Matrix Completion from a Few Entries

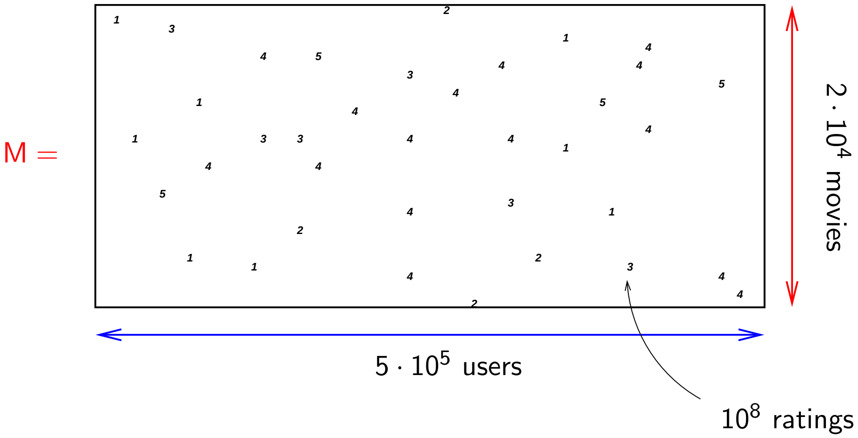
Raghunandan Keshavan, Sewoong Oh and Andrea Montanari

Stanford University

International Symposium on Information Theory
Seoul - June 29, 2009

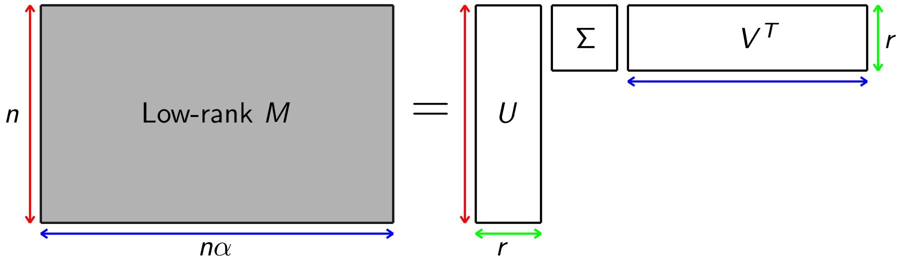
Motivating Example : Recommender System

- Netflix Challenge



The Model

Matrix Completion Problem

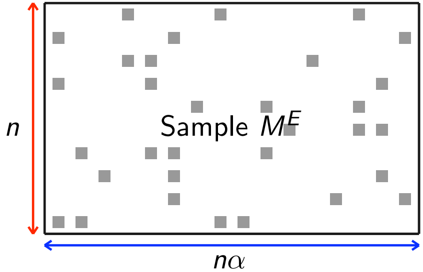


1. Low-rank matrix M.
2. Uniformly random sample E.

Goal : Estimation $\hat{M}(E, M^E)$ that minimizes

$$RMSE \equiv \left(\frac{1}{n^2 M_{\max}^2} \sum_{i,j} (M_{ij} - \hat{M}_{ij})^2 \right)^{1/2} .$$

Matrix Completion Problem

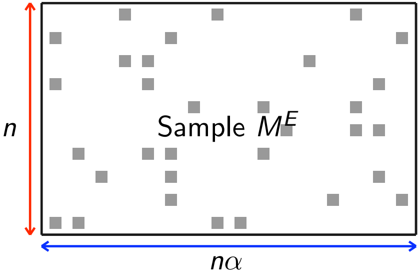


1. Low-rank matrix M.
2. Uniformly random sample E.

Goal : Estimation $\hat{\mathsf{M}}(E, \mathsf{M}^E)$ that minimizes

$$RMSE \equiv \left(\frac{1}{n^2 \mathsf{M}_{\max}^2} \sum_{i,j} (\mathsf{M}_{ij} - \hat{\mathsf{M}}_{ij})^2 \right)^{1/2} .$$

