Provable Matrix Completion using Alternating Minimization

Praneeth Netrapalli

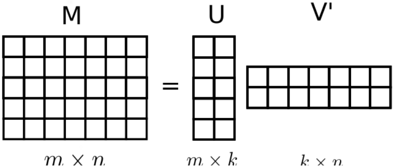
The University of Texas at Austin

Joint work with Prateek Jain and Sujay Sanghavi

Dec 8, 2012

Alternating Minimization (AltMin)

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

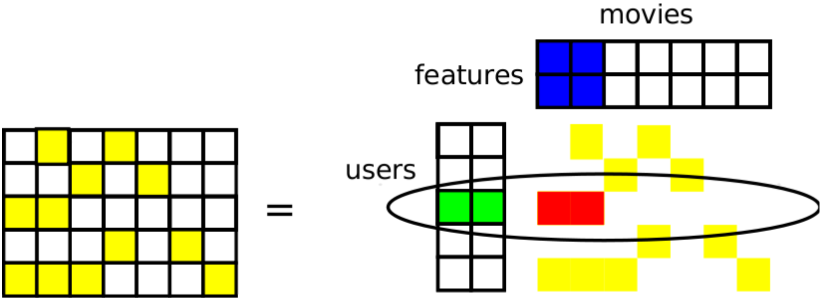


General Algorithm

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

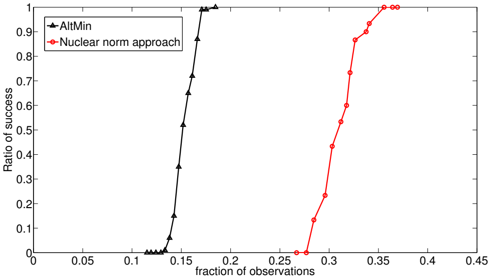
Matrix Completion



Natural decoupling:

- no need to store large matrices
- easily parallelizable: m least squares problems over k variables

A Comparision



- Nuclear norm approach : a leading theoretical approach.
- Empirically, AltMin has
  - similar sample complexity and
  - better computational complexity.

Matrix Sensing

$$y = \mathcal{A}(X)$$

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:
   - $\widehat{V}^t \leftarrow \text{argmin}_{V \in \mathbb{R}^{n \times k}} \left\| y - \mathcal{A}(\widehat{U}^{t-1} V^\dagger) \right\|_2$
   - $\widehat{U}^t \leftarrow \text{argmin}_{U \in \mathbb{R}^{m \times k}} \left\| y - \mathcal{A}(U (\widehat{V}^t)^\dagger) \right\|_2$

Solving a non-convex problem

- no known analysis for convergence.

Existing Results

Existing results require RIP assumptions.

Restricted Isometry Property (RIP) [RFP10]

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

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

- [RFP10] (Nuclear norm approach): $\delta_{5k} < \frac{1}{10}$
- [JMD10] (Projected gradient descent): $\delta_{2k} < \frac{1}{3}$
- **Drawback**: Need to compute many SVDs during execution - very slow in practice

Our Results

Theorem

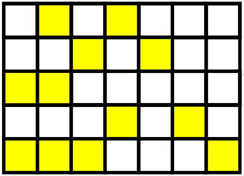
Let $\sigma_1^* \geq \sigma_2^* \cdots \geq \sigma_k^*$ be the singular values of M. If $\mathcal{A}(\cdot)$ satisfies $2k$-RIP with constant $\delta_{2k} < \left(\frac{\sigma_k^*}{\sigma_1^*}\right)^2 \frac{1}{100k}$, then after $T > 2\log\frac{\|M\|_F}{\epsilon}$ iterations of AltMinSense, we have:

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

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

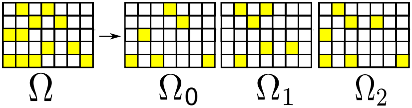


Problem

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

- **Natural approach**: do AltMin on $\left\|P_\Omega(M) - P_\Omega(UV^\dagger)\right\|_F$.
- Major component of [Kor09] (winning entry of the Netflix prize).
- Analysis is harder: Ω does not in general satisfy RIP.

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 $P_{\Omega_0}(M)$

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

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

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

ENDFOR

Conjecture: Do not need this partition.

Existing Results

Existing results assume uniform sampling and incoherence of M.

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

- [CR09, CT09] (Nuclear norm): $Ckn \log n$
  - need many SVD calculations
- [KMO10] (Opt. on Grassman manifold): $C \; f\left(\frac{\sigma_1^*}{\sigma_k^*}\right) \; kn \log n$
  - slower in practice; rate of convergence not known

Our Results

Theorem

Let M be incoherent with parameter μ and $\sigma_1^* \geq \cdots \geq \sigma_k^*$ be its singular values. If

$$p > C \left(\frac{\sigma_1^*}{\sigma_k^*} \right)^4 \frac{\mu^2 k^{4.5} \log n \log \frac{k\|M\|_F}{\epsilon}}{m},$$

then in $T = C' \log \frac{\|M\|_F}{\epsilon}$ iteration of AltMinComplete, we have:

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

Weakness: Dependence on

- condition number
- required accuracy
- k

Advantages:

- linear convergence : $\log \frac{1}{\epsilon}$ vs $\frac{1}{\sqrt{\epsilon}}$
- low rank intermediate matrices

.

Main Idea of the Proof

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

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

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

Summary

- Showed AltMin is globally optimal for
    - Matrix sensing and
    - Matrix completion
- First such result for AltMin
- Linear convergence: ϵ error in $\log \frac{1}{\epsilon}$ iterations
- However, requires more samples

Future Work

Future Work

- **Completion**: remove the dependence on condition number and desired accuracy?
- **Other Problems**: application of the current analysis to obtain guarantees on AltMin for clustering, sparse PCA etc.

Full version: `http://arxiv.org/abs/1212.0467`

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

Yehuda Koren.
The BellKor solution to the Netflix grand prize, 2009.

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.