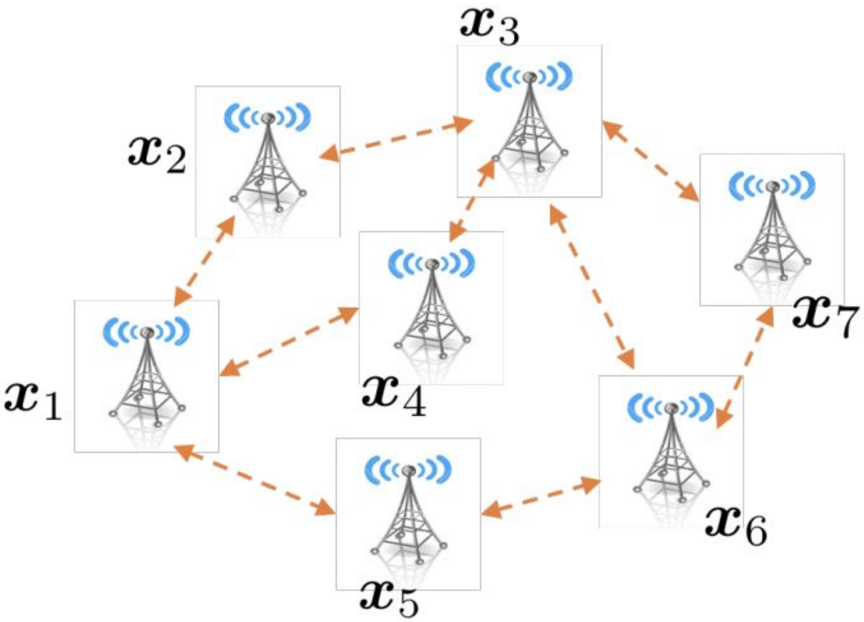
Matrix Completion

Yunjuan Wang: Background, theory, algorithm
Alex Bai: algorithm
Nilay Thakor: Experiments

Background



Convex Relaxation

The problem we are considering is to recover matrix M. The goal is to find the lowest rank matrix X which matches the observed entries Ω from M.

Define $P_\Omega(X) = \begin{cases} X_{i,j}(i,j) \in \Omega \\ 0(i,j) \notin \Omega \end{cases}$, which is the projection operator on the observed index set Ω.

$$\min_X \ \text{rank}(X) \quad s.t. X_{i,j} = M_{i,j} \forall (i,j) \in \Omega$$

$$\Rightarrow \quad \min_X \|X\|_* \quad s.t. P_\Omega(X) = P_\Omega(M)$$

$$\Rightarrow \quad \min_X \max_Y \langle X, Y \rangle \quad s.t. \begin{bmatrix} I & Y \\ Y^\top & I \end{bmatrix} \succeq 0, \quad P_\Omega(X) = P_\Omega(M)$$

Define $X = U\Sigma V^\top$.

$$W_1 = U\Sigma U^\top$$

$$W_2 = V\Sigma V^\top$$

$$\min_{W_1, W_2} \frac{1}{2}(Tr(W_1) + Tr(W_2)) \quad s.t. \begin{bmatrix} W_1 & X \\ X^\top & W_2 \end{bmatrix} \succeq 0, \quad P_\Omega(X) = P_\Omega(M)$$

Basic Theory

Theorem 2.1. Vershynin [2018] Consider fixed $n \times n$ matrix X with $rank(X) = r$, where $r \ll n$. Each entry X_{ij} is revealed to us independently with probability $p \in (0, 1)$. We only observe matrix Y, $Y_{ij} = \delta_{ij} X_{ij}$, $\delta_{i,j} \sim Ber(p)$. Choose $p = \frac{m}{n^2}$. Let $\hat{X}$ be a best rank r approximation to $p^{-1}Y$. Then

$$\mathbb{E}\frac{1}{n}\|\hat{X} - X\|_F \leq C\sqrt{\frac{rn\log n}{m}}\|X\|_\infty$$

