Exact Matrix Completion via Convex Optimization

Emmanuel J. Candès and Benjamin Recht

Minsu Kim[1], Joonyoung Yi[1]

[1]School of Computing, KAIST

CS592, Advanced Machine Learning:
High-Dimensional Data Analysis

April, 2018

Table of Contents

1. Introduction
2. Results of Paper
3. Rationality of Incoherence Assumption
4. Proofs
5. Experiments
6. Discussion

Table of Contents

1. Introduction
2. Results of Paper
3. Rationality of Incoherence Assumption
4. Proofs
5. Experiments
6. Discussion

Matrix Completion Problem(Informal)

- Given: Matrix M with some entries are missing
- Goal: Complete the matrix M

	1	2	3	4	5	6	7	8	9	10
1		4		2	4					
2	3		3	1			3		3	
3		3	2	3					4	
4			2			4		1	2	5
5	3			3			1			
6			2					3	2	

Matrix Completion Problem(Informal)

- Given: Matrix M with some entries are missing
- Goal: Complete the matrix M

	1	2	3	4	5	6	7	8	9	10
1		4		2	4					
2	3		3	1			3		3	
3		3	2	3					4	
4			2			4		1	2	5
5	3			3			1			
6			2					3	2	

- In general, it is impossible.
- But, it can be possible with *low-rank assumption*.

Low-rank Assumption

Low-rank assumption

For $n_1 \times n_2$ matrix M of rank r, assume the $\min(n_1, n_2) \gg r$.

- Why this assumption is needed?

Low-rank Assumption

Low-rank assumption

For $n_1 \times n_2$ matrix M of rank r, assume the $\min(n_1, n_2) \gg r$.

- Why this assumption is needed?
- For simplicity, think about $n \times n$ matrix M of rank r.
- It has $(2n - r)r$ degrees of freedom.
  - Degree of freedom is calculated by counting parameters in the SVD.
    - (The number of singular values)
      $+$ (degree of freedom of left singular vectors)
      $+$ (degree of freedom of right singular vectors)
      $= r + ((2n - r - 1) \times r)/2 + ((2n - r - 1) \times r)/2$
  - Considerably smaller than n^2.
    - With low-rank assumption, degree of freedom is reduced by about $2r/n$.

Low-rank Assumption

Low-rank assumption

For $n_1 \times n_2$ matrix M of rank r, assume the $\min(n_1, n_2) \gg r$.

- Why this assumption is needed?
- For simplicity, think about $n \times n$ matrix M of rank r.
- It has $(2n - r)r$ degrees of freedom.
  - Degree of freedom is calculated by counting parameters in the SVD.
    - (The number of singular values)
      $+$ (degree of freedom of left singular vectors)
      $+$ (degree of freedom of right singular vectors)
      $= r + ((2n - r - 1) \times r)/2 + ((2n - r - 1) \times r)/2$
  - Considerably smaller than n^2.
    - With low-rank assumption, degree of freedom is reduced by about $2r/n$.
- This assumption is similar to sparsity assumption in the Lasso paper we've learned.

Which Matrices?

- Consider the rank-1 matrix M:

$$M = e_1 e_n^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

  - Let m be the number of observed entries of M.

Which Matrices?

- Consider the rank-1 matrix M:

$$M = e_1 e_n^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

  - Let m be the number of observed entries of M.
  - Then, we can see only 0 with probability $1 - m/n^2$.

Which Matrices?

- Consider the rank-1 matrix M:

$$M = e_1 e_n^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

  - Let m be the number of observed entries of M.
  - Then, we can see only 0 with probability $1 - m/n^2$.
  - If sample set doesn't contain 1, we can not complete matrix exactly.

Which Matrices?

- Consider the rank-1 matrix M:

$$M = e_1 e_n^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

  - Let m be the number of observed entries of M.
  - Then, we can see only 0 with probability $1 - m/n^2$.
  - If sample set doesn't contain 1, we can not complete matrix exactly.
- So, impossible to recover all low-rank matrices.

Which Matrices?

- Consider the SVD of a matrix $M = \sum_{k=1}^{r} \sigma_k u_k v_k^*$ where u_k's and v_k's are the left and right singular vectors, and the σ_k's are the singular values.

Random orthogonal model

The family $\{u_k\}_{1 \le k \le r}$ is selected uniformly at random among all families of r orthonormal vectors, and similarly for the family $\{v_k\}_{1 \le k \le r}$.

- With this model, we can see, intuitively, small probability occuring exception as we mentioned.
  - Formal proofs are provided by Lemma 2.1 and 2.2.

Which Sampling Sets?

- Clearly, we can not recover M_{ij} if the sampling set avoids i-th row or j-th column.
- So, we have to define sampling set to recover.

Which Sampling Sets?

- Clearly, we can not recover M_{ij} if the sampling set avoids i-th row or j-th column.
- So, we have to define sampling set to recover.

Definition of sampling set Ω

If M_{ij} is observed, $(i, j) \in \Omega$. The cardinality of Ω is m.

Uniform sampling assumption

The set Ω sampled uniformly at random.

Which Sampling Sets?

- Clearly, we can not recover M_{ij} if the sampling set avoids i-th row or j-th column.
- So, we have to define sampling set to recover.

Definition of sampling set Ω

If M_{ij} is observed, $(i, j) \in \Omega$. The cardinality of Ω is m.

Uniform sampling assumption

The set Ω sampled uniformly at random.

- Additionally, this paper introduce sampling operator $\mathcal{R}_\Omega$ for convenience.

Sampling Operator $\mathcal{R}_\Omega$

Let $\mathcal{R}_\Omega : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^{\|\Omega\|}$ be the sampling operator which extracts the observed entries, $\mathcal{R}_\Omega(X) = (X_{ij})_{ij \in \Omega}$.

Which Algorithm?

Problem 1.3: Formal Matrix Completion Problem

$$\begin{aligned} &\text{minimize } \operatorname{rank}(X) \\ &\text{subject to } \mathcal{R}_\Omega(X) = \mathcal{R}_\Omega(M) \end{aligned}$$

Which Algorithm?

Problem 1.3: Formal Matrix Completion Problem

$$\begin{aligned}&\text{minimize } \operatorname{rank}(X)\\&\text{subject to } \mathcal{R}_\Omega(X) = \mathcal{R}_\Omega(M)\end{aligned}$$

- Unfortunately, this problem is NP-hard and non-convex.
  - rank function is not convex.
- All known algorithms require exponential time to n.

Which Algorithm?

Problem 1.3: Formal Matrix Completion Problem

$$\text{minimize } \operatorname{rank}(X)$$
$$\text{subject to } \mathcal{R}_\Omega(X) = \mathcal{R}_\Omega(M)$$

- Unfortunately, this problem is NP-hard and non-convex.
  - rank function is not convex.
- All known algorithms require exponential time to n.

