Matrix Completion from Fewer Entries

Raghunandan Keshavan, Andrea Montanari and Sewoong Oh

Stanford University

March 30, 2009

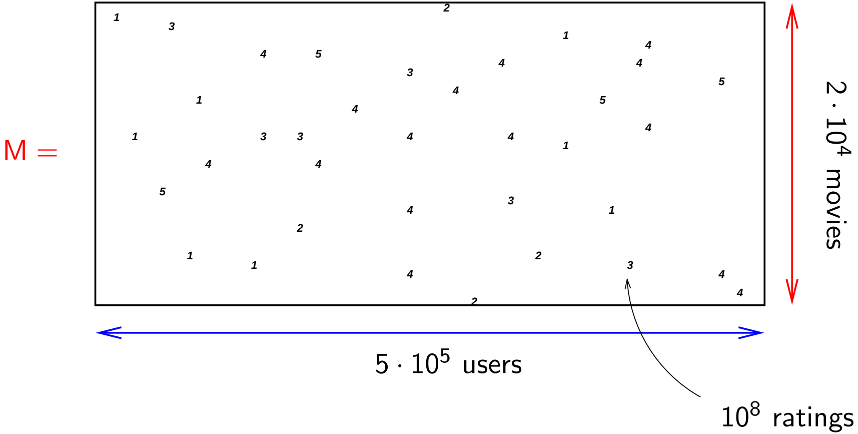
Outline

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

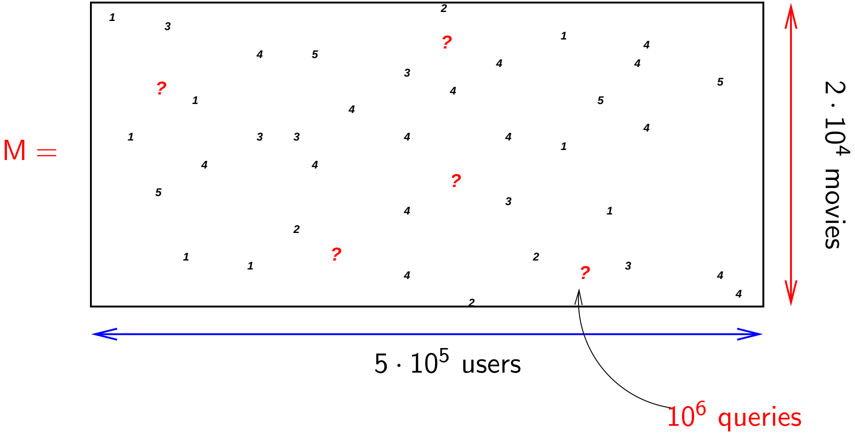
arXiv:0901.3150

The problem, a look at the data, and some results

Netflix dataset: A big (!) matrix



A big (!) matrix



You get a prize if...

$\text{RMSE} < 0.8563$; −)

Is this possible?

You get a prize if. . .

$$\text{RMSE} < 0.8563 \qquad\qquad ;-)$$

Is this possible?

You get a prize if. . .

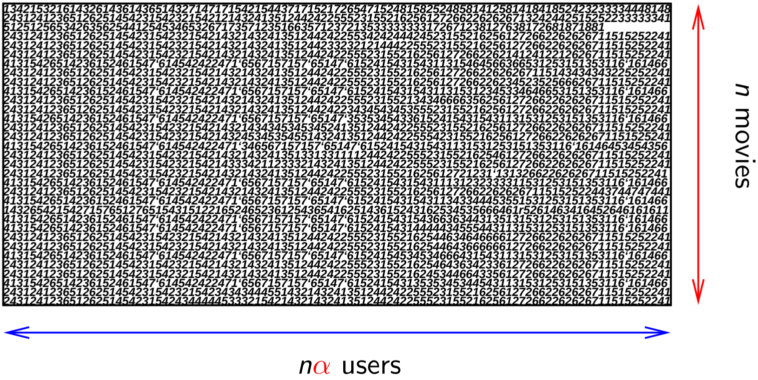
$$\mathrm{RMSE} < 0.8563 \qquad\qquad ;-)$$

Is this possible?

