Matrix Completion from a Few Entries

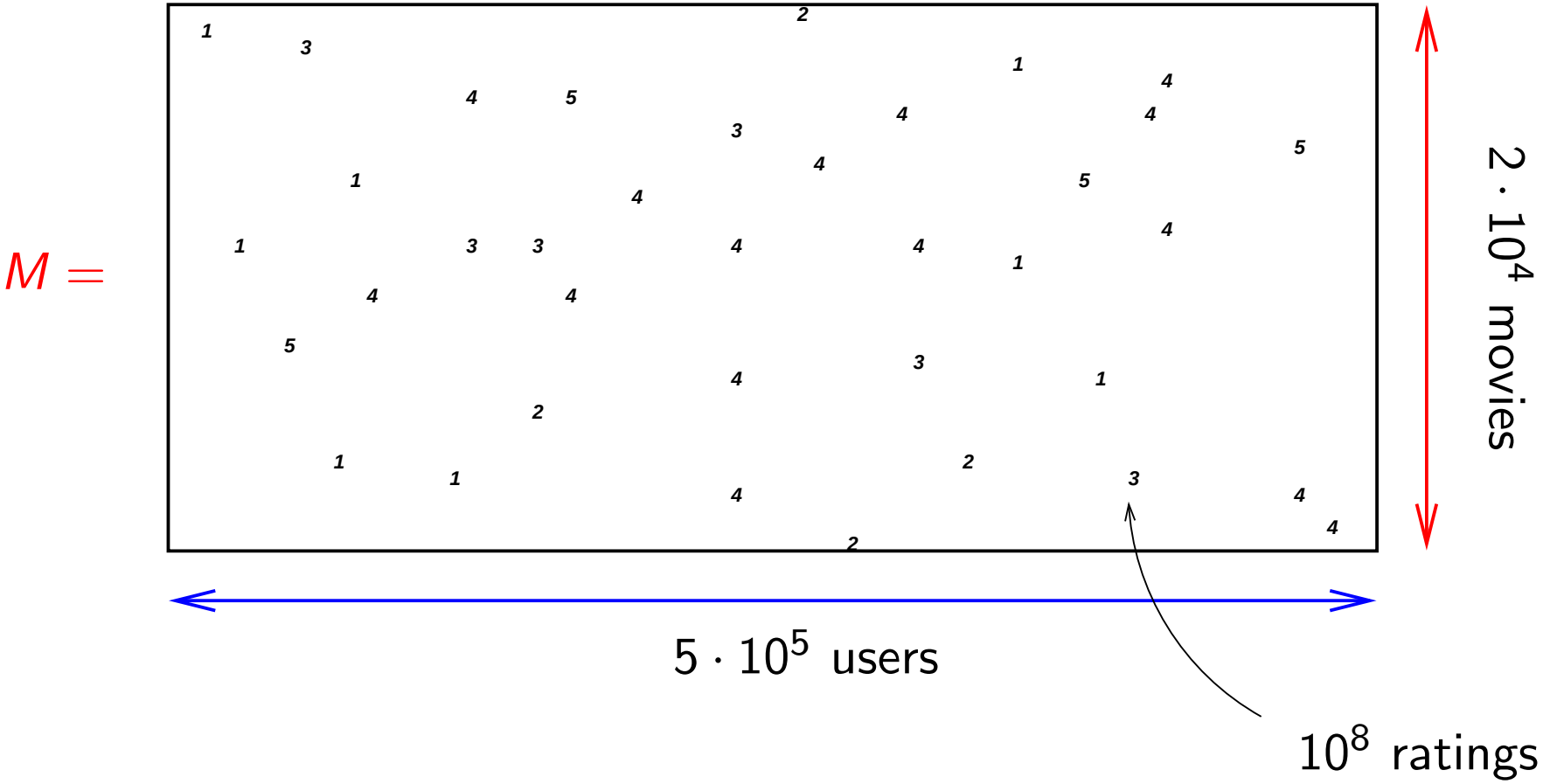
Raghunandan Keshavan, Andrea Montanari and Sewoong Oh

