ISIT 2009, Seoul, Korea, June 28 - July 3, 2009

Matrix Completion from a Few Entries

Raghunandan H. Keshavan and Sewoong Oh
EE Department
Stanford University, Stanford, CA 94304

Andrea Montanari
EE and Statistics Departments
Stanford University, Stanford, CA 94304

***Abstract*—Let M be an $n\alpha \times n$ matrix of rank $r \ll n$, and assume that a uniformly random subset E of its entries is observed. We describe an efficient algorithm that reconstructs M from $|E| = O(r\,n)$ observed entries with relative root mean square error $\mathrm{RMSE} \le C(\alpha)\ (nr/|E|)^{1/2}$. Further, if $r = O(1)$ and M is sufficiently unstructured, then it can be reconstructed *exactly* from $|E| = O(n \log n)$ entries.**

This settles (in the case of bounded rank) a question left open by Candès and Recht and improves over the guarantees for their reconstruction algorithm. The complexity of our algorithm is $O(|E| r \log n)$, which opens the way to its use for massive data sets. In the process of proving these statements, we obtain a generalization of a celebrated result by Friedman-Kahn-Szemerédi and Feige-Ofek on the spectrum of sparse random matrices.

I. Introduction

Imagine that each one of m customers watches and rates a subset of the n movies available through a movie rental service. This yields a dataset of customer-movie pairs[1] $(i, j) \in E \subseteq [m] \times [n]$ and, for each such pair, a rating $M_{ij} \in \mathbb{R}$. The objective of *collaborative filtering* is to predict the rating for missing pairs in such a way to provide targeted suggestions. As an example, in 2006, NETFLIX made public such a dataset with $m \approx 5 \cdot 10^5$, $n \approx 2 \cdot 10^4$ and $|E| \approx 10^8$ and challenged the research community to predict the missing ratings with root mean square error below 0.8563 [1].

The general question we address here is: under which conditions do the known ratings provide sufficient information to efficiently infer the unknown ones?

A. Model definition

A simple mathematical model for such data assumes that the (unknown) matrix of ratings has rank $r \ll m, n$. More precisely, we denote by M the matrix whose entry $(i, j) \in [m] \times [n]$ corresponds to the rating user i would assign to movie j. We assume that there exist matrices U, of dimensions $m \times r$, and V, of dimensions $n \times r$, and a diagonal matrix Σ, of dimensions $r \times r$ such that

$$M = U \Sigma V^T . \tag{1}$$

For justification of these assumptions and background on the use of low rank matrices in information retrieval, we refer to [2]. Motivated by the massive size of actual datasets, we shall focus on the limit of large m, n with $m/n = \alpha$ of order 1.

We further assume that the factors U, V are unstructured. This notion is formalized by the *incoherence condition* [3] as defined in Section II. In particular the incoherence condition is satisfied with high probability if $M = U'V'^T$ with U' and V' uniformly random orthogonal matrices.

Out of the $m \times n$ entries of M, a subset $E \subseteq [m] \times [n]$ (the user/movie pairs for which a rating is available) is revealed. We let M^E be the $m \times n$ matrix that contains the revealed entries of M, and is filled with 0's in the other positions

$$M^E_{i,j} = \begin{cases} M_{i,j} & \text{if } (i,j) \in E\,, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The set E will be uniformly random given its size $|E|$.

B. Algorithm and guarantees

A naive algorithm consists of the following operation.

Projection. Compute the singular value decomposition (SVD) of M^E (with $\sigma_1 \ge \sigma_2 \ge \cdots \ge 0$)

$$M^E = \sum_{i=1}^{\min(m,n)} \sigma_i x_i y_i^T\,, \tag{3}$$

And return the matrix $\mathsf{T}_r(M^E) = (mn/|E|) \sum_{i=1}^{r} \sigma_i x_i y_i^T$ obtained by setting to 0 all but the r largest singular values. Notice that, apart from the rescaling factor $(mn/|E|)$, $\mathsf{T}_r(M^E)$ is the orthogonal projection of M^E onto the set of rank-r matrices. The rescaling factor compensates the smaller average size of the entries of M^E with respect to M.