A model: Incoherent low-rank matrices

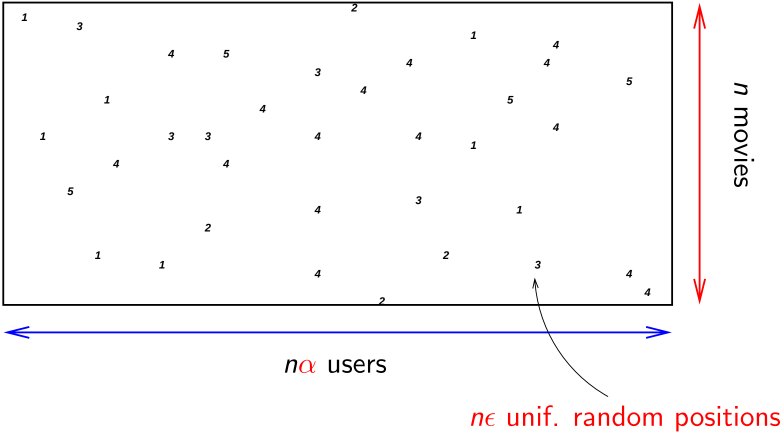
The observations

M =

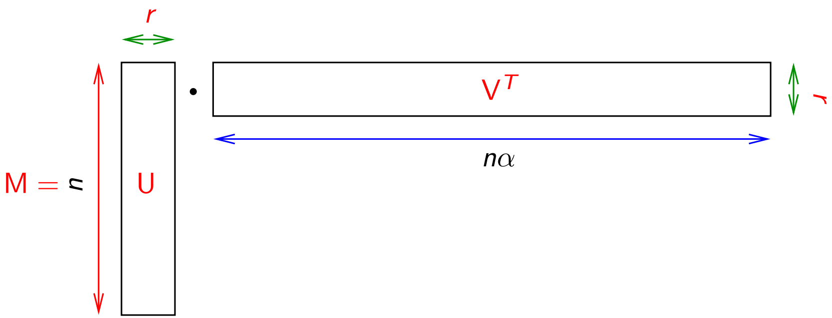


The observations

$M^E =$

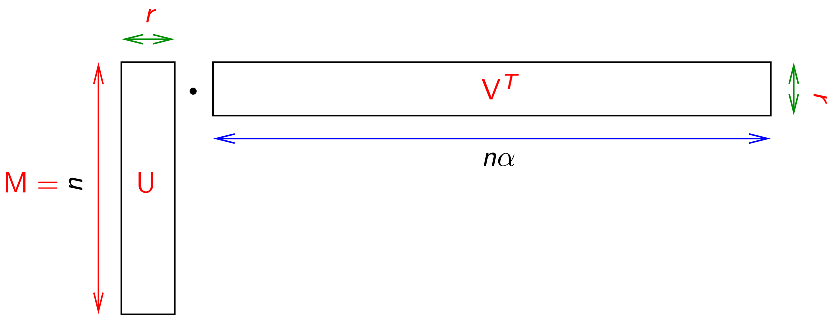


You need some structure!



$r \ll n$

You need some structure!



$r \ll n$

Unstructured factors

A1. Bounded entries

$$|\mathsf{M}_{ia}| \le M_{\max} = \mu_0 \sqrt{r}\,.$$

A2. Incoherence

$$\sum_{k=1}^{r} \mathsf{U}_{ik}^2 \le \mu_1\, r\,, \qquad \sum_{k=1}^{r} \mathsf{V}_{ak}^2 \le \mu_1\, r\,.$$

[Candés, Recht 2008]

Metric (RMSE)

$$D(\mathsf{M}, \hat{\mathsf{M}}) \equiv \left\{ \frac{1}{n^2 \mathsf{M}_{\max}^2} \sum_{i,a} |\mathsf{M}_{ia} - \hat{\mathsf{M}}_{ia}|^2 \right\}^{1/2}$$

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]

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]

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]

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]

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]

Great, but. . .

1. $n^{1/5}$ observations for 1 bit of information?
2. $\mathrm{RMSE} = 0$?
3. SDP $=$ $O(n^{4\cdots6})$. Substitute $n = 10^5$. . .

Great, but...

1. $n^{1/5}$ observations for 1 bit of information?
2. $\mathrm{RMSE} = 0$?
3. SDP $= \ O(n^{4\cdots6})$. Substitute $n = 10^5$...

Great, but. . .

1. $n^{1/5}$ observations for 1 bit of information?
2. $\mathrm{RMSE} = 0$?
3. SDP = $O(n^{4\cdots 6})$. Substitute $n = 10^5$. . .

