A Deterministic Theory of Low Rank Matrix Completion

Sourav Chatterjee

Abstract—The problem of completing a large low rank matrix using a subset of revealed entries has received much attention in the last ten years. The main result of this paper gives a necessary and sufficient condition, stated in the language of graph limit theory, for a sequence of matrix completion problems with arbitrary missing patterns to be asymptotically solvable. It is then shown that a small modification of the Candès–Recht nuclear norm minimization algorithm provides the required asymptotic solution whenever the sequence of problems is asymptotically solvable. The theory is fully deterministic, with no assumption of randomness. A number of open questions are listed.

Index Terms—Matrix completion, low rank matrix, graph limit, graphon.

I. Introduction

THE problem of reconstructing a large low rank matrix from a subset of revealed entries has attracted widespread attention in the statistics and machine learning literatures in the last ten years. For a recent survey of this vast body of work, see [15]. Notice that the problem itself is a problem in linear algebra, with nothing random in it. However, matrix completion in classical linear algebra is restricted to matrices with special structure, such as positive definite matrices [10].

In the literature on low rank matrix completion, randomness enters into the picture through the assumption that the set of missing entries is random. In most papers, the randomness is uniform over all subsets of a given size. This assumption, while unrealistic, allows researchers to prove many beautiful theorems. There are a handful of papers that strive to work with deterministic missing patterns or missing patterns that depend on the matrix, using spectral gap conditions [1], [9], rigidity theory [17], algebraic geometry [12] and other methods [6], [13], [16], [18]. These are discussed in some detail in Section V.

However, a complete characterization of missing patterns that allow approximate completion of large low rank matrices has remained an open question. The aim of this paper is to give such a characterization. Here 'approximate completion' means that the missing entries are required to be recovered

Fig. 1. A pattern of missing entries that cannot be completed (even approximately) even if the rank is known to be small.

approximately and not exactly (a precise definition is given later). The analogous criterion for exact recovery is left as an open question. Our result is an asymptotic statement involving limits; proving a non-asymptotic version of the result is also left as an open question.

Right away, it is important to note that not all patterns of revealed entries allow low rank matrix completion (even in an approximate sense), even if a substantial fraction of entries are revealed. For example, if we have a large square matrix of order n, and only the top $n/2$ rows are revealed, the matrix cannot be completed even if it is known to have rank 1 (see Figure 1). When we say 'cannot be completed', what we really mean is that there multiple very different ways to complete, even under the low rank assumption. This means that any particular completion cannot be a reliable estimate of the true matrix.

This example suggests that the set of revealed entries has to be in some sense 'dense' in the set of all entries for the matrix to be recoverable. However, one has to be cautious about this intuition. Consider a second counterexample: Let n be even, and consider an $n \times n$ matrix whose $(i, j)^{\text{th}}$ entry is revealed if and only if i and j have the same parity (that is, both even or both odd). This set of revealed entries looks sufficiently 'dense' (see Figure 2). Yet, we will now argue that recovery is not possible even if the rank of the matrix is as small as three.

To see this, note that the rows and columns can be relabeled such that the even numbered rows and columns in the original matrix are renumbered from 1 to $n/2$ and the odd numbered rows and columns are renumbered from $n/2 + 1$ to n. Then in this new arrangement of rows and columns, the $(i, j)^{\text{th}}$ entry is revealed if and only if either both i and j are between 1
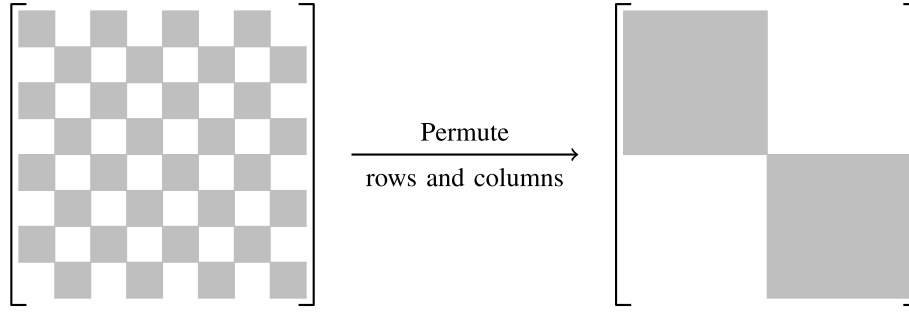
Fig. 2.   A pattern of missing entries that cannot be completed. Shaded regions denote available entries and white regions are missing. Although the pattern of available entries on the left looks 'dense', permuting rows and columns in a particular way gives the pattern on the right, which is clearly not completable (even approximately) even if the rank is known to be small.

and $n/2$, or both i and j are between $n/2+1$ and n. In other words, the matrix is a 2×2 block matrix with blocks of order $n/2 \times n/2$, where only the top-left and bottom-right blocks are revealed (again, see Figure 2). Clearly, the other two blocks cannot be recovered using this information if the rank is three or higher.

The problem with the above counterexample is that the rows and columns could be relabeled so that the pattern of revealed entries is no longer 'dense'. This suggests that for recoverability of low rank matrices, it is necessary that the pattern of revealed entries *remains 'dense' under any relabeling of rows and columns.*

It turns out that this condition is also sufficient. This is the main theorem of this paper (Theorem 2). The precise statement is given in the language of graph limit theory [14]. It is then proved that a modification of a popular method of low rank matrix completion by nuclear norm minimization [2]–[4] succeeds in approximately recovering the full matrix whenever the above condition holds (Theorem 3). In other words, this algorithm does the job whenever the job is doable.

The modification is as follows. The usual Candès–Recht algorithm finds the matrix with minimum nuclear norm among all matrices that agree with the unknown matrix on the set of revealed entries. In the modification, we assume that an upper bound L on the magnitudes of the entries of the unknown matrix is known to the user (which is usually true), and then find the matrix that minimizes the nuclear norm subject to the usual constraint, plus the constraint that the magnitudes of all entries are bounded by L. Note that this is still a convex optimization problem, just like the original algorithm.

The rest of the paper is organized as follows. Some necessary notations are introduced in Section II. Several definitions needed for stating our results in the language of graph limit theory are given in Section III. The main results are presented in Section IV. A brief discussion of the existing literature on matrix completion with non-uniform missing patterns is given in Section V. Some open problems are stated in Section VI. The remaining sections are devoted to proofs.

II. Notations

All our matrices will have real entries. We will denote the $(i, j)^{\text{th}}$ entry of a matrix A by a_{ij}, of B by b_{ij}, and so on. The transpose of a matrix A will be denoted by A^T, and the trace

by $\text{Tr}(A)$, and the rank by $\text{rank}(A)$. Vectors will be treated as matrices with one column.

Let A be an $m \times n$ matrix. We will have occasions to use the following matrix norms. The *Frobenius norm* of A is defined as

$$\|A\|_F := \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}.$$

More frequently, we will use the following *averaged* version of the Frobenius norm:

$$\|A\|_{\overline{F}} := \frac{\|A\|_F}{\sqrt{mn}} = \left(\frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}.$$

For us, the average Frobenius norm will be more useful than the usual Frobenius norm because it is a measure of the size of a typical entry of A.

If $\sigma_1, \ldots, \sigma_r$ are the non-zero singular values of A, the *nuclear norm* of A is defined as

$$\|A\|_* := \sum_{i=1}^{r} \sigma_i.$$

The ℓ^∞ *norm* of A is simply

$$\|A\|_\infty := \max_{i,j} |a_{ij}|.$$

We will also use a somewhat non-standard matrix norm, called the *cut norm*, defined as

$$\|A\|_\square := \frac{1}{mn} \max\{|x^T A y| : x \in \mathbb{R}^m, y \in \mathbb{R}^n,$$
$$\|x\|_\infty \le 1, \|y\|_\infty \le 1\}.$$

In the usual definition of the cut norm for matrices, the maximum is not divided by mn. We divide by mn because it will be more convenient for us to work with this version, and also because this is the custom in graph limit theory.

