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. Yet, a characterization of missing patterns that allow completion has remained an open question. The main result of this paper gives such a characterization in the language of graph limit theory. It is then shown that a modification of the Candès–Recht nuclear norm minimization algorithm succeeds in completing the matrix whenever completion is possible. The theory is fully deterministic, with no assumption of randomness.

1. 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 very recent survey of this vast body of work, see [13]. 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 [8]. 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], rigidity theory [15], algebraic geometry [10] and other methods [11, 14, 16]. 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.

Right away, it is important to note that not all patterns of revealed entries allow low rank matrix completion, 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.

2010 Mathematics Subject Classification. 15A83, 65F30.
Key words and phrases. Matrix completion, low rank matrix, graph limit, graphon.
Research partially supported by NSF grant DMS-1855484.
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 \((i, j)\)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’. Yet, recovery is not possible even if the rank of the matrix is as small as 3. 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 relabeled from 1 to \( n/2 \) and the odd numbered rows and columns are relabeled from \( n/2 + 1 \) to \( n \). Then in this new arrangement of rows and columns, the \((i, j)\)th entry is revealed if and only if either both \( i \) and \( j \) are between 1 and \( n/2 \), or both \( i \) and \( j \) are between \( n/2 + 1 \) and \( n \). In other words, the matrix is a \( 2 \times 2 \) block matrix with blocks of order \( n/2 \times n/2 \), where only the top-left and bottom-right blocks are revealed. Clearly, the other two blocks cannot be recovered using this information if the rank is 3 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 4.1). The precise statement is given in the language of graph limit theory [12]. 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 4.3). In other words, this algorithm does the job whenever the job is doable.

2. Notations

All our matrices will have real entries. We will denote the \((i, j)\)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:

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

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 \( \ell^\infty \) 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
\[
\|A\|_\square := \frac{1}{mn} \max \{ |x^T A y| : x \in \mathbb{R}^m, y \in \mathbb{R}^n, \|x\|_\infty \leq 1, \|y\|_\infty \leq 1 \}.
\]
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 \((i,j)\)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
\[
\delta_\square(A, B) := \min_{\pi \in S_m, \tau \in S_n} \|A^{\pi,\tau} - B\|_\square. \tag{2.1}
\]
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 \( A \odot B \), is the \( m \times n \) matrix whose \((i,j)\)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 \( A \odot P \) is the matrix whose entries equal the entries of \( A \) wherever they are revealed, and zero elsewhere.

3. Definitions

As mentioned earlier, certain patterns of revealed entries may not suffice for 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, 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 3.1.** 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}, \text{ where } A_k \text{ and } B_k \text{ have the same dimensions as } P_k. \text{ Suppose that there are numbers } K \text{ and } L \text{ such that } \text{rank}(A_k) \text{ and } \text{rank}(B_k) \text{ are bounded by } K \text{ and } \|A_k\|_\infty \text{ and } \|B_k\|_\infty \text{ are bounded by } L \text{ for each } k. \text{ Then for any } \varepsilon > 0 \text{ there is some } \delta > 0, \text{ depending only on } \varepsilon, \text{ } K \text{ and } L, \text{ such that if } \limsup_{k \to \infty} \|(A_k - B_k) \circ P_k\|_F \leq \delta, \text{ then } \limsup_{k \to \infty} \|A_k - B_k\|_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 \( \lfloor k/2 \rfloor \) 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 \( \lfloor k/2 \rfloor \) rows and 1 elsewhere. Then \( \|A_k\|_\infty \text{ and } \|B_k\|_\infty \) are bounded by 1 for all \( k \), and \( \text{rank}(A_k) \text{ and } \text{rank}(B_k) \) are bounded by 1 for all \( k \). Now clearly \( \lim_{k \to \infty} \|(A_k - B_k) \circ P_k\|_F = 0 \), but a simple calculation shows that \( \lim_{k \to \infty} \|A_k - B_k\|_F = \frac{1}{\sqrt{2}} \).