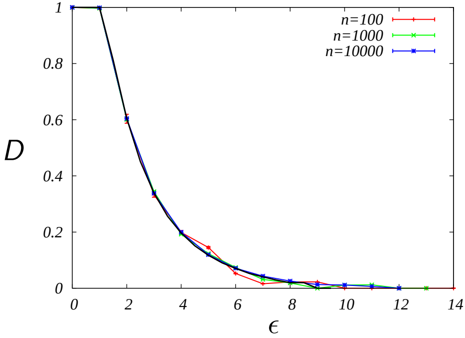
Great, but. . .

1. $n^{1/5}$ observations for 1 bit of information?
2. $\mathrm{RMSE} = 0$?
3. SDP = $O(n^{4\cdots6})$. Substitute $n = 10^5$. . .

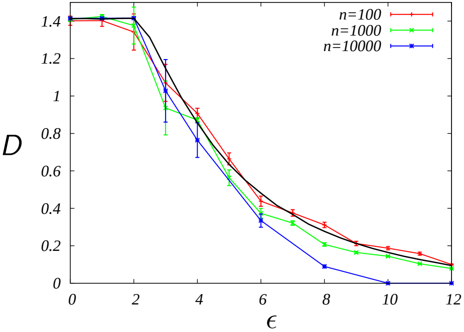
$O(n)$ entries are enough (practice)

A movie

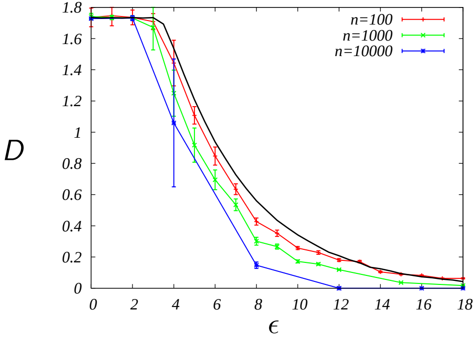
Rank = 1: Bayes optimal vs. Belief Propagation



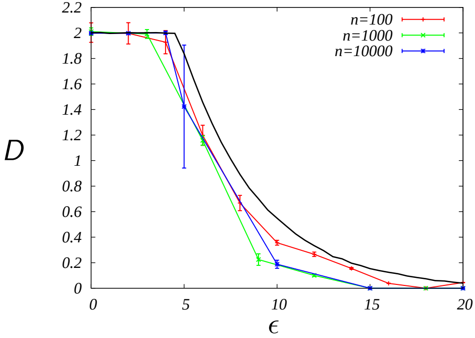
Rank = 2: Belief Propagation



Rank = 3: Belief Propagation



Rank = 4: Belief Propagation



$O(n)$ entries are enough (theory)

Naive spectral algorithm

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

Projection

$$\mathrm{M}^E = \sum_{i=1}^{n} \sigma_i x_i y_i^T, \qquad \sigma_1 \geq \sigma_2 \geq \dots$$

$$\mathrm{T}_r(\mathrm{M}_E) = \frac{n\sqrt{\alpha}}{\epsilon} \sum_{i=1}^{r} \sigma_i x_i y_i^T.$$

Naive spectral algorithm

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

Projection

$$\begin{aligned} \mathrm{M}^E &= \sum_{i=1}^{n} \sigma_i x_i y_i^T \,, \qquad \sigma_1 \geq \sigma_2 \geq \dots \\ \mathrm{T}_r(\mathrm{M}_E) &= \frac{n\sqrt{\alpha}}{\epsilon} \sum_{i=1}^{r} \sigma_i x_i y_i^T \,. \end{aligned}$$

Troubles and solution

If $\epsilon = O(1)$, 'spurious' singular values $\Omega(\sqrt{\log n/(\log \log n)})$.

Trimming

$$\widetilde{M}^E_{ia} = \begin{cases} M^E_{ia} & \text{if } \deg(i) \leq 2\,\mathbb{E}\deg(i), \;\; \deg(a) \leq 2\,\mathbb{E}\deg(a)\,, \\ 0 & \text{otherwise.} \end{cases}$$

Troubles and solution

If $\epsilon = O(1)$, 'spurious' singular values $\Omega(\sqrt{\log n/(\log\log n)})$.