Stanford University

Physics of Algorithms
Santa Fe - Aug 31, 2009

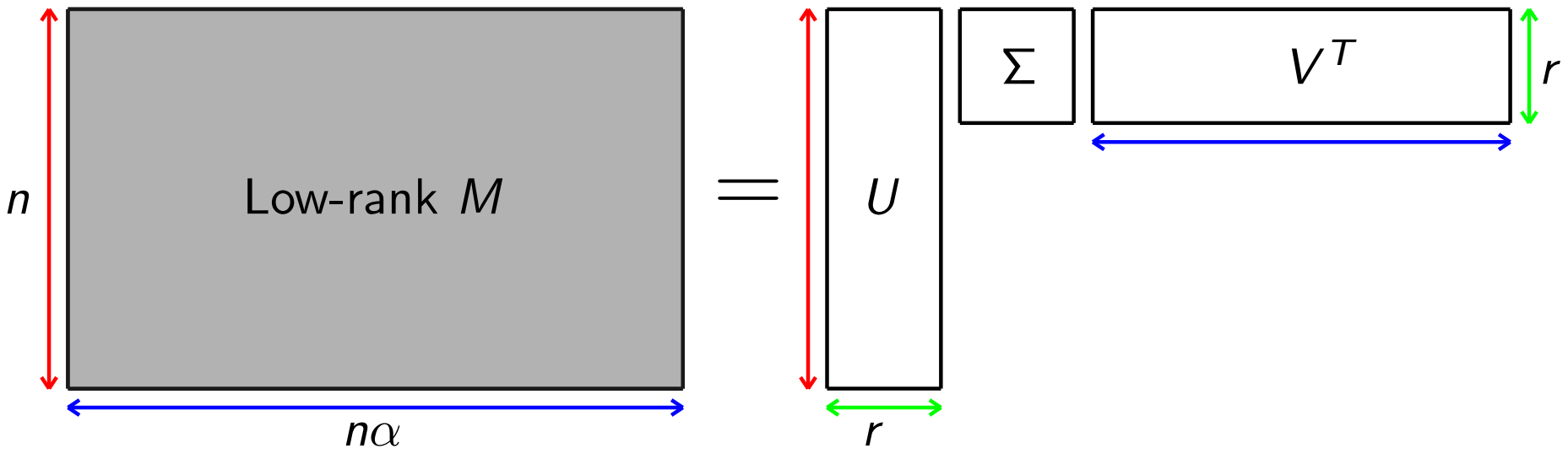
Motivating Example : Recommender System

- Netflix Challenge



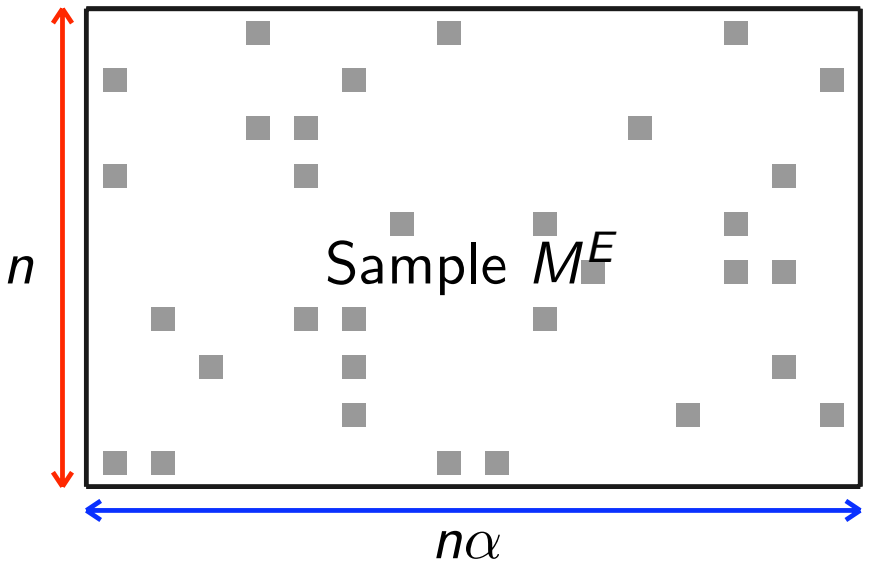
The Model

Matrix Completion Problem



1. Low-rank matrix $M = U\Sigma V^T$.
   $U^T U = n, \quad V^T V = n\alpha, \quad \Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2, \ldots, \Sigma_r)$
2. Uniformly random sample $E \subset [n] \times [n\alpha]$ given its size $|E|$.
   $[k] = \{1, 2, \ldots, k\}$

Matrix Completion Problem



1. Low-rank matrix $M = U\Sigma V^T$.
   $U^T U = n, \quad V^T V = n\alpha, \quad \Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2, \ldots, \Sigma_r)$
2. Uniformly random sample $E \subset [n] \times [n\alpha]$ given its size $|E|$.
   $[k] = \{1, 2, \ldots, k\}$

Matrix Completion Problems

For any estimation $\widehat{M}$, let $\mathrm{RMSE} = \left(\frac{1}{\sqrt{mn}} ||M - \widehat{M}||_F \right)$

$||A||_F^2 = \sum_{ij} A_{ij}^2$

Matrix Completion Problems

For any estimation $\widehat{M}$, let $\mathrm{RMSE} = \left(\frac{1}{\sqrt{mn}}||M - \widehat{M}||_F\right)$

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

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

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

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

Q3. What if the samples are corrupted by noise?

Matrix Completion Problems

For any estimation $\widehat{M}$, let $\mathrm{RMSE} = \left(\frac{1}{\sqrt{mn}}||M - \widehat{M}||_F\right)$

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

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

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

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

Q3. What if the samples are corrupted by noise?

Matrix Completion Problems

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

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

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

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

Q3. What if the samples are corrupted by noise?

Matrix Completion Problems

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

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

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

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

Q3. What if the samples are corrupted by noise?

