while the probabilistic generalized restricted isometry condition (2) requires the inequalities above hold for majority of the subsets $S_n^1 \subset [n]$ of cardinality $|S_n^1| = d$ and for majority of the subsets $S_n^2 \subset [n]$ of cardinality $|S_n^2| = d$.

3. Main Results

We are about to present a new matrix completion algorithm and give recovery guarantees of the proposed algorithm for two scenarios: matrix completion under Assumption 1, and matrix completion under both Assumption 1 and Assumption 2. Furthermore, we will assume that Assumption 1 always holds, and that the rank k, the condition number σ_1^*/σ_k^*, and the incoherence parameter μ_0 of the matrix M are bounded from above by a constant, as $n \to \infty$.

Now we formally describe the matrix completion algorithm we propose in this paper and state our main results. For the statement of our algorithm, we first introduce two operators acting on the matrices. Define $\mathcal{T}_1 : \mathbb{R}^{k\times 1} \to \mathbb{R}^{1\times k}$ by

$$\mathcal{T}_1(u) \triangleq \begin{cases} \sqrt{\frac{\mu_0 k}{n}} \frac{u^T}{\|u\|_2} & \|u\|_2 \geq 2\sqrt{\frac{\mu_0 k}{n}}, \\ u^T & \|u\|_2 < 2\sqrt{\frac{\mu_0 k}{n}}. \end{cases} \tag{4}$$

Specifically, the operator $\mathcal{T}_1$ normalizes the vector u of length at least $2\sqrt{\mu_0 k/n}$ to the vector of the same direction and of length $\sqrt{\mu_0 k/n}$. For the convenience of notation we extend $\mathcal{T}_1$ to the one acting on matrix $U = (u_i^T, i \in [n]) \in \mathbb{R}^{n\times k}$ by

$$\mathcal{T}_1(U) \triangleq \begin{pmatrix} \mathcal{T}_1(u_1) \\ \vdots \\ \mathcal{T}_1(u_n) \end{pmatrix}.$$

Then it follows from the definition of $\mathcal{T}_1(\cdot)$ in (4) that any row vector of $\mathcal{T}_1(U)$ has length at most $2\sqrt{\mu_0 k/n}$.

For $A \in \mathbb{R}^{d\times k}$, let the SVD of A be

$$A = U_A \Sigma_A (V_A)^T.$$

We write Σ_A in the form $\sqrt{d/n}$ diag$(\sigma_1, \cdots, \sigma_k)$ where the diagonal entries $\sigma_1, \sigma_2 \ldots, \sigma_k$ ($\sigma_1 \geq \sigma_2 \ldots \geq \sigma_k$) are the singular values of A divided by $\sqrt{d/n}$. For a given $a \in (0, 1)$ and $\forall i \in [k]$, let

$$\sigma_{i,a} = \begin{cases} \sigma_i & \text{if } \sigma_i \in [\sqrt{a}, \sqrt{2-a}], \\ \sqrt{a} & \text{if } \sigma_i < \sqrt{a}, \\ \sqrt{2-a} & \text{if } \sigma_i > \sqrt{2-a}. \end{cases}$$

Define $\mathcal{T}_2(A, a)$ by

$$\mathcal{T}_2(A, a) \triangleq U_A \hat{\Sigma}_A (V_A)^T \tag{5}$$

where $\hat{\Sigma}_A = \sqrt{d/n}$ diag$(\sigma_{1,a}, \cdots, \sigma_{k,a})$ and hence the entire $\sigma_{1,a}, \cdots, \sigma_{k,a}$ satisfy

$$\sqrt{2-a} \geq \sigma_{1,a} \geq \sigma_{2,a} \ldots \geq \sigma_{k,a} \geq \sqrt{a}.$$

Specifically, the operator $\mathcal{T}_2$ lifts the normalized singular values in Σ_A less than $\sqrt{a}$ to $\sqrt{a}$ and truncates the normalized singular values in Σ_A more than $\sqrt{2-a}$ to $\sqrt{2-a}$.

