A deterministic theory of low rank matrix completion

Sourav Chatterjee

Low rank matrix completion

- The problem of reconstructing a large low rank matrix from a subset of revealed entries has attracted widespread attention in statistics and machine learning in the last ten years.

Low rank matrix completion

- The problem of reconstructing a large low rank matrix from a subset of revealed entries has attracted widespread attention in statistics and machine learning in the last ten years.
- Matrix completion in classical linear algebra is restricted to matrices with special structure, such as positive definite matrices.

Low rank matrix completion

- The problem of reconstructing a large low rank matrix from a subset of revealed entries has attracted widespread attention in statistics and machine learning in the last ten years.
- Matrix completion in classical linear algebra is restricted to matrices with special structure, such as positive definite matrices.
- Much of the modern literature on low rank matrix completion starts with the assumption that a certain fraction of entries are missing uniformly at random.

Low rank matrix completion

- The problem of reconstructing a large low rank matrix from a subset of revealed entries has attracted widespread attention in statistics and machine learning in the last ten years.
- Matrix completion in classical linear algebra is restricted to matrices with special structure, such as positive definite matrices.
- Much of the modern literature on low rank matrix completion starts with the assumption that a certain fraction of entries are missing uniformly at random.
- This assumption, while unrealistic, allows researchers to prove many beautiful theorems.

Low rank matrix completion

- The problem of reconstructing a large low rank matrix from a subset of revealed entries has attracted widespread attention in statistics and machine learning in the last ten years.
- Matrix completion in classical linear algebra is restricted to matrices with special structure, such as positive definite matrices.
- Much of the modern literature on low rank matrix completion starts with the assumption that a certain fraction of entries are missing uniformly at random.
- 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.

Content of this talk

- In this talk, I will give a complete characterization of missing patterns that allow approximate completion of large low rank matrices. This is from the following paper:

  Sourav Chatterjee (2020). A deterministic theory of low rank matrix completion. *IEEE Trans. Inf. Theory,* **66** no. 12, 8046–8055.

Content of this talk

- In this talk, I will give a complete characterization of missing patterns that allow approximate completion of large low rank matrices. This is from the following paper:

  Sourav Chatterjee (2020). A deterministic theory of low rank matrix completion. *IEEE Trans. Inf. Theory,* **66** no. 12, 8046–8055.

- The characterization will be in the language of graph limit theory.

When completion is not possible

- 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.

When completion is not possible

- 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.

When completion is not possible

- 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.

- By 'cannot be completed', we mean that there are multiple very different ways to complete, even under the low rank assumption.

When completion is not possible

- 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.

- By 'cannot be completed', we mean that there are 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.

Another example

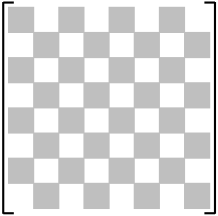
- The previous example suggests that the set of revealed entries has to be in some sense 'dense' in the set of all entries for the matrix to be recoverable.

Another example

- The previous 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).

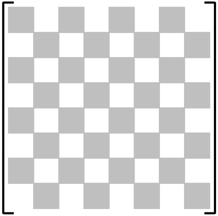
Another example

- The previous 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’:



Another example

- The previous 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':



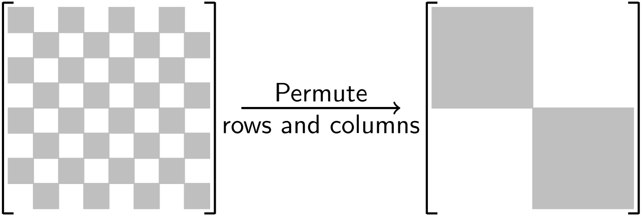
- Yet, we will now argue that recovery is not possible even if the rank of the matrix is as small as three.

Why not?

- To see this, relabel the rows and columns 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.