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 3.1, one needs to verify the stated condition for all sequences \( \{A_k\}_{k \geq 1} \text{ and } \{B_k\}_{k \geq 1}. \) 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 [12], 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 3.2.** 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 \((i,j)\)th entry is the average value of \( W \) in the rectangle \( \left[ \frac{i-1}{m}, \frac{i}{m} \right] \times \left[ \frac{j-1}{n}, \frac{j}{n} \right] \), that is,

\[
mn \int_{(i-1)/m}^{i/m} \int_{(j-1)/n}^{j/n} W(x,y) dy dx.
\]

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 (2.1).
Definition 3.3. We will say that a sequence of matrices \( \{A_k\}_{k \geq 1} \) converges to an asymmetric graphon \( W \) if \( \delta_{\mathbb{E}}(A_k, W) \to 0 \) 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 3.4. 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 \([12, \text{Theorem 9.23}]\). It is probable that the asymmetric version already exists in the literature, but since the proof is not difficult, it is presented in Section 8.

4. 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 3.4, it suffices to only consider convergent sequences.

Theorem 4.1. 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 3.1) 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 \([0, 1]\). If \( p > 0 \), Theorem 4.1 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 4.1 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 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 4.1 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 4.2.** Let $A$ be a matrix whose entries are partially revealed. Suppose that $\|A\|_\infty \leq L$ 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 $\|B\|_\infty \leq L$.

The assumption of a known upper bound on the $\ell^\infty$ 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 4.3.** Let $\{P_k\}_{k \geq 1}$ be a sequence of binary matrices with dimensions tending to infinity that admits stable recovery of low rank matrices. Let $\{A_k\}_{k \geq 1}$ 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 4.2) when the locations of the revealed entries are defined by $P_k$. Then $\lim_{k \to \infty} \|\widehat{A}_k - A_k\|_F = 0$.

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 [3, 7, 9]) 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 4.3 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 rest of the paper is devoted to proofs. The proof of Theorem 4.1 is divided between Sections 5 and 6. Theorem 4.3 is proved in Section 7, and Theorem 3.4 is proved in Section 8.

5. Towards the proof of Theorem 4.1

The goal of this section is to prove a quantitative result that underlies the proof of Theorem 4.1. We need to prove a number of lemmas before arriving at this theorem.

Lemma 5.1. Let $X$ be an $m \times n$ matrix with $\|X\|_{\infty} \leq 1$ and singular value decomposition

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

Then for each $i$,

\[ \sigma_i \leq \sqrt{mn}, \quad \|u_i\|_{\infty} \leq \frac{\sqrt{n}}{\sigma_i} \quad \text{and} \quad \|v_i\|_{\infty} \leq \frac{\sqrt{m}}{\sigma_i}. \]

Proof. Let $u_{ij}$ denote the $j^{th}$ component of $u_i$. Since $X v_i = \sigma_i u_i$ and $\|X\|_{\infty} \leq 1$, we get

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

Dividing throughout by $\sigma_i$ 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 $\sigma_i$, notice that since $\sum_j u_{ij}^2 = 1$ there is some $j$ such that $|u_{ij}| \geq m^{-1/2}$, and use this information in the above display. \qed

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 5.2. Let $X$ and $Y$ be $m \times n$ matrices with $\|X\|_{\infty} \leq 1$ and $\|Y\|_{\infty} \leq 1$. Let $q \geq 1$ 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 $(20000q^6 \varepsilon^{-10})^{5q \varepsilon^{-2}}$ blocks, and permutations $\pi \in S_m$ and $\tau \in S_n$, such that $\|X^{\pi, \tau} - A\|_{\mathcal{F}} \leq \varepsilon$ and $\|Y^{\pi, \tau} - B\|_{\mathcal{F}} \leq \varepsilon$. Moreover, it can be ensured that $\|A\|_{\infty} \leq 1$ and $\|B\|_{\infty} \leq 1$.

