Universal Matrix Completion

Srinadh Bhojanapalli BSRINADH@UTEXAS.EDU
The University of Texas at Austin

Prateek Jain PRAJAIN@MICROSOFT.COM
Microsoft Research, India

Abstract

The problem of low-rank matrix completion has recently generated a lot of interest leading to several results that offer exact solutions to the problem. However, in order to do so, these methods make assumptions that can be quite restrictive in practice. More specifically, the methods assume that: a) the observed indices are sampled uniformly at random, and b) for every new matrix, the observed indices are sampled *afresh*. In this work, we address these issues by providing a universal recovery guarantee for matrix completion that works for a variety of sampling schemes. In particular, we show that if the set of sampled indices come from the edges of a bipartite graph with large spectral gap (i.e. gap between the first and the second singular value), then the nuclear norm minimization based method exactly recovers all low-rank matrices that satisfy certain incoherence properties. Moreover, we also show that under certain stricter incoherence conditions, $O(nr^2)$ uniformly sampled entries are enough to recover any rank-r $n \times n$ matrix, in contrast to the $O(nr \log n)$ sample complexity required by other matrix completion algorithms as well as existing analyses of the nuclear norm method.

1. Introduction

In this paper, we study the problem of *universal* low-rank matrix completion. Low-rank matrix completion is an important problem with several applications in areas such as recommendation systems, sketching, and quantum tomography (Recht et al., 2010; Candès & Recht, 2009; Gross et al., 2010). The goal in matrix completion is to recover a rank-r matrix, given a small number of entries of the matrix. That is, to find a matrix M given values $\{M_{ij}, (i,j) \in \Omega\}$, where Ω is the set of observed indices.

Recently, several methods with provable guarantees have been proposed for solving the problem under the following two assumptions: a) M is incoherent, b) Ω is sampled *uniformly* and $|\Omega| \geq cnr \log n$. Moreover, Ω needs to be re-sampled for each matrix M that is to be recovered, i.e., the same Ω cannot be re-used without worsening the guarantees significantly.

While the first assumption can be shown to be necessary for any matrix oblivious sampling, the second assumption is relatively restrictive and might not hold in several practical settings. The main goal of this work is to develop a general result that can handle other sampling schemes as well. Moreover, we aim to develop a universal method where one fixed Ω would be enough to recover any low-rank matrix M. Such a universal recovery result is highly desirable in several signal processing applications, where the goal is to design one Ω that can recover any low-rank signal matrix M by observing M over Ω alone.

To this end, we reduce the problem of recoverability using an index set Ω to the spectral gap (gap between largest and second largest singular values) of $\mathcal{G}$, where $\mathcal{G}$ is a bipartite graph whose biadjacency matrix $G \in \mathbb{R}^{n \times n}$ is given by: $G_{ij} = 1$ iff $(i,j) \in \Omega$ and $G_{ij} = 0$ otherwise. In particular, we show that if $\mathcal{G}$ has a large enough spectral gap and if the rank-r matrix M satisfies the standard incoherence property, then the best rank-r approximation of $P_\Omega(M)$ (see (1)) itself is enough to get a "reasonable" approximation to M (see Theorem 4.1 for details).

Note that our approximation result is similar to Theorem 1.1 of (Keshavan et al., 2010), but our result holds for any Ω with large spectral gap unlike (Keshavan et al., 2010) that requires uniform sampling. On the other hand, we require explicit incoherence condition on singular vectors of M, while the result of (Keshavan et al., 2010) only requires a bound on $M_{max} = \max_{ij} M_{ij}$. The later assumption is strictly weaker assumption; the assumptions coincide for PSD matrices.

Proceedings of the 31^{st} International Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP volume 32.

Next, we show that by assuming certain stronger incoherence properties, the number of samples required by the popular nuclear-norm minimization method (Candès & Recht, 2009) to recover back M depends *only* on n, r and the spectral gap of the d-regular bipartite graph $\mathcal{G}$. In particular, we require $d \geq \sigma_2(G) \cdot r$, where $\sigma_2(G)$ is the second largest singular value of G. Hence, if $\sigma_2(G) = O(\sqrt{d})$, i.e., if $\mathcal{G}$ is an expander, then $|\Omega| = nd = O(nr^2)$ samples suffice for exact recovery.

Our recovery results applies to *any* low rank matrix M that satisfies the stronger incoherence property, given that the fixed graph $\mathcal{G}$ has a large spectral gap. To the best of our knowledge, this is the *first universal guarantee for matrix completion*. Furthermore, using recent results by (Feige & Ofek, 2005) we show that for the standard uniform sampling of Ω, only $O(nr^2)$ samples suffice for exact recovery of a rank-r matrix M that satisfies a stronger incoherence condition (see $A2$ in Section 3).

Next, we discuss the stronger incoherence property that we require for our universal recovery guarantees. In particular, we show that the standard incoherence condition alone cannot provide universal recovery with any graph $\mathcal{G}$ and hence a stronger incoherence property is required.

Finally, we empirically demonstrate our observation that, instead of the number of samples, the spectral gap of $\mathcal{G}$ is what really governs recoverability of the true matrix. In particular, we construct a family of graphs based on the stochastic block model and show that the probability of success grows linearly with the spectral gap, irrespective of the number of samples.

Notation: We denote matrices by capital letters (e.g. U) and vectors by small letters (e.g. u). U^T denotes the transpose of the matrix U. U_{ij} represents the (i,j)-th element of U. U_i represents the i-th column of U and U^i represents the i-th row of U (but in column format). $\|u\|$ represents the L_2 norm of u and $\|U\|$ represents the spectral norm of U, i.e., $\|U\| = \max_{x:\|x\|\leq 1} \|Ux\|$. $\|U\|_F$ represents the Frobenius norm and $\|U\|_\infty$ is the absolute maximum element of U. $C = A.B$ represents the Hadamard product of A, B, i.e., $C_{ij} = A_{ij}B_{ij}$. Similarly, $(u.v)_i = u_i v_i$. $\mathbf{1}$ denotes the all 1's vector. $\mathbf{1}_\perp$ represents a *unit* vector that is perpendicular to $\mathbf{1}$ and is determined appropriately by the context.