This algorithm fails if $|E| = \Theta(n)$. The reason is that, in this regime, the matrix M^E contains columns and rows with $\Omega(\log n/ \log\log n)$ non-zero (revealed) entries. The largest singular values of M^E are an artifact of these high weight columns/rows and do not provide useful information about the hidden entries of M. This motivates the definition of the following operation (hereafter the *degree* of a column or of a row is the number of its revealed entries).

Trimming. Set to zero all columns in M^E with degree larger that $2|E|/n$. Set to 0 all rows with degree larger than $2|E|/m$.

In terms of the above routines, our algorithm has the following structure.

SPECTRAL MATRIX COMPLETION (matrix M^E)	
1:	Trim M^E, and let $\widetilde{M}^E$ be the output;
2:	Project $\widetilde{M}^E$ to $\mathsf{T}_r(\widetilde{M}^E)$;
3:	Clean residual errors by minimizing $F(X,Y)$.

[1]Throughout this paper we denote by $[N] = \{1, 2, \ldots, N\}$ the set of first N integers.

The last step of the above algorithm allows to reduce (or eliminate) small discrepancies between $\mathsf{T}_r(\widetilde{M}^E)$ and M, and is described below.

Cleaning. Various implementations are possible, but we found the following one particularly appealing. Given $X \in \mathbb{R}^{m\times r}$, $Y \in \mathbb{R}^{n\times r}$ with $X^TX = m\mathbf{1}$ and $Y^TY = n\mathbf{1}$, we define

$$F(X,Y) \equiv \min_{S\in\mathbb{R}^{r\times r}} \mathcal{F}(X,Y,S)\,, \tag{4}$$

$$\mathcal{F}(X,Y,S) \equiv \frac{1}{2}\sum_{(i,j)\in E}(M_{ij}-(XSY^T)_{ij})^2\,. \tag{5}$$

The cleaning step consists in writing $\mathsf{T}_r(\widetilde{M}^E) = X_0S_0Y_0^T$ and minimizing $F(X,Y)$ locally with initial condition $X = X_0$, $Y = Y_0$.

Notice that $F(X,Y)$ is easy to evaluate since it is defined by minimizing the quadratic function $S \mapsto \mathcal{F}(X,Y,S)$ over the low-dimensional matrix S. Further it depends on X and Y only through their column spaces. In geometric terms, F is a function defined over the cartesian product of two Grassmann manifolds (we refer to the journal version of this paper for background and references). Optimization over Grassmann manifolds is a well understood topic [4] and efficient algorithms (in particular Newton and conjugate gradient) can be applied. To be definite, we assume that gradient descent with line search is used to minimize $F(X,Y)$.

Our main result establishes that this simple procedure achieves arbitrarily small root mean square error $||M - \mathsf{T}_r(\widetilde{M}^E)||_{\rm F}/\sqrt{mnr}$ with $O(nr)$ revealed entries.

Theorem I.1. *Assume M to be a rank $r \le n^{1/2}$ matrix with $|M_{ij}| \le M_{\max}$ for all i,j. Then with high probability*

$$\frac{1}{mnM_{\max}^2}||M-\mathsf{T}_r(\widetilde{M}^E)||_{\rm F}^2 \le C(\alpha)\,\frac{nr}{|E|}\,. \tag{6}$$

The proof is provided in Section IV (the proofs of several technical remarks can be found in the journal version [5]).

Theorem I.2. *Assume M to be a rank $r \le n^{1/2}$ matrix that satisfies the incoherence conditions A1 and A2. Further, assume $\Sigma_{\min} \le \Sigma_1,\dots,\Sigma_r \le \Sigma_{\max}$ with $\Sigma_{\min},\Sigma_{\max}$ bounded away from 0 and ∞. Then there exists $C'(\alpha)$ such that, if*

$$|E| \ge C'(\alpha)nr\,\max\{\log n, r\}\,, \tag{7}$$

then the cleaning procedure in SPECTRAL MATRIX COMPLETION *converges, with high probability, to the matrix M.*

The proof will appear in the journal version of this paper [5]. The basic intuition is that, for $|E| \ge C'(\alpha)rn\max\{\log n, r\}$, $T_r(\widetilde{M}^E)$ is so close to M that the cost function is well approximated by a quadratic function.