Problem 1.5: Convex Relaxation to Matrix Completion Problem

$$\text{minimize } \|X\|_*$$
$$\text{subject to } \mathcal{R}_\Omega(X) = \mathcal{R}_\Omega(M)$$

Which Algorithm?

Problem 1.3: Formal Matrix Completion Problem

$$\text{minimize } \operatorname{rank}(X)$$
$$\text{subject to } \mathcal{R}_\Omega(X) = \mathcal{R}_\Omega(M)$$

- Unfortunately, this problem is NP-hard and non-convex.
  - rank function is not convex.
- All known algorithms require exponential time to n.

Problem 1.5: Convex Relaxation to Matrix Completion Problem

$$\text{minimize } \|X\|_*$$
$$\text{subject to } \mathcal{R}_\Omega(X) = \mathcal{R}_\Omega(M)$$

- Nuclear norm $\|X\|_*$ is surrogate of rank.
  - Also, nuclear norm is convex function.
  - This heuristic is introduced by [Fazel et al. 2002].

Nuclear Norm Relaxation is Reasonable?(Informal)

- We already know that l_1-norm is surrogate to l_0-norm in the Lasso paper.

Nuclear Norm Relaxation is Reasonable?(Informal)

- We already know that l_1-norm is surrogate to l_0-norm in the Lasso paper.

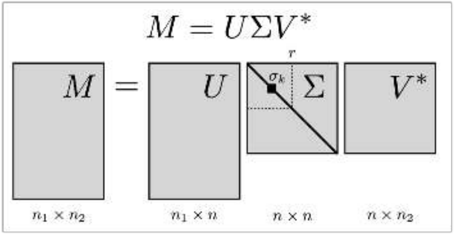


Figure: http://public.lanl.gov/mewall/kluwer2002.html

- Let vector $\sigma = (\sigma_1, \sigma_2, \cdots, \sigma_n)$, σ_i is i-th singular value of matrix M.
- The rank function is l_0-norm of σ vector.
- The nuclear norm function is l_1-norm of σ vector.

How to Solve Nuclear Norm Minimization?

- This method is also suggested by Fazel.
- Convex Relaxation to Matrix Completion Problem(Nuclear Norm Minimization) can be solved by Semi-definite Programming(SDP).

How to Solve Nuclear Norm Minimization?

- This method is also suggested by Fazel.
- Convex Relaxation to Matrix Completion Problem(Nuclear Norm Minimization) can be solved by Semi-definite Programming(SDP).

- Let SVD of matrix $X = U\Sigma V^*$.
- Define $W_1 = U\Sigma U^*, W_2 = V\Sigma V^*, X' = \begin{bmatrix} W_1 & X \\ X^* & W_2 \end{bmatrix}$.

How to Solve Nuclear Norm Minimization?

- This method is also suggested by Fazel.
- Convex Relaxation to Matrix Completion Problem(Nuclear Norm Minimization) can be solved by Semi-definite Programming(SDP).

- Let SVD of matrix $X = U\Sigma V^*$.
- Define $W_1 = U\Sigma U^*, W_2 = V\Sigma V^*, X' = \begin{bmatrix} W_1 & X \\ X^* & W_2 \end{bmatrix}$.

- Under above settings, the following properties satisfy:
- $\|X\|_* = \|W_1\|_* = \|W_2\|_*$
- $X' \succeq 0$ (positive semidefinite matrix)
  - $\because X' = \begin{bmatrix} U \\ V \end{bmatrix} \Sigma \begin{bmatrix} U \\ V \end{bmatrix}^* = \left(\begin{bmatrix} U \\ V \end{bmatrix} \sqrt{\Sigma} \right) \left(\begin{bmatrix} U \\ V \end{bmatrix} \sqrt{\Sigma} \right)^* \succeq 0$
    - $\because \forall$ matrix $A, AA^* \succeq 0$
- $\forall$ symmetric matrix X, trace$(X) = \|X\|_*$
  - `https://chrischoy.github.io/research/matrix-norms/`

How to Solve Nuclear Norm Minimization?

- Therefore, the Problem 1.5 can be reduced to

Reduced form of nuclear norm minimization

$$\begin{aligned} &\text{minimize } \operatorname{trace}(X') \\ &\text{subject to } \mathcal{R}_\Omega(X) = \mathcal{R}_\Omega(M), X' \succeq 0 \end{aligned}$$

- $X' = \begin{bmatrix} W_1 & X \\ X^* & W_2 \end{bmatrix}$
- $\operatorname{trace}(X') = \operatorname{trace}(W_1) + \operatorname{trace}(W_2) = \|X\|_* + \|X\|_* = 2\|X\|_*$

How to Solve Nuclear Norm Minimization?

- Now, reduced form can be solved by SDP.
  - Many algorithms solving SDP already exist.

Semidefinite Programming(SDP)

$$\begin{aligned} \text{minimize}_{X' \in \mathbb{S}^n} \quad & \langle C, X' \rangle_{\mathbb{S}^n} \\ \text{subject to} \quad & \langle A_k, X' \rangle_{\mathbb{S}^n} = b_k, \quad k = 1, \ldots, m \\ & X' \succeq 0 \end{aligned}$$

- $\mathbb{S}^n$: $n \times n$ symmetric matrix.
- $\langle X, Y \rangle = \text{trace}(X^* Y)$ is element-wise product.
- A_k is all 0s without one entry which is 1 for $(i, j) \in \Omega$.

Table of Contents

1. Introduction
2. Results of Paper
3. Rationality of Incoherence Assumption
4. Proofs
5. Experiments
6. Discussion

Nuclear Norm Relaxation is Reasonable?(Formal)

- A first typical result:
  - M: $n_1 \times n_2$ matrix of rank r, $n = \max(n_1, n_2)$.

Theorem 1.1(General)

- M obeys the random orthogonal model.
- With uniformly random sampling assumption.

Then, $\exists$ constants C, c such that if

$$m \geq Cn^{5/4} r \log n$$

the minimizer to the problem 1.5 is unique and equal to M with probability at least $1 - cn^{-3}$.

Nuclear Norm Relaxation is Reasonable?(Formal)

- A first typical result:
  - M: $n_1 \times n_2$ matrix of rank r, $n = \max(n_1, n_2)$.

Theorem 1.1(General)

- M obeys the random orthogonal model.
- With uniformly random sampling assumption.

Then, $\exists$ constants C, c such that if

$$m \geq C n^{5/4} r \log n$$

the minimizer to the problem 1.5 is unique and equal to M with probability at least $1 - cn^{-3}$.