Proof. Fix some $\varepsilon > 0$. Let $\delta, \gamma$ and $\eta$ 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 $\sigma_1 \geq \cdots \geq \sigma_k > 0$ and $\lambda_1 \geq \cdots \geq \lambda_l > 0$. Choose two numbers $k_1$ and $l_1$ such that $\sigma_{k_1} > \delta \geq \sigma_{k_1+1}$ and $\lambda_{l_1} > \delta \geq \lambda_{l_1+1}$. If $\sigma_i \leq \delta$ for all $i$, let $k_1 = 0$, and if $\sigma_i > \delta$ for all $i$, let $k_1 = k$. Similarly, if $\lambda_i \leq \delta$ 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 $k_1$,

$$\|X - X_1\|_F^2 = \frac{1}{mn} \sum_{i=k_1+1}^{k} \sigma_i^2 \leq \frac{\delta}{mn} \sum_{i=1}^{k} \sigma_i = \frac{\|X\|_\infty \delta}{mn} \leq \frac{q \delta}{\sqrt{mn}}. \tag{5.1}$$

Similarly, the same bound holds for $\|Y - Y_1\|_F^2$.

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

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

Now take any $1 \leq i \leq k_1$. By Lemma 5.1 and the choice of $k_1$, we have

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

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

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

where the inequality was obtained under the assumption that

$$\frac{2\sqrt{n}}{\delta \gamma} \geq 1. \tag{5.2}$$

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

Declare that two rows $a$ and $a'$ are ‘equivalent’ if $\tilde{u}_{ia} = \tilde{u}_{ia'}$ and $\tilde{w}_{i'a} = \tilde{w}_{i'a'}$ for all $1 \leq i \leq k_1$ and $1 \leq i' \leq l_1$. Similarly declare that two columns $b$ and $b'$ are equivalent if $\tilde{v}_{ib} = \tilde{v}_{ib'}$ and $\tilde{z}_{i'b} = \tilde{z}_{i'b'}$ for all $1 \leq i \leq k_1$ and $1 \leq i' \leq l_1$. 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 $\pi$ be a permutation of the rows that ‘clumps together’ equivalent rows, and let $\tau$ be a permutation of the columns that clumps together equivalent columns. Then it is clear that $\tilde{X}^\pi_1$ and $\tilde{Y}^\pi_1$ 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 $k_1 \sigma_{k_1} \leq \sum_{i=1}^{k_1} \sigma_i = \|X\|_* \leq q\sqrt{mn}$.

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

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

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 [5.1] and the definition of $\tilde{X}_1$,

$$\|X_1 - \tilde{X}_1\|_F \leq \|X_1 - \tilde{X}_1\|_\infty$$

$$\leq \sum_{i=1}^{k_1} \sigma_i(\|u_i - \tilde{u}_i\|_\infty \|v_i\|_\infty + \|\tilde{u}_i\|_\infty \|v_i - \tilde{v}_i\|_\infty)$$

$$\leq q\sqrt{mn}\left(\frac{\sqrt{m}\gamma}{\delta} + \frac{\sqrt{n}\eta}{\delta}\right).$$

By a similar argument, the same bound holds for $\|Y_1 - \tilde{Y}_1\|_F$. Combining with [5.1], we see that if $A := \tilde{X}^\pi_1$ and $B := \tilde{Y}^\pi_1$, then $A$ and $B$ have a simultaneous block structure with at most $b$ blocks, and $\|X^\pi - A\|_F$ and $\|Y^\pi - B\|_F$ are bounded by

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

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 $\beta$, 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 $\beta$ to make this equal to $\varepsilon$, which ensures that $\|X^\pi - A\|_F$ and $\|Y^\pi - B\|_F$ are bounded by $\varepsilon$. With these choices of $\alpha$ and $\beta$, an easy calculation gives

$$b = (16\alpha^{-2} \beta^{-2})^{2q/\alpha} \leq (20000q^6 \varepsilon^{-10})^{5q\varepsilon^{-2}}.$$
Also, it is easy to check (using \( q \geq 1 \) and \( \varepsilon \in (0,1) \)) that with these choices of \( \alpha \) and \( \beta \), the inequality \([5.2]\) holds. Thus, the proof is complete except that we have not ensured that \( \|A\|_\infty \leq 1 \) and \( \|B\|_\infty \leq 1 \) 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 \(|x_{\pi(i)\tau(j)}| \leq 1 \) and \(|y_{\pi(i)\tau(j)}| \leq 1 \).

