Provable Matrix Completion using Alternating Minimization

Praneeth Netrapalli

The University of Texas at Austin

Joint work with Prateek Jain and Sujay Sanghavi

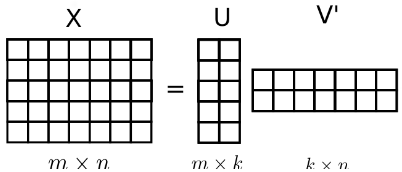
Jun 4, 2013

Alternating Minimization (AltMin)

General Algorithm

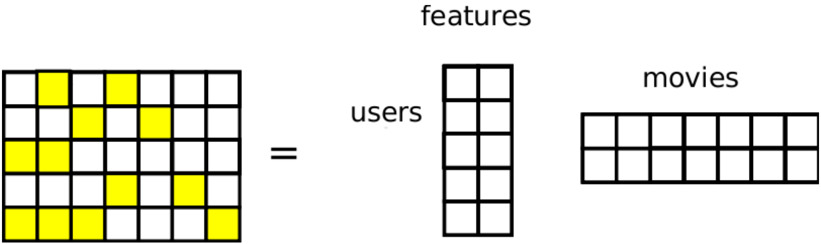
To minimize $f(X)$ over rank-k matrices X, repeat the following:

- fix U and minimize $f(UV^\dagger)$over V
- fix V and minimize $f(UV^\dagger)$over U



- A popular Empirical approach to solve low rank matrix problems eg. matrix completion, clustering etc.
- **Challenge**: few theoretical guarantees

Matrix Completion

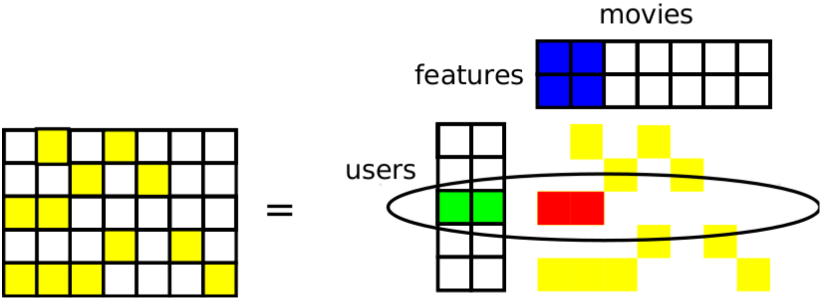


- Given some elements, fill in the rest
- Not possible in general; what if low rank?
- Metrics: Sample complexity and Computational complexity

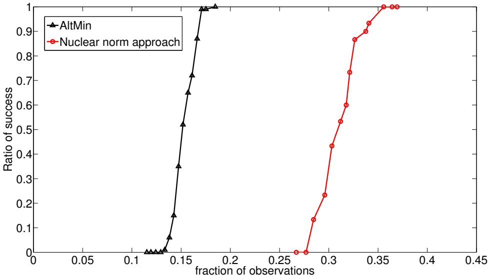
Matrix Completion via Alternating Minimization

$$\min \sum_{(i,j)\in \text{known set}} (M_{ij} - X_{ij})^2 \text{ s.t. } \mathrm{rank}(X) \leq k$$

$$= \min \sum_{(i,j)\in \text{known set}} \left(M_{ij} - U_i^{\dagger} V_j\right)^2 \text{ s.t. } U \in \mathbb{R}^{m\times k}, V \in \mathbb{R}^{n\times k}$$



A Comparision



- Nuclear norm / Trace norm approach : convex relaxation.
- Empirically, AltMin has
  - similar sample complexity and
  - better computational complexity.

Challenge: AltMin formulation is non-convex.

Our Results

- First theoretical guarantees for AltMin in any low rank setting
- We prove results for
  - matrix sensing
  - matrix completion

Matrix Sensing

$$y = \mathcal{A}(X) = \begin{bmatrix} < A_1, X > \\ \vdots \\ < A_d, X > \end{bmatrix}$$

vector

linear operator

Low rank matrix

Problem: Given y and $\mathcal{A}$, recover X.

Natural Algorithm (AltMinSense)

1. (Initialization) $\widehat{U}^0 \leftarrow$ top k-left s.v. of $\sum y_i A_i$
2. In iteration t:
   - $\widehat{V}^t \leftarrow \text{argmin}_{V \in \mathbb{R}^{n \times k}} \left\| y - \mathcal{A}(\widehat{U}^{t-1} V^\dagger) \right\|_2$
   - $\widehat{U}^t \leftarrow \text{argmin}_{U \in \mathbb{R}^{m \times k}} \left\| y - \mathcal{A}(U (\widehat{V}^t)^\dagger) \right\|_2$

Restricted Isometry Property (RIP)

Existing results require RIP assumptions.