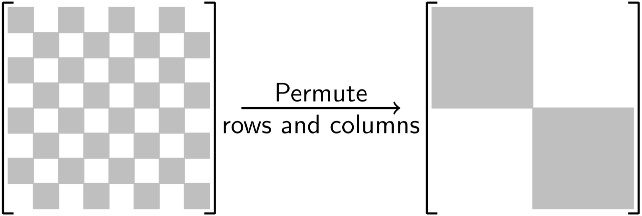
Why not?

- To see this, relabel the rows and columns 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.
- The missing entries in the relabeled matrix look like the following:



Why not?

- To see this, relabel the rows and columns 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.
- The missing entries in the relabeled matrix look like the following:



- Clearly, the missing blocks cannot be recovered reliably if the rank is three or higher.

So, when is completion possible?

- The problem with the second example is that the rows and columns could be relabeled so that the pattern of revealed entries is no longer ‘dense’.

So, when is completion possible?

- The problem with the second example 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.*

So, when is completion possible?

- The problem with the second example 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 first main theorem of this talk.

So, when is completion possible?

- The problem with the second example 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 first main theorem of this talk.
- The precise statement is given in the language of graph limit theory.

So, when is completion possible?

- The problem with the second example 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 first main theorem of this talk.
- The precise statement is given in the language of graph limit theory.
- The second main result is that a modification of a popular method of low rank matrix completion by nuclear norm minimization (due to Candès and Recht) succeeds in approximately recovering the full matrix *whenever the above condition holds.*

Some matrix norms

- Let A be an $m \times n$ matrix. We define the averaged Frobenius norm of A as

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

Some matrix norms

- Let A be an $m \times n$ matrix. We define the averaged Frobenius norm of A as
$$\|A\|_{\bar{F}} := \left(\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right)^{1/2}.$$
- If $\sigma_1, \ldots, \sigma_r$ are the non-zero singular values of A, the nuclear norm of A is defined as
$$\|A\|_* := \sum_{i=1}^{r} \sigma_i.$$

Some matrix norms

- Let A be an $m \times n$ matrix. We define the averaged Frobenius norm of A as

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

- 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

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

Some matrix norms

- Let A be an $m \times n$ matrix. We define the averaged Frobenius norm of A as
$$\|A\|_{\bar{F}} := \left(\frac{1}{mn}\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2\right)^{1/2}.$$
- 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
$$\|A\|_\infty := \max_{i,j} |a_{ij}|.$$
- Finally, the cut norm of A is defined as
$$\|A\|_\square := \frac{1}{mn}\max\{|x^T Ay| : x \in \mathbb{R}^m,\ y \in \mathbb{R}^n,\ \|x\|_\infty \le 1,\ \|y\|_\infty \le 1\}.$$

Cut distance

- Let S_n be the group of all permutations of $\{1, \ldots, n\}$.

Cut distance

- Let S_n be the group of all permutations of $\{1, \ldots, n\}$.
- For $\pi \in S_m$ and $\tau \in S_n$, let $A^{\pi,\tau}$ be the matrix whose $(i,j)^{\text{th}}$ entry is $a_{\pi(i)\tau(j)}$.

Cut distance

- Let S_n be the group of all permutations of $\{1, \ldots, n\}$.
- 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

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

Binary matrices

- We will say that a matrix is a binary matrix if each of its entries is either 0 or 1.

Binary matrices

- 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.

Hadamard product

- 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 product

- 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 $a_{ij} b_{ij}$.
- 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.

Recoverability

- As explained earlier, certain patterns of revealed entries may not suffice for approximately recovering the full matrix, whereas other patterns may suffice.

Recoverability

- As explained 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.

Recoverability

- As explained 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.
- 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.

Recoverability

- As explained 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.
- 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. This is done in the next slide.

Stable recovery

Definition

Let $\{P_k\}_{k\geq 1}$ be a sequence of binary matrices, possibly with different dimensions. 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, for each k, A_k and B_k have the same dimensions as P_k. Suppose that there are numbers K and L such that $\mathrm{rank}(A_k)$ and $\mathrm{rank}(B_k)$ 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

$$\limsup_{k\to\infty} \|(A_k - B_k) \circ P_k\|_{\bar{F}} \leq \delta,$$

then

$$\limsup_{k\to\infty} \|A_k - B_k\|_{\bar{F}} \leq \varepsilon.$$

Remarks

- 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.