Let $\Omega_t \subseteq [n] \times [n]$, $t = 0, 1, \ldots, 2N$, be the index sets associated with $2N+1$ independent random bipartite d-regular graphs $\mathbb{G}_d(n,n)$. Define $\mathcal{RRG}(d,n,N)$ as the random d-regular graph model of Ω, that is,

$$\mathcal{RRG}(d,n,N) \triangleq \{\Omega_0, \Omega_1, \cdots, \Omega_{2N}\}. \tag{6}$$

Let D be a subset of $[n]$ with d entries, namely, $D = \{i_1, i_2, \ldots, i_d\}$. For a matrix $U = (u_i^T, i \in [n]) \in \mathbb{R}^{n\times k}$, let its submatrix with the row indices in D and the column indices the same as U be

$$U_D = \begin{pmatrix} u_{i_1}^T \\ \vdots \\ u_{i_d}^T \end{pmatrix}.$$

Let $S_j^{t,L} = \{i \in [n] : (i,j) \in \Omega_t\}$, $\forall j \in [n]$. Then $|S_j^{t,L}| = d$. Namely, $S_j^{t,L}$ consists of all the left neighbors of vertex j on the right in the random bipartite d-regular graph associated with the index set Ω_t. Correspondingly given any $a \in (0,1)$ and any $j \in [n]$, we denote $\hat{U}_{S_j^{t,L}} = \mathcal{T}_2(U_{S_j^{t,L}}, a)$ and the row in $\hat{U}_{S_j^{t,L}}$ associated with the index $i \in S_j^{t,L}$ by $\hat{u}_i^{t,T}$. Similarly, let $S_i^{t,R} = \{j \in [n] : (i,j) \in \Omega_t\}$, $\forall i \in [n]$, that is, $S_i^{t,R}$ consists of all the right neighbors of vertex i on the left in the random bipartite d-regular graph associated with the index set Ω_t. Also, we have $|S_i^{t,R}| = d$. For a matrix $V \in \mathbb{R}^{n\times k}$ and a given $a \in (0,1)$, denote similarly $\hat{V}_{S_i^{t,R}} = \mathcal{T}_2(V_{S_i^{t,R}}, a)$ and the row in $\hat{V}_{S_i^{t,R}}$ associated with the index $j \in S_i^{t,R}$ by $\hat{v}_j^{t,T}$.

Now we introduce the algorithm $\mathcal{TAM}$ for matrix completion in the sparse regime. For the algorithm below we fix arbitrary $\delta \in (0,1)$ and we let β be any constant in $(0, 1-\delta)$.

Thresholded Alternating Minimization algorithm ($\mathcal{TAM}$)

Input: *Observed index sets $\mathcal{RRG}(d,n,N)$ and values $P_{\cup_{t=0}^{2N}\Omega_t}(M)$.*
Initialize: *$\bar{U}^0 = SVD(\frac{n}{d}P_{\Omega_0}(M), k)$, i.e. top-k left singular vectors of $\frac{n}{d}P_{\Omega_0}(M)$.*
Truncation step: first apply $\mathcal{T}_1$ on $\bar{U}^0$ then orthonormalize the columns of $\mathcal{T}_1(\bar{U}^0)$. Denote the resultant orthonormal matrix by $U^0 = (u_i^{0,T}, 1 \le i \le n)$.
Loop: *For $t = 0$ to $N-1$*
For each $j \in [n]$:
If $\frac{n}{d}\sigma_l(\sum_{i\in[n]:(i,j)\in\Omega_{t+1}} u_i^t u_i^{t,T}) \in [\beta, 2-\beta]$ for all $l \in [k]$, then set

$$\tilde{v}_j^{t+1} = \left(\sum_{i\in[n]:(i,j)\in\Omega_{t+1}} u_i^t u_i^{t,T}\right)^{-1} \sum_{i\in[n]:(i,j)\in\Omega_{t+1}} u_i^t M_{ij}. \tag{7}$$

Otherwise let $\hat{U}^t_{S_j^{t+1,L}} = \mathcal{T}_2(U^t_{S_j^{t+1,L}}, \beta)$ and

$$\tilde{v}_j^{t+1} = \left(\sum_{i\in[n]:(i,j)\in\Omega_{t+1}} \hat{u}_i^t \hat{u}_i^{t,T}\right)^{-1} \sum_{i\in[n]:(i,j)\in\Omega_{t+1}} \hat{u}_i^t M_{ij}. \tag{8}$$