- For low-rank case, tighter bound is given.

Theorem 1.1(Low-rank)

if $r \leq n^{1/5}$, the recovery is exact with same probability provided that

$$m \geq C n^{6/5} r \log n$$

Nuclear Norm Relaxation is Reasonable?(Formal)

- Asymptotically,
  - Theorem 1.1(General): $m = \Omega(n^{1.25} r \log n)$
  - Theorem 1.1(Low-rank): $m = \Omega(n^{1.2} r \log n)$
  - This bound is not tight(Plug $r \leftarrow n$ into Theorem 1.1 general case).

Nuclear Norm Relaxation is Reasonable?(Formal)

- Asymptotically,
  - Theorem 1.1(General): $m = \Omega(n^{1.25} r \log n)$
  - Theorem 1.1(Low-rank): $m = \Omega(n^{1.2} r \log n)$
  - This bound is not tight(Plug $r \leftarrow n$ into Theorem 1.1 general case).
- This theorem tells:
  - With some conditions, we can recover matrix exactly.
  - Problem 1.3 and Problem 1.5 are formally equivalent with some conditions.
    - Rationality of nuclear norm relaxation.

Relaxation of Random Orthogonal Model

- Recall why we introduce random orthogonal model:

$$M = e_1 e_n^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

Relaxation of Random Orthogonal Model

- Recall why we introduce random orthogonal model:

$$M = e_1 e_n^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

- More generally, it is hard to recover if the singular vectors of the matrix M are similar to standard basis.
  - Because, information is highly concentrated on specific region.
  - Hence, the singular vectors need to be sufficiently spread.

Relaxation of Random Orthogonal Model

- Recall why we introduce random orthogonal model:

$$M = e_1 e_n^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

- More generally, it is hard to recover if the singular vectors of the matrix M are similar to standard basis.
  - Because, information is highly concentrated on specific region.
  - Hence, the singular vectors need to be sufficiently spread.
- Random orthogonal model is one of the models to tackle this problem.
  - We want to relax this assumption of model. Why?

Relaxation of Random Orthogonal Model

- Recall why we introduce random orthogonal model:

$$M = e_1 e_n^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

- More generally, it is hard to recover if the singular vectors of the matrix M are similar to standard basis.
  - Because, information is highly concentrated on specific region.
  - Hence, the singular vectors need to be sufficiently spread.
- Random orthogonal model is one of the models to tackle this problem.
  - We want to relax this assumption of model. Why?
  - To get more general theorem that covers a much larger set of matrices M.

Coherence

Definition 1.2

Let U be a subspace or $\mathbb{R}^n$ of dimension r and $\boldsymbol{P}_U$ be the orthogonal projection onto U. Then the coherence of U is defined to be

$$\mu(U) \equiv \frac{n}{r} \max_{1 \le i \le n} \|\mathbf{P}_U \mathbf{e}_i\|^2$$

Coherence

Definition 1.2

Let U be a subspace or $\mathbb{R}^n$ of dimension r and $\boldsymbol{P}_U$ be the orthogonal projection onto U. Then the coherence of U is defined to be

$$\mu(U) \equiv \frac{n}{r} \max_{1 \le i \le n} \|\mathbf{P}_U \mathbf{e}_i\|^2$$

Examples

Incoherence Example(Minimum Case) If U is spanned by vectors whose entries all have magnitude $1/\sqrt{n}$, $\mu(U) = 1$.

Examples

Coherence Example(Maximum Case) If U contains a standard basis element, $\mu(U) = n/r$.

- The more similar to standard basis element, $\mu(U)$ is larger.
- The more far from standard basis element, $\mu(U)$ is smaller.

Coherence

- From definition 1.2, we define two assumptions $\boldsymbol{A0}$ and $\boldsymbol{A1}$.

Assumption $\boldsymbol{A0}$

The coherences obey $\max(\mu(U), \mu(V)) \leq \mu_0$ for some positive μ_0.

Assumption $\boldsymbol{A1}$

The $n_1 \times n_2$ matrix $\Sigma_{1 \leq k \leq r} u_k v_k^*$ has maximum entry bounded by $\mu_1 \sqrt{r/(n_1 n_2)}$ in absolute value for some positive μ_1.

- A1 is always hold with $\mu_1 = \mu_0 \sqrt{r}$.

Coherence

- From definition 1.2, we define two assumptions $\boldsymbol{A0}$ and $\boldsymbol{A1}$.

Assumption $\boldsymbol{A0}$

The coherences obey $\max(\mu(U), \mu(V)) \le \mu_0$ for some positive μ_0.

Assumption $\boldsymbol{A1}$

The $n_1 \times n_2$ matrix $\Sigma_{1 \le k \le r} u_k v_k^*$ has maximum entry bounded by $\mu_1 \sqrt{r/(n_1 n_2)}$ in absolute value for some positive μ_1.

- A1 is always hold with $\mu_1 = \mu_0 \sqrt{r}$.
- This paper insists that A0, A1 are more general compared to random orthogonal model.
  - Lemma 2.2 proves that random orthogonal model is special case of A0, A1.

Coherence

- Let's investigate an quite natural assumption.
  - The maximum entries of left and right singular matrix are bounded.

Coherence

- Let's investigate an quite natural assumption.
    - The maximum entries of left and right singular matrix are bounded.

Assumption 1.12

Assume that the u_j and v_j's obey

$$\max_{ij} \|\langle e_i, u_j\rangle\|^2 \leq \mu_B/n, \max_{ij} \|\langle e_i, v_j\rangle\|^2 \leq \mu_B/n,$$

for some value of $\mu_B = O(1)$.

Coherence

- Let's investigate an quite natural assumption.
  - The maximum entries of left and right singular matrix are bounded.

Assumption 1.12

Assume that the u_j and v_j's obey

$$\max_{ij} \|\langle e_i, u_j\rangle\|^2 \leq \mu_B/n, \max_{ij} \|\langle e_i, v_j\rangle\|^2 \leq \mu_B/n,$$

for some value of $\mu_B = O(1)$.

- With this assumption, assumptions $\boldsymbol{A0}, \boldsymbol{A1}$ hold.
  - $\mu(U), \mu(V) \leq \mu_B$
  - $\mu_1 = O(\sqrt{\log n})$ (Proved by Lemma 2.1)

Main Result

With incoherence conditions, relax Theorem 1.1 to Theorem 1.3.

Theorem 1.3(General)

- M obeys $\boldsymbol{A0}$ and $\boldsymbol{A1}$.
- With uniformly random sampling assumption.