Theorem I.1 is optimal: the number of degrees of freedom in M is of order nr, without the same number of observations is impossible to fix them. The extra $\log n$ factor in Theorem I.2 is due to a coupon-collector effect [3], [6], [5]: it is necessary that E contains at least one entry per row and one per column and this happens only for $|E| \ge Cn\log n$. As a consequence, for rank r bounded, Theorem I.2 is optimal. It is suboptimal by a polylogarithmic factor for $r = O(\log n)$.

C. Related work

Beyond collaborative filtering, low rank models are used for clustering, information retrieval, machine learning, and image processing. In [7], the NP-hard problem of finding a matrix of minimum rank satisfying a set of affine constraints was addressed through convex relaxation. This problem is analogous to the problem of finding the sparsest vector satisfying a set of affine constraints, which is at the heart of *compressed sensing* [8], [9]. The connection with compressed sensing was emphasized in [10], that provided performance guarantees under appropriate conditions on the constraints.

In the case of collaborative filtering, we are interested in finding a matrix M of minimum rank that matches the known entries $\{M_{ij} : (i,j) \in E\}$. Each known entry thus provides an affine constraint. Candès and Recht [3] proved that, if E is random, the convex relaxation correctly reconstructs M as long as $|E| \ge C\,r\,n^{6/5}\log n$. On the other hand, from a purely information theoretic point of view (i.e. disregarding algorithmic considerations), it is clear that $|E| = O(n\,r)$ observations should allow to reconstruct M with arbitrary precision. Indeed this point was raised in [3] and proved in [6], through a counting argument.

The present paper fills this gap. We describe an efficient algorithm that reconstructs a rank-r matrix from $O(n\,r)$ random observations. The most complex component of our algorithm is the SVD in step 2. Generic routines accomplish this task with $O(n^3)$ operations. Thanks to the sparsity of $\widetilde{M}^E$, this step can be implemented using the Lanczos procedure with $O(|E|r\log n)$ complexity. We were able to treat realistic data sets with $n \approx 10^5$. This must be compared with the $O(n^4)$ complexity of [3] (but see [11] for an iterative implementation of the latter).

After this paper was submitted to ISIT, Candés and Tao [12] proved a guarantee for the convex relaxation algorithm, that is comparable with Theorem I.2. A longer version of the present paper was submitted to IEEE Transactions on Information Theory [5].

II. INCOHERENCE PROPERTY

In order to formalize the notion of incoherence, we write $U = [u_1, u_2, \dots, u_r]$ and $V = [v_1, v_2, \dots, v_r]$ for the columns of the two factors, with $||u_i|| = \sqrt{m}$, $||v_i|| = \sqrt{n}$ and $u_i^Tu_j = 0$, $v_i^Tv_j = 0$ for $i \ne j$ (there is no loss of generality in this, since normalizations can be adsorbed by redefining Σ). We shall further write $\Sigma = \mathrm{diag}(\Sigma_1,\dots,\Sigma_r)$ with $\Sigma_1 \ge \Sigma_2 \ge \dots \ge \Sigma_r \ge 0$.

The matrices U, V and Σ will be said to be (μ_0,μ_1)-*incoherent* if they satisfy the following properties:

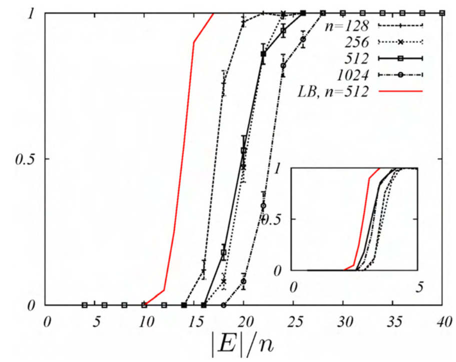

Fig. 1. Probability of successfully reconstructing a rank 4 matrix using our algorithm, for different matrix sizes. The leftmost curve is bound proved in [13]. In the inset the same data are plotted vs. $|E|/n \log n$.