Matrix Completion Problem



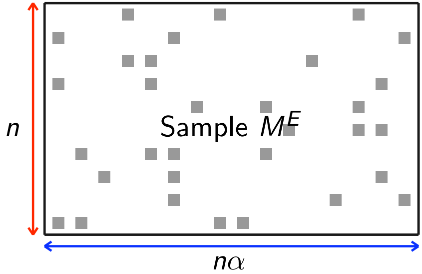
Q1. How many samples do we need to get $RMSE \leq \delta$?

$$(1+\alpha)rn \lesssim |E| = O(n)$$

Q2. How many samples do we need to recover M *exactly*?

$$n \log n \lesssim |E| = O(n \log n)$$

Matrix Completion Problem



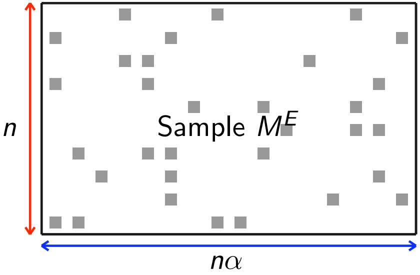
Q1. How many samples do we need to get $RMSE \leq \delta$?

$$(1+\alpha)rn \lesssim |E| = O(n)$$

Q2. How many samples do we need to recover M *exactly*?

$$n \log n \lesssim |E| = O(n \log n)$$

Matrix Completion Problem



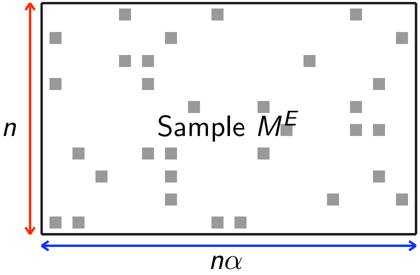
Q1. How many samples do we need to get $RMSE \leq \delta$?

$$(1+\alpha)rn \lesssim |E| = O(n)$$

Q2. How many samples do we need to recover M *exactly*?

$$n \log n \lesssim |E| = O(n \log n)$$

Matrix Completion Problem



Q1. How many samples do we need to get $RMSE \leq \delta$?

$$(1+\alpha)rn \;\lesssim\; |E| = O(n)$$

Q2. How many samples do we need to recover M *exactly*?

$$n \log n \;\lesssim\; |E| = O(n \log n)$$

Pathological Example

$$\mathrm{M} \;=\; e_1 e_1^T$$

$$\begin{array}{cc} n & \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \end{array}$$

$$\mathbb{P}(\text{observing } \mathrm{M}_{11}) = \frac{|E|}{n^2}$$

Pathological Example

$$\mathsf{M} \;=\; e_1 e_1^T$$

$$n \updownarrow \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$\mathbb{P}(\text{observing } \mathsf{M}_{11}) = \frac{|E|}{n^2}$$

Incoherence Property

M is (μ_0, μ_1)-incoherent if

$$\textcolor{red}{A1.} \qquad M_{\max} \le \mu_0 \sqrt{r}\,,$$

$$\textcolor{red}{A2.} \qquad \sum_{a=1}^{r} U_{ia}^2 \le \mu_1 \frac{r}{n}\,, \ \ \sum_{a=1}^{r} V_{ja}^2 \le \mu_1 \frac{r}{n}\,.$$

[Candés, Recht 2008]

Previous Work

Theorem (Candés, Recht 2008)

Let M *be an* $n \times n\alpha$ *matrix of rank* r *satisfying* (μ_0, μ_1)*-incoherence condition. If*

$$|E| \geq C(\alpha, \mu_0, \mu_1) r n^{6/5} \log n \,,$$

then w.h.p. SEMIDEFINITE PROGRAMMING *reconstructs* M *exactly.*

Main Contribution

Open Questions	Main Results
1. Complexity?	Low complexity
2. $RMSE \leq \delta$?	$\|E\| = O(n)$
3. Optimality?	$\|E\| = O(n \log n)$