\[ \square \]

**Lemma 5.3.** 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

\[
\|(A - B) \circ Q\|_F \leq \|(A - B) \circ P\|_F + \sqrt{b\|P - Q\|_\square} \|A - B\|_\infty.
\]

**Proof.** Let each block be represented by the set of pairs of indices \((i, j)\) that belong to the block. Let \( D \) be the set of all blocks. Take any block \( D \in D \).

By the definition the cut norm,

\[
\frac{1}{mn} \left| \sum_{(i,j) \in D} (p_{ij} - q_{ij}) \right| \leq \|P - Q\|_\square. \tag{5.3}
\]

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 \( i, j \),

\[
\|(A - B) \circ Q\|_F^2 = \frac{1}{mn} \sum_{i,j} (a_{ij} - b_{ij})^2 q_{ij}^2 \\
\leq \frac{1}{mn} \sum_{i,j} (a_{ij} - b_{ij})^2 q_{ij} \\
= \sum_{D \in D} (a(D) - b(D))^2 \left( \frac{1}{mn} \sum_{(i,j) \in D} q_{ij} \right).
\]

Therefore by \((5.3)\),

\[
\|(A - B) \circ Q\|_F^2 \leq \sum_{D \in D} (a(D) - b(D))^2 \left( \frac{1}{mn} \sum_{(i,j) \in D} p_{ij} + \|P - Q\|_\square \right).
\]

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

\[
\|(A - B) \circ Q\|_F^2 \leq \frac{1}{mn} \sum_{D \in D} \sum_{(i,j) \in D} (a_{ij} - b_{ij})^2 p_{ij}^2 \\
+ \|P - Q\|_\square \|A - B\|_\infty |D| \\
= \|(A - B) \circ P\|_F^2 + \|P - Q\|_\square \|A - B\|_\infty^2 b.
\]

The proof is not completed by applying the inequality \( \sqrt{x+y} \leq \sqrt{x} + \sqrt{y} \) 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 5.4.** Let $X$ and $Y$ be $m \times n$ matrices with $\ell^\infty$ 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

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

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

**Proof.** Without loss of generality, assume that $q \geq 1$. Take any $\varepsilon > 0$. Let $A, B, \pi$ and $\tau$ be as in Lemma 5.2. Let

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

be the upper bound on the number of blocks given by Lemma 5.2. Note that

$$\|(X - Y) \circ Q\|_F = \|(X^{\pi, \tau} - Y^{\pi, \tau}) \circ Q^{\pi, \tau}\|_F$$

$$\leq \|(X^{\pi, \tau} - A) \circ Q^{\pi, \tau}\|_F + \|(A - B) \circ Q^{\pi, \tau}\|_F + \|(B - Y^{\pi, \tau}) \circ Q^{\pi, \tau}\|_F$$

$$\leq \|X^{\pi, \tau} - A\|_F + \|(A - B) \circ Q^{\pi, \tau}\|_F + \|B - Y^{\pi, \tau}\|_F$$

$$\leq 2\varepsilon + \|(A - B) \circ Q^{\pi, \tau}\|_F.$$  

By Lemma 5.3,

$$\|(A - B) \circ Q^{\pi, \tau}\|_F$$

$$\leq \|(A - B) \circ P^{\pi, \tau}\|_F + \sqrt{b}\|P^{\pi, \tau} - Q^{\pi, \tau}\|_F \|A - B\|_\infty$$

$$\leq \|(A - B) \circ P^{\pi, \tau}\|_F + 2\sqrt{b}\|P - Q\|_F.$$  

But

$$\|(A - B) \circ P^{\pi, \tau}\|_F$$

$$\leq \|(A - X^{\pi, \tau}) \circ P^{\pi, \tau}\|_F + \|(X^{\pi, \tau} - Y^{\pi, \tau}) \circ P^{\pi, \tau}\|_F$$

