Outline

1. The problem, a look at the data, and some results (slides)

2. Proofs (blackboard)

arXiv:0901.3150
The problem, a look at the data, and some results
Netflix dataset: A big (!) matrix

\[ M = \]

2 \cdot 10^4 \text{ movies}

5 \cdot 10^5 \text{ users}

10^8 \text{ ratings}
A big (!) matrix

\[ M = \]

\[
\begin{array}{cccccccc}
1 & 3 & 4 & 5 & ? & 1 & 4 & 4 \\
2 & ? & 3 & 4 & 4 & 1 & 4 & 4 \\
1 & 1 & 4 & ? & 4 & 4 & 4 & 2 \\
3 & 3 & 4 & 1 & 4 & 1 & 2 & 3 \\
4 & 1 & 5 & 3 & ? & 3 & 4 & 3 \\
\end{array}
\]

5 \cdot 10^5 \text{ users}

2 \cdot 10^4 \text{ movies}

10^6 \text{ queries}
You get a prize if...

\[ \text{RMSE} < 0.8563 \]

Is this possible?
You get a prize if...

\[ \text{RMSE} < 0.8563 \]

Is this possible?
You get a prize if...

\[ \text{RMSE} < 0.8563 \]

Is this possible?
A model: Incoherent low-rank matrices
The observations

\[ M = \]

\( n \alpha \) users

\( n \) movies

Raghunandan Keshavan, Andrea Montanari and Sewoong Oh

Matrix Completion
The observations

\[ M^E \]

\[ n\alpha \text{ users} \]

\[ n\epsilon \text{ unif. random positions} \]

n movies

\[ M^E = \]

Raghunandan Keshavan, Andrea Montanari and Sewoong Oh

Matrix Completion
You need some structure!

\[ M = n \]

\[ M = n \]

\[ r \ll n \]

\[ U \]

\[ \mathbf{V}^{T} \]

\[ n \alpha \]
You need some structure!

\[ r \ll n \]

\[ \begin{align*}
M &= n \\
U \quad \sqrt{\mathbf{V}^T} \\
\end{align*} \]
Unstructured factors

A1. Bounded entries

$$|M_{ia}| \leq M_{\text{max}} = \mu_0 \sqrt{r}.$$ 

A2. Incoherence

$$\sum_{k=1}^{r} U_{ik}^2 \leq \mu_1 r, \quad \sum_{k=1}^{r} V_{ak}^2 \leq \mu_1 r.$$ 

[Candés, Recht 2008]
Metric (RMSE)

\[ D(M, \hat{M}) \equiv \left\{ \frac{1}{n^2 M_{\text{max}}^2} \sum_{i,a} |M_{ia} - \hat{M}_{ia}|^2 \right\}^{1/2} \]
Previous work

**Theorem (Candés, Recht, 2008)**

If

$$\epsilon \geq C r n^{1/5} \log n$$

then whp

1. $M$ is unique given the observed entries.
2. $M$ is the unique minimum of a SDP.

cf. also [Recht, Fazel, Parrilo 2007]
Theorem (Candés, Recht, 2008)

If 

\[ \epsilon \geq C r n^{1/5} \log n \]

then whp

1. \( M \) is unique given the observed entries.
2. \( M \) is the unique minimum of a SDP.

cf. also [Recht, Fazel, Parrilo 2007]
Previous work

Theorem (Candés, Recht, 2008)

If

$$\epsilon \geq C r n^{1/5} \log n$$

then whp

1. $M$ is unique given the observed entries.
2. $M$ is the unique minimum of a SDP.

cf. also [Recht, Fazel, Parrilo 2007]
Theorem (Candés, Recht, 2008)

If

\[ \epsilon \geq C r n^{1/5} \log n \]

then whp

1. M is unique given the observed entries.
2. M is the unique minimum of a SDP.

cf. also [Recht, Fazel, Parrilo 2007]
Theorem (Candés, Recht, 2008)

If

$$\epsilon \geq C r n^{1/5} \log n$$

then whp

1. $M$ is unique given the observed entries.
2. $M$ is the unique minimum of a SDP.

cf. also [Recht, Fazel, Parrilo 2007]
1. $n^{1/5}$ observations for 1 bit of information?