Paper Organization: In the next section, we discuss some related works. Then, in Section 3, we define the problem of matrix completion and the bipartite graph $\mathcal{G}$ that we use. We present our main results in Section 4 and discuss the additional incoherence assumption in Section 5. In Section 6, we present observations from our empirical study. Finally in Section 7, we provide the proof of our exact recovery result.

2. Related work

Matrix completion: In a seminal paper on matrix completion, (Candès & Recht, 2009) showed that any $n \times n$ incoherent matrix of rank r can be recovered from $Cn^{1.2}r\log(n)$ uniform random samples using nuclear norm minimization. Later, assuming the matrix to be *strongly incoherent*, (Candès & Tao, 2010) improved the sample complexity for nuclear norm minimization method to $O(nr\log^6(n))$. Subsequently, (Recht, 2009; Gross, 2011) generalized this result for any incoherent matrix using matrix Bernstein inequalities and presented significantly simpler proofs. Concurrently other algorithms were shown to recover incoherent matrices using $O(nr\log(n))$ (or worse) samples such as: SVD followed by descent on Grassmanian manifold (Keshavan et al., 2010), alternating minimization (Jain et al., 2012). We note that all the above mentioned results need to assume a rather restrictive sampling scheme, i.e., each entry is sampled uniformly at random and furthermore require a fresh set of samples for each new matrix. Moreover, the number of samples required is at least $O(nr\log n)$.

Other sampling schemes: Recently, there has been some results for different type of sampling schemes such as power-law distributions (Meka et al., 2009), but here again universal results are not known and furthermore the proposed algorithms are not robust to noise. Another line of work has been to devise sampling schemes *dependent* on the data matrix (Chen et al., 2013), (Király & Tomioka, 2012). Naturally, these schemes cannot be universal as the sampling scheme itself is dependent on the data matrix. Furthermore, practicality of such schemes is not clear a priori.

Universality: Universality is an important property in signal processing or sketching applications, as the goal there is to have one fixed sampling operator that performs well for all the given signals. While, universality results are well known for several other sensing problems, such as sparse vector recovery (Candes & Tao, 2005), one-bit compressive sensing (Gopi et al., 2013), similar results for low-rank matrix sensing are mostly restricted to RIP-type operators (Recht et al., 2010; Liu, 2011). Unfortunately, RIP-type of operators are typically dense, requiring knowledge of all elements of matrix to get observations, and also require large storage/computational complexity. On the other hand sampling individual elements is a sparse operator and hence computationally efficient. Hence, for several signal processing applications, universal matrix completion results are critical.

In fact, several recent works have studied problems similar to that of universal matrix completion. For example, (Király & Tomioka, 2012; Heiman et al., 2013) and (Lee & Shraibman, 2013). However, there are critical differ-

ences in our results/approaches that we now highlight. In (Király & Tomioka, 2012) authors consider an algebraic approach to analyze sufficient conditions for matrix completion. While they propose interesting deterministic sufficient conditions, the algorithm analyzed in the paper requires solving an NP hard problem. In contrast we analyze the nuclear norm minimization method which is known to have several efficient implementations. In (Heiman et al., 2013) and (Lee & Shraibman, 2013) authors consider sampling based on expanders but only provide generalization error bounds rather than exact recovery guarantee. Moreover, the recovered matrix using their algorithm need not have a low-rank.

3. Problem Definition & Assumptions

Let $M \in \mathbb{R}^{n_1 \times n_2}$ be a rank-r matrix and let $n_1 \geq n_2$. Define $n = \max\{n_1, n_2\} = n_1$. Let $M = U\Sigma V^T$ be the SVD of M and let $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_r$ be the singular values of M. We observe a small number of entries of M indexed by a set $\Omega \in [n_1] \times [n_2]$. That is, we observe $M_{ij}, \forall (i,j) \in \Omega$. Define the sampling operator $P_\Omega : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^{n_1 \times n_2}$ as:

$$P_\Omega(M) = \begin{cases} M_{ij}, & \text{if } (i,j) \in \Omega, \\ 0, & \text{if } (i,j) \notin \Omega. \end{cases} \tag{1}$$

Next, we define a bipartite graph associated with the sampling operator P_Ω. That is, let $\mathcal{G} = (V, E)$ be a bipartite graph where $V = \{1, 2, \ldots, n_1\} \cup \{1, 2, \ldots, n_2\}$ and $(i,j) \in E$ iff $(i,j) \in \Omega$. Let $G \in \mathbb{R}^{n_1 \times n_2}$ be the biadjacency matrix of the bipartite graph $\mathcal{G}$ with $G_{ij} = 1$ iff $(i,j) \in \Omega$. Note that, $P_\Omega(M) = M.G$, where . denotes the Hadamard product.

Now, the goal in *universal* matrix completion is to design a set Ω and a recovery algorithm, s.t., all rank-r matrices M can be recovered using only $P_\Omega(M)$. In the next section, we present two results for this problem. Our first result gives an approximate solution to the matrix completion problem and our second result gives exact recovery guarantees.

For our results, we require $\mathcal{G}$, that is associated with Ω, to be a d-regular bipartite graph with large spectral gap. More concretely, we require the following two properties from the sampling graph $\mathcal{G}$:

Assumptions on G/Ω:

- **G1** Top singular vectors of G are all 1's vector.
- **G2** $\sigma_1(G) = d$ and $\sigma_2(G) \leq C\sqrt{d}$.

Note that as the graph is d-regular, hence $|\Omega| = nd$.