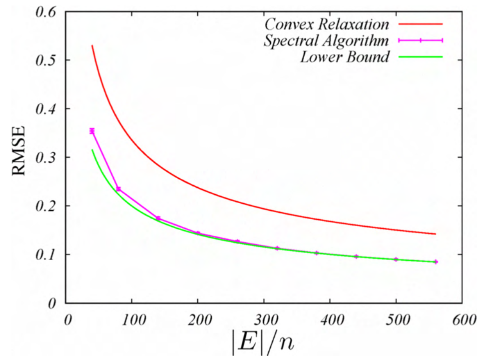

Fig. 2. Reconstructing a rank 2-matrix with dimensions $m = n = 600$ from $|E|$ *noisy* observations of the entries $M_{ij} + Z_{ij}$, with Z_{ij} i.i.d. Normal$(0, 1)$. The root mean square error of our algorithm is compared with the convex relaxation of [14], and an information theoretic lower bound.

A1. For all $i \in [m]$, $j \in [n]$, we have $\sum_{k=1}^{r} U_{i,k}^2 \le \mu_0 r$, $\sum_{k=1}^{r} V_{i,k}^2 \le \mu_0 r$.

A2. For all $i \in [m]$, $j \in [n]$, we have $|\sum_{k=1}^{r} U_{i,k}\Sigma_k V_{j,k}| \le \mu_1 r^{1/2}$.

The first one coincides with one of the incoherence assumptions in [3]. The second one is easier to verify than the analogous one in [3], in that it concerns the matrix elements themselves.

Notice that assumption A2 implies the bounded entry condition in Theorem I.1 with $M_{\max} = \mu_1 r^{1/2}$. In the following, whenever we write that a property A holds with high probability (w.h.p.), we mean that there exists a function $f(n) = f(n; \alpha)$ such that $\mathbb{P}(A) \ge 1 - f(n)$ and $f(n) \to 0$.

Define a constant $\epsilon \equiv |E|/\sqrt{mn}$. Then it is convenient to work with a model in which each entry is revealed independently with probability $\epsilon/\sqrt{mn}$. Since, w.h.p., $|E| = \epsilon\sqrt{\alpha}\, n + A\sqrt{n \log n}$, it will be sufficient to prove that our algorithm is successful for $\epsilon \ge Cr$. Finally, we will use C, C' etc. to denote generic constants that depend uniquely on $\alpha, \Sigma_{\min}, \Sigma_{\max}, \mu_0, \mu_1$.

Given a vector $x \in \mathbb{R}^n$, $||x||$ will denote its Euclidean norm. For a matrix $X \in \mathbb{R}^{n \times n'}$, $||X||_F$ is its Frobenius norm, and $||X||_2$ its operator norm.

III. Algorithm Implementation and Simulations

A MATLAB implementation of our algorithm is available from http://www.stanford.edu/~raghuram. In Fig. 1, we plot the probability that Spectral Matrix Completion exactly reconstructs M as a function of the number of revealed entries $|E|$. The algorithm is evaluated on random matrices of rank $r = 4$. As predicted by Theorem I.2, the success probability presents a sharp threshold for $|E| = C\, n \log n$. The location of the threshold is surprisingly close to the lower bound proved in [13], below which the problem admits more than one solution.

In Fig. 2 we apply our algorithm to 'approximately' low-rank matrices as defined in [14]. The resulting root mean square error is smaller by roughly 50 % with respect to the one obtained with the convex relaxation of [3], [14].

IV. Proof of Theorem I.1 and Technical Results

As explained in the previous section, the crucial idea is to consider the singular value decomposition of the trimmed matrix $\widetilde{M}^E$ instead of the original matrix M^E, as in Eq. (3). We shall then redefine $\{\sigma_i\}$, $\{x_i\}$, $\{y_i\}$, by letting

$$\widetilde{M}^E = \sum_{i=1}^{\min(m,n)} \sigma_i x_i y_i^T \, . \tag{8}$$

Here $||x_i|| = ||y_i|| = 1$, $x_i^T x_j = y_i^T y_j = 0$ for $i \neq j$ and $\sigma_1 \ge \sigma_2 \ge \cdots \ge 0$. Our key technical result is that, apart from a trivial rescaling, these singular values are close to the ones of the full matrix M.

