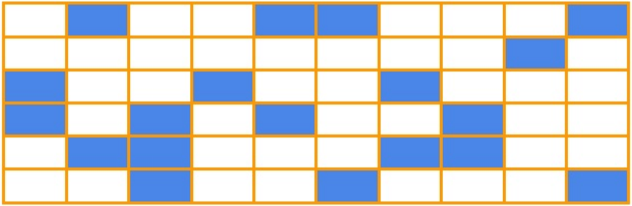
Noisy Matrix Completion Using Alternating Minimization

Suriya Gunasekar, **Ayan Acharya**, Neeraj Gaur, Joydeep Ghosh

UT Austin, Dept. of ECE

September 23, 2013

- **Problem:** Given a subset of entries of the matrix, reconstruct the original matrix.
- The observation matrix is assumed to be of low rank.
- Applications: recommender systems, multitask learning, remotesensing, image inpainting



Noisy Matrix Completion

- **Problem:** Given a noisy matrix $\widetilde{M} = M + N$ and an observation set Ω, reconstruct M.

Alternating Minimization (AltMin) Algorithm

- A popular empirical approach to solve low rank matrix problems
- Predict the unobserved entries of M using low rank approximation, $\widehat{M} = \widehat{U}\widehat{V}^\dagger$
- $P_\Omega(X)_{ij} = X_{ij}$ if $(i,j) \in \Omega$ and 0 otherwise.
- Solve the following non–convex problem:

$$\widehat{U}, \widehat{V} = \underset{U,V}{argmin} \|P_\Omega(M - UV^\dagger)\|_F$$

Repeat till convergence:

- Step 1: $V \leftarrow \text{argmin}_V \|P_\Omega(M - UV^\dagger)\|_F^2$
- Step 2: $U \leftarrow \text{argmin}_U \|P_\Omega(M - UV^\dagger)\|_F^2$

Algorithm Analyzed

Algorithm 1 Alternating Least Square Minimization (ALSM)

1: Create $(2T+1)$ subsets of Ω by sampling $|\Omega|$ elements uniformly with replacement (independence of iterations).

2: Set $\widetilde{U}^0 = \text{SVD}(P_{\Omega_0}(\widetilde{M})/p, k)$ (initialization).

3: Set all elements of $\widetilde{U}^0$ that have magnitude greater than $\frac{2\mu\sqrt{k}}{\sqrt{n}}$ to zero and orthonormalize the columns to get $\widehat{U}^0$ (clipping).

4: **for** $t = 0, \cdots, (T-1)$ do

$$\widehat{V}^{(t+1)} \leftarrow \underset{V \in \mathbb{R}^{n \times k}}{\text{argmin}} \|P_{\Omega^{(t+1)}}(\widehat{U}^t V^\dagger - \widetilde{M})\|_F.$$

$$\widehat{U}^{(t+1)} \leftarrow \underset{U \in \mathbb{R}^{m \times k}}{\text{argmin}} \|P_{\Omega^{(T+t+1)}}\left(U(\widehat{V}^{(t+1)})^\dagger - \widetilde{M}\right)\|_F.$$

end

Main Idea of the Proof

- The initialization step of the Algorithm described provides a good starting point.
- The space spanned by AltMin estimates of $\widehat{U}$ and $\widehat{V}$ converge towards U^* and V^* respectively.
- Combine the above two results to prove the main result
- Also use the following theorem by [3]:

Theorem

If N is a matrix from the worst case model, then for any realization of N, $\|P_\Omega(N)\|_2 \leq \frac{2|\Omega|}{m\sqrt{\alpha}} N_{max}$.