The eigenvalues of the adjacency matrix of the bipartite graph $\mathcal{G}$ are $\{\sigma_i(G), -\sigma_i(G)\}, i = 1, ..n$. We state all the definitions in terms of singular values of G instead of the eigenvalues of the adjacency matrix. The above two properties are satisfied by a class of expander graphs called Ramanujan graphs; in fact, Ramanujan graphs are defined by using this spectral gap property:

Definition 3.1 (Ramanujan graph (Hoory et al., 2006))**.** *Let $\sigma_1(G), \sigma_2(G), ..., \sigma_n(G)$ be the singular values of G in decreasing order. Then, a d-regular bipartite graph $\mathcal{G}$ is a Ramanujan graph if $\sigma_2(G) \leq 2\sqrt{d-1}$.*

Ramanujan graphs $\mathcal{G}$ are well-studied in literature and there exists several randomized/deterministic methods to generate such graphs. We briefly discuss a couple of popular constructions in Section 4.2.

Incoherence assumptions: Now, we present incoherence assumptions that we impose on M:

$$\mathbf{A}1 \; ||U^i||^2 \leq \frac{\mu_0 r}{n_1}, \forall i \quad \text{and} \quad ||V^j||^2 \leq \frac{\mu_0 r}{n_2}, \forall j \tag{2}$$

$$\mathbf{A}2 \; \| \sum_{k \in S} \frac{n_1}{d} U^k U^{k^T} - I \| \leq \delta_d, \forall S \subset [n_1], |S| = d \quad \text{and} \quad \| \sum_{k \in S} \frac{n_2}{d'} V^k V^{k^T} - I \| \leq \delta_d, \forall S \subset [n_2], |S| = d'. \tag{3}$$

$d' = dn_2/n_1$. Note that $A1$ is the standard incoherence assumption required by most of the existing matrix completion results. However, $A2$ is a stricter assumption than $A1$ and is similar to the stronger incoherence property introduced by (Candès & Tao, 2010). We discuss necessity of such assumption for universal matrix completion in Section 5.

4. Main Results

We now present our main results for the matrix completion problem. We assume that Ω is generated using a bipartite d-regular expander and satisfies $G1$ and $G2$ (see Section 3). Our first result shows that, if M satisfies $A1$, then the best rank-r approximation of $P_\Omega(M)$ is "close" to M and hence serves as a good approximation for M that can also be used for initialization of other methods like alternating least squares. Our second results shows that if M satisfies both $A1$ and $A2$, then using nuclear-norm minimization based method, $P_\Omega(M)$ can be used to recover back M exactly.

4.1. Matrix approximation

Theorem 4.1. *Let $\mathcal{G}$ be a d-regular bipartite graph satisfying $G1$ and $G2$. Let M be a rank-r matrix that satisfies assumption $A1$. Then,*

$$\left\| \frac{n}{d} P_\Omega(M) - M \right\| \leq \frac{C\mu_0 r}{\sqrt{d}} ||M||.$$

That is, $\|\frac{n}{d}P_k(P_\Omega(M)) - M\| \leq \frac{2C\mu_0 r}{\sqrt{d}}||M||$, *for any* $k \geq r$, *where* $P_k(A)$ *is the best rank-*k *approximation of* A *and can be obtained using top-*k *singular vectors of* A.

Now, if M is a PSD matrix then the above result is exactly same as the Theorem 1.1 of (Keshavan et al., 2010). For non-PSD matrices, our result requires a bound on norm of each row of singular vectors of M, while the result of (Keshavan et al., 2010) only requires a bound on the largest element of M, hence is similar to our requirement but is strictly weaker as well.

On the other hand, our result holds for all M for a given Ω, if Ω's associated graph $\mathcal{G}$ satisfies both $G1$ and $G2$. Moreover, if $\mathcal{G}$ is generated using an Erdos-Renyi graph then, after a standard trimming step, the above theorem directly implies(for PSD matrices) Theorem 1.1 of (Keshavan et al., 2010). Finally, we would like to stress that our proof is significantly simpler and is able to exploit the fact that Erdös-Rényi graphs have good spectral gap in a fairly straightforward and intuitive manner.

We now present a detailed proof of the above theorem.

Proof. Let $M = U\Sigma V^T$, $U, V \in \mathbb{R}^{n\times r}$. Note that,

$$\|\frac{n}{d}P_\Omega(M) - M\| = \max_{\substack{\{x,y:\|x\|_2=1,\\ \|y\|_2=1\}}} y^T(\frac{n}{d}P_\Omega(M) - M)x.$$

Now,

$$y^T(\frac{n}{d}P_\Omega(M) - M)x = \sum_{i=1}^{r}\Big(\frac{n}{d}\sigma_i(y.U_i)^T G(x.V_i) - \sigma_i(y^TU_i)(x^TV_i)\Big). \quad (4)$$

Let $y.U_i = \alpha_i\mathbf{1} + \beta_i\mathbf{1}_\perp^{\mathbf{i}}$. Then, $\alpha_i = \frac{\mathbf{1}^T(y.U_i)}{n} = \frac{y^TU_i}{n}$. Hence,

$$\begin{aligned} &y^T(\frac{n}{d}P_\Omega(M) - M)x \\ &= \sum_{i=1}^{r}\Big(\sigma_i(y^TU_ix^TV_i + \frac{n}{d}\beta_i\mathbf{1}_\perp^{\mathbf{i}\,T}G(x.V_i)) - \sigma_iy^TU_ix^TV_i\Big) \\ &\overset{\zeta_1}{\leq}\frac{Cn}{\sqrt{d}}\sum_i\sigma_i\beta_i\|x.V_i\|_2 \overset{\zeta_2}{\leq}\sigma_1\frac{Cn}{\sqrt{d}}\sqrt{\sum_i\beta_i^2}\sqrt{\sum_i\|x.V_i\|_2^2}. \end{aligned} \quad (5)$$

where ζ_1 follows from assumption $G2$ and ζ_2 follows from the Cauchy-Schwarz inequality. Now,

$$\sum_{i=1}^{r}\beta_i^2 \leq \sum_{i=1}^{r}\|y.U_i\|^2 = \sum_{j=1}^{n}\sum_{i=1}^{r}y_j^2U_{ji}^2 \overset{\zeta_1}{\leq}\sum_{j=1}^{n}y_j^2\frac{\mu_0 r}{n} \overset{\zeta_2}{=}\frac{\mu_0 r}{n}, \quad (6)$$

where ζ_1 follows from $A1$ and ζ_2 follows by using $\|y\|_2 = 1$. Using similar argument as above, $\sum_{i=1}^{r}\|x.V_i\|^2 \leq \frac{\mu_0 r}{n}$. Theorem now follows by using (5), (6), and the above inequality. The proof of the second part is given in appendix A. □

4.2. Nuclear norm minimization

We now present our result for exact recovery of the matrix M using $P_\Omega(M)$ alone. For recovery, we use the standard nuclear norm minimization algorithm, i.e., we obtain a matrix X by solving the following convex optimization problem:

$$\begin{aligned} &\min \quad \|X\|_* \\ &\text{s. t. } P_\Omega(X) = P_\Omega(M), \end{aligned} \quad (7)$$

where $\|X\|_*$ denotes the nuclear norm of X; nuclear norm of X is equal to the sum of its singular values.