Lemma IV.1. *There exists a constant $C > 0$ such that, with high probability*

$$\left|\frac{\sigma_q}{\epsilon} - \Sigma_q\right| \le \frac{C M_{\max}}{\sqrt{\epsilon}} \, , \tag{9}$$

where it is understood that $\Sigma_q = 0$ for $q > r$.

This result generalizes a celebrated bound on the second eigenvalue of random graphs [15], [16] and is illustrated in Fig. 3: the spectrum of $\widetilde{M}^E$ clearly reveals the rank-4 structure of M.

As shown in Section VI, Lemma IV.1 is a direct consequence of the following estimate.

Lemma IV.2. *There exists a constant $C > 0$ such that, with high probability*

$$\left|\left|\frac{\epsilon}{\sqrt{mn}} M - \widetilde{M}^E\right|\right|_2 \le C M_{\max}\sqrt{\epsilon} \, . \tag{10}$$

The proof of this lemma is given in Section V.

We will now prove Theorem I.1.

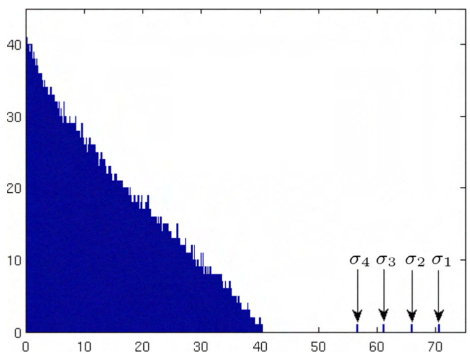

Fig. 3. Singular value distribution of $\widetilde{M}_E$ for $10^4 \times 10^4$ random rank 4 matrix M with $\epsilon = 50$ and $\Sigma = \text{diag}(1.3, 1.2, 1.1, 1)$.

Proof: (Theorem I.1) By triangular inequality

$$\begin{aligned} \left|\left|M - \mathsf{T}_r(\widetilde{M}^E)\right|\right|_2 &\le \left|\left|\frac{\sqrt{mn}}{\epsilon}\widetilde{M}^E - \mathsf{T}_r(\widetilde{M}^E)\right|\right|_2 \\ &\quad + \left|\left|M - \frac{\sqrt{mn}}{\epsilon}\widetilde{M}^E\right|\right|_2 \\ &\le \sqrt{mn}\sigma_{r+1}/\epsilon + CM_{\max}\sqrt{mn}/\sqrt{\epsilon} \\ &\le 2CM_{\max}\sqrt{mn}/\sqrt{\epsilon}\,, \end{aligned}$$

where we used Lemma IV.2 for the second inequality and Lemma IV.1 for the last inequality. Now, for any matrix A of rank at most $2r$, $||A||_F \le \sqrt{2r}||A||_2$, whence

$$\begin{aligned} \frac{1}{\sqrt{rmn}}\left|\left|M - \mathsf{T}_r(\widetilde{M}^E)\right|\right|_F &\le \frac{\sqrt{2}}{\sqrt{mn}}\left|\left|M - \mathsf{T}_r(\widetilde{M}^E)\right|\right|_2 \\ &\le C' M_{\max}/\sqrt{\epsilon}\,. \end{aligned}$$

■

V. PROOF OF LEMMA IV.2

We want to show that $|x^T(\widetilde{M}^E - \frac{\epsilon}{\sqrt{mn}}M)y| \le CM_{\max}\sqrt{\epsilon}$ for any $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$ such that $||x|| = ||y|| = 1$. Our basic strategy (inspired by [15]) will be the following:
(1) Reduce to x, y belonging to discrete sets T_m, T_n;
(2) Apply union bound to these sets, with a large deviation estimate on the random variable $x^T(\widetilde{M}^E - \frac{\epsilon}{\sqrt{mn}}M)y$.
The technical challenge is that a worst-case bound on the tail probability of $x^T(\widetilde{M}^E - \frac{\epsilon}{\sqrt{mn}}M)y$ is not good enough, and we must keep track of its dependence on x and y.

