Matrix Completion with Implicit Clustering

Md Enayat Ullah[1]

SUPERVISORS:

Dr. Purushottam Kar [1]

Dr. Debassis Kundu [1]

Dr. Prateek Jain[2]

November 8, 2016

[1]Indian Institute of Technology Kanpur

[2]Microsoft Research, India

Table of Contents

Matrix Completion

Matrix Completion

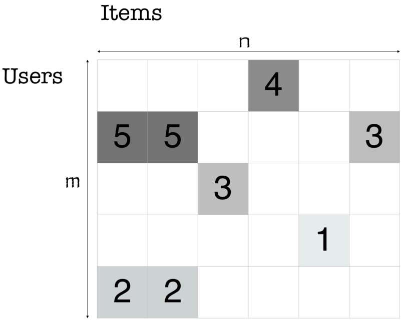


Figure 1: User-Items Rating Matrix (Incomplete)

Matrix Completion

Items

n

Users

m

1	1	3	4	3	4
5	5	4	5	4	3
4	3	3	2	3	2
3	4	3	5	1	4
2	2	1	3	5	4

Figure 2: User-Items Rating Matrix (Completed!)

Low rank Matrix Completion

Ill-posed Problem

Constraints

1. Low rank ($\text{rank(M)} = k \ll \min(m, n)$)

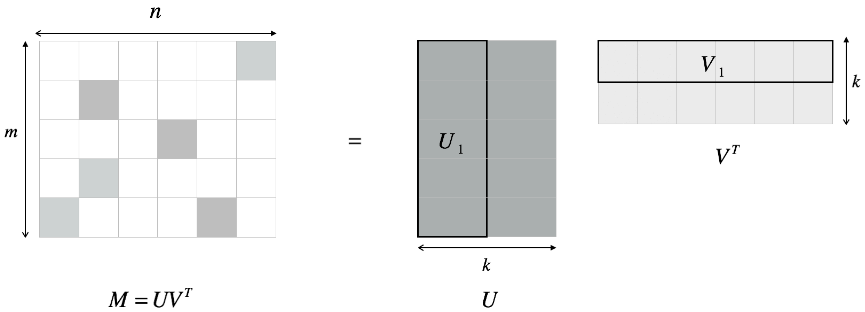


Figure 3: Low rank Decomposition

Matrix Completion

Applications

- Collaborative Filtering (Recommendation Systems)
- System Identification

Algorithms

- Convex Relaxation (CR09)
- Alternating Minimization (KOM09), (JNS13)

Projection Operator

Given:

			4		
5	5				3
		3			
				1	
2	2				

Define:

$$\Omega := \{(i,j) \in [m] \times [n] \text{ s.t } M_{ij} \text{ is observed}\}$$

$$P_{\Omega}(M)_{ij} := \begin{cases} M_{ij} & \forall (i,j) \in \Omega \\ 0 & \text{otherwise} \end{cases}$$

Optimization

Optimization Problem

$$\min_{U,V} f(U,V) = \left\| P_\Omega\left(UV^T\right) - P_\Omega\left(M\right)\right\|_{\mathcal{F}}^2$$

where $U \in \mathbb{R}^{m\times r}$, $V \in \mathbb{R}^{n\times r}$

Non-Convex Objective

1. Exact-minimization is NP-Hard.
2. Exponentially large number of local minima.
3. Saddle points

Matrix Completion

Constraints

1. Low rank ($\text{rank}(M) = k \ll \min(m, n)$)
2. Uniformly and Independently sampled entries.

Sample complexity $= \mathcal{O}(k^4 n \log n)$.

What if?

0	0	0	0	0	0
0	0	0	0	0	0
0	0	3	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0

Incoherence

Incoherence

A matrix $M \in \mathcal{R}^{m \times n}$ of rank k is incoherent with parameter μ if

$$\left\|U^i\right\|_2 \leq \frac{\mu\sqrt{k}}{\sqrt{m}} \;\forall\; i \in [m]$$
$$\left\|V^j\right\|_2 \leq \frac{\mu\sqrt{k}}{\sqrt{n}} \;\forall\; j \in [n]$$

where $M = U\Sigma V$ is the SVD of M and U^i and V^j are the i^{th} and j^{th} row of U and V respectively.