2. RMSE = 0?

3. SDP = $O(n^{4\ldots6})$. Substitute $n = 10^5\ldots$
1. $n^{1/5}$ observations for 1 bit of information?

2. RMSE = 0?

3. SDP = $O(n^{4\cdots6})$. Substitute $n = 10^5\ldots$
1. $n^{1/5}$ observations for 1 bit of information?

2. $\text{RMSE} = 0$?

3. $\text{SDP} = O(n^{4\ldots 6})$. Substitute $n = 10^5\ldots$
1. \( n^{1/5} \) observations for 1 bit of information?

2. RMSE = 0?

3. SDP = \( O(n^{4\ldots6}) \). Substitute \( n = 10^5 \ldots \)
\( O(n) \) entries are enough (practice)
A movie
Rank = 1: Bayes optimal vs. Belief Propagation

Matrix Completion
Rank = 2: Belief Propagation

\[ D \]

\[ \epsilon \]

Matrix Completion

Raghunandan Keshavan, Andrea Montanari and Sewoong Oh
Rank = 3: Belief Propagation

![Graph showing the relationship between $\epsilon$ and $D$ for different values of $n$.]
Rank = 4: Belief Propagation

Matrix Completion

Raghunandan Keshavan, Andrea Montanari and Sewoong Oh
$O(n)$ entries are enough (theory)
Naive spectral algorithm

\[ M_{ia}^E = \begin{cases} M_{ia} & \text{if } (i, a) \in E, \\ 0 & \text{otherwise.} \end{cases} \]

Projection

\[ M^E = \sum_{i=1}^{n} \sigma_i x_i y_i^T, \quad \sigma_1 \geq \sigma_2 \geq \ldots \]

\[ \text{Tr}(M_E) = \frac{n\sqrt{\alpha}}{\epsilon} \sum_{i=1}^{r} \sigma_i x_i y_i^T. \]
Naive spectral algorithm

\[ M_{ia}^E = \begin{cases} M_{ia} & \text{if } (i, a) \in E, \\ 0 & \text{otherwise.} \end{cases} \]

Projection

\[ M^E = \sum_{i=1}^{n} \sigma_i x_i y_i^T, \quad \sigma_1 \geq \sigma_2 \geq \ldots \]

\[ \text{Tr}(M_E) = \frac{n\sqrt{\alpha}}{\epsilon} \sum_{i=1}^{r} \sigma_i x_i y_i^T. \]
If $\epsilon = O(1)$, ‘spurious’ singular values $\Omega(\sqrt{\log n/(\log \log n)})$.

Trimming

$$\tilde{M}^E_{ia} = \begin{cases} M^E_{ia} & \text{if } \deg(i) \leq 2 \mathbb{E}\deg(i), \ deg(a) \leq 2 \mathbb{E}\deg(a), \\ 0 & \text{otherwise.} \end{cases}$$
If $\epsilon = O(1)$, ‘spurious’ singular values $\Omega(\sqrt{\log n/(\log \log n)})$.

Trimming

$$\tilde{M}_{ia}^E = \begin{cases} M_{ia}^E & \text{if } \deg(i) \leq 2 \mathbb{E}\deg(i), \deg(a) \leq 2 \mathbb{E}\deg(a), \\ 0 & \text{otherwise.} \end{cases}$$
Not-as-naive spectral algorithm

Spectral Matrix Completion (matrix $M^E$)

1: Trim $M^E$, and let $\tilde{M}^E$ be the output;
2: Project $\tilde{M}^E$ to $\text{Tr}(\tilde{M}^E)$;
3: Clean residual errors by gradient descent in the factors.
Theorem (Keshavan, M, Oh, 2009)

Assume \( r \leq n^{1/2} \) and bounded entries. Then

\[
\frac{1}{nM_{\max}} \|M - \text{Tr}(\widetilde{M}^E)\|_F = \text{RMSE} \leq C \sqrt{r/\epsilon}.
\]

with probability larger than \( 1 - \exp(-Bn) \).

Theorem (Keshavan, M, Oh, 2009)

Assume \( r = O(1) \), bounded entries and incoherent factors, with \( \Sigma_{\min}, \Sigma_{\max} \) uniformly bounded away from \( 0 \) and \( \infty \).