Theorem 2.3 (Candès and Recht [2009]). Let M be a $n_1 \times n_2$ matrix of rank r satisfying assumption 2 and 3, choose $n = \max(n_1, n_2)$. Suppose we observe m entries of M with locations sampled uniformly at random. Then there exist constants C, c, such that if

$$m \geq C \max(\mu_1^2, \mu_0^{0.5}\mu_1, \mu_0 n^{0.25}) nr(\beta \log n)$$

for some $\beta > 2$, then the optimal solution to the norm minimization problem is unique and equal to M with probability at least $1 - cn^{-\beta}$. Furthrermore, for low rank case $r \leq \mu_0^{-1} n^{0.2}$, with same probability provided that

$$m \geq C\mu_0 n^{1.2} r(\beta \log n).$$

Singular Value Thresholding

Given matrix M, only observe $M_{i,j}, (i,j) \in \Omega$

In order to recover matrix M, we need to solve

$$\min_{X \in \mathbb{R}^{m \times n}} F(X) = \frac{1}{2}\|P_\Omega(M) - P_\Omega(X)\|_F^2 + \lambda\|X\|_*$$

Introduce soft-thresholding operator,

$S_\lambda(X) = U\Sigma_\lambda V^\top$, where $X = U\Sigma V^\top$ is an SVD

$(\Sigma_\lambda)_{ii} = \max\{\Sigma_{ii} - \lambda, 0\}$

Thm: Suppose the sequence of step sizes obeys

$0 < \inf \delta_t \leq \sup \delta_t \leq 2$, then the {Xk} converges to

the unique solution of

$$\min_X \lambda\|X\|_* + \frac{1}{2}\|X\|_F^2$$
$$s.t. P_\Omega(X) = P_\Omega(M)$$

Algorithm 1 SVT Algorithm

Input: observed entries $P_\Omega(M)$,
a sequence of positive step sizes $\{\delta_k\}_{k\geq 0}$,
a regularization parameter λ,
termination criteria
Initialize: $Y_0 = 0_{n_1 \times n_2}$
while *termination criteria is not reached* **do**
  $[U_k, \Sigma_k, V_k] = \text{svd}(Y_{k-1})$
  $X_{k+1} = U_k \text{diag}(\{\sigma_i(\Sigma_k) - \lambda)_+\}_i)V_k^\top$
  $Y_{k+1} = Y_k + \delta_k(P_\Omega(M) - P_\Omega(X_{k+1}))$
  $k = k + 1$
end
Output: X_k

Alternating Least Squares Minimization

Assumptions:

1. M is of rank k << min{m,n}

   $X = UV^T$ where $U \in \mathbb{R}^{mxk}$, $V \in \mathbb{R}^{nxk}$

2. M is μ-incoherent

Definition (μ-incoherence). A matrix $M \in \mathbb{R}^{m\times n}$ is *incoherent with parameter* μ if:

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

where $M = U\Sigma V^\dagger$ is the SVD of M and $u^{(i)}, v^{(j)}$ denote the i^{th} row of U and j^{th} row of V respectively.

$$\min_{U\in\mathbb{R}^{m\times k}, V\in\mathbb{R}^{n\times k}} ||P_\Omega(UV^\dagger) - P_\Omega(M)||_F^2$$

Alternating Least Squares Minimization

Algorithm 3 AltMinComplete

Input: observed set Ω, values $P_\Omega(M)$

Partition Ω into $2T+1$ subsets $\Omega_0, \ldots, \Omega_{2T}$ with each element of Ω belonging to one of the Ω_t with equal probability (sampling with replacement)

Initialize: $\hat{U}^0 = SVD(\frac{1}{p}P_{\Omega_0}(M), k)$ i.e., top-k left singular vectors of $\frac{1}{p}P_{\Omega_0}(M)$