Algorithm

Algorithm 1 Alternating Minimization for Matrix Completion

1: **Input**: $P_\Omega(M)$, Ω

2: **Output**: $\widehat{U}^T, \widehat{V}^T$

3: Initialize $U^0 = \text{LSVD}(P_\Omega(M), r)$

4: **for** $t = 1, 2, \ldots T$ **do**

5: $\quad \widehat{V}^{t+1} = \arg\min_V \left\| P_\Omega\left(\widehat{U}^t V^T - M\right)\right\|_{\mathcal{F}}^2$

6: $\quad \widehat{U}^{t+1} = \arg\min_U \left\| P_\Omega\left(U\left(\widehat{V}^{t+1}\right)^T - M\right)\right\|_{\mathcal{F}}^2$

7: **end for**

Therorem (JNS13)

Let $M = U^* \Sigma^* {V^*}^T (n \geq m) \in \mathcal{R}^{m \times n}$ be a rank-k μ-incoherent matrix. Also, let each entry of M be observed uniformly and independently with probability

$$p > C \frac{\left(\frac{\sigma_1^*}{\sigma_r^*}\right)^2 \mu^4 k^{2.5} \log n \log\left(\frac{r\|M\|_{\mathcal{F}}}{\epsilon}\right)}{m\delta_{2k}^2}$$

where $\delta_{2k} \leq \frac{\sigma_r^*}{12k\sigma_1^*}$ and $C > 0$ is a global constant. Then with high probability, for $T = C' \log \frac{\|M\|_{\mathcal{F}}}{\epsilon}$, the outputs $\widehat{U}_T$ and $\widehat{V}_T$ with input $(\Omega, P_\Omega(M))$ satisfy $\left\| M - \widehat{U}_T \widehat{V}_T^T \right\|^{\mathcal{F}} \leq \epsilon$.