Matrix Completion Problems

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

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

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

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

Q3. What if the samples are corrupted by noise?

Matrix Completion Problems

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

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

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

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

Q3. What if the samples are corrupted by noise?

Pathological Example

$$M = e_1 e_1^T$$

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

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

Pathological Example

$$M = e_1 e_1^T$$

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

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

Incoherence Property

M is (μ_0, μ_1)-incoherent if

A1. $M_{\max} \leq \mu_0 \Sigma_1 \sqrt{r};,$

A2. $\sum_{a=1}^{r} U_{ia}^2 \leq \mu_1 r \ , \ \sum_{a=1}^{r} V_{ja}^2 \leq \mu_1 r \ .$

[Candés, Recht 2008 [1]]

Previous Work

Theorem (Candés, Recht 2008 [1])

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 Contributions

Questions	Main Results
1. RMSE	$\leq C(\alpha)\left(\frac{nr}{\lvert E\rvert}\right)^{\frac{1}{2}}$
2. Exact Reconstruction	$\lvert E\rvert = O(n \log n)$
3. Noisy Reconstruction	$\lvert E\rvert = O(n \log n)$
$(N = M + Z)$	$\frac{1}{\sqrt{mn}}\lvert\lvert M - \widehat{M}\rvert\rvert_F \leq C\frac{n\sqrt{\alpha r}}{\lvert E\rvert}\lvert\lvert Z^E\rvert\rvert_2$
4. Complexity?	$O(\lvert E\rvert r \log n)$

Main Contributions

Questions	Main Results
1. RMSE	$\leq C(\alpha)\left(\frac{nr}{\lvert E\rvert}\right)^{\frac{1}{2}}$
2. Exact Reconstruction	$\lvert E\rvert = O(n \log n)$
3. Noisy Reconstruction	$\lvert E\rvert = O(n \log n)$
$(N = M + Z)$	$\frac{1}{\sqrt{mn}}\lvert\lvert M - \widehat{M}\rvert\rvert_F \leq C\frac{n\sqrt{\alpha r}}{\lvert E\rvert}\lvert\lvert Z^E\rvert\rvert_2$
4. Complexity?	$O(\lvert E\rvert r \log n)$

Main Contributions

Questions	Main Results
1. RMSE	$\leq C(\alpha)\left(\frac{nr}{\lvert E\rvert}\right)^{\frac{1}{2}}$
2. Exact Reconstruction	$\lvert E\rvert = O(n \log n)$
3. Noisy Reconstruction	$\lvert E\rvert = O(n \log n)$
$(N = M + Z)$	$\frac{1}{\sqrt{mn}}\lvert\lvert M - \widehat{M}\rvert\rvert_F \leq C\frac{n\sqrt{\alpha r}}{\lvert E\rvert}\lvert\lvert Z^E\rvert\rvert_2$
4. Complexity?	$O(\lvert E\rvert r \log n)$

Main Contributions

Questions	Main Results
1. RMSE	$\leq C(\alpha)\left(\frac{nr}{\lvert E\rvert}\right)^{\frac{1}{2}}$
2. Exact Reconstruction	$\lvert E\rvert = O(n\log n)$
3. Noisy Reconstruction	$\lvert E\rvert = O(n\log n)$
$(N = M + Z)$	$\frac{1}{\sqrt{mn}}\lvert\lvert M - \widehat{M}\rvert\rvert_F \leq C\frac{n\sqrt{\alpha r}}{\lvert E\rvert}\lvert\lvert Z^E\rvert\rvert_2$
4. Complexity?	$O(\lvert E\rvert r\log n)$

Main Contributions

Questions	Main Results
1. RMSE	$\leq C(\alpha)\left(\frac{nr}{\lvert E\rvert}\right)^{\frac{1}{2}}$
2. Exact Reconstruction	$\lvert E\rvert = O(n\log n)$
3. Noisy Reconstruction	$\lvert E\rvert = O(n\log n)$
$(N = M + Z)$	$\frac{1}{\sqrt{mn}}\lvert\lvert M - \widehat{M}\rvert\rvert_F \leq C\frac{n\sqrt{\alpha r}}{\lvert E\rvert}\lvert\lvert Z^E\rvert\rvert_2$
4. Complexity?	$O(\lvert E\rvert r\log n)$