Remarks

- 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.
- The two examples discussed earlier do not admit stable recovery of low rank matrices.

What is a criterion for stable recovery?

- To verify that a sequence $\{P_k\}_{k\geq 1}$ admits stable recovery of low rank matrices, one needs to verify the stated condition for all sequences $\{A_k\}_{k\geq 1}$ and $\{B_k\}_{k\geq 1}$.

What is a criterion for stable recovery?

- To verify that a sequence $\{P_k\}_{k\geq 1}$ admits stable recovery of low rank matrices, 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 $\{P_k\}_{k\geq 1}$.

What is a criterion for stable recovery?

- To verify that a sequence $\{P_k\}_{k\geq 1}$ admits stable recovery of low rank matrices, 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 $\{P_k\}_{k\geq 1}$.
- Our first main result gives such a criterion. To state this result, we need the language of graph limit theory.

Asymmetric graphons

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

Asymmetric graphons

- In graph limit theory, 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.

Asymmetric graphons

- In graph limit theory, 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

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

Discretization of asymmetric graphons

Definition

If W is an asymmetric graphon and m and n are two positive integers, we define the $m \times n$ discretization 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,

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

Convergence of matrices to asymmetric graphons

- 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

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

where $W_{m,n}$ is the $m \times n$ discretization of W.

Convergence of matrices to asymmetric graphons

- 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 $W_{m,n}$ is the $m \times n$ discretization of W.

- 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 $k \to \infty$.

Convergence of matrices to asymmetric graphons

- 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 $W_{m,n}$ is the $m \times n$ discretization of W.
- 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 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. We can do the same for asymmetric graphons.

Subsequential limits of binary matrices

- We will use asymmetric graphons to represent limits of binary matrices.

Subsequential limits of binary matrices

- We will use asymmetric graphons to represent limits of binary matrices.
- Not every sequence has a limit, but the following result shows that subsequential limits always exist.

Subsequential limits of binary matrices

- We will use asymmetric graphons to represent limits of binary matrices.
- Not every sequence has a limit, but the following result shows that subsequential limits always exist.

Theorem (C., 2020)

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

Subsequential limits of binary matrices

- We will use asymmetric graphons to represent limits of binary matrices.
- Not every sequence has a limit, but the following result shows that subsequential limits always exist.

Theorem (C., 2020)

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, due to Lovász and Szegedy.

First main result

- We want to give a necessary and sufficient condition for a sequence of binary matrices to admit stable recovery of low rank matrices.

First main result

- We want to give a necessary and sufficient condition for a sequence of binary matrices to admit stable recovery of low rank matrices.
- Because of the compactness theorem displayed in the previous slide, it suffices to only consider convergent sequences of binary matrices.

First main result

- We want to give a necessary and sufficient condition for a sequence of binary matrices to admit stable recovery of low rank matrices.
- Because of the compactness theorem displayed in the previous slide, it suffices to only consider convergent sequences of binary matrices.

Theorem (C., 2020)

A sequence of binary matrices with dimensions tending to infinity and converging to an asymmetric graphon W admits stable recovery of low rank matrices if and only if W is nonzero almost everywhere (w.r.t. Lebesgue measure).

Understanding the result

- To understand this result, first consider the case of entries missing uniformly at random.

Understanding the result

- To understand this result, first consider the case of entries missing uniformly at random.
- Suppose that each entry is revealed with probability p, independently of each other.

Understanding the result

- To understand this result, first consider the case of entries missing uniformly 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 $[0,1]^2$.

Understanding the result

- To understand this result, first consider the case of entries missing uniformly 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 $p > 0$, the theorem tells us that this sequence of revelation patterns admits stable recovery of low rank matrices.

Understanding the result

- To understand this result, first consider the case of entries missing uniformly 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 $p > 0$, the theorem tells us that this sequence of revelation patterns admits stable recovery of low rank matrices.
- On the other hand, consider the example where only the top half of the rows are revealed.