Clipping step: Set all elements of $\hat{U}^0$ that have magnitude greater than $\frac{2\mu\sqrt{k}}{\sqrt{n}}$ to zero and orthonormalize the columns of $\hat{U}^0$

for $t = 0, \ldots, T\text{-}1$ **do**

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

  $\hat{U}^{t+1} \longleftarrow \mathrm{argmin}_{U \in \mathbb{R}^{m\times k}} ||P_{\Omega_{T+t+1}}(U(\hat{V}^{t+1})^\dagger - M)||_F^2$

end

Return $X = \hat{U}^T(\hat{V}^T)^\dagger$

Alternating Least Squares Minimization

Theorem 3.3. Let $M = U^* \Sigma^* V^{*\dagger} \in \mathbb{R}^{m \times n}$ $(n \geq m)$ be a rank-k incoherent matrix, i.e., both U^* and V^* are μ-incoherent. Also, let each entry of M be observed uniformly and independently with probability,

$$p > C \frac{(\frac{\sigma_1^*}{\sigma_k^*})^2 \mu^2 k^{2.5} \log n \log \frac{k||M||_F}{\epsilon}}{m \delta_{2k}^2},$$

where $\delta_{2k} \leq \frac{\sigma_k^*}{12k\sigma_1^*}$ and $C > 0$ is a global constant. Then with high probability for $T = C' \log \frac{||M||_F}{\epsilon}$, the outputs $\hat{U}^T$ and V^T of AltMinComplete, with input $(\Omega, P_\Omega(M))$ satisfy: $||M - \hat{U}^T (V^T)^\dagger||_F \leq \epsilon$

- Requires $O((\frac{\sigma_1^*}{\sigma_k^*})^4 \mu^2 k^{4.5} n \log n \log \frac{k||M||_F}{\epsilon})$ matrix entries to be revealed
- Can recover M in $O(\log \frac{1}{\epsilon})$ steps

Numerical Experiments

Simulation Process

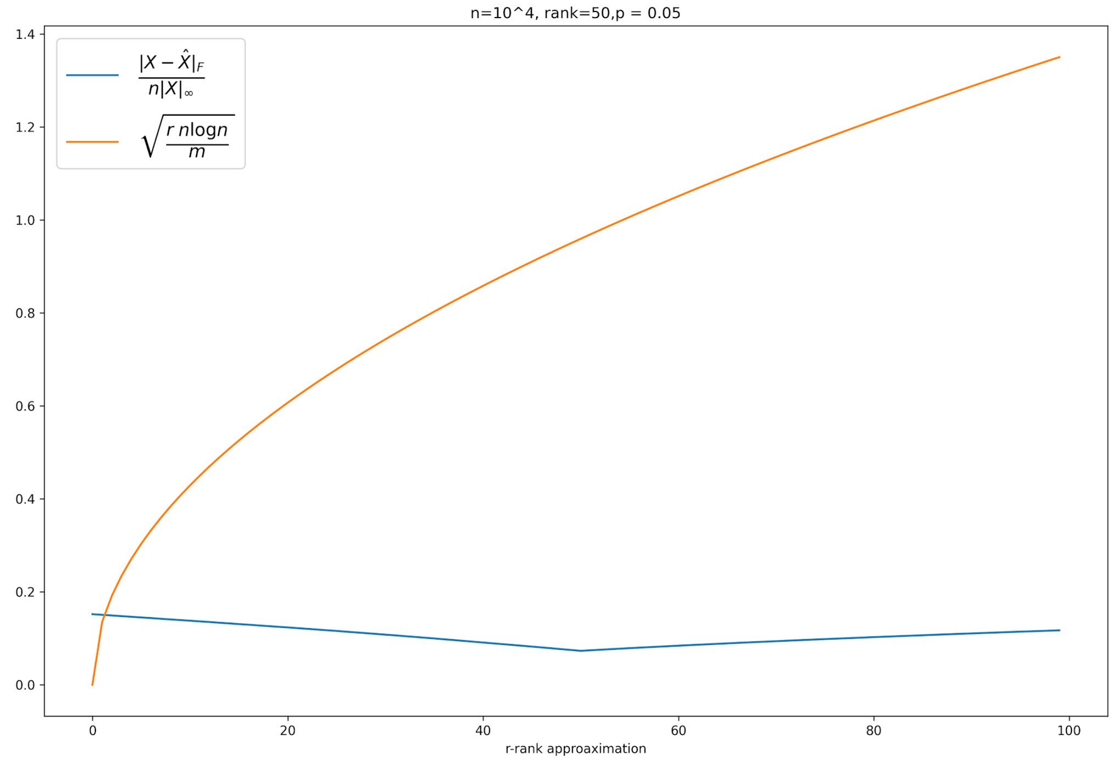
- K - rank matrix :
  - Sample $U \sim R^{(m,k)}$, $U_{ij} \sim$ Normal(0,1)
  - Sample $V \sim R$^$^{(k,n)}$, $V_{ij} \sim$ Normal(0,1)
  - Return L*R
- $P_{\Omega}(p) = \triangle^{(m,n)}$, $\triangle_{ij}$ = 1 with probability p
- Reconstruction Error = $||P_{\Omega}(X-X_k)||_F$