RIP [RFP10]

A linear operator $\mathcal{A}(\cdot) : \mathbb{R}^{m\times n} \rightarrow \mathbb{R}^d$ satisfies k-RIP with δ_k, if for all $X \in \mathbb{R}^{m\times n}$ s.t. $\text{rank}(X) \leq k$, the following holds:

$$(1-\delta_k)\left\|X\right\|_F^2 \leq \left\|\mathcal{A}(X)\right\|_2^2 \leq (1+\delta_k)\left\|X\right\|_F^2 .$$

- $\delta_k = 0 \Rightarrow$ Identity map
- $\delta_k = 1 \Rightarrow$ No information

Existing Results

Trace norm approach [RFP10]

$$\begin{array}{c} \min \|y - \mathcal{A}(X)\|_2 \\ \text{s.t. } \operatorname{rank}(X) \le k \end{array} \rightarrow \begin{array}{c} \min \|y - \mathcal{A}(X)\|_2 \\ \text{s.t. } \|X\|_* \le \sqrt{k} \end{array}$$

- $\delta_{5k} < \frac{1}{10}$

Singular Value Projection [JMD10]

- $\delta_{2k} < \frac{1}{3}$

Drawback

- Need to compute many SVDs during execution - very slow in practice

Our Results

Theorem

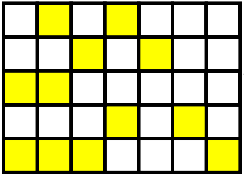
If $\delta_{2k} < \left(\frac{\sigma_k^*}{\sigma_1^*}\right)^2 \frac{1}{100k}$, then

$$\left\| M - \widehat{U}^T (\widehat{V}^T)^\dagger \right\|_F < \left(\frac{1}{2}\right)^T$$

Remarks

1. δ_{2k} depends on the condition number unlike in existing work
   - modified algorithm: $\delta_{2k} < \frac{1}{3200k^2}$
2. Linear convergence: $log\frac{1}{\epsilon}$ iterations for ϵ error.

Matrix Completion



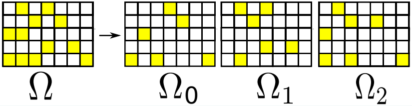
Problem

Given elements in Ω, find the low rank matrix M.

Analysis is harder

- Ω does not in general satisfy RIP.
- Dependence between iterates.

Our Algorithm



- Divide Ω into $2T+1$ subsets $\Omega_0, \cdots \Omega_{2T}$ by uniform sampling.
- Use Ω_i for the i^{th} iteration of AltMin.

AltMinComplete

(Initialization) $\widehat{U}^0 \leftarrow$ top k-left s.v. of $(M)_{\Omega_0}$

FOR $t = 0, \cdots, T-1$

$$\widehat{V}^{t+1} \leftarrow \text{argmin}_{V \in \mathbb{R}^{n \times k}} \left\| \left(\widehat{U}^t V^\dagger - M \right)_{\Omega_{t+1}} \right\|_F^2$$

$$\widehat{U}^{t+1} \leftarrow \text{argmin}_{U \in \mathbb{R}^{m \times k}} \left\| \left(U \left(\widehat{V}^{t+1} \right)^\dagger - M \right)_{\Omega_{T+t+1}} \right\|_F^2$$

ENDFOR

- **Conjecture**: Do not need this partition.
- Same algorithm proposed and analyzed independently by [Kes12]

A Hard Case

?	0			0		0
		0	0		0	
0	0			0		
0			0		0	
	0	0				0

Incoherence

Incoherence [CR09]

$M = U^* \Sigma^* (V^*)^\dagger$ is incoherent with parameter μ if

- $\left\| u^{(i)} \right\|_2 \leq \frac{\mu \sqrt{k}}{\sqrt{m}}\ \forall\, i \in [m]$ and
- $\left\| v^{(j)} \right\|_2 \leq \frac{\mu \sqrt{k}}{\sqrt{n}}\ \forall\, j \in [n]$.

coherent

1	0
0	1
0	0
0	0
0	0
0	0

incoherent

0	$\frac{-1}{\sqrt{n}}$
$\frac{1}{\sqrt{n}}$	$\frac{-1}{\sqrt{n}}$
$\frac{1}{\sqrt{n}}$	$\frac{-1}{\sqrt{n}}$
$\frac{1}{\sqrt{n}}$	0
$\frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{n}}$
$\frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{n}}$

Existing Results

Existing results assume uniform sampling and incoherence of M.