As mentioned in Section 2, nuclear norm minimization technique is a popular technique for the low-rank matrix completion problem and has been shown to provably recover the true matrix, assuming that Ω is sampled uniformly at random and $|\Omega| \geq cnr\log n$ (Candès & Tao, 2010).

Below, we provide a universal recovery result for the nuclear-norm minimization method as long as the samples Ω come from G that satisfies $G1$ and $G2$.

Theorem 4.2. *Let* M *be an* $n_1 \times n_2$ *matrix of rank* r *satisfying assumptions (A1) and (A2) with* $\delta_d \leq \frac{1}{6}$, *and* Ω *is generated from a* d*-regular graph* $\mathcal{G}$ *that satisfies the assumptions (G1) and (G2). Also, let* $d \geq 36C^2\mu_0^2r^2$, *i.e.,* $|\Omega| = nd \geq 36C^2\mu_0^2r^2\max\{n_1, n_2\}$. *Then* M *is the unique optimum of problem* (7).

Note that the above result requires only deterministic constraints on the sampling operator P_Ω and guarantees exact recovery for *any* matrix M that satisfies $A1, A2$. As mentioned earlier, $A2$ is a stronger assumption than $A1$. But as we show in Section 5, universal recovery is not possible with assumption $A1$ alone.

We can use the above theorem to derive results for several interesting sampling schemes such as random d-regular graphs. Using Theorem 1 in (Friedman, 2003), the second singular value of a *random* d*-regular graph* is $\leq 2\sqrt{d-1}+\epsilon$, for every $\epsilon > 0$, with high probability. Hence, a random d-regular graph, with high probability, obeys $G1$ and $G2$ which implies the following exact recovery result:

Corollary 4.3. *Let* M *be an* $n_1 \times n_2$ *matrix of rank* r *satisfying assumptions (A1) and (A2) with* $\delta_d \leq \frac{1}{6}$, *and* Ω *is generated from a random* d*-regular graph, then* M *is the unique optimum of program* (7) *when* $d \geq 36 * 4\mu_0^2r^2$, *with high probability.*

Note that the standard completion results such as (Candès & Recht, 2009), (Keshavan et al., 2010) generate Ω using Erdös-Rényi graph, which are slightly different than the random d-regular graph we considered above. However, (Feige & Ofek, 2005) showed that the second largest singular value of the Erdös-Rényi graph, $\mathcal{G}(n_1, n_2, p)$, is $O(\sqrt{d})$ when p is $\Theta(\log(n_1)/n_2)$. Interestingly, even when $p = c/n_2$, i.e. $n_2 \cdot p$ is a constant, trimming the graph (i.e., removing few nodes with high degree) gives a graph G s.t. $\sigma_2(G) = O(\sqrt{d})$. Hence, we can again apply Theorem 4.2 to obtain the following result:

Corollary 4.4. *Let M be an $n_1 \times n_2$ matrix of rank r satisfying assumptions (A1) and (A2) with $\delta_d \leq \frac{1}{6}$, and Ω is generated from a $\mathcal{G}(n, p)$ graph after trimming, then M is the unique optimum of program* (7) *when $p \geq \frac{36c\mu_0^2 r^2}{\min\{n_1,n_2\}}$, with high probability.*

While the above two results exploit the fact that a random graph is almost a Ramanujan expander and hence our general recovery result can be applied, the graph construction is still randomized. Interestingly, (Lubotzky et al., 1988; Margulis, 1988; Morgenstern, 1994) proposed explicit deterministic constructions of Ramanujan graphs when $d - 1$ is a prime power. Moreover, (Marcus et al., 2013) showed that bipartite Ramanujan graphs exist for all n and d. However, explicit construction for all n and d still remains an open problem.

5. Discussion

In this section, we discuss the two assumptions $A1$ and $A2$ that are mentioned in Section 3.

Note that $A1$ is a standard assumption that is used by most of the existing approaches (Candès & Recht, 2009), (Keshavan et al., 2010). Moreover, it is easy to show that for any matrix "oblivious" sampling approach, this assumption is necessarily required for exact recovery. For example, if $G_{ij} = 0$, i.e., (i, j)-th element is not observed then we cannot recover $M = \mathbf{e}_i \mathbf{e}_j^T$.

However, $A2$ is a slightly non-standard assumption and intuitively it requires the singular vectors of M to satisfy RIP. Note that $A2$ is similar to the strong incoherence property introduced by (Candès & Tao, 2010). Below we show the connection between strong incoherence property (SIP) assumed in (Candès & Tao, 2010) and assumption $A2$.

Claim 5.1. *Let $M \in \mathbb{R}^{n_1 \times n_2}$ be a rank-r matrix. Let $M = U\Sigma V^T$ satisfy SIP i.e.,*

$$|\langle \mathbf{e}_i, UU^T \mathbf{e}_j \rangle - \frac{r}{n_1} \mathbf{1}_{i=j}| \leq \mu_1 \frac{\sqrt{r}}{n_1}, \forall 1 \leq i, j \leq n_1,$$

$$|\langle \mathbf{e}_i, VV^T \mathbf{e}_j \rangle - \frac{r}{n_1} \mathbf{1}_{i=j}| \leq \mu_1 \frac{\sqrt{r}}{n_2}, \forall 1 \leq i, j \leq n_2. \quad (8)$$

Then, M satisfies $A2$ for all $d \geq r$ and $\delta_d \leq \mu_1 \sqrt{r}$.