Understanding the result

- To understand this result, first consider the case of entries missing uniformly 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 $p > 0$, the theorem tells us that this sequence of revelation patterns admits stable recovery of low rank matrices.
- On the other hand, consider the example 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.

Resolving a paradox

- At this point one may be puzzled by the fact that the theorem implies that stable 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.

Resolving a paradox

- At this point one may be puzzled by the fact that the theorem implies that stable 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’.

Resolving a paradox

- At this point one may be puzzled by the fact that the theorem implies that stable 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.

Resolving a paradox

- At this point one may be puzzled by the fact that the theorem implies that stable 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.

How to recover the matrix

- Our theorem gives an intrinsic characterization of recoverability in terms of the locations of revealed entries.

How to recover the matrix

- Our theorem gives 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.

How to recover the matrix

- Our theorem gives 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.

How to recover the matrix

- Our theorem gives 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.
- 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.

How to recover the matrix

- Our theorem gives 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.
- 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.

Definition

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 $\|B\|_\infty \leq L$.

Remarks

- The assumption of a known upper bound on the ℓ^∞ norm is not unrealistic. Usually such upper bounds are known.

Remarks

- The assumption of a known upper bound on the ℓ^∞ norm is not unrealistic. Usually such upper bounds are known.
- The modified estimator is the solution of a convex optimization problem, just like the original estimator, and should therefore be easily computable if the dimensions are not too large.

Second main result

- The following theorem shows that the modified Candès–Recht algorithm is able to approximately recover the full matrix whenever the pattern of revealed entries allows stable recovery.

Second main result

- The following theorem shows that the modified Candès–Recht algorithm is able to approximately recover the full matrix whenever the pattern of revealed entries allows stable recovery.

Theorem (C., 2020)

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 $\operatorname{rank}(A_k)$ and $\|A_k\|_\infty$ are uniformly bounded over k, and a uniform upper bound on $\|A_k\|_\infty$ is known. Let $\hat{A}_k$ be the modified Candès–Recht estimate of A_k when the locations of the revealed entries are given by P_k. Then $\lim_{k\to\infty} \|\hat{A}_k - A_k\|_{\bar{F}} = 0$.

An application

- Consider a setting where the probability that an entry is missing is an unknown function of the value of the entry.

An application

- Consider a setting where the probability that an entry is missing is an unknown function of the value of the entry.
- Under a Lipschitz assumption on the function, this class of completion problems was shown to be solvable by the modified Candès–Recht estimator in the following paper:

  Sohom Bhattacharya and Sourav Chatterjee (2022). Matrix completion with data-dependent missingness probabilities. *IEEE Trans. Inf. Theory.*, **68** no. 10, 6762–6773.

An application

- Consider a setting where the probability that an entry is missing is an unknown function of the value of the entry.
- Under a Lipschitz assumption on the function, this class of completion problems was shown to be solvable by the modified Candès–Recht estimator in the following paper:

  Sohom Bhattacharya and Sourav Chatterjee (2022). Matrix completion with data-dependent missingness probabilities. *IEEE Trans. Inf. Theory.*, **68** no. 10, 6762–6773.

- The results presented in this talk were key to the proof.

Open problems

- 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?

Open problems

- 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?
- Developing non-asymptotic versions of the theorems 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.

Open problems, contd.

- Developing a version of the theory that works for recovery of sparsely revealed matrices is an important open question.

Open problems, contd.

- Developing a version of the theory that works for recovery of sparsely revealed matrices is an important open question.
- It is not clear if the Candès–Recht algorithm indeed needs to be modified, or if the original version is good enough in our setting.

Open problems, contd.

- Developing a version of the theory that works for recovery of sparsely revealed matrices is an important open question.
- It is not clear if the Candès–Recht algorithm indeed needs to be modified, or if the original version is good enough in our setting.
- The Candès–Recht algorithm is rather slow for very large matrices. Is there a faster algorithm that can take its place in our setting?