A. Discretization

We define

$$T_n = \left\{ x \in \left\{\frac{\Delta}{\sqrt{n}}\mathbb{Z}\right\}^n \ : \ ||x|| \le 1 \right\} ,$$

Notice that $T_n \subseteq S_n \equiv \{x \in \mathbb{R}^n : ||x|| \le 1\}$. The next two remarks are proved in [15], [16], and relate the original problem to the discretized one.

Remark V.1. *Let $R \in \mathbb{R}^{m\times n}$ be a matrix. If $|x^T R y| \le B$ for all $x \in T_m$ and $y \in T_n$, then $|x'^T R y'| \le (1-\Delta)^{-2}B$ for all $x' \in S_m$ and $y' \in S_n$.*

Hence it is enough to show that, with high probability, $|x^T(\widetilde{M}^E - \frac{\epsilon}{\sqrt{mn}}M)y| \le CM_{\max}\sqrt{\epsilon}$ for all $x \in T_m$ and $y \in T_n$.

A naive approach would be to apply concentration inequalities directly to the random variable $x^T(\widetilde{M}^E - \frac{\epsilon}{\sqrt{mn}}M)y$. This fails because the vectors x, y can contain entries that are much larger than the typical size $O(n^{-1/2})$. We thus separate two contributions. The first contribution is due to *light couples* $L \subseteq [m] \times [n]$, defined as

$$L = \left\{(i,j) \ : \ |x_i M_{ij} y_j| \le M_{\max}\,(\epsilon/mn)^{1/2}\right\} .$$

The second contribution is due to its complement $\overline{L}$, which we call *heavy couples*. We have

$$\begin{aligned} &\left|x^T\left(\widetilde{M}^E - \frac{\epsilon}{\sqrt{mn}}M\right)y\right| \le \\ &\left|\sum_{(i,j)\in L} x_i \widetilde{M}^E_{ij} y_j - \frac{\epsilon}{\sqrt{mn}}x^T M y\right| + \left|\sum_{(i,j)\in \overline{L}} x_i \widetilde{M}^E_{ij} y_j\right| . \end{aligned}$$

In the next subsection, we will prove that the first contribution is upper bounded by $C_1 M_{\max}\sqrt{\epsilon}$ for all $x \in T_m$, $y \in T_n$. The analogous proof for heavy couples can be found in the journal version [5]. Applying Remark V.1 to $|x^T(\widetilde{M}^E - \frac{\epsilon}{\sqrt{mn}}M)y|$, this proves the thesis.

B. Bounding the contribution of light couples

Let us define the subset of row and column indices which have not been trimmed as $\mathcal{A}_l$ and $\mathcal{A}_r$:

$$\begin{aligned} \mathcal{A}_l &= \{i \in [m] \ : \ \deg(i) \le \frac{2\epsilon}{\sqrt{\alpha}}\}\,, \\ \mathcal{A}_r &= \{j \in [n] \ : \ \deg(j) \le 2\epsilon\sqrt{\alpha}\}\,, \end{aligned}$$

where $\deg(\cdot)$ denotes the degree (number of revealed entries) of a row or a column. Notice that $\mathcal{A} = (\mathcal{A}_l, \mathcal{A}_r)$ is a function of the random set E. It is easy to get a rough estimate of the sizes of $\mathcal{A}_l$, $\mathcal{A}_r$.

Remark V.2. *There exists C_1 and C_2 depending only on α such that, with probability larger than $1 - 1/n^3$, $|\mathcal{A}_l| \ge m - \max\{e^{-C_1\epsilon}m, C_2\alpha\}$, and $|\mathcal{A}_r| \ge n - \max\{e^{-C_1\epsilon}n, C_2\}$.*

