Noise-Tolerant Life-Long Matrix Completion via Adaptive Sampling

Hongyang Zhang
Joint work with Maria-Florina Balcan
Machine Learning Department, CMU

Machine Learning Lunch
Life-Long Matrix Completion

Real-world applications:
- Recommendation System
- Compressed Sensing

What if the Signal is 2D?
- Trivial solution: make it into 1D
- May not work

Netflix Challenge:
We have to consider the correlation among rows and columns!

Matrix Completion

What if the data comes online?

Arrived Columns

Coming Columns

Real-world applications:
Recommendation System
Compressed Sensing
Outline

• Motivation and examples
• Our goal and approach
• Matrix completion background
• Robustness Analysis
• Experimental Results
Our Sampling Model

• Adaptive Sampling
  • Scheme 1: Uniformly take the samples *randomly* (smaller sample complexity)
  • Scheme 2: Request all entries of column from *oracle* (larger sample complexity)

• Sampling scheme in the real world
  • Network Tomography
  • Gene Expression Analysis
  • Recommendation System

Goal: Keep Sample Complexity as small as possible

Krishnamurthy and Singh, Low-Rank Matrix and Tensor Completion via Adaptive Sampling, NIPS 2013
Our Approach

Real-world applications:
- Recommendation System
- Compressed Sensing
- Approx Representable?

Matrix Completion

What if the data comes online?
What if the data is noisy?

Basis Set

Arrived Columns

Coming Columns

Krishnamurthy and Singh, Low-Rank Matrix and Tensor Completion via Adaptive Sampling, NIPS 2013
Noisy Life-Long Matrix Completion

- Challenges in the noisy setting
  - Noise might be adversarial
  - Noise propagates as the data comes along online
Outline

• Motivation and examples
• Our goal and approach
• **Matrix completion background**
• Robustness Analysis
• Experimental Results
Matrix Completion Background

What if the Signal is 2D?

• Trivial solution: make it into 1D
• May not work

Netflix Challenge:

We have to consider the correlation among rows and columns!

Effectiveness of Low Rankness

• It is a global constraint
• Data compression: $mn \rightarrow r(m+n-r)$
• Significantly reduces the degrees of freedom: $mn \rightarrow r(m+n-r)$

We have to consider the correlation among rows and columns.
Matrix Completion Background (cont’d)

- Incoherence is necessary

\[ X = \begin{pmatrix} \vdots \\ \end{pmatrix} \quad (= e_1 e_1^*) \]

- Any subset of entries that misses the (1,1) component tells you nothing!

\[ X = U \Sigma V^T, \quad r \]

- Still need to see the entire first row

\[ \|U^T e_i\|_2 \leq \sqrt{\frac{\mu_0 r}{m}} \quad \text{(left incoherence)} \]

\[ \|V^T e_i\|_2 \leq \sqrt{\frac{\mu_0 r}{n}} \quad \text{(right incoherence)} \]

- Want each entry to provide nearly the same amount of information.
Related Work

• Matrix completion by nuclear norm.

• Matrix completion by alternating minimization

• Matrix completion by adaptive sampling
  • Krishnamurthy & Singh, 2013 & 2014

Very little analysis of noise-tolerant online matrix completion algorithm
Outline

• Motivation and examples
• Our goal and approach
• Matrix completion background
• Robustness Analysis
• Experimental Results
Noise Model

- Bounded Deterministic Noise

\[ \| E_x \| = \leq \varepsilon_{\text{noise}} \]

Basis Set

Arrived Columns

Coming Columns
Noise Model (cont’d)

- Sparse Random Noise

![Diagram of Sparse Random Noise]

- Basis Set
- Arrived Columns
- Coming Columns

Sparse Random Noise Model

Noise Model (cont’d)
Main Results --- Bounded Deterministic Noise

**Theorem (Bounded Deterministic Noise)**

**Sample Complexity:** $O((\mu_0nr + mnk\epsilon_{noise}) \log^2(\frac{r}{\delta}))$

**Output Error:** $\|\hat{M}:t - M:t\|_2 \leq \Theta \left( \frac{m}{d} \sqrt{k\epsilon_{noise}} \right)$

Parameter: $k$ number of bases, $\epsilon_{noise}$ noise magnitude, $r$ rank, $\mu_0$ incoherence, $\delta$ failure prob., $d$ unif($d$)
Discussion --- Bounded Deterministic Noise