Algorithm not based on the Convex Relaxation

Main Contribution

Open Questions	Main Results
1. Complexity?	Low complexity
2. $RMSE \leq \delta$?	$\lvert E \rvert = O(n)$
3. Optimality?	$\lvert E \rvert = O(n \log n)$

Algorithm not based on the Convex Relaxation

Main Contribution

Open Questions	Main Results
1. Complexity?	Low complexity
2. $RMSE \leq \delta$?	$\|E\| = O(n)$
3. Optimality?	$\|E\| = O(n \log n)$

Algorithm not based on the Convex Relaxation

The Algorithm and Main Theorems

Naïve Approach

$$\mathrm{M}^E_{ij} = \begin{cases} \mathrm{M}_{ij} & \text{if } (i,j) \in E \ , \\ 0 & \textit{otherwise}. \end{cases}$$

$$\mathrm{M}^E = \sum_{k=1}^{n} x_k \sigma_k y_k^T$$

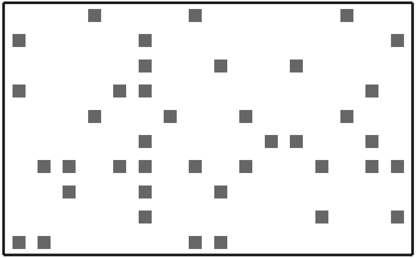
Rank-r projection :

$$\mathcal{P}_r(\mathrm{M}^E) \equiv \frac{n^2 \alpha}{|E|} \sum_{k=1}^{r} x_k \sigma_k y_k^T$$

Naïve Approach Fails

- Define : $deg(row_i) \equiv \#$ of samples in row i.
- For $|E| = O(n)$, *spurious* singular values of $\Omega(\sqrt{\log n/(\log\log n)})$.
- Solution : Trimming

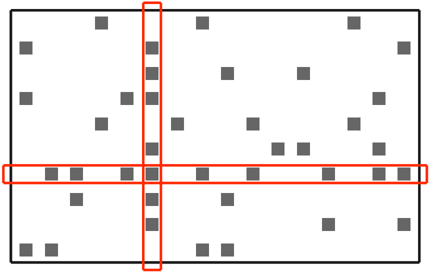
$$\widetilde{M}^E_{ij} = \begin{cases} 0 & \text{if } deg(row_i) > 2\mathbb{E}[deg(row_i)]\ , \\ 0 & \text{if } deg(col_j) > 2\mathbb{E}[deg(col_i)]\ , \\ M^E_{ij} & otherwise. \end{cases}$$



Naïve Approach Fails

- Define : $deg(row_i) \equiv \#$ of samples in row i.
- For $|E| = O(n)$, *spurious* singular values of $\Omega(\sqrt{\log n/(\log\log n)})$.
- Solution : Trimming

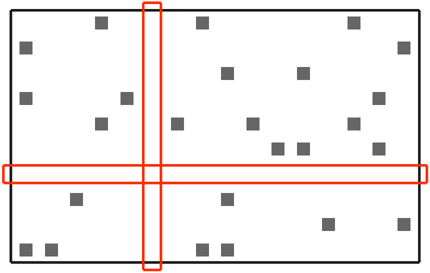
$$\widetilde{\mathrm{M}}^E_{ij} = \begin{cases} 0 & \text{if } deg(row_i) > 2\mathbb{E}[deg(row_i)]\ , \\ 0 & \text{if } deg(col_j) > 2\mathbb{E}[deg(col_i)]\ , \\ \mathrm{M}^E_{ij} & otherwise. \end{cases}$$



Naïve Approach Fails

- Define : $deg(row_i) \equiv \#$ of samples in row i.
- For $|E| = O(n)$, *spurious* singular values of $\Omega(\sqrt{\log n/(\log\log n)})$.
- Solution : Trimming

$$\widetilde{\mathrm{M}}^E_{ij} = \begin{cases} 0 & \text{if } deg(row_i) > 2\mathbb{E}[deg(row_i)] \ , \\ 0 & \text{if } deg(col_j) > 2\mathbb{E}[deg(col_i)] \ , \\ \mathrm{M}^E_{ij} & otherwise. \end{cases}$$