For each k, let S_k be the group of all permutations of $\{1, \ldots, k\}$. For $\pi \in S_m$ and $\tau \in S_n$, let $A^{\pi,\tau}$ be the matrix whose $(i, j)^{\text{th}}$ entry is $a_{\pi(i)\tau(j)}$. The cut norm is used to define the *cut distance* between two $m \times n$ matrices A and B as

$$\delta_\square(A, B) := \min_{\pi \in S_m, \tau \in S_n} \|A^{\pi,\tau} - B\|_\square. \tag{1}$$

We will say that a matrix is a *binary matrix* if each of its entries is either 0 or 1. We will use binary matrices to denote the locations of revealed entries in matrix completion problems.

If A and B are two $m \times n$ matrices, the *Hadamard product* of A and B, denoted by $A \circ B$, is the $m \times n$ matrix whose $(i,j)^{\text{th}}$ entry is $a_{ij}b_{ij}$. Hadamard products will be useful for us in the following way. If A is a matrix which is partially revealed, and P is a binary matrix indicating the locations of the revealed entries, then $A \circ P$ is the matrix whose entries equal the entries of A wherever they are revealed, and zero elsewhere.

III. DEFINITIONS

As mentioned earlier, certain patterns of revealed entries may not suffice for approximately recovering the full matrix, whereas other patterns may suffice. While this makes intuitive sense, we need to give a precise mathematical definition of the notion of recoverability before proceeding further with this. Roughly speaking, approximate recoverability should mean that if two low rank matrices are approximately equal on the revealed entries, they should also be approximately equal everywhere. To make this fully precise, we need to state it in terms of sequences of matrices rather than a single matrix.

Definition 1: Let $\{P_k\}_{k \geq 1}$ be a sequence of binary matrices. We will say that this sequence *admits stable recovery of low rank matrices* if it has the following property. Take any two sequences of matrices $\{A_k\}_{k \geq 1}$ and $\{B_k\}_{k \geq 1}$, where A_k and B_k have the same dimensions as P_k. Suppose that there are numbers K and L such that $\operatorname{rank}(A_k)$ and $\operatorname{rank}(B_k)$ are bounded by K and $\|A_k\|_\infty$ and $\|B_k\|_\infty$ are bounded by L for each k. Then for any $\varepsilon > 0$ there is some $\delta > 0$, depending only on ε, K and L, such that if $\limsup_{k \to \infty} \|(A_k - B_k) \circ P_k\|_{\overline{F}} \leq \delta$, then $\limsup_{k \to \infty} \|A_k - B_k\|_{\overline{F}} \leq \varepsilon$.

The word 'stable' is added in the above definition to emphasize that we only need approximate equality of the revealed entries, rather than exact equality.

To understand the essence of the above definition, it is probably helpful to revisit the counterexample mentioned earlier. For each k, let P_k be the $k \times k$ binary matrix whose entries are 1 in the first $[k/2]$ rows, and 0 elsewhere. Let A_k be the $k \times k$ matrix of all zeros, and B_k be the $k \times k$ matrix whose entries are 0 in the top $[k/2]$ rows and 1 elsewhere. Then $\|A_k\|_\infty$ and $\|B_k\|_\infty$ are bounded by 1 for all k, and $\operatorname{rank}(A_k)$ and $\operatorname{rank}(B_k)$ are bounded by 1 for all k. Now clearly $\lim_{k \to \infty} \|(A_k - B_k) \circ P_k\|_{\overline{F}} = 0$, but a simple calculation shows that

$$\lim_{k \to \infty} \|A_k - B_k\|_{\overline{F}} = \frac{1}{\sqrt{2}}.$$

Thus, the sequence $\{P_k\}_{k \geq 1}$ does not admit stable recovery of low rank matrices.

To verify that a sequence $\{P_k\}_{k \geq 1}$ admits stable recovery of low rank matrices according to Definition 1, one needs to verify the stated condition for all sequences $\{A_k\}_{k \geq 1}$ and $\{B_k\}_{k \geq 1}$. It would however be much more desirable to have an equivalent criterion in terms of some intrinsic property of the sequence $\{P_k\}_{k \geq 1}$. The main result of this paper gives such a criterion. To state this result, we need to introduce some further definitions.

In graph limit theory [14], a *graphon* is a Borel measurable function from $[0,1]^2$ into $[0,1]$, which is symmetric in its

arguments. Since we are dealing with matrices that need not be symmetric, we need to generalize this definition by dropping the symmetry condition.

Definition 2: An *asymmetric graphon* is a Borel measurable function from $[0,1]^2$ into $[0,1]$.

If W is an asymmetric graphon and m and n are two positive integers, we define the $m \times n$ discrete approximation of W to be the $m \times n$ matrix $W_{m,n}$, whose $(i,j)^{\text{th}}$ entry is the average value of W in the rectangle $[\frac{i-1}{m}, \frac{i}{m}] \times [\frac{j-1}{n}, \frac{j}{n}]$, that is,

$$mn \int_{(i-1)/m}^{i/m} \int_{(j-1)/n}^{j/n} W(x,y)dydx.$$

If A is an $m \times n$ matrix and W is an asymmetric graphon, we define the cut distance between A and W to be

$$\delta_\square(A, W) := \delta_\square(A, W_{m,n}),$$

where the right side is as defined in equation (1).

Definition 3: We will say that a sequence of matrices $\{A_k\}_{k \geq 1}$ converges to an asymmetric graphon W if $\delta_\square(A_k, W) \to 0$ as $k \to \infty$.

Note that the limit defined in the above sense may not be unique. The same sequence may converge to many different limits. In graph limit theory, all of these different limits are considered to be equivalent by defining an equivalence relation on the space of graphons. It is possible to do a similar thing for asymmetric graphons, but that is not needed for this paper.

We will use asymmetric graphons to represent limits of binary matrices. Not every sequence has a limit, but subsequential limits always exist.

Theorem 1: Any sequence of binary matrices with dimensions tending to infinity has a subsequence that converges to an asymmetric graphon.

The above theorem is the asymmetric analog of a fundamental compactness theorem in graph limit theory [14, Theorem 9.23]. It is probable that the asymmetric version already exists in the literature, but since the proof is not difficult, it is presented in Section X.

IV. RESULTS

Our main objective is to give a necessary and sufficient condition for a sequence of binary matrices to admit stable recovery of low rank matrices. Because of Theorem 1, it suffices to only consider convergent sequences.

Theorem 2: A sequence of binary matrices with dimensions tending to infinity and converging to an asymmetric graphon W admits stable recovery of low rank matrices (in the sense of Definition 1) if and only if W is nonzero almost everywhere.

To understand this result, first consider the familiar case of entries missing at random. Suppose that each entry is revealed with probability p, independently of each other. Then the corresponding sequence of binary matrices converges to the graphon that is identically equal to p on $[0,1]^2$. If $p > 0$, Theorem 2 tells us that this sequence of revelation patterns admits stable recovery of low rank matrices. On the other hand, consider our running counterexample, where only the top half of the rows are revealed. The corresponding sequence of binary matrices converges to the graphon that is 1 in $[0, 1/2] \times [0,1]$

and 0 in $(1/2, 1] \times [0, 1]$. Therefore this sequence does not admit stable recovery of low rank matrices, as we observed before.

At this point the reader may be slightly puzzled by the fact that Theorem 2 implies that recovery is impossible if the set of revealed entries is sparse (because then the limit graphon is identically zero), whereas there are many existing results about recoverability of low rank matrices from a sparse set of revealed entries. The reason is that we are not assuming randomness and at the same time demanding that the recovery is 'stable'. Suppose that most entries are the same for two matrices, but the entries that differ are the only ones that are revealed. Then there is no way to tell that the matrices are mostly the same. Thus, stable recovery is impossible from a small set of revealed entries if there is no assumption of randomness.

Theorem 2 succeeds in giving an intrinsic characterization of recoverability in terms of the locations of revealed entries. However, it does not tell us how to actually recover a matrix from a set of revealed entries when recovery is possible. Fortunately, it turns out that this is doable by a small modification of an algorithm that is already used in practice, namely, the Candès–Recht algorithm for matrix completion by nuclear norm minimization [2]–[4]. The Candès–Recht estimator of a partially revealed matrix A is the matrix with minimum nuclear norm among all matrices that agree with A at the revealed entries. The modified estimator is the following.

Definition 4: Let A be a matrix whose entries are partially revealed. Suppose that $\|A\|_\infty \le L$ for some known constant L. We define the *modified Candès–Recht estimator* of A as the matrix that minimizes nuclear norm among all B that agree with A at the revealed entries and satisfy $\|B\|_\infty \le L$.