Theorem (Bounded Deterministic Noise)

Sample Complexity: \( O((\mu_0 nr + mk\epsilon_{\text{noise}}) \log^2(\frac{r}{\delta})) \)

Output Error: \( ||\hat{M}_{t} - M_{t}||_2 \leq \Theta \left( \frac{m}{d} \sqrt{k\epsilon_{\text{noise}}} \right) \)

Parameter: \( k \) number of bases, \( \epsilon_{\text{noise}} \) noise magnitude, \( r \) rank, \( \mu_0 \) incoherence, \( \delta \) failure prob., \( d \) \( \text{unif}(d) \)

- The error propagates only in the speed of \( \sqrt{k} \), low propagation rate
- Sample Complexity \( O(\mu_0 nr \log^2 n) \), if \( \epsilon_{\text{noise}} \leq O(\mu_0 r/mk) \)
- Incoherence assumption in only one direction
Why left incoherence is enough?

Avoid by \( \left\| U^T e_i \right\|_2 \leq \sqrt{\frac{\mu_0 r}{m}} \) (left incoherence)

Not an Issue:

Arrived Columns

Avoid by \( \left\| V^T e_i \right\|_2 \leq \sqrt{\frac{\mu_0 r}{n}} \) (right incoherence)
**Proof Sketch --- Bounded Deterministic Noise**

**Fact 1**

\[ U^k = \text{span}\{u_1, u_2, \ldots, u_k\} \]

\[ \tilde{U}^k = \text{span}\{\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_k\} \]

If \( \theta(u_i, \tilde{u}_i) \leq \varepsilon_{\text{noise}} \) and \( \theta(\tilde{U}^{i-1}, \tilde{u}_i) \geq 20i\varepsilon_{\text{noise}} \)

Then \( \theta(U^k, \tilde{U}^k) \leq \gamma_k / 2 \)

Proof idea: Reduction on \( k \)

**Fact 2**

\[ U^k \approx \tilde{U}^k \]

\[ \theta(M_{\Omega t}, U^k) \approx \theta(M_{\Omega t}, \tilde{U}^k) \]  
\[ \left\| M_{\Omega t} - P_{\tilde{U}^k_{\Omega t}} M_{\Omega t} \right\|_2 \approx \frac{d}{m} \left\| M_{\Omega t} - P_{\tilde{U}^k_{\Omega t}} M_{\Omega t} \right\|_2 \]

\[ = f(\theta(M_{\Omega t}, \tilde{U}^k)) \]

\[ \theta(M_{\Omega t}, U^k) \text{ is determined by } \left\| M_{\Omega t} - P_{\tilde{U}^k_{\Omega t}} M_{\Omega t} \right\|_2 \]

Far away from subspace

Appro Representable?
Main Results --- Sparse Random Noise

Theorem (Sparse Random Noise, Upper Bound)

Noise Sparsity: \( s_0 \leq O(m) \)

Noise Magnitude: Arbitrarily large

Sample Complexity: \( O(\mu_0 rn \log \left( \frac{r}{\delta} \right)) \)

Output Error: Exact Recovery

Theorem (Sparse Random Noise, Lower Bound)

Sample Complexity: \( \Omega \left( \mu_0 rn \log \left( \frac{r}{\delta} \right) \right) \)

Output Error: Exact Recovery

<table>
<thead>
<tr>
<th>Complexity Lower bound</th>
<th>Passive Sampling</th>
<th>Adaptive Sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathcal{O} \left( \mu_0 nr \log^2 \left( \frac{n}{\delta} \right) \right) [22] )</td>
<td>( \mathcal{O} \left( \mu_0 nr \log^2 \left( \frac{r}{\delta} \right) \right) [19] )</td>
<td>( \mathcal{O} \left( \mu_0 nr \log (r/\delta) \right) ) (Ours)</td>
</tr>
<tr>
<td>( \mathcal{O} \left( \mu_0 nr \log(n/\delta) \right) [10] )</td>
<td>( \mathcal{O} \left( \mu_0 nr \log(r/\delta) \right) ) (Ours)</td>
<td></td>
</tr>
</tbody>
</table>
Main Results --- Mixture of Subspaces