For any $E \subseteq [m] \times [n]$ and $A = (A_l, A_r)$ with $A_l \subseteq [m]$, $A_r \subseteq [n]$, we define $M^{E,A}$ by setting to zero the entries of M that are not in E, those whose row index is not in A_l, and those whose column index not in A_r. Consider the event

$$\mathcal{H}(E,A) = \left\{\exists x,y : \left|\sum_L x_i M^{E,A}_{ij} y_j - \frac{\epsilon}{\sqrt{mn}}x^T M y\right| > C_1 M_{\max}\sqrt{\epsilon}\right\}, \tag{11}$$

where it is understood that x and y belong, respectively, to T_m and T_n. Note that $\widetilde{M}^E = M^{E,\mathcal{A}}$, and hence we want to bound $\mathbb{P}\{\mathcal{H}(E,\mathcal{A})\}$. We proceed as follows

$$\begin{aligned}\mathbb{P}\{\mathcal{H}(E,\mathcal{A})\} &= \sum_A \mathbb{P}\{\mathcal{H}(E,A),\, \mathcal{A}=A\} \\ &\leq \sum_{\substack{|A_l|\geq m(1-\delta),\\ |A_r|\geq n(1-\delta)}} \mathbb{P}\{\mathcal{H}(E,A),\, \mathcal{A}=A\} + \frac{1}{n^3} \\ &\leq 2^{(n+m)H(\delta)} \max_{\substack{|A_l|\geq m(1-\delta),\\ |A_r|\geq n(1-\delta)}} \mathbb{P}\{\mathcal{H}(E;A)\} + \frac{1}{n^3}, \quad (12)\end{aligned}$$

with $\delta \equiv \max\{e^{-C_1\epsilon}, C_2/n\}$ and $H(x)$ the entropy function.

We are now left with the task of bounding $\mathbb{P}\{\mathcal{H}(E;A)\}$ uniformly over A where $\mathcal{H}$ is defined as in Eq. (11). The key step consists in proving the following tail estimate

Lemma V.3. *Let $x \in S_m$, $y \in S_n$, $Z = \sum_{(i,j)\in L} x_i M_{ij}^{E,A} y_j - \frac{\epsilon}{\sqrt{mn}} x^T M y$, and assume $|A_l| \geq m(1-\delta)$, $|A_r| \geq n(1-\delta)$ with δ small enough. Then*

$$\mathbb{P}\left(Z > LM_{\max}\sqrt{\epsilon}\right) \leq \exp\Big\{ -\frac{n\alpha^{1/2}(L-3)}{2} \Big\}.$$

Proof: It is shown in [5] that $|\mathbb{E}[Z]| \leq 2M_{\max}\sqrt{\epsilon}$. For $A = (A_l, A_r)$, let M^A be the matrix obtained from M by setting to zero those entries whose row index is not in A_l, and those whose column index not in A_r. Define the potential contribution of the light couples a_{ij} and independent random variables Z_{ij} as

$$a_{ij} = \begin{cases} x_i M_{ij}^A y_j & \text{if } |x_i M_{ij}^A y_j| \leq M_{\max}\left(\epsilon/mn\right)^{1/2}, \\ 0 & \text{otherwise,}\end{cases}$$

$$Z_{ij} = \begin{cases} a_{i,j} & \text{w.p. } \epsilon/\sqrt{mn}, \\ 0 & \text{w.p. } 1-\epsilon/\sqrt{mn},\end{cases}$$

Let $Z_1 = \sum_{i,j} Z_{ij}$ so that $Z = Z_1 - \frac{\epsilon}{\sqrt{mn}} x^T M y$. Note that $\sum_{i,j} a_{ij}^2 \leq \sum_{i,j}\left(x_i M_{ij}^A y_j\right)^2 \leq M_{\max}^2$. Fix $\lambda = \sqrt{mn}/2M_{\max}\sqrt{\epsilon}$ so that $|\lambda a_{i,j}| \leq 1/2$, whence $e^{\lambda a_{ij}} - 1 \leq \lambda a_{ij} + 2(\lambda a_{ij})^2$. It then follows that

$$\begin{aligned}&\mathbb{E}[e^{\lambda Z}] = \\ &\exp\Big\{\frac{\epsilon}{\sqrt{mn}}\Big(\sum_{i,j}\lambda a_{i,j} + 2\sum_{i,j}(\lambda a_{i,j})^2\Big) - \frac{\lambda\,\epsilon}{\sqrt{mn}} x^T M y\Big\} \\ &\leq \exp\Big\{\lambda\mathbb{E}[Z] + \sqrt{mn}/2\Big\}.\end{aligned}$$