Main Result

Theorem

With high probability, after $O(\log \frac{\|M\|_F}{\epsilon})$ iterations of AltMin, the outputs $\widehat{U}$ and $\widehat{V}$ satisfy,

$$\frac{1}{\sqrt{mn}}\|M - \widehat{U}\widehat{V}^{\dagger}\|_F \leq \epsilon + 40\mu\kappa^2 k^{1.5} N_{max},$$

under the following conditions:

- $M \in \mathbb{R}^{m \times n}$ *is rank* $k \ll \{m, n\}$ *and* μ*–incoherent,*
- $N_{max} \leq C\kappa^{-2}\, k^{-1.5} \frac{\|M\|_F}{\sqrt{mn}}$ *(κ being the condition number of M),*
- *Each entry of* $\widetilde{M} = M + N$ *is observed uniformly and independently with probability*

$$p > C(\mu)\kappa^6 k^7 m^{-1} \log n \log \frac{\|M\|_F}{\epsilon}.$$

Comparison of Similar Results I: Noise Free Case

Common assumptions:

- Low rank M
- μ–incoherence of M

Algorithm	Number of Iterations	Sample Complexity
AltMin [1]	$O\left(\log \frac{\|M\|_F}{\epsilon}\right)$	$O\left(\kappa^6 k^7 n \log n \frac{\log \|M\|_F}{\epsilon}\right)$
OptSpace [5]	Asymptotic	$O(\kappa^2 kn \log n)$
Nuclear Norm Minimization [4]	$O\left(\frac{1}{\sqrt{\epsilon}}\right)$	$O(kn \log n)$

Table : Noiseless Matrix Completion

Comparison of Similar Results II: Noisy Recovery

Algorithm	Additional Assumptions	$\frac{1}{\sqrt{mn}}\|M - \widehat{M}\|_F$
AltMin	$\frac{\|P_\Omega(N)\|_2}{p} \leq C\frac{\sigma_k^*}{\kappa k}$, $N_{max} \leq \frac{C'\|M\|_F}{\kappa k\sqrt{mn}}$	$40\kappa^2 k^{1.5}\frac{\|P_\Omega(N)\|_2}{\|\Omega\|}$
OptSpace	$\frac{\|P_\Omega(N)\|_2}{p} \leq C\frac{\sigma_k^*}{\kappa^2\sqrt{k}}$	$C\kappa^2 k^{0.5}\frac{\|P_\Omega(N)\|_2}{\|\Omega\|}$
Nuclear Norm Minimization	Strong Incoherence	$C\frac{\sqrt{pn}\|P_\Omega(N)\|_F}{\|\Omega\|} + 2\frac{\|P_\Omega(N)\|_F}{\sqrt{mn}}$

Table : Noisy Matrix Completion

- σ_k^* : k^{th} largest singular value of M
- The requirements on the noise matrix for recovery guarantees by OptSpace is similar.
- The error in the recovered matrix in our analysis is off by a small factor of k as compared to the analysis in [3].
- The sample complexity required by ALSM is much higher than that in [3].

Modification in ALSM for Analysis

- Each step of AltMin is followed by a QR decomposition. The distance between subspaces spanned by the matrices before and after QR decomposition does not change.

$$\widehat{V}^{(t+1)} \leftarrow \underset{\widehat{V}\in\mathbb{R}^{n\times k}}{\operatorname{argmin}} \|P_{\Omega^{(t+1)}}(U^t \widehat{V}^\dagger - \widetilde{M})\|_F$$

$$V^{(t+1)} R_V^{(t+1)} = \widehat{V}^{(t+1)} \quad \text{(QR decomposition)}$$

$$\widehat{U}^{(t+1)} \leftarrow \underset{\widehat{U}\in\mathbb{R}^{m\times k}}{\operatorname{argmin}} \|P_{\Omega^{(T+t+1)}}\left(\widehat{U} V^{(t+1)^\dagger} - \widetilde{M}\right)\|_F$$

$$U^{(t+1)} R_U^{(t+1)} = \widehat{U}^{(t+1)} \quad \text{(QR decomposition)}$$

Future Work

- Sample complexity for well conditioned matrices is $O(k^7 n \log n)$ which can further be improved.
- Analysis with regularized ALSM algorithm – cost function modified to include regularization on the factors U and V

Questions?

References:

1. Low-rank Matrix Completion using Alternating Minimization, Jain *et al.* [Link].
2. Matrix Completion With Noise, Candes & Plan [Link].
3. Matrix Completion from Noisy Entries, Keshavan *et al.* [Link].
4. Exact Matrix Completion via Convex Optimization, Candes & Recht [Link].
5. Matrix Completion from a Few Entries, Keshavan *et al.* [Link].