Then, $\exists$ constants C, c such that if

$$m \geq C\text{max}(\mu_1^2, \mu_0^{1/2}\mu_1, \mu_0 n^{1/4})nr(\beta \log n)$$

for some $\beta > 2$, then the minimizer to the problem (1.5) is unique and equal to M with probability at least $1 - cn^{-\beta}$.

Main Result

With incoherence conditions, relax Theorem 1.1 to Theorem 1.3.

Theorem 1.3(General)

- M obeys $\boldsymbol{A0}$ and $\boldsymbol{A1}$.
- With uniformly random sampling assumption.

Then, $\exists$ constants C, c such that if

$$m \geq C\text{max}(\mu_1^2, \mu_0^{1/2}\mu_1, \mu_0 n^{1/4})nr(\beta \log n)$$

for some $\beta > 2$, then the minimizer to the problem (1.5) is unique and equal to M with probability at least $1 - cn^{-\beta}$.

Theorem 1.3(Low-rank)

If $r \leq \mu_0^{-1} n^{1/5}$, the recovery is exact with same prob. provided that

$$m \geq C\mu_0 n^{6/5} r(\beta \log n).$$

Main Result

With incoherence conditions, relax Theorem 1.1 to Theorem 1.3.

Theorem 1.3(General)

- M obeys $\boldsymbol{A0}$ and $\boldsymbol{A1}$.
- With uniformly random sampling assumption.

Then, $\exists$ constants C, c such that if

$$m \geq C\max(\mu_1^2, \mu_0^{1/2}\mu_1, \mu_0 n^{1/4})nr(\beta \log n)$$

for some $\beta > 2$, then the minimizer to the problem (1.5) is unique and equal to M with probability at least $1 - cn^{-\beta}$.

Theorem 1.3(Low-rank)

If $r \leq \mu_0^{-1} n^{1/5}$, the recovery is exact with same prob. provided that

$$m \geq C\mu_0 n^{6/5} r(\beta \log n).$$

- Asymptotically similar to Theorem 1.1

Connections to Lasso

- Recall: Lasso Paper

Compressive sampling(or compressed sensing, matrix sensing)

$A\boldsymbol{x} = \boldsymbol{b}$ (A is a design matrix, $\boldsymbol{b}$ is a observation vector)

- In the Lasso paper, they add the sparsity assumption with $\boldsymbol{x}$ to tackle the problem that problem dimension is much larger than observations$(n \ll p)$.
- $\boldsymbol{x}$ is k-sparse in the Fourier domain, it can be perfectly recovered with high probability by l_1 minimization when $m = \Omega(k \log n)$.

Connections to Lasso

- Recall: Lasso Paper

Compressive sampling(or compressed sensing, matrix sensing)

$A\boldsymbol{x} = \boldsymbol{b}$ (A is a design matrix, $\boldsymbol{b}$ is a observation vector)

  - In the Lasso paper, they add the sparsity assumption with $\boldsymbol{x}$ to tackle the problem that problem dimension is much larger than observations$(n \ll p)$.
  - $\boldsymbol{x}$ is k-sparse in the Fourier domain, it can be perfectly recovered with high probability by l_1 minimization when $m = \Omega(k \log n)$.

- Also, this paper claim that they generalizes the notion of incoherence to problems beyond the setting of sparse signal recovery.

Connections to Matrix Sensing Problem

- Original Fazel's problem was
  - Solving matrix sensing problem with low-rank assumption(nuclear norm heuristic) and other assumptions.

Connections to Matrix Sensing Problem

- Original Fazel's problem was
  - Solving matrix sensing problem with low-rank assumption(nuclear norm heuristic) and other assumptions.
- Contribution of this paper compared to original Fazel's work.
  - Extend Fazel's work to matrix completion problem.
  - Define some conditions(including incoherence) to complete matrix exactly.
  - Serve theoretical bound of convex relaxation(nuclear norm minimization).

Table of Contents

1. Introduction
2. Results of Paper
3. Rationality of Incoherence Assumption
4. Proofs
5. Experiments
6. Discussion

Which Matrices are Incoherent?

- Theorem 1.3 says completing matrix is possible if the incoherence conditions are hold.
- Now, we want to find out how incoherence assumptions are reasonable.

Which Matrices are Incoherent?

- Theorem 1.3 says completing matrix is possible if the incoherence conditions are hold.
- Now, we want to find out how incoherence assumptions are reasonable.
- In this section, we'll show that most random matrices are incoherent.
  - Lemma 2.1 shows incoherence assumptions are reasonable.
  - Lemma 2.2 shows a matrix with random orthogonal model satisfies incoherence conditions.

Incoherent Bases Span Incoherent Subspaces

Recall: Assumption 1.12

Assume that the u_j and v_j's obey

$$\max_{ij} \|\langle e_i, u_j\rangle\|^2 \leq \mu_B/n, \max_{ij} \|\langle e_i, v_j\rangle\|^2 \leq \mu_B/n,$$

for some value of $\mu_B = O(1)$.

Incoherent Bases Span Incoherent Subspaces

Recall: Assumption 1.12

Assume that the u_j and v_j's obey

$$\max_{ij} \|\langle e_i, u_j \rangle\|^2 \leq \mu_B/n, \max_{ij} \|\langle e_i, v_j \rangle\|^2 \leq \mu_B/n,$$

for some value of $\mu_B = O(1)$.

- If the assumption 1.12 holds,
  - $\boldsymbol{A0}$ holds with $\mu_0 = \mu_B$(It is trivial).
  - $\boldsymbol{A1}$ holds with $\mu_1 = C\mu_B\sqrt{\log n}$(It is not trivial).
    - $\boldsymbol{A1}$ hold with $\mu_1 = \mu_B\sqrt{r}$, but we can't assume that $r = O(\log n)$.
    - So, we'll prove that $\boldsymbol{A1}$ holds with $\mu_1 = C\mu_B\sqrt{\log n}$.

Incoherent Bases Span Incoherent Subspaces

Lemma 2.1

If assumption 1.12 holds,

$$\boldsymbol{A1} \text{ holds with } \mu_1 = C\mu_B\sqrt{\log n} \text{ with high prob.}$$

Incoherent Bases Span Incoherent Subspaces

Lemma 2.1

If assumption 1.12 holds,

$$\boldsymbol{A1} \text{ holds with } \mu_1 = C\mu_B\sqrt{\log n} \text{ with high prob.}$$

(Proof) Consider the matrix $M = \Sigma_{k=1}^{r} \epsilon_k u'_k v_k^*$ where $\{\epsilon_k\}_{1 \le k \le r}$ is an arbitrary sign sequence and $\epsilon_k u'_k = u_k$.