Trimming

$$\widetilde{\mathsf{M}}^E_{ia} = \begin{cases} \mathsf{M}^E_{ia} & \text{if } \deg(i) \le 2\,\mathbb{E}\deg(i),\ \ \deg(a) \le 2\,\mathbb{E}\deg(a)\,, \\ 0 & \text{otherwise.} \end{cases}$$

Not-as-naive spectral algorithm

Spectral Matrix Completion(matrix M^E)

1: Trim M^E, and let $\widetilde{\mathrm{M}}^E$ be the output;
2: Project $\widetilde{\mathrm{M}}^E$ to $\mathrm{T}_r(\widetilde{\mathrm{M}}^E)$;
3: Clean residual errors by gradient descent in the factors.

Theorem (Keshavan, M, Oh, 2009)

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

$$\frac{1}{n\mathrm{M}_{\max}}||\mathrm{M} - \mathrm{T}_r(\widetilde{\mathrm{M}}^E)||_{\mathrm{F}} = \mathrm{RMSE} \le C\sqrt{r/\epsilon}.$$

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* ∞.

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

SPECTRAL MATRIX COMPLETION *returns, whp, the matrix* M.

Theorem (Keshavan, M, Oh, 2009)

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

$$\frac{1}{n\mathrm{M}_{\max}}||\mathrm{M} - \mathrm{T}_r(\widetilde{\mathrm{M}}^E)||_{\mathrm{F}} = \mathrm{RMSE} \le C\sqrt{r/\epsilon}.$$

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

Theorem (Keshavan, M, Oh, 2009)

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

If $\epsilon \ge C' \log n$ *then*

SPECTRAL MATRIX COMPLETION *returns, whp, the matrix* M.

A comparison

Theorem (Achlioptas, McSherry 2007)

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

$$\frac{1}{n\mathrm{M}_{\max}} ||\mathrm{M} - \mathrm{T}_r(\widetilde{\mathrm{M}}^E)||_{\mathrm{F}} = \mathrm{RMSE} \leq 4\sqrt{r/\epsilon}.$$

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

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

A comparison

Theorem (Achlioptas, McSherry 2007)

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

$$\frac{1}{n\mathrm{M}_{\max}} ||\mathrm{M} - \mathrm{T}_r(\widetilde{\mathrm{M}}^E)||_{\mathrm{F}} = \mathrm{RMSE} \leq 4\sqrt{r/\epsilon}.$$

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

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

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

$$\sum_{k=1}^{r} \mathrm{U}_{ik}^2 \leq \mu_1\, r\,,$$

$$\left| \sum_{k=1}^{r} \mathrm{U}_{ik} \mathrm{U}_{jk} \right| \leq \mu_1 \sqrt{r}\,,$$

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

$$\sum_{k=1}^{r} \mathrm{U}_{ik}^2 \leq \mu_1\, r\,,$$

$$\left|\sum_{k=1}^{r} \mathrm{U}_{ik}\mathrm{U}_{jk}\right| \leq \mu_1 \sqrt{r}\,,$$

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

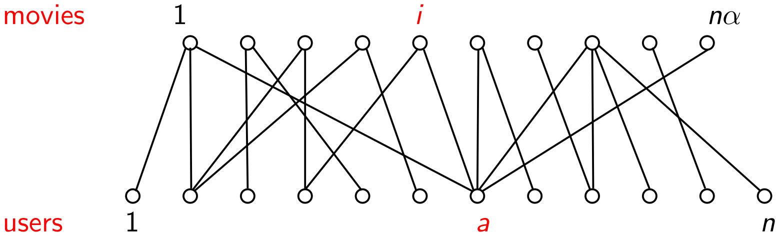
$$\sum_{k=1}^{r} \mathrm{U}_{ik}^2 \leq \mu_1\, r\,,$$

$$\left|\sum_{k=1}^{r} \mathrm{U}_{ik}\mathrm{U}_{jk}\right| \leq \mu_1\, \sqrt{r}\,,$$

Our approach: Graph theory

$(i, a) \in E \quad \Leftrightarrow \quad$ User a rated movie i.