The Algorithm and Main Theorems

Naïve Approach

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

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

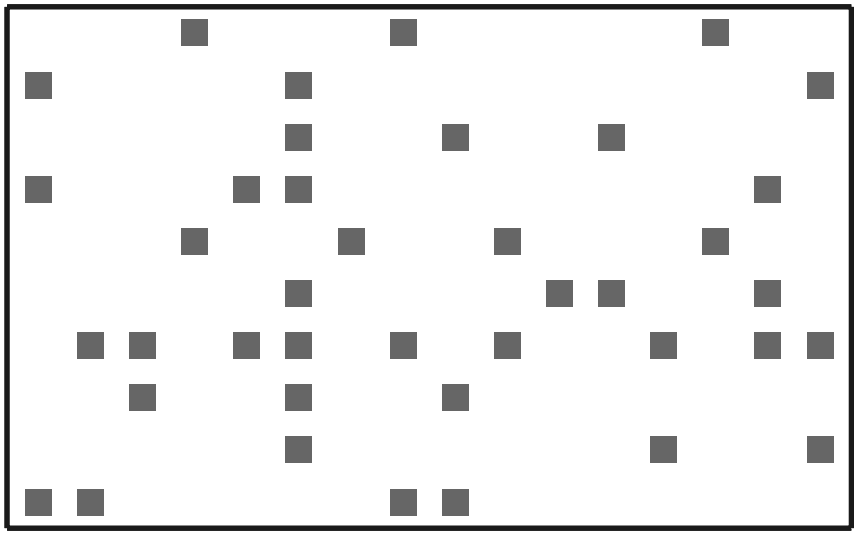
Rank-r projection :

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

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

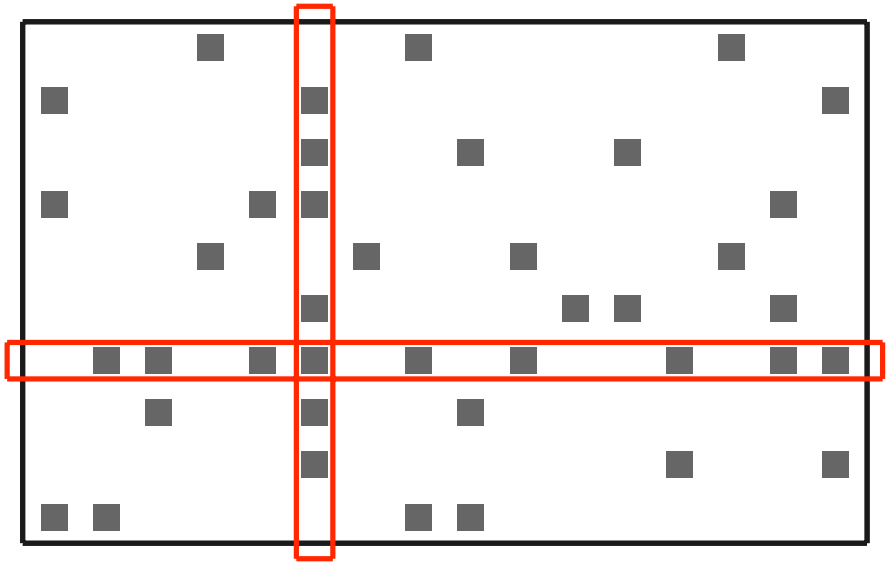
$$\widetilde{M}^E_{ij} = \begin{cases} 0 & \text{if } deg(row_i) > 2\mathbb{E}[deg(row_i)] \,, \\ 0 & \text{if } deg(col_j) \ > 2\mathbb{E}[deg(col_i)] \,, \\ M^E_{ij} & otherwise. \end{cases}$$



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

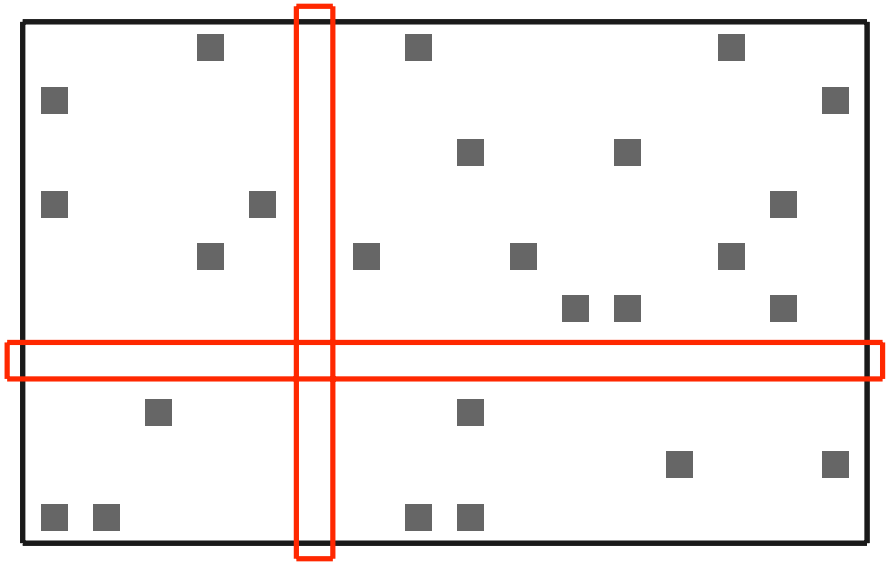
$$\widetilde{M}^E_{ij} = \begin{cases} 0 & \text{if } deg(row_i) > 2\mathbb{E}[deg(row_i)] \,, \\ 0 & \text{if } deg(col_j) \ > 2\mathbb{E}[deg(col_i)] \,, \\ M^E_{ij} & otherwise. \end{cases}$$