Incoherent Bases Span Incoherent Subspaces

Lemma 2.1

If assumption 1.12 holds,

$$\boldsymbol{A1} \text{ holds with } \mu_1 = C\mu_B\sqrt{\log n} \text{ with high prob.}$$

(Proof) Consider the matrix $M = \Sigma_{k=1}^{r} \epsilon_k u_k' v_k^*$ where $\{\epsilon_k\}_{1\leq k\leq r}$ is an arbitrary sign sequence and $\epsilon_k u_k' = u_k$.
From this setting, simple applying Hoeffding's inequality,

$$\mathbb{P}(|M_{ij}| \geq \lambda\mu_B\sqrt{r}/n) \leq 2e^{-\lambda^2/2}$$

With setting $\lambda = \sqrt{2\beta\log n}$ and applying union bound,

$$\mathbb{P}(|M| \geq \mu_B\sqrt{2\beta r\log n}/n) \leq 2n^2e^{-\beta}e^{-\beta\log n} = 2n^2n^{-\beta}e^{-\beta} \leq 2n^2n^{-\beta}$$

Plug $\beta \leftarrow \beta + 2$, then Lemma 2.1 holds.

Random Subspaces Span Incoherent Subspaces

- Now, we prove that random orthogonal model obeys the two assumptions $\boldsymbol{A0}$ and $\boldsymbol{A1}$(with appropriate values for the μ's) with large probability.

Random Subspaces Span Incoherent Subspaces

- Now, we prove that random orthogonal model obeys the two assumptions $\boldsymbol{A0}$ and $\boldsymbol{A1}$(with appropriate values for the μ's) with large probability.

- Lemma 2.3 is one of the the results of [Laurent et al. 2000]. Using the result, we can see that Lemma 2.2 is induced from Lemma 2.3.

Lemma 2.2

Set $\bar{r} = \max(r, \log n)$. Then $\exists$ constants C, c such that the random orthogonal model obeys:

1. $\mu(U) = (n/r)\max_i \|\boldsymbol{P}_U e_i\|^2 \leq C\bar{r}/r = \mu_0$,
2. $(n/r)\|\sum_{1\leq k\leq r} u_k v_k^*\|_\infty \leq C\log n\sqrt{\bar{r}/r} = \mu_1$ with high probability.

Random Subspaces Span Incoherent Subspaces

- Now, we prove that random orthogonal model obeys the two assumptions $\boldsymbol{A0}$ and $\boldsymbol{A1}$(with appropriate values for the μ's) with large probability.

- Lemma 2.3 is one of the the results of [Laurent et al. 2000]. Using the result, we can see that Lemma 2.2 is induced from Lemma 2.3.

Lemma 2.2

Set $\bar{r} = \max(r, \log n)$. Then $\exists$ constants C, c such that the random orthogonal model obeys:
1. $\mu(U) = (n/r) \max_i \|\boldsymbol{P}_U e_i\|^2 \leq C\bar{r}/r = \mu_0$,
2. $(n/r)\| \sum_{1 \leq k \leq r} u_k v_k^* \|_\infty \leq C \log n \sqrt{\bar{r}/r} = \mu_1$ with high probability.

- $\mu_0 = O(1), \mu_1 = O(\log n)$.
- As a result of Lemma 2.2, we can see that random subspace span incoherent subspace with high probability.

Table of Contents

1. Introduction
2. Results of Paper
3. Rationality of Incoherence Assumption
4. Proofs
5. Experiments
6. Discussion

Proof Strategy

Problem 1.5: Convex relaxation to matrix completion problem

$$\begin{aligned} &\text{minimize } \|\boldsymbol{X}\|_* \\ &\text{subject to } \mathcal{R}_\Omega(\boldsymbol{X}) = \mathcal{R}_\Omega(\boldsymbol{M}) \end{aligned}$$

- We will prove the original matrix $\boldsymbol{M}$ is the unique solution to the problem 1.5.
- What are conditions for some matrix $\boldsymbol{X}$ to be a unique minimizer?

Subgradient of Nuclear Norm

- Convex optimization theory says:

$$\boldsymbol{X} \text{ is solution to } (1.5) \Leftrightarrow \exists \boldsymbol{\lambda} \in \mathbb{R}^{|\Omega|} \text{ s.t. } \mathcal{R}_\Omega^* \boldsymbol{\lambda} \in \partial \|\boldsymbol{X}\|_*$$

- We know nuclear norm of a matrix is sum of its singular values.
- We may derive subgradient of nuclear norm from this.

Subgradient of Nuclear Norm

- Convex optimization theory says:

$$\boldsymbol{X} \text{ is solution to } (1.5) \Leftrightarrow \exists \boldsymbol{\lambda} \in \mathbb{R}^{|\Omega|} \text{ s.t. } \mathcal{R}_\Omega^* \boldsymbol{\lambda} \in \partial \|\boldsymbol{X}\|_*$$

- We know nuclear norm of a matrix is sum of its singular values.
- We may derive subgradient of nuclear norm from this.

Preliminary: subgradient of matrix norm (from [36])

$$\partial \|\boldsymbol{A}\|_\square = \text{conv}\{\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^*, \boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*, \mathbf{d} \in \partial\phi(\boldsymbol{\sigma})\}$$

where $\boldsymbol{D}, \boldsymbol{\Sigma}$ is $m \times n$ diagonal matrix with $\mathbf{d}, \boldsymbol{\sigma}$ as diagonal entries.

- Since ϕ should be *symmetric gauge function*,
  we put $\phi(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma}\|_1 = \sum |\sigma_i|$.

Subgradient of Nuclear Norm (cont.)

For rank r, $n_1 \times n_2$ matrix $\boldsymbol{A}$, there are r non-zero singular values. Then, we know

$$\partial \|\boldsymbol{\sigma}\|_1 = \{\mathbf{x} \in \mathbb{R}^{\min(n_1, n_2)} \mid x_i = 1 \text{ for } i = 1..r,\ |x_i| \leq 1 \text{ otherwise}\}$$

Subgradient of Nuclear Norm (cont.)

For rank r, $n_1 \times n_2$ matrix $\boldsymbol{A}$, there are r non-zero singular values. Then, we know

$$\partial\|\boldsymbol{\sigma}\|_1 = \{\mathbf{x} \in \mathbb{R}^{\min(n_1,n_2)} \mid x_i = 1 \text{ for } i = 1..r,\ |x_i| \leq 1 \text{ otherwise}\}$$

If we partition singular vectors $\boldsymbol{U}, \boldsymbol{V}$ like..

$$\boldsymbol{U} = [\boldsymbol{U}^{(1)} | \boldsymbol{U}^{(2)}], \boldsymbol{V} = [\boldsymbol{V}^{(1)} | \boldsymbol{V}^{(2)}] \text{ where } \boldsymbol{U}^{(1)}, \boldsymbol{V}^{(1)} \text{ has } r \text{ columns}$$

we get the result

$$\partial\|\boldsymbol{A}\|_* = \{\boldsymbol{U}^{(1)}\boldsymbol{V}^{(1)*} + \boldsymbol{U}^{(2)}\boldsymbol{R}\boldsymbol{V}^{(2)*} \text{ for all } \boldsymbol{R} \in \mathbb{R}^{(n_1-r)\times(n_2-r)}, \|\boldsymbol{R}\|_2 \leq 1\}$$

using the definition of convex hull(ALL convex combinations of elements)

Subgradient of Nuclear Norm (cont.)