The Algorithm

OptSpace

Input : sample positions E, sample values M^E, rank r
Output : estimation $\hat{\mathrm{M}}$
1: Trim M^E, and let $\widetilde{\mathrm{M}}^E$ be the output;
2: Compute rank-r projection $\mathcal{P}_r(\widetilde{\mathrm{M}}^E) = X_0 S_0 Y_0^T$;
3:

Main Result

Theorem (Keshavan, Montanari, Oh, 2009)

Let M *be an* $n \times n\alpha$ *matrix of rank-*r *bounded by* $\mathsf{M}_{\rm max}$*. Then*

$$\frac{1}{n\mathsf{M}_{\rm max}}||\mathsf{M} - \mathcal{P}_r(\widetilde{\mathsf{M}}^E)||_{\rm F} = \mathrm{RMSE} \leq C(\alpha)\sqrt{\frac{nr}{|E|}}\,,$$

with probability larger than $1 - 1/n^3$*.*

The Algorithm

OptSpace

Input : sample positions E, sample values M^E, rank r

Output : estimation $\hat{\mathsf{M}}$

1: Trim M^E, and let $\widetilde{\mathsf{M}}^E$ be the output;

2: Compute rank-r projection $\mathcal{P}_r(\widetilde{\mathsf{M}}^E) = X_0 S_0 Y_0^T$;

3: Minimize RMSE by gradient descent starting at (X_0, S_0, Y_0).

Main Result

Theorem (Keshavan, Montanari, Oh, 2009)

Assume $r = O(1)$, *and let* M *be an* $n \times n\alpha$ *matrix satisfying* (μ_0, μ_1)*-incoherence with* $\sigma_1(M)/\sigma_r(M) = O(1)$. *If*

$$|E| \geq C' n \log n \; ,$$

then $\textsc{OptSpace}$ *returns, w.h.p., the matrix* M.

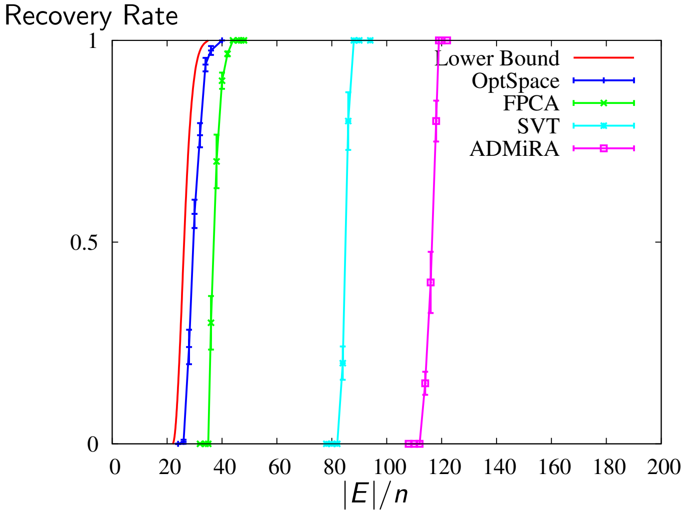
Comparison : Theory

Theorem (Candés, Tao, 2009 March)

Assume strongly incoherent matrix M*.*
If $|E| \geq C\, r\, n\, (\log n)^6$ *then*
SEMIDEFINITE PROGRAMMING *returns, w.h.p., the matrix* M*.*