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

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



The Algorithm

OptSpace

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

Output : estimation $\widehat{M}$

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

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

3:

Main Result

Theorem (Keshavan, Montanari, Oh, 2009 [2])

Let M be an $n \times n\alpha$ matrix of rank-r bounded by $M_{\max}$. Then, w.h.p,

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

The Algorithm

OptSpace

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

Output : estimation $\widehat{M}$

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

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

3: Minimize $\mathsf{F}(X, Y)$ by gradient descent starting at (X_0, Y_0).

$$\mathsf{F}(X, Y) = \min_{S \in \mathbb{R}^{r \times r}} \sum_{(i,j) \in E} \left(M_{ij} - (XSY^T)_{ij} \right)^2$$

Main Result

Theorem (Keshavan, Montanari, Oh, 2009 [2])

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 OptSpace *returns, w.h.p., the matrix M.*

Comparison

Theorem (Candés, Tao, 2009 [3])

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

Corrupted Observations

- $N = M + Z$ for any $n \times n\alpha$ matrix Z
- N^E (defined similar to M^E) input to OptSpace

Corrupted Observations

- $N = M + Z$ for any $n \times n\alpha$ matrix Z
- N^E (defined similar to M^E) input to OptSpace

Corrupted Observations

- $N = M + Z$ for any $n \times n\alpha$ matrix Z
- N^E (defined similar to M^E) input to OptSpace

Theorem (Keshavan, Montanari, Oh, 2009 [4])

Let $N = M + Z$ with M as above and Z any $n \times n\alpha$ matrix. If $|E| \geq C' n \log n$, then OptSpace with input N^E returns $\widehat{M}$ such that w.h.p.,

$$\frac{1}{\sqrt{mn}} ||M - \widehat{M}||_F \leq C \frac{n\sqrt{\alpha r}}{|E|} ||Z^E||_2$$

provided that the right hand side is smaller than Σ_r.

$||A||_2 = \sup_{v \neq 0} \left(\frac{||Av||}{||v||} \right)$

Corrupted Observations

- Z_{ij} are independent, $\mathbb{E}\{Z_{ij}\} = 0$ and $\mathrm{P}\{|Z_{ij}| \geq x\} \leq Ce^{-\frac{x^2}{2\sigma^2}}$

Corrupted Observations

- Z_{ij} are independent, $\mathbb{E}\{Z_{ij}\} = 0$ and $\mathrm{P}\{|Z_{ij}| \geq x\} \leq Ce^{-\frac{x^2}{2\sigma^2}}$

Theorem (Keshavan, Montanari, Oh, 2009 [4])

If Z is a random matrix with entries drawn as above, then

$$||\widetilde{Z}^E||_2 \leq C\sigma \left(\frac{\sqrt{\alpha}|E| \log |E|}{n} \right)^{\frac{1}{2}}$$

Corrupted Observations

- Z_{ij} are independent, $\mathbb{E}\{Z_{ij}\} = 0$ and $\mathrm{P}\{|Z_{ij}| \geq x\} \leq Ce^{-\frac{x^2}{2\sigma^2}}$

Theorem (Keshavan, Montanari, Oh, 2009 [4])

If Z is a random matrix with entries drawn as above, then

$$||\widetilde{Z}^E||_2 \leq C\sigma \left(\frac{\sqrt{\alpha}|E| \log |E|}{n} \right)^{\frac{1}{2}}$$