$\partial\|\boldsymbol{A}\|_* = \{\, \boldsymbol{U}^{(1)}\boldsymbol{V}^{(1)*} + \boldsymbol{U}^{(2)}\boldsymbol{R}\boldsymbol{V}^{(2)*}$ for all $\boldsymbol{R} \in \mathbb{R}^{(n_1-r)\times(n_2-r)}, \|\boldsymbol{R}\|_2 \leq 1\}$ was expressed in the paper as,

$$\boldsymbol{Y} = \sum_{1\leq k\leq r} \mathbf{u}_k \mathbf{v}_k^* + \boldsymbol{W} \tag{3.4}$$

where $\boldsymbol{W}$ obeys two properties:

- the column space of $\boldsymbol{W}$ is orthogonal to $\boldsymbol{U} \equiv \text{span}(\mathbf{u}_1, ..., \mathbf{u}_r)$. and the row space of $\boldsymbol{W}$ is orthogonal to $\boldsymbol{V} \equiv \text{span}(\mathbf{v}_1, ..., \mathbf{v}_r)$;
- the spectral norm of $\boldsymbol{W}$ is less than or equal to 1.

This says we can decompose $\partial\|\boldsymbol{A}\|_*$ into 2 orthogonal spaces, T(blue one) and $T^\perp$(green one). We define $\mathcal{P}_T$, $\mathcal{P}_{T^\perp}$ as projection onto each spaces.

Conditions for Unique Minimizer

Lemma 3.1 conditions for unique minimizer

Consider a matrix $\boldsymbol{X}_0 = \sum_{k=1}^{r} \sigma_k \mathbf{u}_k \mathbf{v}_k^*$ of rank r which is feasible for the problem (1.5), and suppose that the following *two conditions* hold:

1. there exists a dual point $\boldsymbol{\lambda}$ such that $\boldsymbol{Y} = \mathcal{R}_\Omega^* \lambda$ obeys

$$\mathcal{P}_T(\boldsymbol{Y}) = \sum_{k=1}^{r} \mathbf{u}_k \mathbf{v}_k^*, \quad \|\mathcal{P}_{T^\perp}(\boldsymbol{Y})\|_2 < 1$$

2. the sampling operator $\mathcal{R}_\Omega$ restricted to elements in T is injective.

Then $\boldsymbol{X}_0$ is the unique minimizer.

Removing the equality from spectral norm, and adding condition 2 gives you *unique* minimizer! Now, constructing such $\boldsymbol{Y}$ became important.

Construction of Subgradient

Define $\mathcal{P}_\Omega(\boldsymbol{X}) = X_{ij}$ if $(i,j) \in \Omega$, 0 otherwise. Then set matrix $\boldsymbol{Y}$ as the solution to least square problem

$$\begin{aligned} &\text{minimize} \quad \|\boldsymbol{X}\|_F \\ &\text{subject to} \quad (\mathcal{P}_T\mathcal{P}_\Omega)(\boldsymbol{X}) = \textstyle\sum_{k=1}^{r} \mathbf{u}_k\mathbf{v}_k^*. \end{aligned}$$

This will make $\boldsymbol{Y}$ vanish in Ω^C and have smaller $\|\mathcal{P}_{T^\perp}(\boldsymbol{Y})\|_2$ as well.

Statement 4.2 specification of $\boldsymbol{Y}$

Denote $\mathcal{A}_{\Omega T}(\boldsymbol{M}) = \mathcal{P}_\Omega\mathcal{P}_T(\boldsymbol{M})$. Then if $\mathcal{A}_{\Omega T}^*\mathcal{A}_{\Omega T} = \mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T$ has full rank when restricted to T, the minimizer is given by

$$\boldsymbol{Y} = \mathcal{A}_{\Omega T}(\mathcal{A}_{\Omega T}^*\mathcal{A}_{\Omega T})^{-1}(\boldsymbol{E}), \text{ where } \boldsymbol{E} = \sum_{k=1}^{r} \mathbf{u}_k\mathbf{v}_k^*$$

Injectivity

We study the injectivity of $\mathcal{A}_{\Omega T}$. For convenience, we use *Bernoulli model*, not uniform sampling: $\mathcal{P}(\delta_{ij} = 1) = p \equiv \frac{m}{n_1 n_2}, \Omega = \{(i, j) : \delta_{ij} = 1\}$.

Theorem 4.1 Small operator norm

Suppose Ω is sampled according to the Bernoulli model and put $n = \max(n_1, n_2)$. Suppose that the coherences obey $\max(\mu(U), \mu(V)) \leq \mu_0$. Then, there is a numerical constants C_R such that for all $\beta > 1$,

$$p^{-1}\|\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T - p\mathcal{P}_T\|_2 \leq C_R\sqrt{\frac{\mu_0 nr(\beta \log n)}{m}}$$

with probability at least $1 - 3n^{-\beta}$ provided that $C_R\sqrt{\frac{\mu_0 nr(\beta \log n)}{m}} < 1$.

Injectivity (cont.)

With m large enough so that $C_R\sqrt{\mu_0(nr/m)\log n} \le 1/2$,

$$\frac{p}{2}\|\mathcal{P}_T(\boldsymbol{X})\|_F \le \|(\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T)(\boldsymbol{X})\|_F \le \frac{3p}{2}\|\mathcal{P}_T(\boldsymbol{X})\|_F \tag{4.11}$$

Corollary 4.3 Injectivity of $\mathcal{R}_\Omega$ in T

Assume that $C_R\sqrt{\mu_0 nr(\log n)/m} \le 1/2$. With the same probability as in Theorem 4.1, we have

$$\|\mathcal{P}_\Omega\mathcal{P}_T(\boldsymbol{X})\|_F \le \sqrt{3p/2}\|\mathcal{P}_T(\boldsymbol{X})\|_F.$$

This provides the second condition of unique minimizer.