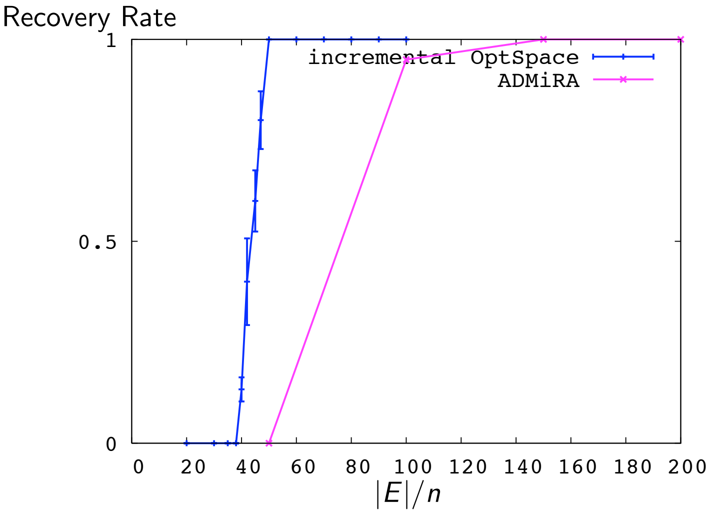
Comparison : Implementation

- $rank = 10$, $\alpha = 1$, $n = 1000$
- M is recovered if $RMSE < 10^{-4}$



Comparison : Ill-conditioned matrices

- $rank = 5$, $\alpha = 1$, $n = 1000$
- condition number $= 10$



Conclusion

Main Results

1. Complexity? $O(r|E|\log n)$
2. $RMSE \leq \delta$? $|E| = O(n)$ [Thm.1]
3. Optimality? $|E| = O(n \log n)$ [Thm.2]

What's left?

1. Prior knowledge of rank?
   RankEstimation is exact for $|E| = O(n)$
2. $r = \Theta(n^{\beta})$?
   Suboptimal bound
3. Noise?
   Order-optimal results [KMO2009b]

Thank you!

Conclusion

Main Results

1. Complexity? $O(r|E|\log n)$
2. $RMSE \leq \delta$? $|E| = O(n)$ [Thm.1]
3. Optimality? $|E| = O(n\log n)$ [Thm.2]

What's left?

1. Prior knowledge of rank?
   RankEstimation is exact for $|E| = O(n)$
2. $r = \Theta(n^{\beta})$?
   Suboptimal bound
3. Noise?
   Order-optimal results [KMO2009b]

Thank you!

Conclusion

Main Results

1. Complexity? $O(r|E|\log n)$
2. $RMSE \leq \delta$? $|E| = O(n)$ [Thm.1]
3. Optimality? $|E| = O(n \log n)$ [Thm.2]

What's left?

1. Prior knowledge of rank?
   RankEstimation is exact for $|E| = O(n)$
2. $r = \Theta(n^{\beta})$?
   Suboptimal bound
3. Noise?
   Order-optimal results [KMO2009b]

Thank you!

Conclusion

Main Results

1. Complexity? $O(r|E|\log n)$
2. $RMSE \leq \delta$? $|E| = O(n)$ [Thm.1]
3. Optimality? $|E| = O(n \log n)$ [Thm.2]

What's left?

1. Prior knowledge of rank?
   RankEstimation is exact for $|E| = O(n)$
2. $r = \Theta(n^{\beta})$?
   Suboptimal bound
3. Noise?
   Order-optimal results [KMO2009b]

Thank you!

Conclusion

Main Results

1. Complexity? $O(r|E|\log n)$
2. $RMSE \leq \delta$? $|E| = O(n)$ [Thm.1]
3. Optimality? $|E| = O(n \log n)$ [Thm.2]

What's left?

1. Prior knowledge of rank?
   $\textsc{RankEstimation}$ is exact for $|E| = O(n)$
2. $r = \Theta(n^\beta)$?
   Suboptimal bound
3. Noise?
   Order-optimal results [KMO2009b]

Thank you!

Conclusion

Main Results

1. Complexity? $O(r|E|\log n)$
2. $RMSE \leq \delta$? $|E| = O(n)$ [Thm.1]
3. Optimality? $|E| = O(n \log n)$ [Thm.2]

What's left?