The thesis follows by Chernoff bound $\mathbb{P}(Z > a) \leq e^{-\lambda a}\mathbb{E}[e^{\lambda Z}]$ after simple calculus. ■

Note that $\mathbb{P}\left(-Z > LM_{\max}\sqrt{\epsilon}\right)$ can also be bounded analogously. We can now finish the upper bound on the light couples contribution. Consider the error event Eq. (11). A simple volume calculation shows that $|T_m| \leq (10/\Delta)^m$. We can therefore apply union bound over T_m and T_n to Eq. (12) to obtain

$$\mathbb{P}\{\mathcal{H}(E,\mathcal{A})\} \leq 2\left(\frac{20}{\Delta}\right)^{n+m} 2^{(n+m)H(\delta)} e^{-\frac{(C_1-3)\sqrt{\alpha}n}{2}} + \frac{1}{n^3},$$

If C_1 is a large enough constant, the first term is of order $e^{-\Theta(n)}$ (for, say, $\epsilon \geq r$) thus finishing the proof.

VI. Proof of Lemma IV.1

Recall the variational principle for the singular values.

$$\begin{aligned}\sigma_q &= \min_{H,\dim(H)=n-q+1}\ \max_{y\in H,||y||=1} ||\widetilde{M}^E y|| \quad (13)\\ &= \max_{H,\dim(H)=q}\ \min_{y\in H,||y||=1} ||\widetilde{M}^E y||. \quad (14)\end{aligned}$$

Here H is understood to be a linear subspace of $\mathbb{R}^n$.

Using Eq. (13) with H the orthogonal complement of $\mathrm{span}(v_1,\dots,v_{q-1})$, we have, by Lemma IV.2,

$$\begin{aligned}\sigma_q &\leq \max_{y\in H,||y||=1} \big|\big|\widetilde{M}^E y\big|\big| \\ &\leq \frac{\epsilon}{\sqrt{mn}}\left(\max_{y\in H,||y||=1}\big|\big|My\big|\big|\right) \\ &\quad + \max_{y\in H,||y||=||x||=1}\left|x^T\left(\widetilde{M}^E - \frac{\epsilon}{\sqrt{mn}}M\right)y\right| \\ &\leq \epsilon\Sigma_q + CM_{\max}\sqrt{\epsilon}\end{aligned}$$

The lower bound is proved analogously, by using Eq. (14) with $H = \mathrm{span}(v_1,\dots,v_q)$.

References

[1] "Netflix prize." [Online]. Available: http://www.netflixprize.com/

[2] M. W. Berry, Z. Drmać, and E. R. Jessup, "Matrices, vector spaces, and information retrieval," *SIAM Review*, vol. 41, no. 2, pp. 335–362, 1999.

[3] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," 2008, arxiv:0805.4471.

[4] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matr. Anal. Appl.*, vol. 20, pp. 303–353, 1999.

[5] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," January 2009, arXiv:0901.3150.

[6] ——, "Learning low rank matrices from $O(n)$ entries," in *Proc. of the Allerton Conf. on Commun., Control and Computing*, September 2008.

[7] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Stanford University, 2002.

[8] D. L. Donoho, "Compressed Sensing," *IEEE Trans. on Inform. Theory*, vol. 52, pp. 1289– 1306, 2006.

[9] E. J. Candes, J. K. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Inform. Theory*, vol. 52, pp. 489– 509, 2006.

[10] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum rank solutions of matrix equations via nuclear norm minimization," 2007, arxiv:0706.4138.

[11] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," 2008, arXiv:0810.3286.

[12] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," 2009, arXiv:0903.1476.

[13] A. Singer and M. Cucuringu, "Uniqueness of low-rank matrix completion by rigidity theory," January 2009, arXiv:0902.3846.

[14] E. J. Candès and Y. Plan, "Matrix completion with noise," 2009, arXiv:0903.3131.

[15] J. Friedman, J. Kahn, and E. Szemerédi, "On the second eigenvalue in random regular graphs," in *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*. Seattle, Washington, USA: ACM, may 1989, pp. 587–598.

[16] U. Feige and E. Ofek, "Spectral techniques applied to sparse random graphs," *Random Struct. Algorithms*, vol. 27, no. 2, pp. 251–275, 2005.