Bound on k-rank approximation

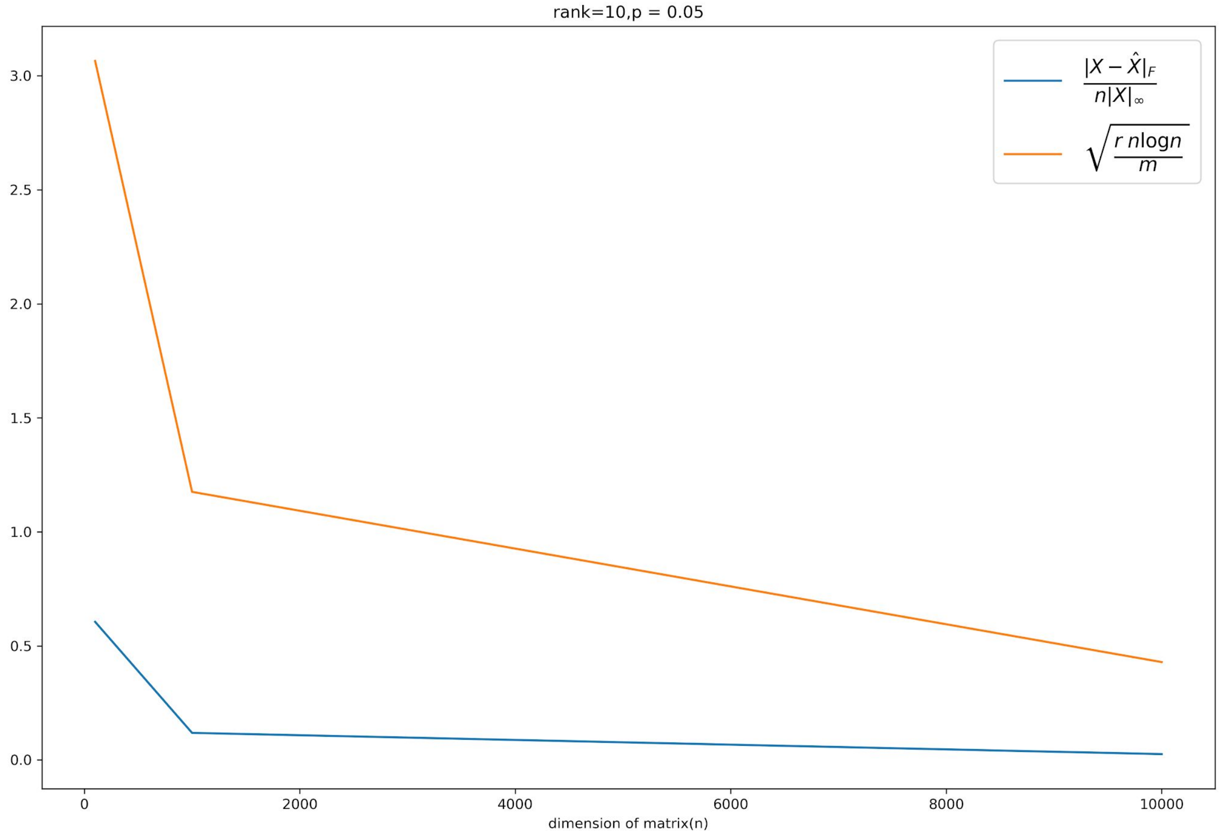
- $U,S,V = svd(p^{-1}Y)$
- Approximate X by $X_k = \sum s_i u_i v_i^T$ (sum upto k terms)