If \( \epsilon \geq C' \log n \) then

Spectral Matrix Completion returns, whp, the matrix \( M \).
Theorem (Keshavan, M, Oh, 2009)

Assume $r \leq n^{1/2}$ and bounded entries. Then

$$\frac{1}{nM_{\max}} \| M - T_r(\tilde{M}^E) \|_F = \text{RMSE} \leq C \sqrt{r/\epsilon}.$$  

with probability larger than $1 - \exp(-Bn)$.

Theorem (Keshavan, M, Oh, 2009)

Assume $r = O(1)$, bounded entries and incoherent factors, with $\Sigma_{\min}, \Sigma_{\max}$ uniformly bounded away from 0 and $\infty$.

If $\epsilon \geq C' \log n$ then

Spectral Matrix Completion returns, whp, the matrix $M$. 

Raghunandan Keshavan, Andrea Montanari and Sewoong Oh

Matrix Completion
Theorem (Achlioptas, McSherry 2007)

Assume $\epsilon \geq (8 \log n)^4$ and bounded entries. Then

$$
\frac{1}{nM_{\text{max}}} \|M - \text{Tr}(\tilde{M}^E)\|_F = \text{RMSE} \leq 4\sqrt{r/\epsilon}.
$$

with probability larger than $1 - \exp(-19(\log n)^4)$.

(For $n = 10^6$, $(8 \log n)^4 \approx 1.5 \cdot 10^8$)
Theorem (Achlioptas, McSherry 2007)

Assume $\epsilon \geq (8 \log n)^4$ and bounded entries. Then

$$\frac{1}{nM_{\max}} \|M - \text{Tr}(\tilde{M}^E)\|_F = \text{RMSE} \leq 4 \sqrt{r/\epsilon}.$$ 

with probability larger than $1 - \exp(-19(\log n)^4)$.

(For $n = 10^6$, $(8 \log n)^4 \approx 1.5 \cdot 10^8$)
Theorem (Candés, Tao, March 8, 2009)

Assume bounded entries and strongly incoherent factors
If $\epsilon \geq C r (\log n)^6$ then
Semidefinite Programming returns, whp, the matrix $M$.

A2'. Strong incoherence

\[
\sum_{k=1}^{r} U_{ik}^2 \leq \mu_1 r, \\
\left| \sum_{k=1}^{r} U_{ik} U_{jk} \right| \leq \mu_1 \sqrt{r},
\]
One more comparison

**Theorem (Candés, Tao, March 8, 2009)**

Assume bounded entries and strongly incoherent factors

If \( \epsilon \geq C r (\log n)^6 \) then

**Semidefinite Programming** returns, whp, the matrix \( M \).

---

A2'. Strong incoherence

\[
\sum_{k=1}^{r} U_{ik}^2 \leq \mu_1 r,
\]

\[
\left| \sum_{k=1}^{r} U_{ik} U_{jk} \right| \leq \mu_1 \sqrt{r},
\]
One more comparison

Theorem (Candés, Tao, March 8, 2009)

Assume bounded entries and strongly incoherent factors
If \( \epsilon \geq C r (\log n)^6 \) then
Semidefinite Programming returns, whp, the matrix \( M \).

A2’. Strong incoherence

\[
\sum_{k=1}^{r} U_{ik}^2 \leq \mu_1 r ,
\]

\[
\left| \sum_{k=1}^{r} U_{ik} U_{jk} \right| \leq \mu_1 \sqrt{r} ,
\]
Our approach: Graph theory

\((i, a) \in E \iff \text{User } a \text{ rated movie } i.\)
Our approach: Graph theory

\[ (i, a) \in E \iff \text{User } a \text{ rated movie } i. \]
Back to the data
Random $r = 4$, $n = 10000$, $\epsilon = 12.5$
Is Netflix a random low-rank matrix?

Compare for coordinate descent (Simon Funk).
Rank = 3

\[ D \]

**fit error**

- **Lowrank**
- **Netflix Data**
- **Random Data \( U[-1 1] \)**

**pred. error**

- **Lowrank**
- **Netflix Data**
- **Random Data \( U[-1 1] \)**

Raghunandan Keshavan, Andrea Montanari and Sewoong Oh

Matrix Completion
Rank = 4

![Graphs showing fit error and pred. error over steps for Lowrank, Netflix Data, and Random Data U[-1 1].]
Rank = 5

fit error

pred. error

D

steps

steps
Proofs (blackboard)