Theorem (Mixture of Subspaces, Sparse Random Noise)

**Noise Sparsity:** \( s_0 \leq O(m) \)

**Noise Magnitude:** Arbitrarily large

**Sample Complexity:** \( O(\mu \tau^2 n \log(\frac{r}{\delta})) \) (Single: \( O(\mu_0 rn \log(\frac{r}{\delta})) \))

**Output Error:** Exact Recovery

*Parameter:* \( \mu, \tau \) incoherence of each subspace, \( \tau \) dimension upper bound of each subspace.
Outline

- Motivation and examples
- Our goal and approach
- Matrix completion background
- Robustness Analysis
- Experimental Results
Experiment Results --- Bounded Deterministic Noise

\[ L = \begin{bmatrix} \mathbf{u}_1 \mathbf{1}_{200}^T, \sum_{i=1}^2 \mathbf{u}_i \mathbf{1}_{200}^T, \sum_{i=1}^3 \mathbf{u}_i \mathbf{1}_{200}^T, \sum_{i=1}^4 \mathbf{u}_i \mathbf{1}_{200}^T, \sum_{i=1}^5 \mathbf{u}_i \mathbf{1}_{1,200}^T \end{bmatrix} \in \mathbb{R}^{100 \times 2000} \]

Noise Magnitude: \( \epsilon_{\text{noise}} = 0.6 \).
Experiment Results --- Sparse Random Noise

White Region: Nuclear norm minimization succeeds.
White and Gray Regions: Our algorithm succeeds.
Black Region: Our algorithm fails.
Experiment Results --- Mixture of Subspaces

Table 2: Life-long Matrix Completion on the first 5 tasks in Hopkins 155 database.

<table>
<thead>
<tr>
<th>#Task</th>
<th>Motion Number</th>
<th>$d = 0.8m$</th>
<th>$d = 0.85m$</th>
<th>$d = 0.9m$</th>
<th>$d = 0.95m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>2</td>
<td>$9.4 \times 10^{-3}$</td>
<td>$6.0 \times 10^{-3}$</td>
<td>$3.4 \times 10^{-3}$</td>
<td>$2.6 \times 10^{-3}$</td>
</tr>
<tr>
<td>#2</td>
<td>3</td>
<td>$5.9 \times 10^{-3}$</td>
<td>$4.4 \times 10^{-3}$</td>
<td>$2.4 \times 10^{-3}$</td>
<td>$1.9 \times 10^{-3}$</td>
</tr>
<tr>
<td>#3</td>
<td>2</td>
<td>$6.3 \times 10^{-3}$</td>
<td>$4.8 \times 10^{-3}$</td>
<td>$2.8 \times 10^{-3}$</td>
<td>$7.2 \times 10^{-4}$</td>
</tr>
<tr>
<td>#4</td>
<td>2</td>
<td>$7.1 \times 10^{-3}$</td>
<td>$6.8 \times 10^{-3}$</td>
<td>$6.1 \times 10^{-3}$</td>
<td>$1.5 \times 10^{-3}$</td>
</tr>
<tr>
<td>#5</td>
<td>2</td>
<td>$8.7 \times 10^{-3}$</td>
<td>$5.8 \times 10^{-3}$</td>
<td>$3.1 \times 10^{-3}$</td>
<td>$1.2 \times 10^{-3}$</td>
</tr>
</tbody>
</table>
Summary

• Life-Long Matrix Completion
  • Online
  • Noise Tolerant

• Sample Complexity
  • Bounded Noise: As small as noiseless case
  • Sparse Noise: Achieve lower bound in the worst case, better than nuclear norm minimization method
  • Mixture of Subspaces: Potential smaller sample complexity
Thank You