Note that the above claim holds with $\delta_d = \mu_1 \sqrt{r}$, $\forall d \geq r$. This bound is independent of d and hence weak; as d becomes close to n_1, $\left\| \frac{n_1}{d} \sum_{k \in S} U^k U^{k^T} - I \right\|$ gets close to 0 since $U^T U = I$. We leave the task of obtaining a stronger bound as an open problem.

In the context of universal recovery, a natural question is if any additional assumption is required or the standard $A1$ assumption alone suffices. Here, we answer this question in negative. Specifically, we show that if M satisfies $A1$ only, then universal recovery guarantee for even rank-2 matrices cannot be provided by using as many as $n_1 n_2/4$ observations.

Claim 5.2. *Let Ω be a fixed set of indices and let P_Ω be the sampling operator as defined in* (1)*. Let $n_1 = n_2 = n$ and let $|\Omega| = n^2/4$. Then, there exists a rank-2 matrix M that cannot be recovered exactly from $P_\Omega(M)$.*

Now, another question is if we require a property as strong as $A2$ and if just a lower-bound on $\|U^i\|_2, \|V^j\|_2$ is enough for universal recovery. The proof of the above claim (in appendix) shows that even if $\|U^i\|_2, \|V^j\|_2$ are lower-bounded, then also exact recovery is not possible.

6. Simulations

In this section, we will present a few empirical results on both synthetic and real data sets. The goal of this section is to demonstrate effect of the spectral gap of the sampling graph $\mathcal{G}$ (associated with Ω) on successful recovery of a matrix.

First, we use synthetic data sets generated in the following manner. We first sample $U, V \in \mathbb{R}^{500 \times 10}$ using standard normal distribution. We then generate rank-10 matrix M, using $M = UV^T$. As U, V are sampled from the normal distribution, hence w.h.p., M satisfies incoherence assumptions $A1, A2$ mentioned in Section 3. Next, we generate a sequence of sampling operators P_Ω (and the associated graph $\mathcal{G}$) with varying (relative) spectral gap$(1 - \sigma_2(G)/\sigma_1(G))$ by using a stochastic block model. In the basic stochastic block model, the nodes can be thought of as being divided into two clusters. Now, each intra-cluster edge is sampled uniformly with probability p and an inter-cluster edge is sampled uniformly with probability q. Note that, when $p = q$, then the spectral gap is largest and when $p = 1, q = 0$, then spectral gap is smaller as there are two distinct clusters in that case (Nadakuditi & Newman, 2012).

Note that number of samples generated in this model depends only on the value of $p + q$. To generate Ω (i.e., G), we first fix a value for $p + q$, hence fixing the number of samples, and then vary p, q, which gives graphs of different spectral gap. As value of p goes from 0 to $\frac{p+q}{2}$, the spectral gap goes up and from $\frac{p+q}{2}$ to $p + q$, the spectral gap goes

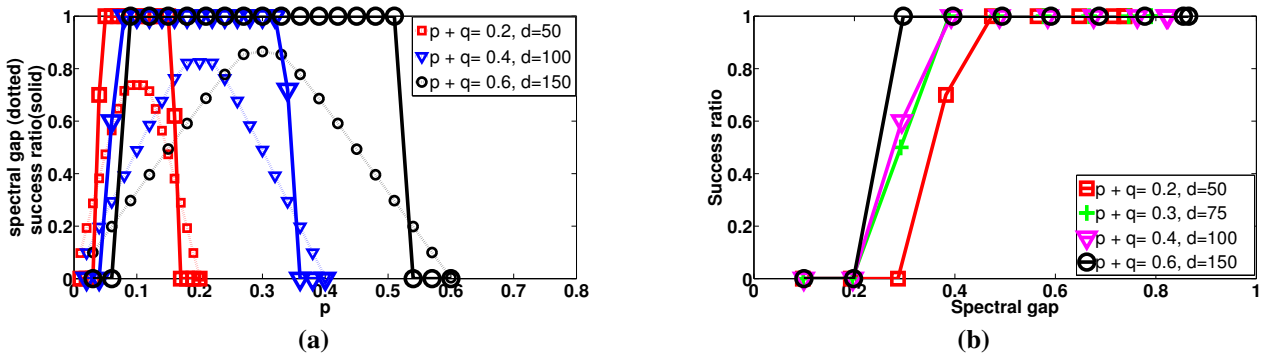

Figure 1. **(a)**: Figure plots the spectral gap (dotted lines) as well as the the fraction of successful recoveries (solid lines) with varying p, the parameter in stochastic block model. Value of $p+q$ dictates the number of observed entries, i.e., $|\Omega|$. **(b):** Fraction of successful recovery vs spectral gap of sampling operator. Clearly, success ratio for matrix recovery show a phase transition type phenomenon w.r.t. the spectral gap. Also, different values of $p+q$, i.e, number of samples do not affect success ratio too much.

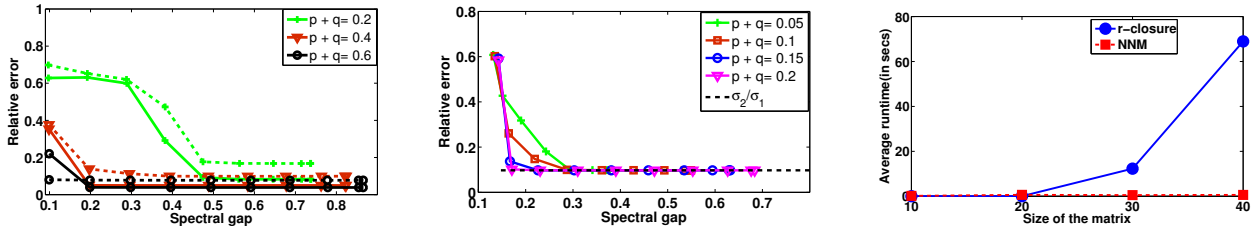

Figure 2. **(a)**: *Noisy samples case:* Relative error in Frobenius norm vs spectral gap of sampling operator. Perturbed samples of matrix $M+Z$ are observed, where Z is Gaussian noise matrix with $\sigma = \|Z\|_F/\|M\|_F$. Solid lines correspond to $\sigma = 0.1$ and dotted for $\sigma = 0.2$. **(b):** *Temperature prediction:* Figure plots relative error in spectral norm vs spectral gap of the sampling operator for different number of samples$(p+q)$. **(c):** *Comparison with r-closure algorithm:* Figure compares running time of r-closure algorithm with nuclear norm minimization algorithm.

down. Figure 1(a) clearly demonstrates this trend.