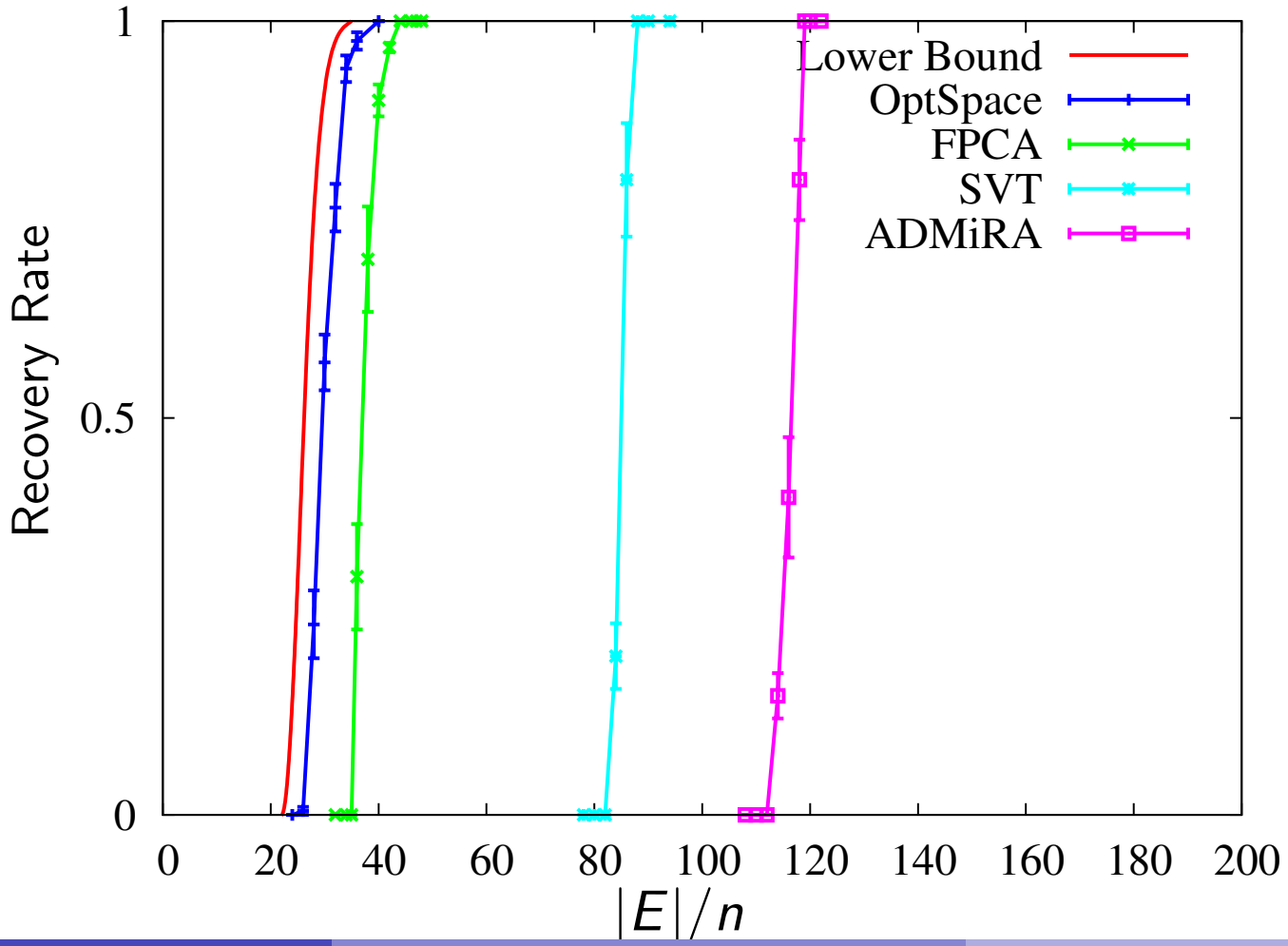
Corollary

Let $N = M + Z$, with Z distributed as above. Then, OptSpace *with input N^E returns $\widehat{M}$ such that*

$$\frac{1}{\sqrt{mn}} ||M - \widehat{M}||_F \leq C\alpha\sigma \left(\frac{nr}{|E|} \right)^{\frac{1}{2}} \sqrt{\log |E|}$$