Modification

Modifications

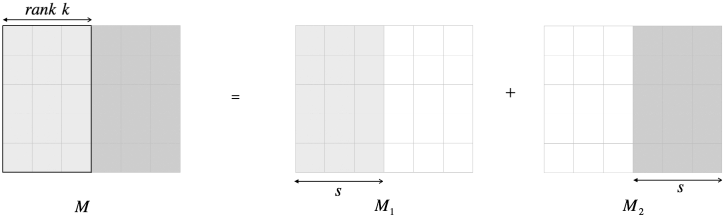


Figure 4: Two rank k components of M

Rank k components

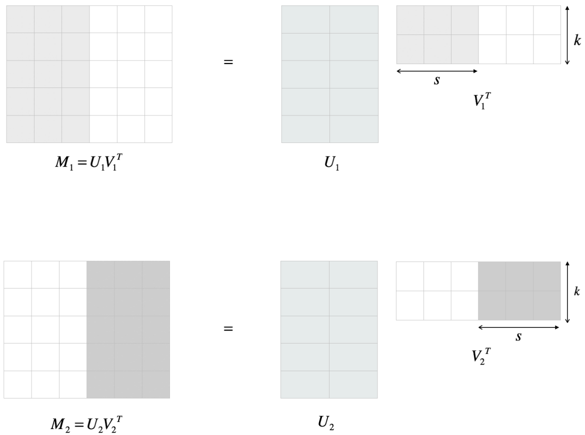


Figure 5: Matrix Factorization of M_1 and M_2

Matrix Factorization

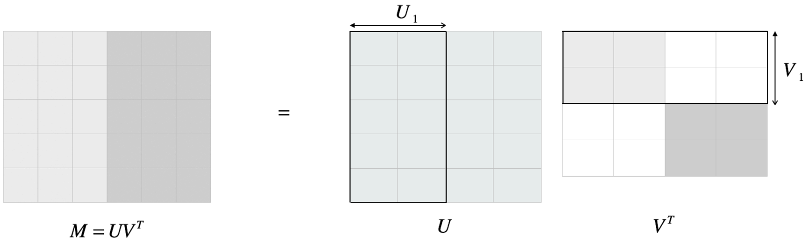


Figure 6: Matrix Completion with r rank-k components

Clustering

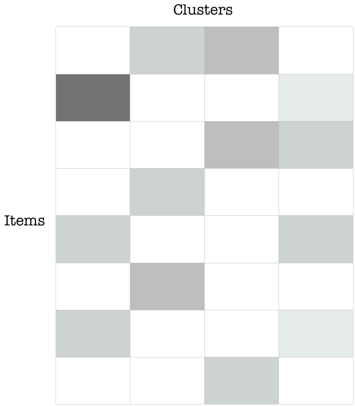


Figure 7: Clusters

Dictionary Learning

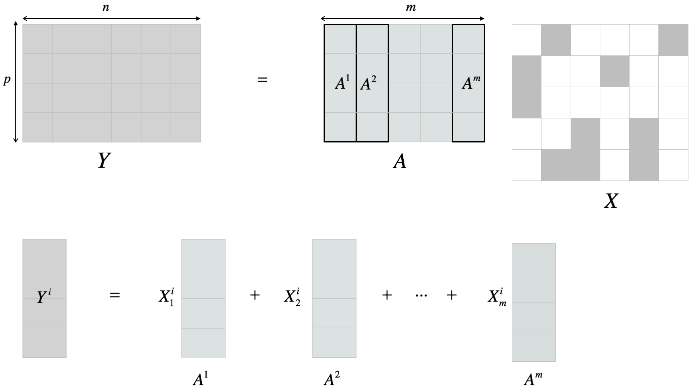


Figure 8: Dictionary Learning

Algorithm 2 Alternating Minimization for Matrix Completion

1: Initialize $U^0 = \text{LSVD}\left(P_\Omega\left(M\right), r\right)$
2: **for** $t = 1, 2, \ldots$ **do**
3: **for** $i = 1, 2 \ldots n$ **do**
4: $\widehat{V}_i^t = \arg\min_{V_i} \left\| P_\Omega \left(M^i - \widehat{U}^t V_i \right) \right\|$ s.t. $\|V_i\|_0 \leq k$ `//Sparse Recovery`
5: **end for**
6: **for** $i = 1, 2 \ldots m$ **do**
7: $\widehat{U}_i^{t+1} = \left(\sum_{i:(i,j)\in\Omega} \left(\widehat{V}_j^t\right)^T \widehat{V}_j^t \right) \left(\sum_{i:(i,j)\in\Omega} M_i^j \widehat{V}_j^t \right)$ `//LS Estimate`
8: **end for**
9: **end for**

Sparse Recovery

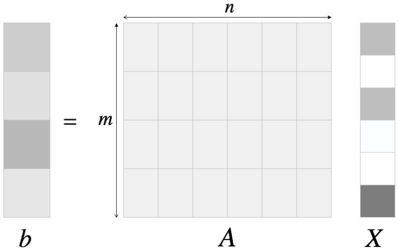


Figure 9: Sparse Recovery ($m \ll n, \ \|X\|_0 \leq k$)

Constraint

A should satisfy Restricted Isometric Property.

Restricted Isometry Property (RIP)

Restricted Isometry Property (k, δ_k) (CT06)

A matrix $A \in \mathbb{R}^{m \times n}$ is said to satisfy k-RIP with RIP constant δ_k if $\forall\ X \in \mathbb{R}^n$ with $\|X\|_0 \leq k$, the following hold:

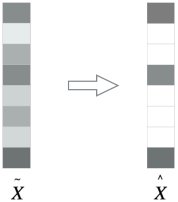
$$(1 - \delta_k) \|X\|_2 \leq \|AX\|_2 \leq (1 + \delta_k) \|X\|_2$$

Optimization Problem

$$\widehat{X} = \arg\min_X \|X\|_0 \text{ s.t. } AX = b$$

Sparse Recovery

ISTA(BT09)	**IHT**(BD09)
A satistfies $(2k, \delta)$-RIP for $\delta < \dfrac{1}{1+\sqrt{2}}$	A satisfies $(3k, \delta)$-RIP for $\delta \leq \dfrac{1}{2}$
Convergence: $\dfrac{1}{\epsilon}$ or $\dfrac{1}{\sqrt{\epsilon}}$	Convergence: $\log\left(\dfrac{1}{\epsilon}\right)$
Convex Relaxation: $\Vert.\Vert_0 \to \Vert.\Vert_1$ LASSO: $\min\limits_{X} \Vert AX - b \Vert_2^2 + \lambda \Vert X \Vert_1$	$X^{t+1} = \Pi_{\mathcal{B}_0(k)}\left(X^t - A^T\left(AX^t - b\right)\right)$