We use an Augmented Lagrangian Method (ALM) based method (Lin et al., 2010) to solve the nuclear norm minimization (7) problem. A trial is considered to be successful if the relative error (in Frobenius norm) is less than 0.01. We average over 50 such trials to determine the success ratio.

Figure 1(a) plots the (relative) spectral gap (dotted lines) and the success ratio (solid lines) as p varies. Lines of different colors indicate different number of samples $(p+q)$. As expected, the spectral gap increases initially, as p varies from 0 to $\frac{p+q}{2}$ and then it decreases. Moreover, the trend of successful recovery also follows a similar trajectory and hence, is more or less independent of the number of samples (given a particular spectral gap).

Figure 1(b) shows fraction of successful recoveries as the spectral gap increases. Here again, lines of different colors indicate different number of samples $(p+q)$. Clearly, success ratio is positively correlated with the spectral gap of the sampling operator and in fact exhibits a phase transition type of phenomenon. We expect the difference in success ratio for different $p+q$ values to decrease with increasing problem size, i.e, dimensionality of M.

Now, we conduct an experiment to show that spectral gap of G helps in reducing effect of noise as well. To this end, we generated noisy input matrix $(M+Z)$, where Z is a random Gaussian matrix and let $\sigma = ||Z||_F/||M||_F$. We consider two values of σ, i.e., $\sigma = 0.1$ and 0.2. Figure 2(a) plots the error(in Frobenius norm) in the recovered matrix against the relative spectral gap in the noisy setting. Solid lines represent $\sigma = 0.1$ and dotted lines represent $\sigma = 0.2$. Clearly, larger spectral gap leads to smaller error in recovery. Moreover, the "matrix completion denoising" effect (Candes & Plan, 2010) can also be observed. For example, when $||Z||_F = 0.1$ and $p+q = 0.4$, output error is less than 0.05, and when $||Z||_F = 0.2$, error is less than 0.1

Temperature prediction: Finally we take a real dataset of temperature values(T) for 365 days at 316 different locations from (NCDC), which has been used to test matrix completion algorithms (Candes & Plan, 2010) before. Note that T is approximately rank-1 matrix with

$\sigma_1(T)/||T||_F = 0.98(\sigma_i(T)$ are singular values of $T)$. We use the block model sampling scheme to sample entries from T, and let the output of (7) be $\hat{T}$. In figure 2(b) we plot the error $||\hat{T} - T||/||T||$ for different values of spectral gap of G and for different number of samples $(p + q)$. Note that $||X - T||/||T|| \geq \sigma_2(T)/\sigma_1(T)$ for any rank-1 matrix X, and we see that for large enough spectral gap we achieve this bound.

Finally in figure 2(c) we compare running times of the r-closure algorithm proposed in (Király & Tomioka, 2012) with the nuclear norm minimization algorithm. While it is noted that this algorithm has better error guarantees, it is combinatorial and takes exponential time to compute.

7. Proof of Theorem 4.2

In this section, we present the proof of our main result (Theorem 4.2). The main steps of our proof are similar to the proof given by (Recht, 2009). The main difference is that the bounds in the existing proof assume that Ω is independent of M and hence is not adversarial and holds with high probability. In contrast, for our proofs, bounds are deterministic and our proofs have to work under the assumption that M is adversarially selected for a given Ω.

The key steps in the proof are: a) provide conditions that an optimal dual solution (or dual certificate) of problem (7) should satisfy, so that the true matrix M is the *unique* optimum of (7), b) construct such a dual certificate and hence guarantee that M is the unique optimum of (7).

We first introduce a few notations required by our proof. For simplicity, we assume that $n_1 = n_2 = n$. Note that, our proof easily generalizes to case when $n_1 \neq n_2$. Define T which is a subspace of $\mathbb{R}^{n\times n}$, and is span of all matrices of form UX^T and YV^T, i.e. all matrices with either row space in V or column space in U. Hence, the projection operator $\mathcal{P}_T$ is defined as follows:

$$\begin{aligned}\mathcal{P}_T(Z) =& UU^TZ + ZVV^T - UU^TZVV^T \\ =& UU^TZ + (I - UU^T)ZVV^T.\end{aligned}$$

Hence any matrix in T can be written as $UX^T + YV^T$, for some X and Y such that Y and U are orthogonal to each other. Similarly, we can define the projection operator onto $T^\perp$, the orthogonal complement of T:

$$\mathcal{P}_{T^\perp}(Z) = (I - UU^T)Z(I - VV^T).$$

Now, before presenting conditions on the dual certificate and construction of the dual certificate, we provide a few structural lemmas that show "goodness" of operators $\mathcal{P}_T$ and P_Ω. We would like to stress that the key differences of our proof from that of (Recht, 2009) is in fact proofs of these structural lemmas and also in the way we apply Lemma 7.3. We specifically show (using Lemma 7.3) that each matrix in the series used in the construction of dual certificate Y(discussed later in the section) is incoherent and has small infinity norm.