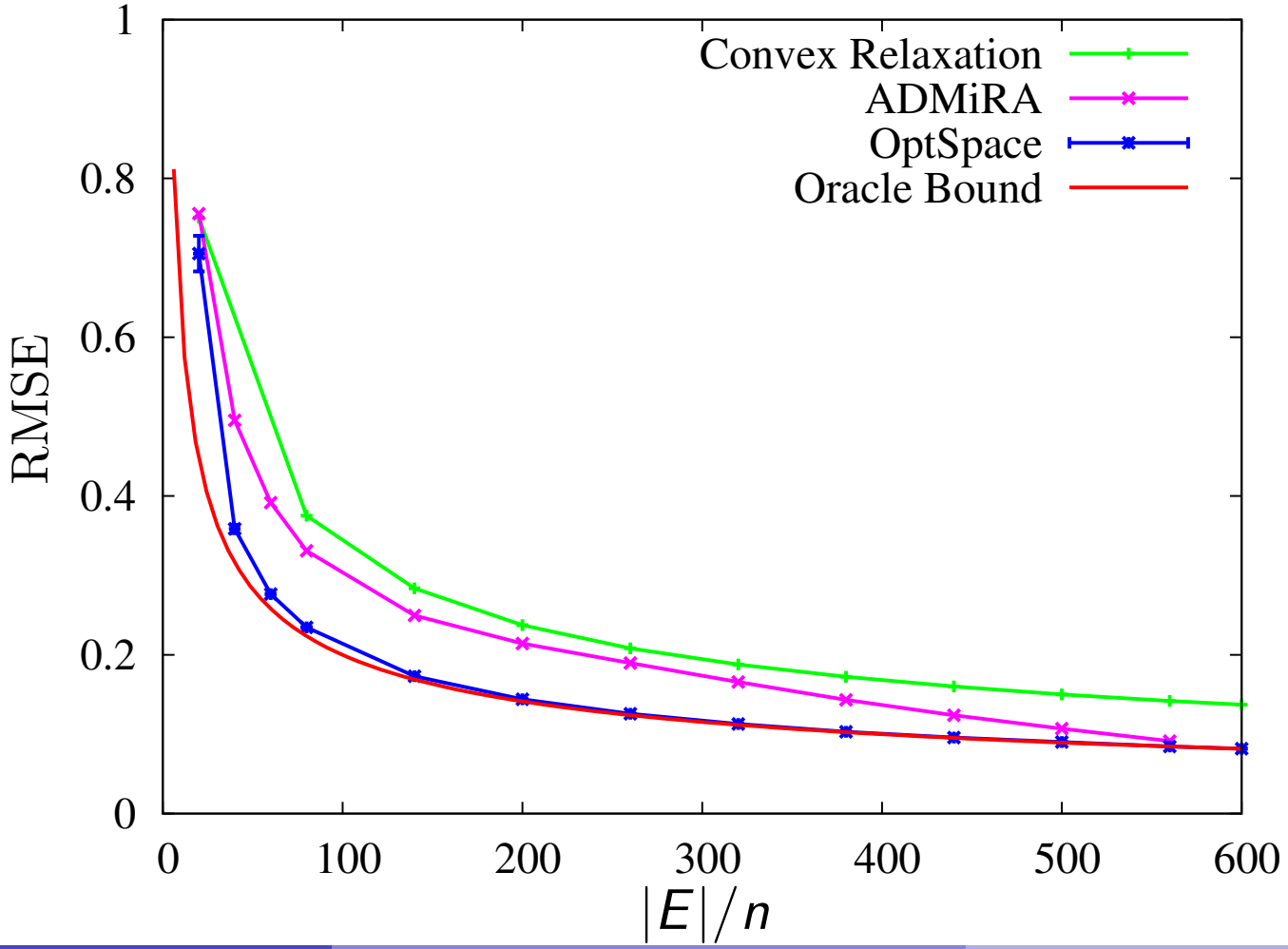
Simulation Results

- $M = UV^T$, $U_{ij} \sim N(0,1)$, $V_{ij} \sim N(0,1)$
- $r = 10$, $\alpha = 1$, $n = 1000$
- M is recovered if $||M - \widehat{M}||_F / ||M||_F < 10^{-4}$



Simulation Results

- $U_{ij} \sim N(0, \sigma^2 = 20/\sqrt{n})$[5]
- $Z_{ij} \sim N(0, 1)$
- $r = 2,\ \alpha = 1,\ n = 600$



Conclusion

Main Results

1. $\mathrm{RMSE} \leq \delta$: $|E| = O(nr)$
2. Exact : $|E| = O(n \log n)$
3. Noisy : $\leq C \frac{n\sqrt{\alpha r}}{|E|} ||Z^E||_2$
4. Complexity : $O(|E| r \log n)$

What's left?

1. Prior knowledge of rank?
   RankEstimation is exact w.h.p for $|E| = O(n)$
2. $r = \Theta(n^\beta)$?
   Suboptimal bound

Conclusion

Main Results

1. $\mathrm{RMSE} \leq \delta$: $|E| = O(nr)$
2. Exact : $|E| = O(n \log n)$
3. Noisy : $\leq C \frac{n\sqrt{\alpha r}}{|E|} ||Z^E||_2$
4. Complexity : $O(|E| r \log n)$

What's left?

1. Prior knowledge of rank?
   RankEstimation is exact w.h.p for $|E| = O(n)$
2. $r = \Theta(n^\beta)$?
   Suboptimal bound

Conclusion

Main Results

1. $\mathrm{RMSE} \leq \delta$: $|E| = O(nr)$
2. Exact : $|E| = O(n \log n)$
3. Noisy : $\leq C \frac{n\sqrt{\alpha r}}{|E|} ||Z^E||_2$
4. Complexity : $O(|E| r \log n)$

What's left?

1. Prior knowledge of rank?
   RankEstimation is exact w.h.p for $|E| = O(n)$
2. $r = \Theta(n^\beta)$?
   Suboptimal bound

Conclusion

Main Results

1. $\mathrm{RMSE} \leq \delta$: $|E| = O(nr)$
2. Exact : $|E| = O(n \log n)$
3. Noisy : $\leq C \frac{n\sqrt{\alpha r}}{|E|} ||Z^E||_2$
4. Complexity : $O(|E| r \log n)$

What's left?

1. Prior knowledge of rank?
   RANKESTIMATION is exact w.h.p for $|E| = O(n)$
2. $r = \Theta(n^\beta)$?
   Suboptimal bound

Conclusion

Main Results

1. $\mathrm{RMSE} \leq \delta$: $|E| = O(nr)$
2. Exact : $|E| = O(n \log n)$
3. Noisy : $\leq C \frac{n\sqrt{\alpha r}}{|E|} ||Z^E||_2$
4. Complexity : $O(|E| r \log n)$

What's left?

1. Prior knowledge of rank?
   RankEstimation is exact w.h.p for $|E| = O(n)$
2. $r = \Theta(n^\beta)$?
   Suboptimal bound

Conclusion

Main Results

1. $\mathrm{RMSE} \leq \delta$: $|E| = O(nr)$
2. Exact : $|E| = O(n \log n)$
3. Noisy : $\leq C \frac{n\sqrt{\alpha r}}{|E|} ||Z^E||_2$
4. Complexity : $O(|E| r \log n)$

What's left?