Size of Spectral Norm

We will investigate the probability of such $\boldsymbol{Y}$ will satisfy $\|\mathcal{P}_{T^\perp}(\boldsymbol{Y})\|_2 < 1$. Denote $\mathcal{H} \equiv \mathcal{P}_T - p^{-1}\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T$, then we can decompose $\mathcal{P}_{T^\perp}(\boldsymbol{Y})$ as

$$\mathcal{P}_{T^\perp}(\boldsymbol{Y}) = p^{-1}(\mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T)(\boldsymbol{E} + \mathcal{H}(\boldsymbol{E}) + \mathcal{H}^2(\boldsymbol{E}) + \cdots), \quad \boldsymbol{E} = \sum_{1 \le k \le r} \mathbf{u}_k \mathbf{v}_k^*.$$

Then lemma 4.4-4.8 will give us the upper bound of these terms!

$$\begin{aligned}
&p^{-1}\|(\mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T)\boldsymbol{E}\|_2 \\
&p^{-1}\|(\mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T)\mathcal{H}^2(\boldsymbol{E})\|_2 \\
&p^{-1}\|(\mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T)\mathcal{H}^3(\boldsymbol{E}))\|_2 \\
&\cdots \\
&p^{-1}\|(\mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T)\sum_{k \ge k_0}\mathcal{H}^k(\boldsymbol{E})\|_2
\end{aligned}$$

Size of Spectral Norm (cont.)

If we set $k_0 = 3$ in lemma 4.8, we can bound the spectral norm of $\|\mathcal{P}_{T^\perp}(\boldsymbol{Y})\|_2 < 1$ with probability at least $1 - cn^{-\beta}$ provided

$$m \geq C \max(\mu_1^2, \mu_0^{1/2}\mu_1, \mu_0^{4/3} r^{1/3}, \mu_0 n^{1/4}) nr\beta \log n$$

If $\mu_0^{4/3} r^{1/3}$ is maximum, it leads to trivial case(greater than n^2). Without this, it is the first result of Theorem 1.3.

If we set $k_0 = 4$ in lemma 4.8, we can bound m

$$m \geq C \max(\mu_0^2 r, \mu_0 n^{1/5}) nr\beta \log n$$

which is the second result of Thoerem 1.3.

vspace0.3cm

A0 (maximum coherence) | **A1** (maximum entry bound of $\sum_{1\leq k\leq r} u_k v_k^*$)

Lemma 3.1 (conditions of uniqueness of the minimizer)

Existence of subgraident Y with $\mathcal{P}_T(Y) = \sum_{k=1}^{r} u_k v_k^*, \quad \|\mathcal{P}_{T^\perp}(Y)\|_2 < 1;$

Injectivity of sampling operator

Lemma 4.4, 4.5, 4.6, 4.8 ($\|\mathcal{P}_{T^\perp}(Y)\|_2 < 1$ with large probability) | **Corollary 4.3** (small operator norm)

Theorem 1.3 (complete recovery)

Table of Contents

1 Introduction

2 Results of Paper

3 Rationality of Incoherence Assumption

4 Proofs

5 Experiments

6 Discussion

Experiment Overview

- 1. Experiment of Theorem 1.3
- 2. Experiment of Theorem 1.3 with the assumption that M is semi-definite matrix.
- 3. Experiment of Fazel's work.

Experiment Overview

- 1. Experiment of Theorem 1.3
- 2. Experiment of Theorem 1.3 with the assumption that M is semi-definite matrix.
- 3. Experiment of Fazel's work.
- Matrix M($n \times n$, rank r) generating method:
  - Sample two $n \times r$ factors M_L and M_R with i.i.d. Gaussian entries.
  - Setting M $= M_L M_R^*$
  - Sample a subset Ω of m entries uniformly at random.

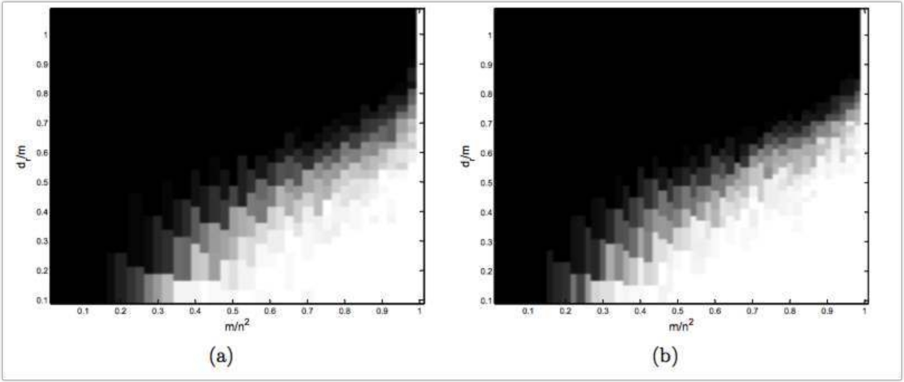
Experiment Overview

- 1. Experiment of Theorem 1.3
- 2. Experiment of Theorem 1.3 with the assumption that M is semi-definite matrix.
- 3. Experiment of Fazel's work.
- Matrix M($n \times n$, rank r) generating method:
  - Sample two $n \times r$ factors M_L and M_R with i.i.d. Gaussian entries.
  - Setting M $= M_L M_R^*$
  - Sample a subset Ω of m entries uniformly at random.
- Determine exactness:
  - M to be recovered if the solution returned by SDP, X_{opt} satisfied $\|X_{opt} - M\|_F / \|M\|_F < 10^{-3}$

Experiment Overview

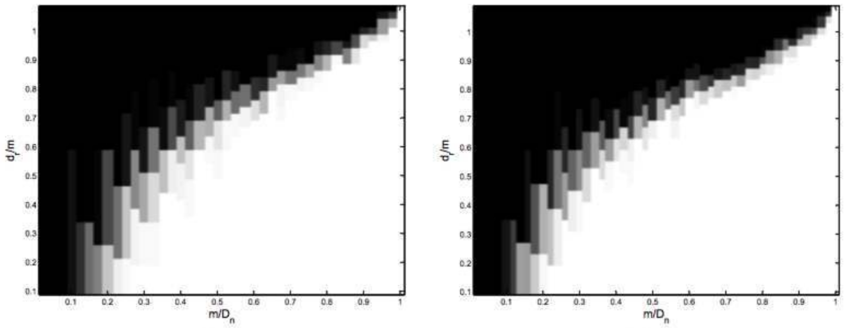
- 1. Experiment of Theorem 1.3
- 2. Experiment of Theorem 1.3 with the assumption that M is semi-definite matrix.
- 3. Experiment of Fazel's work.
- Matrix M($n \times n$, rank r) generating method:
  - Sample two $n \times r$ factors M_L and M_R with i.i.d. Gaussian entries.
  - Setting M $= M_L M_R^*$
  - Sample a subset Ω of m entries uniformly at random.
- Determine exactness:
  - M to be recovered if the solution returned by SDP, X_{opt} satisfied $\|X_{opt} - M\|_F / \|M\|_F < 10^{-3}$
- Repeated same procedure 50 times for 1, 2. 10 for 3.