$$+ \|(Y^{\pi, \tau} - B) \circ P^{\pi, \tau}\|_F$$

$$\leq \|A - X^{\pi, \tau}\|_F + \|(X - Y) \circ P\|_F + \|Y^{\pi, \tau} - B\|_F$$

$$\leq 2\varepsilon + \|(X - Y) \circ P\|_F.$$  

Adding up, we get

$$\|(X - Y) \circ Q\|_F \leq \|(X - Y) \circ P\|_F + 4\varepsilon + 2\sqrt{b}\|P - Q\|_F$$

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

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


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 $4\varepsilon$ term. The required bound is now obtained by choosing

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

for some sufficiently large constant $C(q)$ depending only on $q$.

6. Proof of Theorem 4.1

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 4.1.

**Theorem 6.1.** 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 $\varepsilon$, $q$ and $L$, such that if $\limsup_{k \to \infty} \|(A_k - B_k) \circ P_k\|_F \leq \delta$, then $\limsup_{k \to \infty} \|A_k - B_k\|_F \leq \varepsilon$.

**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 $\pi_k$ and $\tau_k$ are the identity permutations, so that

$$
\lim_{k \to \infty} \|P_k - W(k)\|_\square = 0. \tag{6.1}
$$

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

$$
\limsup_{k \to \infty} \|(A_k - B_k) \circ P_k\|_F \leq \delta. \tag{6.2}
$$

Without loss of generality, $L = 1$. Then by (6.1), (6.2) and Theorem 5.4

$$
\limsup_{k \to \infty} \|(A_k - B_k) \circ W(k)\|_F \leq \delta. \tag{6.3}
$$

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

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

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 $4\varepsilon$ term. The required bound is now obtained by choosing

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

for some sufficiently large constant $C(q)$ depending only on $q$. □
and \( g(x) := (f(x) - x)/\eta \). Let \( U^{(k)} \) be the matrix whose \((i, j)\)th element is \( f(w_{ij}^{(k)}) \) and let \( V^{(k)} \) be the matrix whose \((i, j)\)th element is \( g(w_{ij}^{(k)}) \). Since \( f(x) \geq \eta \) for all \( x \),

\[
\|A_k - B_k\|_F \leq \frac{1}{\eta} \|A_k - B_k\|_F \|U^{(k)}\|_F
\]

\[
\leq \frac{1}{\eta} \|A_k - B_k\|_F \|U^{(k)} - W^{(k)}\|_F + \frac{1}{\eta} \|A_k - B_k\|_F \|W^{(k)}\|_F
\]

\[
= \|A_k - B_k\|_F \|V^{(k)}\|_F + \frac{1}{\eta} \|A_k - B_k\|_F \|W^{(k)}\|_F.
\]

Therefore by (6.3),

\[
\limsup_{k \to \infty} \|A_k - B_k\|_F \leq \limsup_{k \to \infty} \|A_k - B_k\|_F \|V^{(k)}\|_F + \frac{\delta}{\eta}, \tag{6.4}
\]

Since \( L = 1 \),

\[
\|A_k - B_k\|_F \|V^{(k)}\|_F \leq \frac{4}{m_kn_k} \sum_{i,j} g(w_{ij}^{(k)})^2
\]

\[
= 4 \int \int g(W^{(k)}(x, y))^2 dx dy,
\]

where \( W^{(k)} \) now denotes the function which equals \( w_{ij}^{(k)} \) for all \((x, y)\) in the rectangle \([i-\frac{1}{m_k}, i] \times [j-\frac{1}{n_k}, j]\). In other words, \( W^{(k)} \) is obtained by averaging \( W \) within each such rectangle. Since \( m_k \) and \( n_k \) tend to \( \infty \) and \( W \) is measurable, it follows by a standard result from analysis (see, for example, [12, 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 continuous function, this shows that

\[
\limsup_{k \to \infty} \|A_k - B_k\|_F \|V^{(k)}\|_F \leq 4 \int \int g(W(x, y))^2 dx dy,
\]