1. Prior knowledge of rank?
   RankEstimation is exact w.h.p for $|E| = O(n)$
2. $r = \Theta(n^\beta)$?
   Suboptimal bound

Thank you!

References

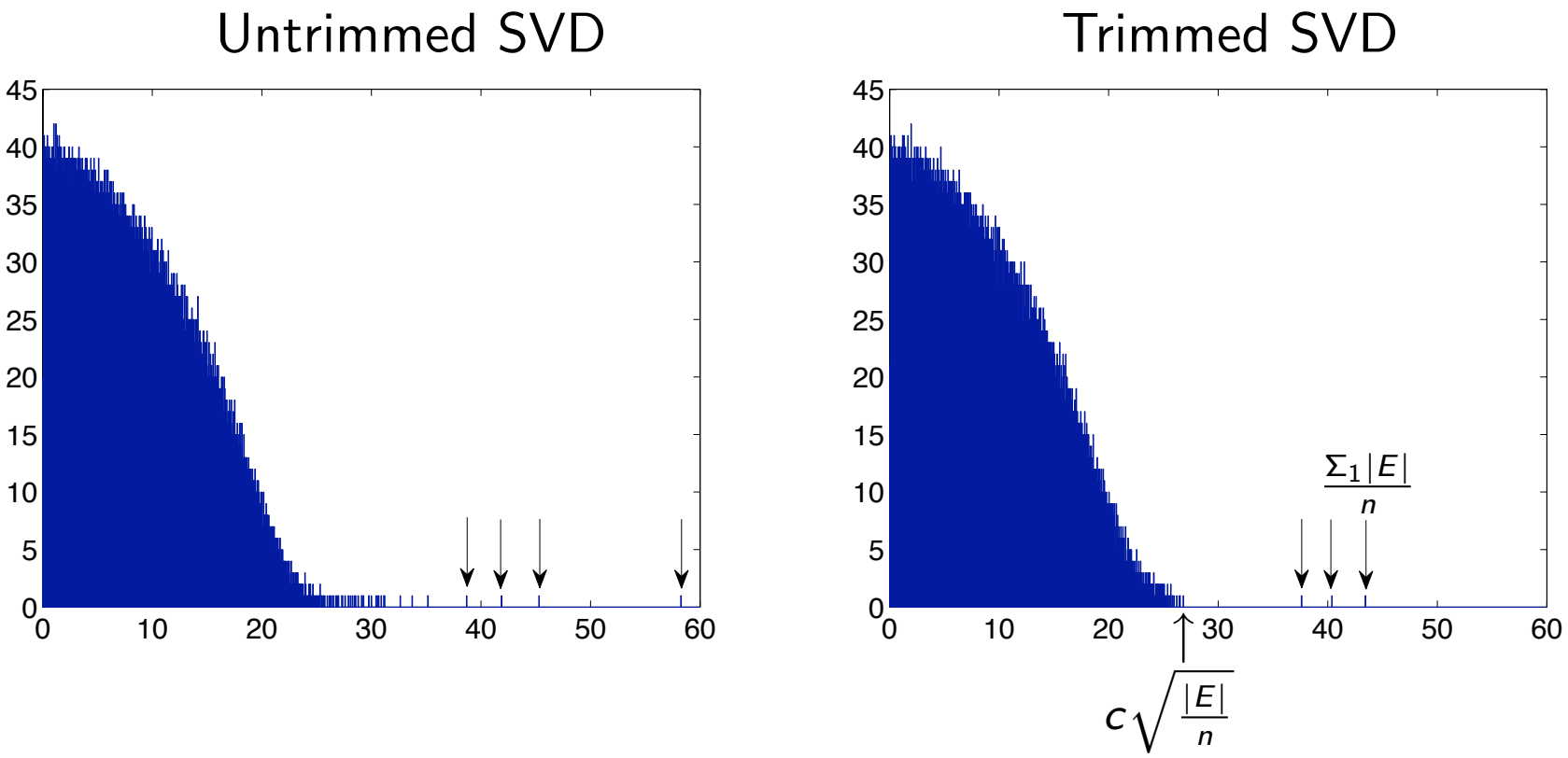
E. J. Candès and B. Recht.
Exact matrix completion via convex optimization.
arxiv:0805.4471, 2008.

R. H. Keshavan, A. Montanari, and S. Oh.
Matrix completion from a few entries.
arXiv:0901.3150, January 2009.

E. J. Candès and T. Tao.
The power of convex relaxation: Near-optimal matrix completion.
arXiv:0903.1476, 2009.

R. H. Keshavan, A. Montanari, and S. Oh.
Matrix completion from noisy entries.
arXiv:0906.2027, June 2009.

E. J. Candès and Y. Plan.
Matrix completion with noise.
arXiv:0903.3131, 2009.


Untrimmed SVD
Trimmed SVD
$\frac{\Sigma_1 |E|}{n}$
$c\sqrt{\frac{|E|}{n}}$