The assumption of a known upper bound on the ℓ^∞ norm is not unrealistic. Usually such upper bounds are known, for example in recommender systems. The modified estimator is the solution of a convex optimization problem, just like the original estimator, and should therefore be computable on a computer if the dimensions are not too large. The following theorem shows that this algorithm is able to approximately recover the full matrix whenever the pattern of revealed entries allows stable recovery.

Theorem 3: Let $\{P_k\}_{k \ge 1}$ be a sequence of binary matrices with dimensions tending to infinity that admits stable recovery of low rank matrices. Let $\{A_k\}_{k \ge 1}$ be a sequence of matrices such that for each k, A_k has the same dimensions as P_k. Suppose that $\text{rank}(A_k)$ and $\|A_k\|_\infty$ are uniformly bounded over k. Let $\widehat{A}_k$ be the modified Candès–Recht estimate of A_k (as defined in Definition 4) when the locations of the revealed entries are defined by P_k. Then $\lim_{k \to \infty} \|\widehat{A}_k - A_k\|_{\overline{F}} = 0$.

The modified Candès–Recht estimator, just like the original estimator, will run into computational cost issues for very large matrices. It would be interesting to figure out if there is a faster algorithm (for example, by some kind of singular value thresholding [5], [8], [11]) that also has the above 'universal recovery' feature.

Another interesting and important problem is to develop an analog of the above theory when the set of revealed entries

is sparse. As noted before, the problem is unsolvable in this setting if we demand that the recovery be stable. However, dropping the stability requirement may render it possible to recover the full matrix from a sparse set of revealed entries even in the absence of randomness. In particular, Theorem 3 may have an extension to the sparse setting under appropriate assumptions. The methods of this paper would need to be significantly extended to make this possible.

This concludes the statements of results. The proofs are organized as follows. The proof of Theorem 2 is divided between Sections VII and VIII. Theorem 3 is proved in Section IX, and Theorem 1 is proved in Section X.

V. SOME RELATED LITERATURE

A small number of papers in the literature investigate the problem of matrix completion when the entries are not missing uniformly at random. As mentioned earlier, this list is minuscule in comparison to the vast body of literature on matrix completion under the assumption of missing uniformly at random. The following is a non-exhaustive list of some of the notable contributions.

A general recovery method was developed in [1] when the pattern of revealed entries is the adjacency matrix of a bipartite graph with a large spectral gap. A similar question was investigated in a slightly different setting in [9]. Inhomogeneous — but still independent and random — patterns of missing entries were studied in [6].

The very interesting paper [17] applied rigidity theory to understand whether a partially revealed low rank matrix is completable or not. However, this paper did not give an algorithm for completion when completion is possible. A comparison of the criterion from [17] with our Theorem 2 is an interesting question that merits further investigation.

An attempt at giving a criterion for completability using algebraic geometry was made in [12]. This paper has a criterion for recoverability of specific entries of the matrix. But again, a recovery algorithm was not given. Algebraic criteria have also been investigated in other recent papers, such as [16].

A different approach was taken in [13], which proposed a new way of interpreting the quality of the output of a given matrix completion algorithm under arbitrary patterns of missing entries.

The idea of using the EM algorithm for matrix completion under data-dependent missing patterns was recently studied in [18].

The main advantage of our results over most of the papers mentioned above is that we give a condition that is both necessary and sufficient for completability, and also demonstrate that a small modification of a popular algorithm can do the job when it is doable. The main disadvantage, on the other hand, is that our results are of an asymptotic nature. Developing non-asymptotic versions is an important goal. This is further discussed in the next section.

VI. OPEN PROBLEMS

The results of this paper leave a lot of questions unanswered. The following is a partial list.

1) The definition of 'stable recovery' entails that the revealed entries are only approximately equal to the corresponding entries of the unknown matrix. What if we drop this condition and assume that the revealed entries are exactly equal to the true entries? How should the theory be modified?

2) Developing non-asymptotic versions of Theorems 2 and 3 is extremely desirable. Note that it is not quite clear what should be the proper non-asymptotic statements that one can aspire to prove. A precise formulation of the non-asymptotic problem is itself an open question. The non-asymptotic formulation is needed for dealing with sparse recovery problems, for the following reason. The theorems of this paper have meaningful implications when the fraction of revealed entries remains fixed as the size of the matrix goes to infinity. To properly understand the level of sparsity allowable for a matrix of a given size, one needs a non-asymptotic result.

3) It is not clear if the Candès–Recht algorithm indeed needs to be modified, or if the original version is good enough for Theorem 3. We believe that the modification is necessary, but we do not have a counterexample to show that the original algorithm will not work.

4) As mentioned before, the Candès–Recht algorithm is rather slow for very large matrices. Is there a faster algorithm that can take its place in Theorem 3?

VII. Towards the Proof of Theorem 2

The goal of this section is to prove a quantitative result that underlies the proof of Theorem 2. We need to prove a number of lemmas before arriving at this theorem.

Lemma 1: Let X be an $m \times n$ matrix with $\|X\|_\infty \le 1$ and singular value decomposition

$$X = \sum_{i=1}^{k} \sigma_i u_i v_i^T.$$

Then for each i,

$$\sigma_i \le \sqrt{mn}, \quad \|u_i\|_\infty \le \frac{\sqrt{n}}{\sigma_i} \quad \text{and} \quad \|v_i\|_\infty \le \frac{\sqrt{m}}{\sigma_i}.$$

Proof: Let u_{ij} denote the jth component of u_i. Since $Xv_i = \sigma_i u_i$ and $\|X\|_\infty \le 1$, we get

$$\sigma_i |u_{ij}| \le \sum_{l=1}^{n} |x_{jl} v_{il}| \le \sum_{l=1}^{n} |v_{il}| \le \left(n \sum_{l=1}^{n} v_{il}^2 \right)^{1/2} = \sqrt{n}.$$

Dividing throughout by σ_i and maximizing over j, we get the required bound for $\|u_i\|_\infty$. The bound for $\|v_i\|_\infty$ is obtained similarly. For the bound on σ_i, notice that since $\sum_j u_{ij}^2 = 1$ there is some j such that $|u_{ij}| \ge m^{-1/2}$, and use this information in the above display. $\square$

Recall that a matrix is called a *block matrix* if its entries are constant in rectangular blocks — in other words, if the matrix can be expressed as an array of constant matrices. We will say that two matrices A and B have a simultaneous block structure if they are both block matrices and the rows and columns defining the blocks are the same. Note that block structures may not be uniquely defined, but that will not be a problem for us.

Lemma 2: Let X and Y be $m \times n$ matrices with $\|X\|_\infty \le 1$ and $\|Y\|_\infty \le 1$. Let $q \ge 1$ be a number such that $\|X\|_*$ and $\|Y\|_*$ are bounded by $q\sqrt{mn}$. Take any $\varepsilon \in (0, 1)$. Then, there exist $m \times n$ matrices A and B with a simultaneous block structure with at most $(20000 \, q^6 \varepsilon^{-10})^{5q^2 \varepsilon^{-2}}$ blocks, and permutations $\pi \in S_m$ and $\tau \in S_n$, such that $\|X^{\pi,\tau} - A\|_{\overline{F}} \le \varepsilon$ and $\|Y^{\pi,\tau} - B\|_{\overline{F}} \le \varepsilon$. Moreover, it can be ensured that $\|A\|_\infty \le 1$ and $\|B\|_\infty \le 1$.

Proof: Fix some $\varepsilon > 0$. Let δ, γ and η be three other positive numbers, to be chosen later. Let

$$X = \sum_{i=1}^{k} \sigma_i u_i v_i^T, \quad Y = \sum_{i=1}^{l} \lambda_i w_i z_i^T$$

be the singular value decompositions of X of Y, with $\sigma_1 \ge \cdots \ge \sigma_k > 0$ and $\lambda_1 \ge \cdots \ge \lambda_l > 0$. Choose two numbers k_1 and l_1 such that $\sigma_{k_1} > \delta \ge \sigma_{k_1+1}$ and $\lambda_{l_1} > \delta \ge \lambda_{l_1+1}$. If $\sigma_i \le \delta$ for all i, let $k_1 = 0$, and if $\sigma_i > \delta$ for all i, let $k_1 = k$. Similarly, if $\lambda_i \le \delta$ for all i, let $l_1 = 0$, and if $\lambda_i > \delta$ for all i, let $l_1 = l$. Let

$$X_1 := \sum_{i=1}^{k_1} \sigma_i u_i v_i^T, \quad Y_1 := \sum_{i=1}^{l_1} \lambda_i w_i z_i^T.$$

Then by the definition of k_1,

$$\|X - X_1\|_{\overline{F}}^2 = \frac{1}{mn} \sum_{i=k_1+1}^{k} \sigma_i^2$$

$$\le \frac{\delta}{mn} \sum_{i=1}^{k} \sigma_i = \frac{\|X\|_* \delta}{mn} \le \frac{q\delta}{\sqrt{mn}}. \tag{2}$$

Similarly, the same bound holds for $\|Y - Y_1\|_{\overline{F}}^2$.