1. Prior knowledge of rank?
   RankEstimation is exact for $|E| = O(n)$
2. $r = \Theta(n^\beta)$?
   Suboptimal bound
3. Noise?
   Order-optimal results [KMO2009b]

Thank you!

Conclusion

Main Results

1. Complexity? $O(r|E|\log n)$
2. $RMSE \leq \delta$? $|E| = O(n)$ [Thm.1]
3. Optimality? $|E| = O(n \log n)$ [Thm.2]

What's left?

1. Prior knowledge of rank?
   $\textsc{RankEstimation}$ is exact for $|E| = O(n)$
2. $r = \Theta(n^{\beta})$?
   Suboptimal bound
3. Noise?
   Order-optimal results [KMO2009b]

Thank you!

Conclusion

Main Results

1. Complexity? $O(r|E|\log n)$
2. $RMSE \leq \delta$? $|E| = O(n)$ [Thm.1]
3. Optimality? $|E| = O(n \log n)$ [Thm.2]

What's left?

1. Prior knowledge of rank?
   $\textsc{RankEstimation}$ is exact for $|E| = O(n)$
2. $r = \Theta(n^\beta)$?
   Suboptimal bound
3. Noise?
   Order-optimal results [KMO2009b]

Thank you!

Conclusion

Main Results

1. Complexity? $O(r|E|\log n)$
2. $RMSE \leq \delta$? $|E| = O(n)$ [Thm.1]
3. Optimality? $|E| = O(n \log n)$ [Thm.2]

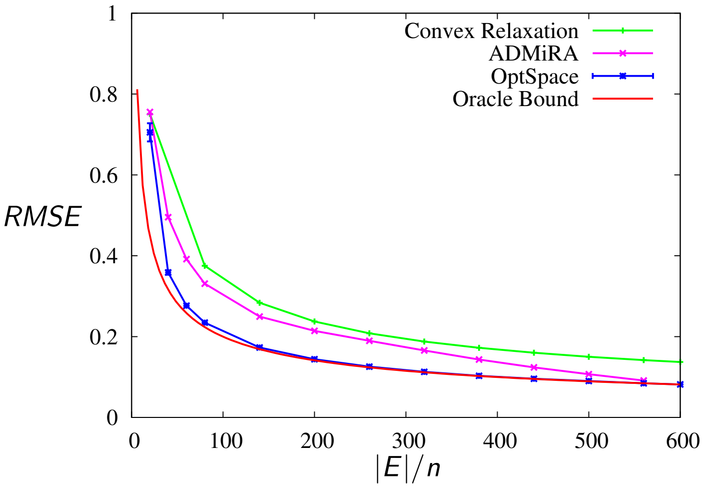
What's left?

1. Prior knowledge of rank?
   $\textsc{RankEstimation}$ is exact for $|E| = O(n)$
2. $r = \Theta(n^\beta)$?
   Suboptimal bound
3. Noise?
   Order-optimal results [KMO2009b]

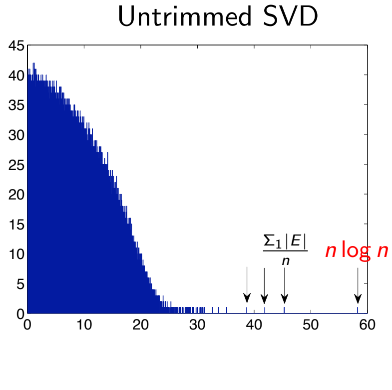
Thank you!

Comparison : Noisy Samples

- $rank = 2$, $\alpha = 1$, $n = 600$
- Noise Variance $= 1$



Proof



Untrimmed SVD
45
40
35
30
25
20
15
10
5
0
0
10
20
30
40
50
60
$\frac{\Sigma_1 |E|}{n}$
$n \log n$
Trimmed SVD
$\frac{\Sigma_3 |E|}{n}$
$C\sqrt{\frac{|E|}{n}}$