Experiment 1



- Success: white, Failure: Black. (a): n = 40, (b): n = 50
  - Two experiments in very similar plots for different n.
  - The results of this paper may be conservative.

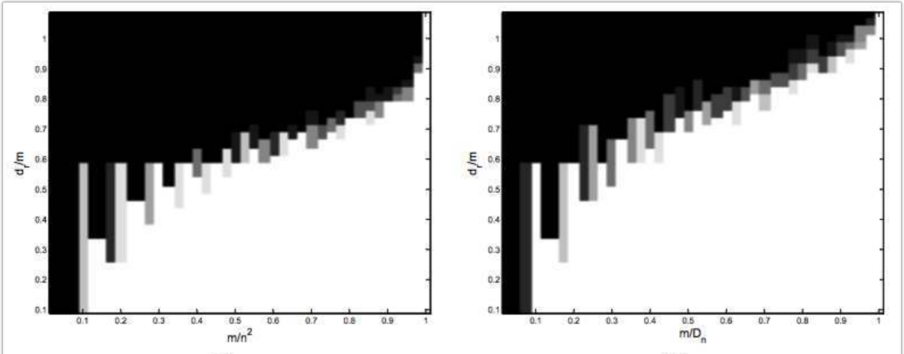
Experiment 2



(a) (b)

- For positive semidefinite matrices case, the recovery region is much larger.
  - Future work is needed to investigate.

Experiment 3



(a) (b)

- To check theoretical bound of matrix sensing problem(Original Fazel's problem).

Reproduce Experiments

- We've reproduced this experiments using `Python`.
  - Using `CVXOPT` package to SDP.
  - `https://github.com/JoonyoungYi/exact-mc`.

Reproduce Experiments

- We've reproduced this experiments using `Python`.
  - Using `CVXOPT` package to SDP.
  - `https://github.com/JoonyoungYi/exact-mc`.
- We plotted the relation between n and m/n.
  - $rank = 2$, $trial = 1$, Exactness metric is same as experiments in this paper:
    X_{opt} satisfied $\|X_{opt} - M\|_F / \|M\|_F < 10^{-3}$

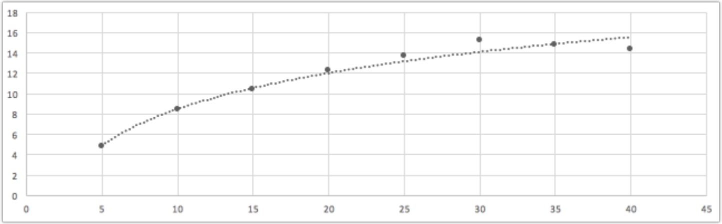


Table of Contents

1. Introduction
2. Results of Paper
3. Rationality of Incoherence Assumption
4. Proofs
5. Experiments
6. Discussion

Improvements

- The main result of this paper.
  - $m = \Omega(n^{1.25} r \log n)$
- With low-rank assumption($r \leq n^{1/5}$),
  - $m = \Omega(n^{1.2} r \log n)$

Improvements

- The main result of this paper.
  - $m = \Omega(n^{1.25} r \log n)$
- With low-rank assumption($r \leq n^{1/5}$),
  - $m = \Omega(n^{1.2} r \log n)$

- Can we find tighter bound?
  - Authors insisted that it would be hard as far as approaching in this way.
  - Bound $\Omega(n^{1.2})$ is mainly determined by $(\mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T\mathcal{H}^k(E))$ for $k = 1, 2, 3$ in the series (4.13).
  - If we k to expand 4, Bound will become $\Omega(n^{7/6})$.
  - If we k to expand 5, Bound will become $\Omega(n^{8/7})$.
  - This can be done to reach k of size about $\log n$, but size of the decoupling constants C_D is varying on k.
  - This is why the stronger results are hard to see.

Improvements

- The main result of this paper.
  - $m = \Omega(n^{1.25} r \log n)$
- With low-rank assumption($r \leq n^{1/5}$),
  - $m = \Omega(n^{1.2} r \log n)$
- Can we find tighter bound?
  - Authors insisted that it would be hard as far as approaching in this way.
  - Bound $\Omega(n^{1.2})$ is mainly determined by $(\mathcal{P}_{T^\perp}\mathcal{P}_\Omega\mathcal{P}_T\mathcal{H}^k(E))$ for $k = 1, 2, 3$ in the series (4.13).
  - If we k to expand 4, Bound will become $\Omega(n^{7/6})$.
  - If we k to expand 5, Bound will become $\Omega(n^{8/7})$.
  - This can be done to reach k of size about $\log n$, but size of the decoupling constants C_D is varying on k.
  - This is why the stronger results are hard to see.
- However, [Candes et al.2010] showed that $m = \Omega(nr \log n)$ with additional assumptions.

Further directions

- (CASE 1) Noise Handling
  - Observation is not M_ij, but Y_ij.
  - $Y_{ij} = M_{ij} + z_{ij}, (i, j) \in \Omega$, where z is a deterministic or stochastic perturbation.

Further directions

- (CASE 2) Low-rank Matrix Fitting
  - Let $M = \Sigma_{1\leq k\leq n}\sigma_k u_k v_k^*$ where $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ and the truncated SVD of the matrix M, $M_r = \Sigma_{1\leq k\leq r}\sigma_k u_k v_k^*$ where the sum extends over the r largest singular values. r is the number of singular value that we can not negligible.
  - This is similar to process of PCA.

Low-rank matrix fitting(or Low-rank matrix approximation)

For general rank M,

$$\begin{aligned} &\text{minimize } \|X\|_* \\ &\text{subject to } \|X - M\|_* \simeq \|X - M_R\|_* \end{aligned}$$

- It is possible to try getting theoretical bound in low-rank matrix fitting problem.

Further directions

- (CASE 3) Slow SDP
  - This direction can be done after (CASE 3).
  - SDP is too slow. So, algorithm described in this paper rarely used in practically.

Further directions

- (CASE 3) Slow SDP
  - This direction can be done after (CASE 3).
  - SDP is too slow. So, algorithm described in this paper rarely used in practically.

  - Practically, SVD(Alternating Minimization) Method is widely used.
  - But, SVD method is really hard to prove theoretical bound.
    - However, [Jain et al. 2012] showed that $m = \Omega(nr^{2.5} logn)$.
    - This bound is slightly higher compared to convex optimization case, but [Jain et al. 2012] argued that this bound is not tight.