Objective

Sample Complexity

- Information Theoretic samples required $= \mathcal{O}(mr + nk)$

Initialization

- $\widehat{U}^0 \, (= LSVD(P_\Omega(M), r))$ is "close" to U^*

Convergence

- Alternating minimization converges to local/global minima

Rank-1 Components

Rank-1

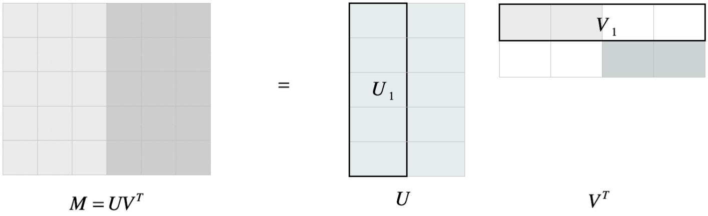


Figure 10: All components of rank 1

Experiments

Experiment

$$\text{Error} = \left\| \widehat{M} - M \right\|_{\mathcal{F}}$$

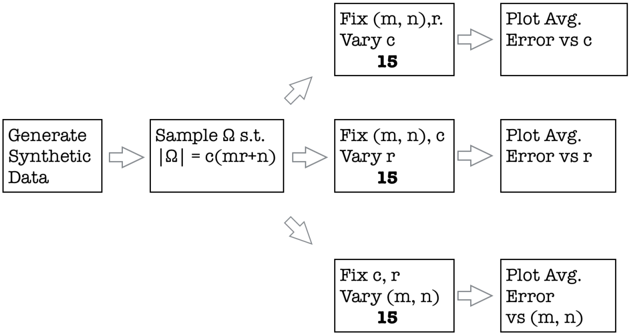


Figure 11: Experiment Flowchart

Error vs c

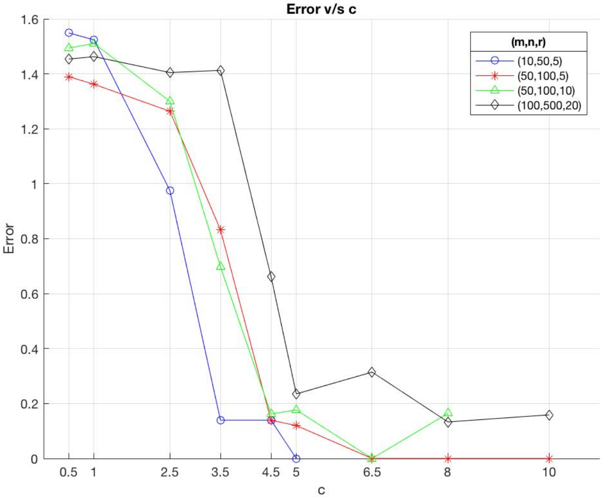


Figure 12: Error vs. c

Error vs rank

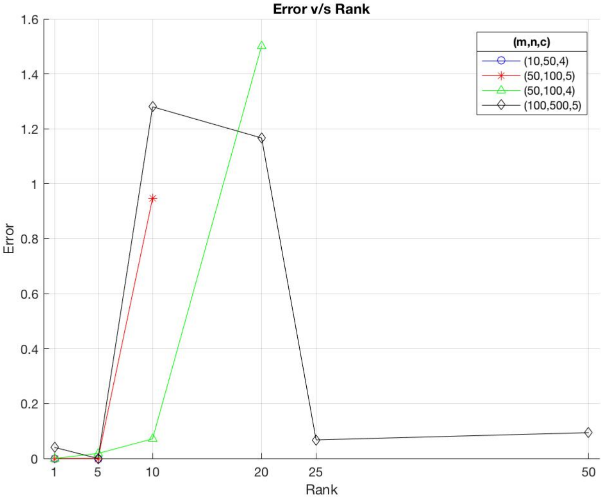


Figure 13: Error vs. rank

Error vs dimensions (m,n)

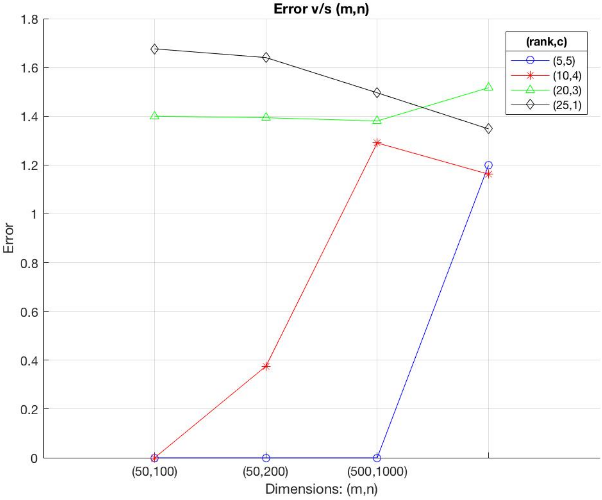


Figure 14: Error vs. dimensions (m,n)

References

[BD09] Thomas Blumensath and Mike E Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, 2009.

[BT09] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[CR09] Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

References II

[CT06] Emmanuel J Candes and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory*, 52(12):5406–5425, 2006.

[JNS13] Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 665–674. ACM, 2013.

[KOM09] Raghunandan H Keshavan, Sewoong Oh, and Andrea Montanari. Matrix completion from a few entries. In *2009 IEEE International Symposium on Information Theory*, pages 324–328. IEEE, 2009.

Questions?