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

\[
\limsup_{k \to \infty} \|A_k - B_k\|_F \|V^{(k)}\|_F \leq 4\phi(\eta),
\]

where \( \phi(\eta) \) is the Lebesgue measure of the set of all \((x, y)\) where \( W(x, y) \leq \eta \). Combining with (6.4), we get

\[
\limsup_{k \to \infty} \|A_k - B_k\|_F \leq 2\sqrt{\phi(\eta)} + \frac{\delta}{\eta}.
\]

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 \( \eta \) so small that \( 2\sqrt{\phi(\eta)} \leq \varepsilon/2 \), and then choose \( \delta \) so small that \( \delta/\eta \leq \varepsilon/2 \). If the sequences \( A_k \) and \( B_k \) satisfy (6.2) with this \( \delta \), then the above display allows us to conclude that \( \limsup_{k \to \infty} \|A_k - B_k\|_F \leq \varepsilon \).

We are now ready to prove Theorem 4.1.
Proof of Theorem 4.1. The ‘if’ part of Theorem 4.1 follows immediately from Theorem 6.1 and the observation, by Lemma 5.1, that
\[
\|X\|_* \leq \text{rank}(X)\|X\|_\infty \sqrt{mn}
\] (6.5)
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 \((i,j)\)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 \(\|A_k - B_k\|_F^2\) equals the fraction of indices \((i,j)\) such that \((\frac{i}{m_k}, \frac{j}{n_k}) \in T\). Therefore as \(k \to \infty\), \(\|A_k - B_k\|_F^2\) tends to \(\lambda(T)\). If \(\varepsilon\) is small enough, this ensures that
\[
\lim_{k \to \infty} \|A_k - B_k\|_F \geq \sqrt{\frac{\lambda(S)}{2}} > 0.
\] (6.6)

On the other hand, \(\|(A_k - B_k) \circ P_k\|_F^2\) equals the fraction of indices \((i,j)\) such that \((\frac{i}{m_k}, \frac{j}{n_k}) \in T\) and the \((i,j)\)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 \((i,j)\)th entry of \(P_k\) is 1. From the definition of cut norm, it follows that
\[
\lim_{k \to \infty} f_k(D) = \int\int_D W(x,y) dxdy.
\]

Summing over all \(D\) as above, we get
\[
\lim_{k \to \infty} \|(A_k - B_k) \circ P_k\|_F^2 = \int\int_T W(x,y) dxdy.
\]

Since \(\lambda(S \Delta T) \leq \varepsilon\), \(W\) takes values in \([0,1]\), and \(W = 0\) on \(S\), this shows that
\[
\lim_{k \to \infty} \|(A_k - B_k) \circ P_k\|_F^2 \leq \varepsilon + \int\int_S W(x,y) dxdy = \varepsilon.
\] (6.7)

Since \(\varepsilon\) is arbitrary, the combination of (6.6) and (6.7) shows that the sequence \(P_k\) does not admit stable recovery of low rank matrices. This completes the proof of Theorem 4.1. \(\square\)

7. Proof of Theorem 4.3

To prove that \(\|A_k - \hat{A}_k\|_F \to 0\), we will show that for any subsequence, there is a further subsequence through which this convergence takes place. By Theorem 3.4, 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 4.1, 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 $\text{rank}(A_k)$. Then by (6.5),

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

where $m_k$ and $n_k$ are the number of rows and number of columns in $A_k$. Consequently, $\|\hat{A}_k\|_*$ is also bounded by $L \sqrt{m_k n_k}$. Moreover, by construction, $\|\hat{A}_k - A_k\|_F = 0$ for all $k$. Therefore by Theorem 6.1, we can now conclude that $\|\hat{A}_k - A_k\|_F \to 0$ as $k \to \infty$.