Let $\tilde{V}^{t+1} = (\tilde{v}_j^{t+1,T}, 1 \le j \le n)$ *and* $\tilde{V}^{t+1} = \bar{V}^{t+1} R^{t+1}$ *be the QR decomposition of* $\tilde{V}^{t+1}$*. Orthonormalize the columns of* $\mathcal{T}_1(\bar{V}^{t+1})$*. Denote the resultant orthonormal matrix by* $V^{t+1} = (v_j^{t+1,T}, 1 \le j \le n)$*.*

For each $i \in [n]$*:*

If $\frac{n}{d}\sigma_l(\sum_{j\in[n]:(i,j)\in\Omega_{N+t+1}} v_j^{t+1} v_j^{t+1,T}) \in [\beta, 2-\beta]$ *for all* $l \in [k]$*, then set*

$$\tilde{u}_i^{t+1} = \left(\sum_{j\in[n]:(i,j)\in\Omega_{N+t+1}} v_j^{t+1} v_j^{t+1,T} \right)^{-1} \sum_{j\in[n]:(i,j)\in\Omega_{N+t+1}} v_j^{t+1} M_{ij}. \tag{9}$$

Otherwise let $\hat{V}^{t+1}_{S_i^{N+t+1,R}} = \mathcal{T}_2(V^{t+1}_{S_i^{N+t+1,R}}, \beta)$ *and*

$$\tilde{u}_i^{t+1} = \left(\sum_{j\in[n]:(i,j)\in\Omega_{N+t+1}} \hat{v}_j^{t+1} \hat{v}_j^{t+1,T} \right)^{-1} \sum_{j\in[n]:(i,j)\in\Omega_{N+t+1}} \hat{v}_j^{t+1} M_{ij}. \tag{10}$$

Let $\tilde{U}^{t+1} = (\tilde{u}_j^{t+1,T}, 1 \le j \le n)$ *and* $\tilde{U}^{t+1} = \bar{U}^{t+1} R^{N+t+1}$ *be the QR decomposition of* $\tilde{U}^{t+1}$*. Orthonormalize the columns of* $\mathcal{T}_1(\bar{U}^{t+1})$*. Denote the resultant orthonormal matrix by* $U^{t+1} = (u_i^{t+1,T}, 1 \le i \le n)$*.*

Output*: Set* $U^{N-1} = (u_i^{N-1,T}, 1 \le i \le n), \tilde{V}^N = (\tilde{v}_j^{N,T}, 1 \le j \le n)$*. Output* $M_N = U^{N-1}\tilde{V}^{N,T}$*.*

Now we provide the intuition behind the proposed algorithm. This algorithm adds two new steps: a thresholding of singular values and a rescaling of singular vectors upon the "vanilla" alternating minimization algorithm. The key idea behind these steps is regularization of the least square estimation in the form of the singular value thresholding. Due to the decreased sample complexity by a logarithmic factor $\log n$, certain matrices inverted in each step of the alternating minimization algorithm may become ill-conditioned. For example, given $j \in [n]$ and a constant d, it is not guaranteed that at the t-th iteration of the alternating minimization algorithm

$$U^{t,T}_{S_j^{t+1,L}} U^t_{S_j^{t+1,L}} = \sum_{i\in[n]:(i,j)\in\Omega_{t+1}} u_i^t u_i^{t,T}$$

concentrates around its expectation

$$\begin{aligned}\mathbb{E}[U^{t,T}_{S_j^{t+1,L}} U^t_{S_j^{t+1,L}}] &= \frac{1}{\binom{n}{d}} \sum_{D\in\{S\subset[n]:|S|=d\}} U_D^{t,T} U_D^t \\ &= \frac{1}{\binom{n}{d}} \frac{\binom{n}{d} d}{n} \sum_{i\in[n]} u_i^t u_i^{t,T} = \frac{d}{n} I.\end{aligned}$$

Some $U^{t,T}_{S_j^{t+1,L}} U^t_{S_j^{t+1,L}}$ might be ill-conditioned, namely, its least singular value is 0 or closed to zero. If the matrix $U^{t,T}_{S_j^{t+1,L}} U^t_{S_j^{t+1,L}}$ is ill-conditioned, the results from the iteration (7) in the "vanilla" alternating minimization algorithm might blow up. To prevent this adversarial scenario, we use