For $1 \le i \le k$ and $1 \le a \le m$, let u_{ia} denote the ath component of u_i. Define $\widetilde{u}_{ia}$ to be the integer multiple of γ that is closest to u_{ia} under the constraint that $|\widetilde{u}_{ia}| \le |u_{ia}|$. Then note that $|u_{ia} - \widetilde{u}_{ia}| \le \gamma$. Let $\widetilde{u}_i$ be the vector whose a^{th} component is $\widetilde{u}_{ia}$. Let $\widetilde{w}_i$ be defined similarly. Define $\widetilde{v}_i$ and $\widetilde{z}_i$ the same way, but using η instead of γ. Let

$$\widetilde{X}_1 := \sum_{i=1}^{k_1} \sigma_i \widetilde{u}_i \widetilde{v}_i^T, \quad \widetilde{Y}_1 := \sum_{i=1}^{l_1} \lambda_i \widetilde{w}_i \widetilde{z}_i^T.$$

Now take any $1 \le i \le k_1$. By Lemma 1 and the choice of k_1, we have

$$\|u_i\|_\infty < \frac{\sqrt{n}}{\delta}.$$

Therefore for any $1 \le a \le m$, the set of possible values of $\widetilde{u}_{ia}$ has size at most

$$\frac{2\sqrt{n}}{\delta\gamma} + 1 \le \frac{4\sqrt{n}}{\delta\gamma},$$

where the inequality was obtained under the assumption that

$$\frac{2\sqrt{n}}{\delta\gamma} \ge 1. \tag{3}$$

We will later choose δ and γ such that this assumption is valid. We can give similar bounds on the sizes of the sets of possible values of the components of $\widetilde{v}_i$, $\widetilde{w}_i$ and $\widetilde{z}_i$.

Declare that two rows a and a' are 'equivalent' if $\widetilde{u}_{ia} = \widetilde{u}_{ia'}$ and $\widetilde{w}_{i'a} = \widetilde{w}_{i'a'}$ for all $1 \le i \le k_1$ and $1 \le i' \le l_1$. Similarly declare that two columns b and b' are equivalent if $\widetilde{v}_{ib} = \widetilde{v}_{ib'}$ and $\widetilde{z}_{i'b} = \widetilde{z}_{i'b'}$ for all $1 \le i \le k_1$ and $1 \le i' \le l_1$. Clearly, these define equivalence relations. By the previous paragraph, there are at most $(4\sqrt{n}/(\delta\gamma))^{k_1+l_1}$ equivalence classes of rows, and at most $(4\sqrt{m}/(\delta\eta))^{k_1+l_1}$ equivalence classes of columns.

Let π be a permutation of the rows that 'clumps together' equivalent rows, and let τ be a permutation of the columns that clumps together equivalent columns. Then it is clear that $\widetilde{X}_1^{\pi,\tau}$ and $\widetilde{Y}_1^{\pi,\tau}$ are block matrices. By the previous paragraph, the number of blocks is at most $(16\sqrt{mn}/(\delta^2\gamma\eta))^{k_1+l_1}$.

Now note that

$$k_1\sigma_{k_1} \le \sum_{i=1}^{k} \sigma_i = \|X\|_* \le q\sqrt{mn}.$$

But $\sigma_{k_1} > \delta$. Thus,

$$k_1 \le \frac{q\sqrt{mn}}{\delta}.$$

Similarly, l_1 is also bounded by the same quantity. Thus, the number of blocks is at most

$$b := \left(\frac{16\sqrt{mn}}{\delta^2\gamma\eta}\right)^{2q\sqrt{mn}/\delta}.$$

Now notice that by Lemma 1 and the definition of $\widetilde{X}_1$,

$$\|X_1 - \widetilde{X}_1\|_{\overline{F}} \le \|X_1 - \widetilde{X}_1\|_\infty$$
$$\le \sum_{i=1}^{k_1} \sigma_i(\|u_i - \widetilde{u}_i\|_\infty\|v_i\|_\infty + \|\widetilde{u}_i\|_\infty\|v_i - \widetilde{v}_i\|_\infty)$$
$$\le q\sqrt{mn}\left(\frac{\sqrt{m}\gamma}{\delta} + \frac{\sqrt{n}\eta}{\delta}\right).$$

By a similar argument, the same bound holds for $\|Y_1 - \widetilde{Y}_1\|_{\overline{F}}$. Combining with (2), we see that if $A := \widetilde{X}_1^{\pi,\tau}$ and $B := \widetilde{Y}_1^{\pi,\tau}$, then A and B have a simultaneous block structure with at most b blocks, and $\|X^{\pi,\tau} - A\|_{\overline{F}}$ and $\|Y^{\pi,\tau} - B\|_{\overline{F}}$ are bounded by

$$\frac{\sqrt{q\delta}}{(mn)^{1/4}} + q\sqrt{mn}\left(\frac{\sqrt{m}\gamma}{\delta} + \frac{\sqrt{n}\eta}{\delta}\right).$$

Now take any $\alpha, \beta > 0$ and define

$$\delta := \alpha\sqrt{mn}, \quad \gamma := \frac{\beta}{\sqrt{m}}, \quad \eta := \frac{\beta}{\sqrt{n}}.$$

Plugging these values into the previous display gives

$$\sqrt{q\alpha} + \frac{2q\beta}{\alpha}.$$

For a given β, the above quantity is minimized by taking $\alpha = 2^{4/3}q^{1/3}\beta^{2/3}$, and the minimum value is $(2^{2/3} + 2^{-1/3})q^{2/3}\beta^{1/3}$. Choose β to make this equal to ε, which ensures that $\|X^{\pi,\tau} - A\|_{\overline{F}}$ and $\|Y^{\pi,\tau} - B\|_{\overline{F}}$ are bounded by ε. With these choices of α and β, an easy calculation gives

$$b = (16\alpha^{-2}\beta^{-2})^{2q/\alpha} \le (20000 \, q^6\varepsilon^{-10})^{5q^2\varepsilon^{-2}}.$$

Also, it is easy to check (using $q \ge 1$ and $\varepsilon \in (0,1)$) that with these choices of α and β, the inequality (3) holds. Thus,

the proof is complete except that we have not ensured that $\|A\|_\infty \le 1$ and $\|B\|_\infty \le 1$ in our construction. To force this, just take any element of either matrix; if it is bigger than 1, replace it by 1; if it is less than -1, replace it by -1. This retains the block structures of the matrices, and it cannot increase $|x_{\pi(i)\tau(j)} - a_{ij}|$ or $|y_{\pi(i)\tau(j)} - b_{ij}|$ for any i, j since $|x_{\pi(i)\tau(j)}| \le 1$ and $|y_{\pi(i)\tau(j)}| \le 1$. $\square$

Lemma 3: Let A and B be $m \times n$ matrices with a simultaneous block structure. Let b be the number of blocks. Let P and Q be $m \times n$ matrices such that P is binary and the entries of Q are all in $[0,1]$. Then

$$\|(A - B) \circ Q\|_{\overline{F}} \le \|(A - B) \circ P\|_{\overline{F}} + \sqrt{b\|P - Q\|_\square} \, \|A - B\|_\infty.$$

Proof: Let each block be represented by the set of pairs of indices (i,j) that belong to the block. Let $\mathcal{D}$ be the set of all blocks. Take any block $D \in \mathcal{D}$. By the definition the cut norm,

$$\frac{1}{mn}\left|\sum_{(i,j)\in D} (p_{ij} - q_{ij})\right| \le \|P - Q\|_\square. \tag{4}$$