8. Proof of Theorem 3.4

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

$$\|A^{\mathcal{P}, \mathcal{Q}}\|_\square \leq \|A\|_\square. \quad (8.1)$$

We need the following lemma.

Lemma 8.1. 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 $(2^{j+2}j)^2$,
3. $\|A - A^{\mathcal{P}_j, \mathcal{Q}_j}\|_\square \leq 2j^{-1} + 6j^32^{-j}$.

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 $\sigma_l > \sqrt{mn}/j$. 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 $a^{th}$ component of $u_i$. Let $\tilde{u}_{ia}^{(j)}$ be the largest integer multiple of $2^{-j}m^{1/2}$ that is $\leq u_{ia}$. Let $\tilde{v}_{ib}^{(j)}$ be the vector whose $a^{th}$ component is $\tilde{u}_{ia}^{(j)}$. Similarly, for $1 \leq b \leq n$, let $\tilde{v}_{ib}^{(j)}$ be the largest integer multiple of $2^{-j}n^{1/2}$ that is $\leq v_{ib}^{(j)}$. Define

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

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

From the above definition, it is clear that if $\tilde{u}_{ia}^{(j+1)} = \tilde{u}_{ia'}^{(j+1)}$, then $\tilde{u}_{ia}^{(j)} = \tilde{u}_{ia'}^{(j)}$. This shows that $P_{j+1}$ is a refinement of $P_j$. Similarly, $Q_{j+1}$ is a refinement of $Q_j$.

Next, note that by Lemma 5.1 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 $\tilde{u}_{ia}$ has size at most

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

Therefore, $|P_j| \leq (2^{j+2}j)^l$. 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 $\sigma_i \geq \sqrt{mn}/j$, this gives

$$
l \leq j^2.
$$

Thus, $|P_j| \leq (2^{j+2}j)^j$. Similarly, $|Q_j| \leq (2^{j+2}j)^j$.

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 \(8.2\), Lemma 5.1 and the (easy) fact that the cut norm is bounded above by the average Frobenius norm,
\[
\|A_1 - \tilde{A}_1\|_\square \leq \|A_1 - \tilde{A}_1\|_F
\]
\[
\leq \sum_{i=1}^l \sigma_i(\|u_i - \tilde{u}_i\|_\infty \|v_i\|_\infty + \|\tilde{u}_i\|_\infty \|v_i - \tilde{v}_i\|_\infty)
\]
\[
\leq \sum_{i=1}^l \sigma_i(\|u_i - \tilde{u}_i\|_\infty \|v_i\|_\infty + (\|\tilde{u}_i - u_i\|_\infty + \|u_i\|_\infty) \|v_i - \tilde{v}_i\|_\infty)
\]
\[
\leq l \sqrt{mn(2^{-j}m^{-1/2}jn^{-1/2} + 2^{-j}m^{-1/2}2^{-j}n^{-1/2} + jm^{-1/2}2^{-j}n^{-1/2})}
\]
\[
\leq 3j^{3/2}2^{-j}.
\]
Combining, we get
\[
\|A - \tilde{A}_1\|_\square \leq j^{-1} + 3j^{3/2}2^{-j}.
\]
Now note that \(\tilde{A}_1\) is constant within the blocks defined by the pair \((P_j, Q_j)\).
Thus, by \(8.1\),
\[
\|A - A^{P_j, Q_j}\|_\square \leq \|A - \tilde{A}_1\|_\square + \|\tilde{A}_1 - A^{P_j, Q_j}\|_\square
\]
\[
\leq \|A - \tilde{A}_1\|_\square + \|A^{P_j, Q_j} - A^{P_j, Q_j}\|_\square \leq 2\|A - \tilde{A}_1\|_\square.
\]
This completes the proof. \(\square\)

We are now ready to prove Theorem 3.4. 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 \((i-1/m, i/m) \times (j-1/m, j/m)\) 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
\[
\|W\|_\square := \sup_{a,b} \left| \int \int a(x)b(y)W(x,y)dxdy \right|
\]
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 will be used in the proof is that the cut norm of an asymmetric graphon is bounded above by its \(L^1\) norm.