the operations $\mathcal{T}_2$ to lift the small singular values and truncate the large singular values of $U^t_{S_j^{t+1,L}}$, $\forall j \in [n]$, before each row vector of $\tilde{V}^{t+1} = (\tilde{v}_j^{t+1,T}, 1 \le j \le n)$ is computed. This singular value thresholding step enforces that all the singular values of the Gramian matrix inverted in (8) deviate from their expected values by at most $1-\beta$ after proper normalization, and as a result, guarantees the nonsingularity of this (adjusted) Gramian matrix. The convergence of the algorithm relies on the fact that w.h.p. the number of times the algorithm applies the operation $\mathcal{T}_2$ in each iteration is a small fraction of n. Also, the operators $\mathcal{T}_1$ are applied at the end of each iteration to guarantee the incoherence of the input V^{t+1} (or U^{t+1}) for the next iteration while maintaining that V^{t+1} (or U^{t+1}) is still close enough to V^* (or U^*).

Our main result concerns the performance of the algorithm $\mathcal{TAM}$ under Assumption 1 and under both Assumptions 1 and 2, respectively. We recall that $\mathcal{TAM}$ is parameterized by δ and β.

Theorem 1 *Suppose $M \in \mathbb{R}^{n\times n}$ is a rank-k matrix satisfying Assumption 1. Suppose the observed index set Ω is sampled according to the model $\mathcal{RRG}(d,n,N)$ in (6). Given any $\delta \in (0,1)$, $\beta \in (0, 1-\delta)$ and $\epsilon \in (0, 2/3)$, there exists a $C(\delta,\beta) > 0$ such that for*

$$d \ge C(\delta,\beta)k^4\mu_0^2\left(\frac{\sigma_1^*}{\sigma_k^*}\right)^2 + \frac{5\mu_0 k(1+\delta/3)}{\delta^2}\log\left(\frac{1}{\epsilon}\right) \tag{11}$$

and $N \ge 1 + \lceil \log(\frac{2}{\epsilon})/\log 4 \rceil$, the $\mathcal{TAM}$ algorithm produces a matrix M_N satisfying $\|M - M_N\|_F \le \epsilon\|M\|_F$ w.h.p.

Furthermore, suppose M satisfies both Assumptions 1 and 2. Then for $\delta \in (0,1)$ as defined in Assumption 2 and $\beta \in (0, 1-\delta)$, the same result holds when

$$d \ge C(\delta,\beta)k^4\mu_0^2\left(\frac{\sigma_1^*}{\sigma_k^*}\right)^2, \tag{12}$$

for the same constant $C(\delta,\beta)$ in (11).

Theorem 1 states that under Assumption 1 the $\mathcal{TAM}$ algorithm produces a rank-k ϵ-approximation of matrix M using $O(dn\log(1/\epsilon))$ samples for d satisfying (11). Furthermore, under both Assumption 1 and Assumption 2 the $\mathcal{TAM}$ algorithm produces a rank-k ϵ-approximation of matrix M using $O(dn\log(1/\epsilon))$ samples for d satisfying (12).

References

Mohsen Bayati, Jeong Han Kim, and Amin Saberi. A sequential algorithm for generating random graphs. *Algorithmica*, 4(58):860–910, 2010.

Srinadh Bhojanapalli and Prateek Jain. Universal matrix completion. In *Proceedings of The 31st International Conference on Machine Learning*, pages 1881–1889, 2014.

Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

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

Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of computational mathematics*, 9(6):717–772, 2009.

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.

David Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, 57(3):1548–1566, 2011.

Moritz Hardt. Understanding alternating minimization for matrix completion. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 651–660. IEEE, 2014.

Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi. Low-rank matrix completion using alternating minimization. *arXiv preprint arXiv:1212.0467*, 2012.

Raghunandan H Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from a few entries. *IEEE Transactions on Information Theory*, 56(6):2980–2998, 2010.

Jeong Han Kim. Poisson cloning model for random graphs. In *Proceedings of the International Congress of Mathematicians*, pages 873–897, 2006.

Franz J Király, Louis Theran, and Ryota Tomioka. The algebraic combinatorial approach for low-rank matrix completion. *Journal of Machine Learning Research*, 16:1391–1436, 2015.

Raghu Meka, Prateek Jain, and Inderjit S Dhillon. Matrix completion from power-law distributed samples. In *Advances in neural information processing systems*, pages 1258–1266, 2009.

Daniel L Pimentel-Alarcón, Nigel Boston, and Robert D Nowak. A characterization of deterministic sampling patterns for low-rank matrix completion. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):623–636, 2016.

Benjamin Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12(Dec):3413–3430, 2011.