Recall that a_{ij} is the same for all (i,j) in a block, and the same holds for b_{ij}. Let $a(D)$ and $b(D)$ denote the values of a_{ij} and b_{ij} in a block D. Since $q_{ij} \in [0,1]$ for all i,j,

$$\|(A - B) \circ Q\|_{\overline{F}}^2 = \frac{1}{mn}\sum_{i,j}(a_{ij} - b_{ij})^2 \, q_{ij}^2$$
$$\le \frac{1}{mn}\sum_{i,j}(a_{ij} - b_{ij})^2 \, q_{ij}$$
$$= \sum_{D\in\mathcal{D}}(a(D) - b(D))^2\left(\frac{1}{mn}\sum_{(i,j)\in D} q_{ij}\right).$$

Therefore by (4),

$$\|(A - B) \circ Q\|_{\overline{F}}^2$$
$$\le \sum_{D\in\mathcal{D}}(a(D) - b(D))^2\left(\frac{1}{mn}\sum_{(i,j)\in D} p_{ij} + \|P - Q\|_\square\right).$$

Since P is binary, $p_{ij} = p_{ij}^2$ for all i,j. Thus, we get

$$\|(A - B) \circ Q\|_{\overline{F}}^2 \le \frac{1}{mn}\sum_{D\in\mathcal{D}}\sum_{(i,j)\in D}(a_{ij} - b_{ij})^2 \, p_{ij}^2$$
$$+ \|P - Q\|_\square\|A - B\|_\infty^2|\mathcal{D}|$$
$$= \|(A - B) \circ P\|_{\overline{F}}^2$$
$$+ \|P - Q\|_\square \, \|A - B\|_\infty^2 \, b.$$

The proof is not completed by applying the inequality $\sqrt{x+y} \le \sqrt{x} + \sqrt{y}$ to the right side. $\square$

We are now ready to prove the main result of this section. The result roughly says the following. Let X and Y be matrices with relatively small nuclear norms (of the same order as that for low rank matrices). Let P be a binary matrix and Q be a matrix with entries in $[0,1]$, such that P is close to Q in the cut norm. Then, the closeness of $X \circ P$ to $Y \circ P$ in average Frobenius norm implies the closeness of $X \circ Q$ to $Y \circ Q$ in average Frobenius norm.

Theorem 4: Let X and Y be $m \times n$ matrices with ℓ^∞ norms bounded by 1. Let q be a number such that the nuclear norms of X and Y are bounded by $q\sqrt{mn}$. Let P and Q be $m \times n$ matrices such that P is binary and the entries of Q are all in $[0,1]$. Then

$$\|(X-Y) \circ Q\|_{\overline{F}} \leq \|(X-Y) \circ P\|_{\overline{F}}$$
$$+ C(q)\sqrt{\frac{\log(-\log\|P-Q\|_{\square})}{-\log\|P-Q\|_{\square}}},$$

where $C(q)$ depends only on q.

Proof: Without loss of generality, assume that $q \geq 1$. Take any $\varepsilon > 0$. Let A, B, π and τ be as in Lemma 2. Let

$$b := (20000\, q^6 \varepsilon^{-10})^{5q^2 \varepsilon^{-2}}$$

be the upper bound on the number of blocks given by Lemma 2. Note that

$$\|(X-Y)\circ Q\|_{\overline{F}} = \|(X^{\pi,\tau} - Y^{\pi,\tau}) \circ Q^{\pi,\tau}\|_{\overline{F}}$$
$$\leq \|(X^{\pi,\tau} - A) \circ Q^{\pi,\tau}\|_{\overline{F}} + \|(A-B)\circ Q^{\pi,\tau}\|_{\overline{F}}$$
$$+ \|(B - Y^{\pi,\tau}) \circ Q^{\pi,\tau}\|_{\overline{F}}$$
$$\leq \|X^{\pi,\tau} - A\|_{\overline{F}} + \|(A-B)\circ Q^{\pi,\tau}\|_{\overline{F}} + \|B - Y^{\pi,\tau}\|_{\overline{F}}$$
$$\leq 2\varepsilon + \|(A-B)\circ Q^{\pi,\tau}\|_{\overline{F}}.$$

By Lemma 3,

$$\|(A-B)\circ Q^{\pi,\tau}\|_{\overline{F}}$$
$$\leq \|(A-B)\circ P^{\pi,\tau}\|_{\overline{F}} + \sqrt{b\|P^{\pi,\tau} - Q^{\pi,\tau}\|_{\square}}\,\|A-B\|_{\infty}$$
$$\leq \|(A-B)\circ P^{\pi,\tau}\|_{\overline{F}} + 2\sqrt{b\|P-Q\|_{\square}}.$$

But

$$\|(A-B)\circ P^{\pi,\tau}\|_{\overline{F}}$$
$$\leq \|(A - X^{\pi,\tau})\circ P^{\pi,\tau}\|_{\overline{F}} + \|(X^{\pi,\tau} - Y^{\pi,\tau})\circ P^{\pi,\tau}\|_{\overline{F}}$$
$$+ \|(Y^{\pi,\tau} - B)\circ P^{\pi,\tau}\|_{\overline{F}}$$
$$\leq \|A - X^{\pi,\tau}\|_{\overline{F}} + \|(X-Y)\circ P\|_{\overline{F}} + \|Y^{\pi,\tau} - B\|_{\overline{F}}$$
$$\leq 2\varepsilon + \|(X-Y)\circ P\|_{\overline{F}}.$$

Adding up, we get

$$\|(X-Y)\circ Q\|_{\overline{F}} \leq \|(X-Y)\circ P\|_{\overline{F}}$$
$$+ 4\varepsilon + 2\sqrt{b\|P-Q\|_{\square}}$$
$$\leq \|(X-Y)\circ P\|_{\overline{F}} + 4\varepsilon$$
$$+ (Cq^6 \varepsilon^{-10})^{3q^2 \varepsilon^{-2}} \sqrt{\|P-Q\|_{\square}},$$

where C is a universal constant. This bound holds for $\varepsilon \in (0,1)$, but it also holds for $\varepsilon \geq 1$ due to the presence of the 4ε term. The required bound is now obtained by choosing

$$\varepsilon = C(q)\sqrt{\frac{\log(-\log\|P-Q\|_{\square})}{-\log\|P-Q\|_{\square}}}$$

for some sufficiently large constant $C(q)$ that depends only on q. □

VIII. PROOF OF THEOREM 2

Let $\{P_k\}_{k\geq 1}$ be a sequence of binary matrices converging to a graphon W. Suppose that W is nonzero everywhere. Let m_k and n_k be the number of rows and the number columns in P_k. Suppose that m_k and n_k tend to infinity as $k \to \infty$. We will first prove the following generalization of the 'if' part of Theorem 2.

Theorem 5: Let P_k and W be as above. Take any two sequences of $m_k \times n_k$ matrices $\{A_k\}_{k\geq 1}$ and $\{B_k\}_{k\geq 1}$. Suppose that there are numbers q and L such that $\|A_k\|_*$ and $\|B_k\|_*$ are bounded by $q\sqrt{m_k n_k}$ and $\|A_k\|_\infty$ and $\|B_k\|_\infty$ are bounded by L for each k. Then for any $\varepsilon > 0$ there is some $\delta > 0$, depending only on ε, q and L, such that if $\limsup_{k\to\infty} \|(A_k - B_k)\circ P_k\|_{\overline{F}} \leq \delta$, then $\limsup_{k\to\infty}\|A_k - B_k\|_{\overline{F}} \leq \varepsilon$.