**Proof of Theorem 3.4** 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 8.1 tells us that for each \(k\) and \(j\), we can find a partition \(P_{k,j}\) of \(\{1, \ldots, m_k\}\) and a partition \(Q_{k,j}\) of \(\{1, \ldots, n_k\}\) such that
1. \(P_{k,j+1}\) is a refinement of \(P_{k,j}\) and \(Q_{k,j+1}\) is a refinement of \(Q_{k,j}\),
2. \(|P_{k,j}|\) and \(|Q_{k,j}|\) are bounded by \((2^{i+2}j)^2\), and
(3) \( \|A_k - A_k^{P_k,j,Q_k,j}\|_\square \leq 2j^{-1} + 6j^32^{-j} \).

For simplicity, let us denote \( A_k^{P_k,j,Q_k,j} \) by \( A_{k,j} \). By permuting rows and columns of \( A_k \), let us assume that the members of \( P_{k,j} \) and \( Q_{k,j} \) 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 \((2j^2 + 2j^2)2^{-j}\). 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 [12, 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 \( 2j^{-1} + 6j^32^{-j} \leq \varepsilon \). Then for any \( k \),

\[
\|W - A_k\|_\square \leq \|W - W_j\|_\square + \|W_j - A_{k,j}\|_\square + \|A_{k,j} - A_k\|_\square \\
\leq \varepsilon + \|W_j - A_{k,j}\|_{L^1} + 2j^{-1} + 6j^32^{-j} \\
\leq 2\varepsilon + \|W_j - A_{k,j}\|_{L^1}.
\]

Since \( A_{k,j} \to W_j \) in \( L^1 \) as \( k \to \infty \) and \( \varepsilon \) is arbitrary, this completes the proof. \( \Box \)

Acknowledgments

I thank Sohom Bhattacharya, Nicholas Cook and Terry Tao for helpful comments.

References

[1] Bhojanapalli, S. and Jain, P. (2014). Universal matrix completion. Internat. Conf. Mach. Learn., 1881–1889.
[2] Candès, E. J. and Plan, Y. (2010). Matrix completion with noise. Proc. of IEEE, 98 no. 6, 925–936.
[3] Candès, E. J. and Recht, B. (2009). Exact matrix completion via convex optimization. Found. Comput. Math., 9 no. 6, 717–772.
[4] Candès, E. J. and Tao, T. (2010). The power of convex relaxation: Near-optimal matrix completion. IEEE Trans. Inf. Theory, 56 no. 5, 2053–2080.
[5] Chatterjee, S. (2015). Matrix estimation by universal singular value thresholding. Ann. Statist., 43 no. 1, 177–214.
[6] Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1 no. 3, 211–218.

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

[8] Johnson, C. R. (1990). Matrix completion problems: A survey. In *Matrix theory and applications*, Vol. 40, pp. 171–198. Providence, RI.

[9] Keshavan, R. H., Montanari, A. and Oh, S. (2010). Matrix completion from a few entries. *IEEE Trans. Inf. Theory*, 56 no. 6, 2980–2998.

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

[11] Lee, T. and Shraibman, A. (2013). Matrix completion from any given set of observations. In *Adv. Neur. Inf. Proc. Sys.*, 1781–1787.

[12] Lovász, L. (2012). *Large networks and graph limits*. American Mathematical Society, Providence, RI.

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

[14] Pimentel-Alarcón, D. L., Boston, N. and Nowak, R. D. (2016). A characterization of deterministic sampling patterns for low-rank matrix completion. *IEEE J. Selected Top. Sig. Proc.*, 10 no. 4, 623–636.

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

[16] Sportisse, A., Boyer, C. and Josse, J. (2018). Imputation and low-rank estimation with Missing Non At Random data. *arXiv preprint arXiv:1812.11409*.

DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY, SEQUOIA HALL, 390 SERRA MALL, STANFORD, CA 94305

E-mail address: souravc@stanford.edu