The first lemma proves injectivity of operator P_Ω on the subspace T:

Lemma 7.1. *Let $M = U\Sigma V^T$ satisfies $A1, A2$ and let the graph $\mathcal{G}$ that generates Ω satisfies $G1, G2$ (see Section 3). Then, for any matrix $Z \in T$,*

$$\|\frac{n}{d}\mathcal{P}_T P_\Omega(Z) - Z\|_F \leq \sqrt{2(\delta_d^2 + \frac{C^2\mu_0^2 r^2}{d})}\|Z\|_F.$$

Next, we provide a lemma that characterizes the "difference" between $P_\Omega(Z)$ and Z, for any incoherent-type $Z \in T$:

Lemma 7.2. *Let $Z \in T$, i.e., $Z = UX^T + YV^T$ and Y is orthogonal to U, and X and Y be incoherent, i.e.,*

$$\|X^i\|^2 \leq \frac{c_1^2\mu_0 r}{n}, \;\; \|Y^j\|^2 \leq \frac{c_2^2\mu_0 r}{n}.$$

Let Ω satisfy the assumptions $G1$ and $G2$, then:

$$\|\frac{n}{d}P_\Omega(Z) - Z\| \leq (c_1 + c_2)\frac{C\mu_0 r}{\sqrt{d}}.$$

The next lemma is a stronger version of Lemma 7.1 for special incoherent-type matrices $Z \in T$.

Lemma 7.3. *Let $Z \in T$, i.e., $Z = UX^T + YV^T$ and Y is orthogonal to U. Let X and Y be incoherent, i.e.,*

$$\|X^i\|^2 \leq \frac{c_1^2\mu_0 r}{n}, \;\; \|Y^j\|^2 \leq \frac{c_2^2\mu_0 r}{n}.$$

Let $\tilde{Z} = Z - \frac{n}{d}\mathcal{P}_T P_\Omega(Z)$. Then, the following holds for M, Ω that satisfy conditions given in Lemma 7.1:

- $\|\tilde{Z}\|_\infty \leq \frac{(c_1+c_2)\mu_0 r}{n}(\delta_d + \frac{C\mu_0 r}{\sqrt{d}})$.
- $\tilde{Z} = U\tilde{X}^T + \tilde{Y}V^T$ *and* $\tilde{X}$ *and* $\tilde{Y}$ *are incoherent.* $\|\tilde{X}^i\|^2 \leq \frac{\mu_0 r}{n}\left(\delta_d c_1 + 2c_2\frac{C\mu_0 r}{\sqrt{d}}\right)^2$ *and* $\|\tilde{Y}^j\|^2 \leq \frac{\mu_0 r}{n}(\delta_d c_2 + (c_1 + c_2)\frac{C\mu_0 r}{\sqrt{d}})^2$.

The proof of the above three lemmas is provided in the appendix.

Conditions on the dual certificate: We now present the lemma that characterizes the conditions a dual certificate should satisfy so that M is the unique optimum of (7):

Lemma 7.4. *Let M, Ω satisfy $A1, A2$ and $G1, G2$, respectively. Then, M is the unique optimum of* (7)*, if there exists a $Y \in \mathbb{R}^{n\times n}$ that satisfies the following:*

- $P_\Omega(Y) = Y$

- $||\mathcal{P}_T(Y) - UV^T||_F \leq \sqrt{\frac{d}{8n}}$
- $||\mathcal{P}_{T^\perp}(Y)|| < \frac{1}{2}$

Having specified the conditions on dual certificate and also the key structural lemma, we are now ready to present the proof of Theorem 4.2.

Proof of Theorem 4.2. We prove the theorem by constructing a dual certificate Y that satisfies conditions in lemma 7.4 and hence guarantee that M is exactly recovered by (7). Our construction of Y is similar to the golfing scheme based construction given in (Gross, 2011; Recht, 2009). In particular, Y is obtained as the p-th term of the series given below:

$$W_{k+1} = W_k - \frac{n}{d}\mathcal{P}_T P_\Omega W_k, \quad Y_k = \sum_{i=0}^{k-1} \frac{n}{d} P_\Omega W_i,$$

where $W_0 = UV^T$. That is, $Y = Y_p$ where $p = \lceil \frac{1}{2}\log_3(\frac{n}{18C^2\mu_0^2 r}) \rceil$. Also, define $\alpha = \frac{C\mu_0 r}{\sqrt{d}}$. As $d \geq 36C^2\mu_0^2 r^2$, we get: $\alpha \leq \frac{1}{6}$.

Now, the *first condition* of Lemma 7.4 is satisfied trivially by construction as Y_p is a sum of $P_\Omega(W_i)$ terms.

Bounding $||\mathcal{P}_T(Y) - UV^T||_F$: By construction:

$$\mathcal{P}_T(Y) - UV^T = \sum_{i=0}^{p-1} \frac{n}{d}\mathcal{P}_T P_\Omega W_i - UV^T = -W_p. \quad (9)$$

Now, note that each $W_k \in T$. Hence, using Lemma 7.1,

$$\|W_{k+1}\|_F = \|W_k - \frac{n}{d}\mathcal{P}_T P_\Omega W_k\|_F \leq \sqrt{2(\delta_d^2 + \frac{C^2\mu_0^2 r^2}{d})}\|W_k\|_F \leq \frac{1}{3}\|W_k\|_F, \quad (10)$$

where the last inequality follows by using assumption on δ_d and by using $\alpha \leq 1/6$. Hence, using (9), (10), we have:

$$\|\mathcal{P}_T(Y) - UV^T\|_F = \|W_p\|_F \leq \left(\frac{1}{3}\right)^p \|W_0\|_F \leq \sqrt{\frac{d}{8n}}, \quad (11)$$

where the last inequality follows by using $p = \lceil \frac{1}{2}\log_3(\frac{n}{18C^2\mu_0^2 r}) \rceil$ and by using $\|W_0\|_F = \sqrt{r}$.