Proof: For simplicity of notation, let us denote the matrix W_{m_k,n_k} by $W^{(k)}$. The convergence of P_k to W means that for each k, there are permutations $\pi_k \in S_{m_k}$ and $\tau_k \in S_{n_k}$ such that

$$\lim_{k\to\infty}\|P_k^{\pi_k,\tau_k} - W^{(k)}\|_{\square} = 0.$$

Rearranging the rows and columns of P_k, A_k and B_k, we may assume without loss of generality that π_k and τ_k are the identity permutations, so that

$$\lim_{k\to\infty}\|P_k - W^{(k)}\|_{\square} = 0. \tag{5}$$

Let $\delta > 0$ be a number such that

$$\limsup_{k\to\infty}\|(A_k - B_k)\circ P_k\|_{\overline{F}} \leq \delta. \tag{6}$$

Without loss of generality, $L = 1$. Then by (5), (6) and Theorem 4,

$$\limsup_{k\to\infty}\|(A_k - B_k)\circ W^{(k)}\|_{\overline{F}} \leq \delta. \tag{7}$$

Now take any $\eta \in (0,1)$. Define two functions $f, g : [0,1] \to [0,1]$ as

$$f(x) := \begin{cases} \eta & \text{if } x \leq \eta, \\ x & \text{if } x > \eta, \end{cases}$$

and $g(x) := (f(x) - x)/\eta$. Let $U^{(k)}$ be the matrix whose $(i,j)^{\text{th}}$ element is $f(w_{ij}^{(k)})$ and let $V^{(k)}$ be the matrix whose $(i,j)^{\text{th}}$ element is $g(w_{ij}^{(k)})$. Since $f(x) \geq \eta$ for all x,

$$\|A_k - B_k\|_{\overline{F}} \leq \frac{1}{\eta}\|(A_k - B_k)\circ U^{(k)}\|_{\overline{F}}$$
$$\leq \frac{1}{\eta}\|(A_k - B_k)\circ (U^{(k)} - W^{(k)})\|_{\overline{F}}$$
$$+ \frac{1}{\eta}\|(A_k - B_k)\circ W^{(k)}\|_{\overline{F}}$$
$$= \|(A_k - B_k)\circ V^{(k)}\|_{\overline{F}}$$
$$+ \frac{1}{\eta}\|(A_k - B_k)\circ W^{(k)}\|_{\overline{F}}.$$

Therefore by (7),

$$\limsup_{k\to\infty}\|A_k - B_k\|_{\overline{F}}$$
$$\leq \limsup_{k\to\infty}\|(A_k - B_k)\circ V^{(k)}\|_{\overline{F}} + \frac{\delta}{\eta}. \tag{8}$$

Since $L = 1$,

$$\|(A_k - B_k) \circ V^{(k)}\|_{\overline{F}}^2 \leq \frac{4}{m_k n_k} \sum_{i,j} g(w_{ij}^{(k)})^2$$

$$= 4 \iint g(W^{(k)}(x,y))^2 \, dxdy,$$

where $W^{(k)}$ now denotes the function which equals $w_{ij}^{(k)}$ for all (x, y) in the rectangle $[\frac{i-1}{m_k}, \frac{i}{m_k}] \times [\frac{j-1}{n_k}, \frac{j}{n_k}]$. In other words, $W^{(k)}$ is obtained by averaging W within each such rectangle. Since m_k and n_k tend to ∞ and W is measurable, it follows by a standard result from analysis (see, for example, [14, Proposition 9.8]) that $W^{(k)}(x, y) \to W(x, y)$ as $k \to \infty$ for almost every (x, y). Since g is a bounded continuous function, this shows that

$$\limsup_{k \to \infty} \|(A_k - B_k) \circ V^{(k)}\|_{\overline{F}}^2 \leq 4 \iint g(W(x,y))^2 \, dxdy,$$

On the other hand, $g(x) \leq 1_{\{x \leq \delta\}}$ for all x. Thus,

$$\limsup_{k \to \infty} \|(A_k - B_k) \circ V^{(k)}\|_{\overline{F}}^2 \leq 4\phi(\eta),$$

where $\phi(\eta)$ is the Lebesgue measure of the set of all (x, y) where $W(x, y) \leq \eta$. Combining with (8), we get

$$\limsup_{k \to \infty} \|A_k - B_k\|_{\overline{F}} \leq 2\sqrt{\phi(\eta)} + \frac{\delta}{\eta}.$$

Note that this holds for any $\eta \in (0, 1)$. Since W is nonzero almost everywhere, $\phi(\eta) \to 0$ as $\eta \to 0$. Thus, given $\varepsilon > 0$, we can first choose η so small that $2\sqrt{\phi(\eta)} \leq \varepsilon/2$, and then choose δ so small that $\delta/\eta \leq \varepsilon/2$. If the sequences A_k and B_k satisfy (6) with this δ, then the above display allows us to conclude that $\limsup_{k \to \infty} \|A_k - B_k\|_{\overline{F}} \leq \varepsilon$. $\qquad\square$

We are now ready to prove Theorem 2.

Proof of Theorem 2: The 'if' part of Theorem 2 follows immediately from Theorem 5 and the observation, by Lemma 1, that

$$\|X\|_* \leq \operatorname{rank}(X)\|X\|_\infty \sqrt{mn} \qquad (9)$$

for any $m \times n$ matrix X.

For the 'only if' part, suppose that W is zero on a set of positive Lebesgue measure. Denote this set by S and let $\lambda(S)$ denote its Lebesgue measure. Take any $\varepsilon > 0$. By a standard measure-theoretic argument, there exists $T \subseteq [0, 1]^2$ such that T is a union of dyadic squares of equal size and $\lambda(S \Delta T) < \varepsilon$. Let $\mathcal{D}$ be the set of all dyadic squares of this size in $[0, 1]^2$.

For each k, let A_k be the zero matrix of order $m_k \times n_k$. Let B_k be the $m_k \times n_k$ matrix whose $(i, j)^{\text{th}}$ entry is 1 if $(\frac{i}{m_k}, \frac{j}{n_k}) \in T$ and 0 otherwise. Since T is a union of elements of $\mathcal{D}$, it is not difficult to see that B_k is a block matrix with at most $|\mathcal{D}|$ blocks. In particular, its rank is bounded above by $|\mathcal{D}|$.

Now note that $\|A_k - B_k\|_{\overline{F}}^2$ equals the fraction of indices (i, j) such that $(\frac{i}{m_k}, \frac{j}{n_k}) \in T$. Therefore as $k \to \infty$, $\|A_k - B_k\|_{\overline{F}}^2$ tends to $\lambda(T)$. If ε is small enough, this ensures that

$$\lim_{k \to \infty} \|A_k - B_k\|_{\overline{F}} \geq \sqrt{\frac{\lambda(S)}{2}} > 0. \qquad (10)$$

On the other hand, $\|(A_k - B_k) \circ P_k\|_{\overline{F}}^2$ equals the fraction of indices (i, j) such that $(\frac{i}{m_k}, \frac{j}{n_k}) \in T$ and the $(i, j)^{\text{th}}$ entry of P_k is 1. Let D be one of the constituent dyadic cubes of T. Let $f_k(D)$ be the fraction of (i, j) such that $(\frac{i}{m_k}, \frac{j}{n_k}) \in D$ and the $(i, j)^{\text{th}}$ entry of P_k is 1. From the definition of cut norm, it follows that

$$\lim_{k \to \infty} f_k(D) = \iint_D W(x,y)dxdy.$$

Summing over all D as above, we get

$$\lim_{k \to \infty} \|(A_k - B_k) \circ P_k\|_{\overline{F}}^2 = \iint_T W(x,y)dxdy.$$

Since $\lambda(S \Delta T) \leq \varepsilon$, W takes values in $[0, 1]$, and $W = 0$ on S, this shows that

$$\lim_{k \to \infty} \|(A_k - B_k) \circ P_k\|_{\overline{F}}^2 \leq \varepsilon + \iint_S W(x,y)dxdy = \varepsilon.$$
$$\text{(11)}$$

Since ε is arbitrary, the combination of (10) and (11) shows that the sequence P_k does not admit stable recovery of low rank matrices. This completes the proof of Theorem 2. $\qquad\square$

IX. PROOF OF THEOREM 3

To prove that $\|A_k - \widehat{A}_k\|_{\overline{F}} \to 0$, we will show that for any subsequence, there is a further subsequence through which this convergence takes place. By Theorem 1, we know that any subsequence has a further subsequence along which P_k converges to a limit graphon. Moreover, it is easy to see that if a sequence of binary matrices admits stable recovery of low rank matrices, then any subsequence also does so. Therefore by Theorem 2, we may assume without loss of generality that $P_k \to W$ for some W that is nonzero almost everywhere.