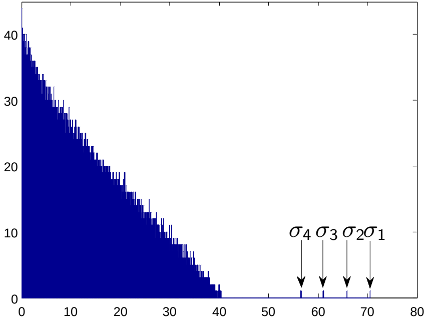
Our approach: Graph theory



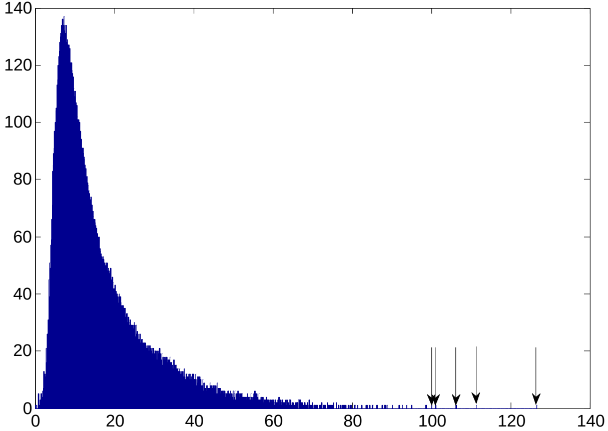
$(i, a) \in E \quad \Leftrightarrow$ User a rated movie i.

Back to the data

Random $r = 4$, $n = 10000$, $\epsilon = 12.5$



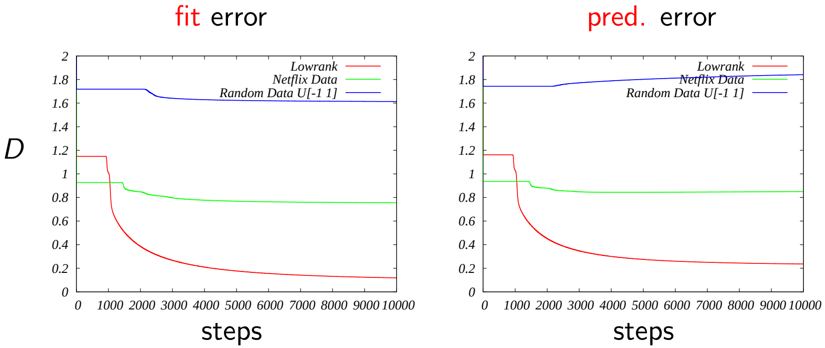
Netflix data (trimmed)



Is Neflix a random low-rank matrix?

Compare for coordinate descent (SimonFunk).

Rank = 3

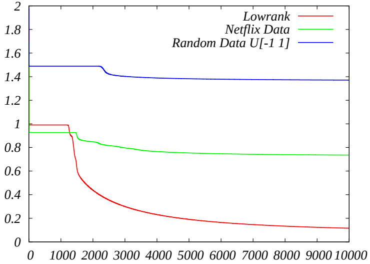


fit error
pred. error
D
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
steps
Lowrank
Netflix Data
Random Data U[-1 1]

Rank $= 4$

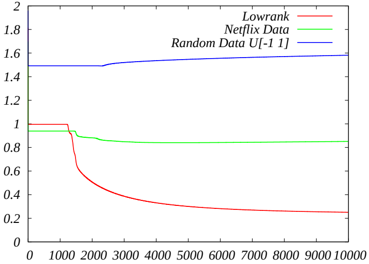
fit error

D



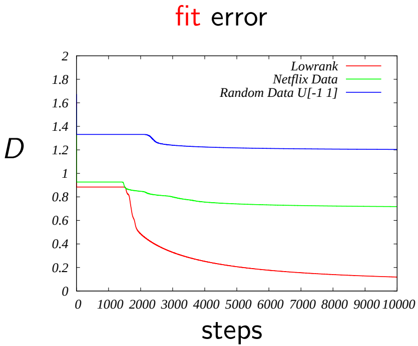
steps

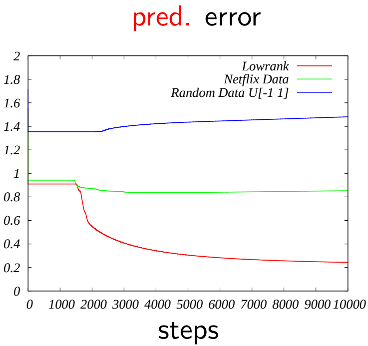
pred. error



steps

Rank = 5


fit error
D
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Lowrank
Netflix Data
Random Data U[-1 1]
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
steps


pred. error
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Lowrank
Netflix Data
Random Data U[-1 1]
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
steps

Proofs (blackboard)