Bounding $\|\mathcal{P}_{T^\perp}(Y)\|$: Recall that, $W_k \in T$. Now, let $W_k = UX_{W_k}^T + Y_{W_k}V^T$ with Y_{W_k} perpendicular to U. Moreover, let,

$$\|X_{W_k}^i\| \leq \frac{c_1^{W_k}\sqrt{\mu_0 r}}{\sqrt{n}}, \ \ \|Y_{W_k}^i\| \leq \frac{c_2^{W_k}\sqrt{\mu_0 r}}{\sqrt{n}}.$$

Note that, for $W_0 = UV^T$, $c_1^{W_0} = 1$ and $c_2^{W_0} = 0$. Hence, using Lemma 7.3: $c_1^{W_1} \leq \frac{1}{6}$, $c_2^{W_1} \leq \frac{1}{6}$. Similarly, applying Lemma 7.3 k-times, we get:

$$c_1^{W_k} = c_2^{W_k} \leq 3^{k-1}\frac{1}{6^k}. \quad (12)$$

Now, by using construction of Y, by triangle inequality, and by using the fact that $\mathcal{P}_{T^\perp}$ is a contraction operator:

$$\|\mathcal{P}_{T^\perp}(Y)\| \leq \sum_{k=1}^{p} \|\frac{n}{d}P_\Omega(W_{k-1}) - W_{k-1}\| \leq \sum_{k=1}^{p}(c_1^{W_{k-1}} + c_2^{W_{k-1}})\frac{C\mu_0 r}{\sqrt{d}}, \quad (13)$$

where the last inequality follows by using Lemma 7.2. Now, using (12), (13), and Lemma 7.3, we have:

$$\|\mathcal{P}_{T^\perp}(Y)\| \leq \frac{C\mu_0 r}{\sqrt{d}}(1 + 2\cdot\frac{1}{6}(\sum_{k=2}^{p-1} 3^{k-2}\frac{1}{6^{k-2}})) \leq \alpha(1 + 2\cdot\frac{1}{6}\cdot\frac{1}{1-\frac{1}{2}}) < 1/2.$$

Hence, proved. □

8. Conclusions

In this paper, we provided the first (to the best of our knowledge) universal recovery guarantee for matrix completion. The main observation of the paper is that the spectral gap of $\mathcal{G}$ (that generates Ω) is the key property that governs recoverability of M using $P_\Omega(M)$ alone. For example, if $\mathcal{G}$ is a Ramanujan expander (i.e., $\sigma_2(G) = O(\sqrt{d})$), then we have universal recovery guarantees for matrices with strong incoherence property.

For uniformly sampled Ω, our main result implies exact recovery of constant rank matrices using $O(n)$ entries, in contrast to the $O(n\log n)$ entries required by the existing analyses. One caveat is that we require stronger incoherence property to obtain the above given sample complexity. Our results also provide a recipe to determine if a given index set Ω is enough to recover a low-rank matrix. That is, given Ω and its associated graph $\mathcal{G}$, we can measure the spectral gap of G and if it is large enough then our results guarantee exact recovery of strongly incoherent matrices.

In Section 5, we showed that the standard incoherence assumption alone is not enough for universal recovery and a property similar to $A2$ (see Section 3) is required. However, it is an open problem to obtain precise information theoretic limits on δ_d (see $A2$) for universal recovery guarantees. Another interesting research direction is to study the alternating minimization method under assumptions given in Section 3.

References

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

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

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

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

Chen, Yudong, Bhojanapalli, Srinadh, Sanghavi, Sujay, and Ward, Rachel. Coherent matrix completion. *arXiv preprint arXiv:1306.2979*, 2013.

Feige, Uriel and Ofek, Eran. Spectral techniques applied to sparse random graphs. *Random Structures & Algorithms*, 27(2):251–275, 2005.

Friedman, Joel. A proof of alon's second eigenvalue conjecture. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pp. 720–724. ACM, 2003.

Gopi, Sivakant, Netrapalli, Praneeth, Jain, Prateek, and Nori, Aditya. One-bit compressed sensing: Provable support and vector recovery. In *Proceedings of the 30th international conference on machine learning (ICML-13)*, pp. 154–162, 2013.

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

Gross, David, Liu, Yi-Kai, Flammia, Steven T, Becker, Stephen, and Eisert, Jens. Quantum state tomography via compressed sensing. *Physical review letters*, 105(15): 150401, 2010.

Heiman, Eyal, Schechtman, Gideon, and Shraibman, Adi. Deterministic algorithms for matrix completion. *Random Structures & Algorithms*, 2013.

Hoory, Shlomo, Linial, Nathan, and Wigderson, Avi. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

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

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

Király, Franz J. and Tomioka, Ryota. A combinatorial algebraic approach for the identifiability of low-rank matrix completion. In *ICML*, 2012.

Lee, Troy and Shraibman, Adi. Matrix completion from any given set of observations. In *Advances in Neural Information Processing Systems*, pp. 1781–1787, 2013.

Lin, Zhouchen, Chen, Minming, and Ma, Yi. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010.

Liu, Yi-Kai. Universal low-rank matrix recovery from pauli measurements. *arXiv preprint arXiv:1103.2816*, 2011.

Lubotzky, Alexander, Phillips, Ralph, and Sarnak, Peter. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

Marcus, Adam, Spielman, Daniel A, and Srivastava, Nikhil. Interlacing families i: Bipartite ramanujan graphs of all degrees. *arXiv preprint arXiv:1304.4132*, 2013.

Margulis, Grigorii Aleksandrovich. Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators. *Problemy peredachi informatsii*, 24(1):51–60, 1988.

Meka, Raghu, Jain, Prateek, and Dhillon, Inderjit S. Matrix completion from power-law distributed samples. In *Advances in Neural Information Processing Systems*, pp. 1258–1266, 2009.

Morgenstern, Moshe. Existence and explicit constructions of q+ 1 regular ramanujan graphs for every prime power q. *Journal of Combinatorial Theory Series B*, 62(1):44–62, 1994.

Nadakuditi, Raj Rao and Newman, Mark EJ. Graph spectra and the detectability of community structure in networks. *Physical review letters*, 108(18):188701, 2012.

NCDC. National climatic data center. URL `http://www.ncdc.noaa.gov/oa/ncdc.html`.

Recht, Benjamin. A simpler approach to matrix completion. *arXiv preprint arXiv:0910.0651*, 2009.

Recht, Benjamin, Fazel, Maryam, and Parrilo, Pablo A. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.