Also without loss of generality, suppose that $\|A_k\|_\infty \leq 1$ for all k. Let L be a uniform upper bound on $\operatorname{rank}(A_k)$. Then by (9),

$$\|A_k\|_* \leq L\sqrt{m_k n_k},$$

where m_k and n_k are the number of rows and number of columns in A_k. Consequently, $\|\widehat{A}_k\|_*$ is also bounded by $L\sqrt{m_k n_k}$. Moreover, by construction, $\|\widehat{A}_k\|_\infty \leq 1$ and $\|(A_k - \widehat{A}_k) \circ P_k\|_{\overline{F}} = 0$ for all k. Therefore by Theorem 5, we can now conclude that $\|\widehat{A}_k - A_k\|_{\overline{F}} \to 0$ as $k \to \infty$.

X. PROOF OF THEOREM 1

Let m and n be two positive integers. Let $\mathcal{P}$ be a partition of $\{1, \ldots, m\}$ and let $\mathcal{Q}$ be a partition of $\{1, \ldots, n\}$. The pair $(\mathcal{P}, \mathcal{Q})$ defines a block structure for $m \times n$ matrices in the natural way: Two pairs of indices (i, j) and (i', j') belong to the same block if and only if i and i' belong to the same member of $\mathcal{P}$ and j and j' belong to the same member of $\mathcal{Q}$.

If A is an $m \times n$ matrix, let $A^{\mathcal{P},\mathcal{Q}}$ be the 'block averaged' version of A, obtained by replacing the entries in each block (in the block structure defined by $(\mathcal{P}, \mathcal{Q})$) by the average value in that block. It is easy to see from the definition of the cut norm that

$$\|A^{\mathcal{P},\mathcal{Q}}\|_\square \leq \|A\|_\square. \qquad (12)$$

We need the following lemma.

Lemma 4: For any $m \times n$ matrix A with $\|A\|_\infty \leq 1$, there is a sequence of partitions $\{\mathcal{P}_j\}_{j\geq 1}$ of $\{1, \ldots, m\}$ and a sequence of partitions $\{\mathcal{Q}_j\}_{j\geq 1}$ of $\{1, \ldots, n\}$ such that for each j,

1) $\mathcal{P}_{j+1}$ is a refinement of $\mathcal{P}_j$ and $\mathcal{Q}_{j+1}$ is a refinement of $\mathcal{Q}_j$,
2) $|\mathcal{P}_j|$ and $|\mathcal{Q}_j|$ are bounded by $(2^{j+2}j)^{j^2}$, and
3) $\|A - A^{\mathcal{P}_j, \mathcal{Q}_j}\|_\square \leq 2j^{-1} + 6j^3 2^{-j}$.

Proof: Let

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

be the singular value decomposition of A, where $\sigma_1 \geq \cdots \geq \sigma_r > 0$ are the nonzero singular values. Take any $j \geq 1$. Let l be the largest number such that $\sigma_i > \sqrt{mn}/j$. If there is no such l, let $l = 0$. Let

$$A_1 := \sum_{i=1}^{l} \sigma_i u_i v_i^T.$$

For $1 \leq i \leq l$ and $1 \leq a \leq m$, let u_{ia} denote the ath component of u_i. Let $\widetilde{u}_{ia}^{(j)}$ be the largest integer multiple of $2^{-j} m^{-1/2}$ that is $\leq u_{ia}$. Let $\widetilde{u}_i^{(j)}$ be the vector whose a^{th} component is $\widetilde{u}_{ia}^{(j)}$. Similarly, for $1 \leq b \leq n$, let $\widetilde{v}_{ib}^{(j)}$ be the largest integer multiple of $2^{-j} n^{-1/2}$ that is $\leq v_{ib}^{(j)}$. Define

$$\widetilde{A}_1 := \sum_{i=1}^{l} \sigma_i \widetilde{u}_i \widetilde{v}_i^T.$$

Declare that two rows a and a' are equivalent if $\widetilde{u}_{ia}^{(j)} = \widetilde{u}_{ia'}^{(j)}$ for all $1 \leq i \leq l$. Similarly declare that two columns b and b' are equivalent if $\widetilde{v}_{ib}^{(j)} = \widetilde{v}_{ib'}^{(j)}$ for all $1 \leq i \leq l$. Let $\mathcal{P}_j$ be the set of equivalence classes of rows and $\mathcal{Q}_j$ be the set equivalence classes of columns.

From the above definition, it is clear that if $\widetilde{u}_{ia}^{(j+1)} = \widetilde{u}_{ia'}^{(j+1)}$, then $\widetilde{u}_{ia}^{(j)} = \widetilde{u}_{ia'}^{(j)}$. This shows that $\mathcal{P}_{j+1}$ is a refinement of $\mathcal{P}_j$. Similarly, $\mathcal{Q}_{j+1}$ is a refinement of $\mathcal{Q}_j$.

Next, note that by Lemma 1 and the definition of l,

$$\|u_i\|_\infty \leq \frac{\sqrt{n}}{\sigma_i} \leq \frac{j}{\sqrt{m}}$$

for $1 \leq i \leq l$. Thus, the set of possible values of $\widetilde{u}_{ia}$ has size at most

$$\frac{2j/\sqrt{m}}{2^{-j}/\sqrt{m}} + 1 = 2^{j+1}j + 1 \leq 2^{j+2}j.$$

Therefore, $|\mathcal{P}_j| \leq (2^{j+2}j)^l$. Now,

$$l\sigma_l^2 \leq \sum_{i=1}^{r} \sigma_i^2 = \|A\|_F^2 \leq mn,$$

where the last inequality holds because $\|A\|_\infty \leq 1$. Since $\sigma_l \geq \sqrt{mn}/j$, this gives

$$l \leq j^2. \tag{13}$$

Thus, $|\mathcal{P}_j| \leq (2^{j+2}j)^{j^2}$. Similarly, $|\mathcal{Q}_j| \leq (2^{j+2}j)^{j^2}$.

Now recall that the operator norm $\|M\|_{op}$ of a matrix M is the maximum of $\|Mx\|$ over all vectors x with $\|x\| \leq 1$. The operator norm of a matrix is equal to its largest singular value. From our definition of the cut norm, it is not difficult to see that for an $m \times n$ matrix M,

$$\|M\|_\square \leq \frac{\|M\|_{op}}{\sqrt{mn}}.$$

Thus,

$$\|A - A_1\|_\square \leq \frac{\|A - A_1\|_{op}}{\sqrt{mn}} = \frac{\sigma_{l+1}}{\sqrt{mn}} \leq \frac{1}{j}.$$

Next, by (13), Lemma 1 and the (easy) fact that the cut norm is bounded above by the average Frobenius norm,

$$\|A_1 - \widetilde{A}_1\|_\square \leq \|A_1 - \widetilde{A}_1\|_{\overline{F}}$$

$$\leq \sum_{i=1}^{l} \sigma_i(\|u_i - \widetilde{u}_i\|_\infty \|v_i\|_\infty + \|\widetilde{u}_i\|_\infty \|v_i - \widetilde{v}_i\|_\infty)$$

$$\leq \sum_{i=1}^{l} \sigma_i \bigg(\|u_i - \widetilde{u}_i\|_\infty \|v_i\|_\infty$$

$$+ (\|\widetilde{u}_i - u_i\|_\infty + \|u_i\|_\infty)\|v_i - \widetilde{v}_i\|_\infty \bigg)$$

$$\leq l\sqrt{mn}\bigg(2^{-j} m^{-1/2} j n^{-1/2}$$

$$+ 2^{-j} m^{-1/2} 2^{-j} n^{-1/2} + j m^{-1/2} 2^{-j} n^{-1/2} \bigg)$$

$$\leq 3j^3 2^{-j}.$$

Combining, we get

$$\|A - \widetilde{A}_1\|_\square \leq j^{-1} + 3j^3 2^{-j}.$$

Now note that $\widetilde{A}_1$ is constant within the blocks defined by the pair $(\mathcal{P}_j, \mathcal{Q}_j)$. Thus, by (12),

$$\|A - A^{\mathcal{P}_j, \mathcal{Q}_j}\|_\square \leq \|A - \widetilde{A}_1\|_\square + \|\widetilde{A}_1 - A^{\mathcal{P}_j, \mathcal{Q}_j}\|_\square$$

$$\leq \|A - \widetilde{A}_1\|_\square + \|\widetilde{A}_1^{\mathcal{P}_j, \mathcal{Q}_j} - A^{\mathcal{P}_j, \mathcal{Q}_j}\|_\square$$

$$\leq 2\|A - \widetilde{A}_1\|_\square.$$

This completes the proof. $\qquad\square$

We are now ready to prove Theorem 1. In this proof, we will use the following scheme to define a graphon using a matrix. Suppose that A is an $m \times n$ matrix. The graphon defined by A, which we will also denote by A, is the function $A : [0,1]^2 \to [0,1]$ which equals a_{ij} in the rectangle $(\frac{i-1}{m}, \frac{i}{m}) \times (\frac{j-1}{m}, \frac{j}{m})$ for each $1 \leq i \leq m$ and $1 \leq j \leq n$. On the boundaries of the rectangles, A can be defined arbitrarily.