Trace norm approach [CR09, CT09]

$$\begin{array}{c}\min \sum_{(i,j)\in\Omega}(X_{ij}-M_{ij})^2 \\ \text{s.t. } \operatorname{rank}(X) \le k\end{array} \rightarrow \begin{array}{c}\min \sum_{(i,j)\in\Omega}(X_{ij}-M_{ij})^2 \\ \text{s.t. } \|X\|_* \le \sqrt{k}\end{array}$$

- $O\left(knlogn\right)$ observations
- Drawback: need many SVD calculations

OptSpace [KMO10]

- Opt. on Grassman manifold: $O\left(f\left(\frac{\sigma_1^*}{\sigma_k^*}\right) \; kn \log n\right)$
  - rate of convergence not known

Our Results

Theorem

Let M be incoherent. If

$$\#\text{ measurements } > C\left(\frac{\sigma_1^*}{\sigma_k^*}\right)^6 k^7 n \log n \log \frac{1}{\epsilon},$$

then after $T = O\left(\log \frac{1}{\epsilon}\right)$ iterations, we have:

$$\left\| M - \widehat{U}^T (\widehat{V}^T)^\dagger \right\|_F < \epsilon.$$

Advantages:

- linear convergence : $\log \frac{1}{\epsilon}$ vs $\frac{1}{\sqrt{\epsilon}}$
- each iteration very fast
- low storage requirement

Weakness: Dependence on

- condition number
- required accuracy
- k

Comparison

	Sample comp. (d)	Comp. comp.
Our Results	$O\left(\left(\frac{\sigma_1^*}{\sigma_k^*}\right)^6 k^7 n \log n \log \frac{1}{\epsilon}\right)$	$O\left(dk^2 \log \frac{1}{\epsilon}\right)$
AltMin [Kes12]	$O\left(\left(\frac{\sigma_1^*}{\sigma_k^*}\right)^8 kn \log n \log \frac{1}{\epsilon}\right)$	$O\left(dk^2 \log \frac{1}{\epsilon}\right)$
Trace norm [CT09]	$O\left(kn \log n\right)$	$O\left(\frac{n^3}{\sqrt{\epsilon}}\right)$

Main Idea of the Proof

- If $\Omega =$ all elements, then AltMin becomes the well-known power method.
- In general, iterates take the form:

$$\widehat{V}^{t+1} = \underbrace{V^* \Sigma^* {U^*}^\dagger U^t}_{\text{Power-method Update}} - \underbrace{F}_{\text{Error Term}} \text{ and}$$

$$\|F\|_2 \downarrow \text{ as } t \uparrow .$$

- Use RIP to show decay.
- Technical difficulty: Establishing incoherence of U^t.

Summary and Beyond

Summary

- First theoretical guarantees for AltMin in any low rank setting
- Results for
  - Matrix sensing
  - Matrix completion

Further Directions

- Recent result for AltMin in Phase retrieval [NJS13]
- Theory for AltMin in clustering, sparse PCA, NMF etc.?

Thank you!

References

Emmanuel J. Candès and Benjamin Recht.
Exact matrix completion via convex optimization.
Foundations of Computational Mathematics, 9(6):717–772, December 2009.

Emmanuel J. Candès and Terence Tao.
The power of convex relaxation: Near-optimal matrix completion.
IEEE Trans. Inform. Theory, 56(5):2053–2080, 2009.

Prateek Jain, Raghu Meka, and Inderjit S. Dhillon.
Guaranteed rank minimization via singular value projection.
In *NIPS*, pages 937–945, 2010.

Raghunandan H. Keshavan.
Efficient algorithms for collaborative filtering.
Phd Thesis, Stanford University, 2012.

Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh.
Matrix completion from a few entries.
IEEE Transactions on Information Theory, 56(6):2980–2998, 2010.

Praneeth Netrapalli, Prateek Jain, and Sujay Sanghavi.
Phase retrieval using alternating minimization.
CoRR, abs/1306.0160, 2013.

Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo.
Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization.
SIAM Review, 52(3):471–501, 2010.