In the proof, we will need to work with cut norms of asymmetric graphons. The cut norm of an asymmetric graphon W is defined as

$$\|W\|_\square := \sup_{a,b} \left| \iint a(x) b(y) W(x,y) dx dy \right|,$$

where the supremum is taken over all Borel measurable $a, b : [0,1] \to [-1,1]$. If the graphon is defined by a matrix as in the previous paragraph, it is easy to see that the cut norm of the graphon equals the cut norm of the matrix. A property of the cut norm that we will use in the proof is that the cut norm of an asymmetric graphon is bounded above by its L^1 norm.

Proof of Theorem 1: Let $\{A_k\}_{k\geq 1}$ be a sequence of matrices with dimensions tending to infinity. Let m_k and n_k be

the number of rows and number of columns in A_k. Lemma 4 tells us that for each k and j, we can find a partition $\mathcal{P}_{k,j}$ of $\{1, \ldots, m_k\}$ and a partition $\mathcal{Q}_{k,j}$ of $\{1, \ldots, n_k\}$ such that

1) $\mathcal{P}_{k,j+1}$ is a refinement of $\mathcal{P}_{k,j}$ and $\mathcal{Q}_{k,j+1}$ is a refinement of $\mathcal{Q}_{k,j}$,
2) $|\mathcal{P}_{k,j}|$ and $|\mathcal{Q}_{k,j}|$ are bounded by $(2^{j+2}j)^{j^2}$, and
3) $\|A_k - A_k^{\mathcal{P}_{k,j}, \mathcal{Q}_{k,j}}\|_\square \leq 2j^{-1} + 6j^3 2^{-j}$.

For simplicity, let us denote $A_k^{\mathcal{P}_{k,j}, \mathcal{Q}_{k,j}}$ by $A_{k,j}$. By permuting rows and columns of A_k, let us assume that the members of $\mathcal{P}_{k,j}$ and $\mathcal{Q}_{k,j}$ are intervals, so that $A_{k,j}$ is a block matrix. As described in the paragraph preceding this proof, the matrix $A_{k,j}$ defines an asymmetric graphon which is also denoted by $A_{k,j}$. This graphon is constant in rectangular blocks, where the number of blocks is bounded by $(2^{j+2}j)^{2j^2}$. Passing to a subsequence if necessary, we may assume that for each fixed j, these blocks tend to limiting blocks as $k \to \infty$, and moreover, that the value of $A_{k,j}$ within each block also tends to a limit. This limit defines an asymmetric graphon; let us call it W_j. Clearly, $A_{k,j} \to W_j$ in the L^1 metric as $k \to \infty$.

Now note that by construction, the block structure for W_{j+1} is a refinement of the block structure for W_j. Moreover, also by construction, the value of W_j in one of its blocks is the average value of W_{j+1} within that block. From this, by a standard martingale argument (for example, as in the proof of [14, Theorem 9.23]) it follows that W_j converges pointwise almost everywhere to an asymmetric graphon W as $j \to \infty$. In particular, $W_j \to W$ in L^1. We claim that $A_k \to W$ in the cut norm as $k \to \infty$. To show this, take any $\varepsilon > 0$. Find j so large that $\|W - W_j\|_{L^1} \leq \varepsilon$ and $2j^{-1} + 6j^3 2^{-j} \leq \varepsilon$. Then for any k,

$$
\begin{aligned}
\|W - A_k\|_\square &\leq \|W - W_j\|_\square + \|W_j - A_{k,j}\|_\square \\
&\quad + \|A_{k,j} - A_k\|_\square \\
&\leq \varepsilon + \|W_j - A_{k,j}\|_{L^1} + 2j^{-1} + 6j^3 2^{-j} \\
&\leq 2\varepsilon + \|W_j - A_{k,j}\|_{L^1}.
\end{aligned}
$$

Since $A_{k,j} \to W_j$ in L^1 as $k \to \infty$ and ε is arbitrary, this completes the proof. $\qquad\square$

ACKNOWLEDGMENT

REFERENCES

[1] S. Bhojanapalli and P. Jain, "Universal matrix completion," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1881–1889.

[2] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.

[3] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, Dec. 2009.

[4] E. J. Candes and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.

[5] S. Chatterjee, "Matrix estimation by universal singular value thresholding," *Ann. Statist.*, vol. 43, no. 1, pp. 177–214, Feb. 2015.

[6] Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward, "Completing any low-rank matrix, provably," *J. Mach. Learn. Res.*, vol. 16, no. 1, pp. 2999–3034, 2015.

[7] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, Sep. 1936.

[8] M. Gavish and D. L. Donoho, "The optimal hard threshold for singular values is $4/\sqrt{3}$," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 5040–5053, Aug. 2014.

[9] E. Heiman, G. Schechtman, and A. Shraibman, "Deterministic algorithms for matrix completion," *Random Struct. Algorithms*, vol. 45, no. 2, pp. 306–317, Sep. 2014.

[10] C. R. Johnson, "Matrix completion problems: A survey," in *Matrix Theory and Applications*, vol. 40. American Mathematical Society: Providence, RI, USA, 1990, pp. 171–198.

[11] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2980–2998, Jun. 2010.

[12] F. J. Király, L. Theran, and R. Tomioka, "The algebraic combinatorial approach for low-rank matrix completion," *J. Mach. Learn. Res.*, vol. 16, no. 1, pp. 1391–1436, 2015.

[13] T. Lee and A. Shraibman, "Matrix completion from any given set of observations," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 1781–1787.

[14] L. Lovász, *Large Networks and Graph Limits*. Providence, RI, USA: American Mathematical Society, 2012,

[15] L. T. Nguyen, J. Kim, and B. Shim, "Low-rank matrix completion: A contemporary survey," *IEEE Access*, vol. 7, pp. 94215–94237, 2019.

[16] D. L. Pimentel-Alarcon, N. Boston, and R. D. Nowak, "A characterization of deterministic sampling patterns for low-rank matrix completion," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 623–636, Jun. 2016.

[17] A. Singer and M. Cucuringu, "Uniqueness of low-rank matrix completion by rigidity theory," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 4, pp. 1621–1641, Jan. 2010.

[18] A. Sportisse, C. Boyer, and J. Josse, "Imputation and low-rank estimation with missing not at random data," 2018, *arXiv:1812.11409*. [Online]. Available: http://arxiv.org/abs/1812.11409

Sourav Chatterjee received the bachelor's and master's degrees in statistics from the Indian Statistical Institute, Calcutta, in 2000 and 2002, respectively, and the Ph.D. degree in statistics from Stanford University, Stanford, CA, USA, in 2005.

He joined UC Berkeley as a Neyman Assistant Professor of statistics in 2005, and started as a tenure-track Assistant Professor with the UC Berkeley Statistics Department in 2006. He moved to New York University as an Associate Professor of mathematics with the Courant Institute of Mathematical Sciences in 2009. Since 2013, he has been a Professor of mathematics and statistics with Stanford University. His areas of interest are probability theory, statistics, and mathematical physics.

Dr. Chatterjee was elected as a Fellow of the Institute of Mathematical Statistics in 2018. He was awarded a Sloan Research Fellowship in mathematics in 2007, the 2008 Tweedie New Researcher Award from the Institute of Mathematical Statistics, the 2010 Rollo Davidson Prize for work in probability theory, the 2012 Doeblin Prize from the Bernoulli Society, the 2012 Young Researcher Award from the International Indian Statistical Association, and the 2013 Line and Michel Loève Prize in probability from UC Berkeley. He gave a Medallion Lecture of the Institute of Mathematical Statistics in 2012 and was an Invited Speaker with the International Congress of Mathematicians in 2014.