$$\mathbb{E}\,\frac{1}{n}\|\hat{X} - X\|_F \le C\sqrt{\frac{rn\log n}{m}}\,\|X\|_\infty,$$

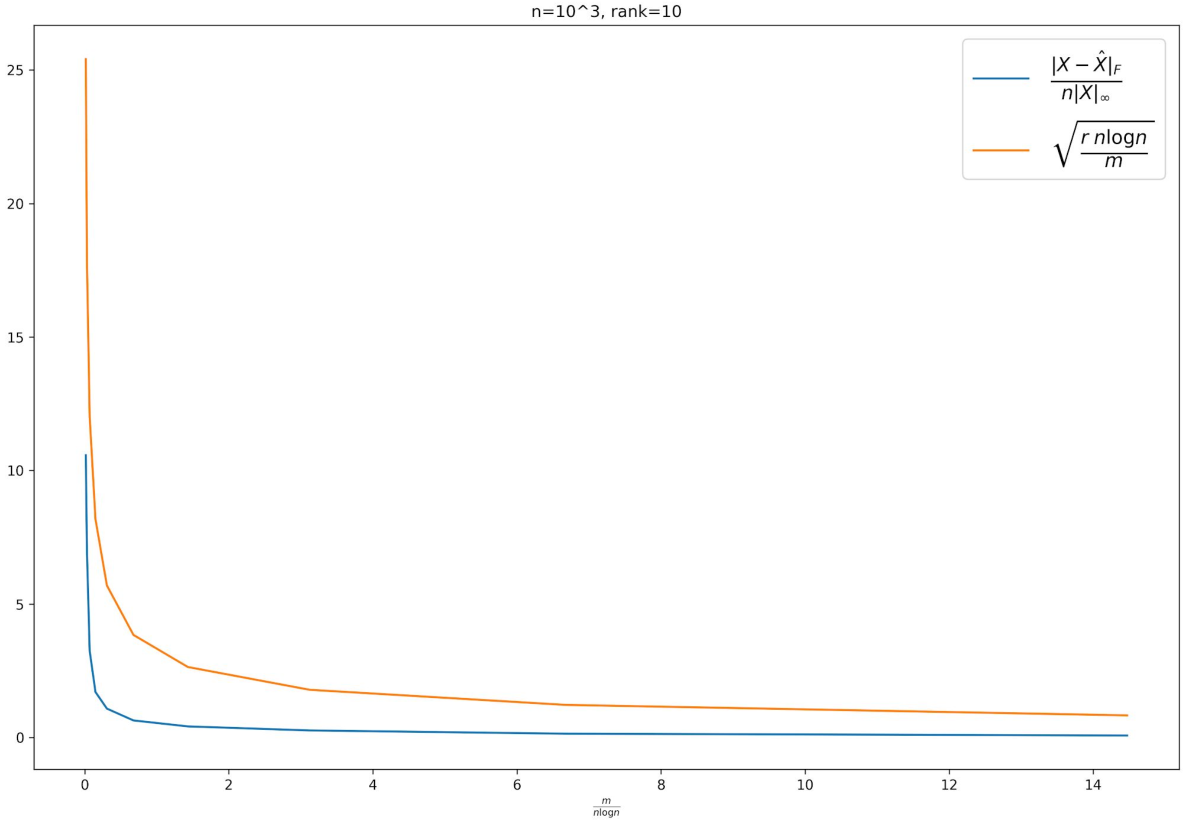
Reconstruction Error for r-rank approaximation



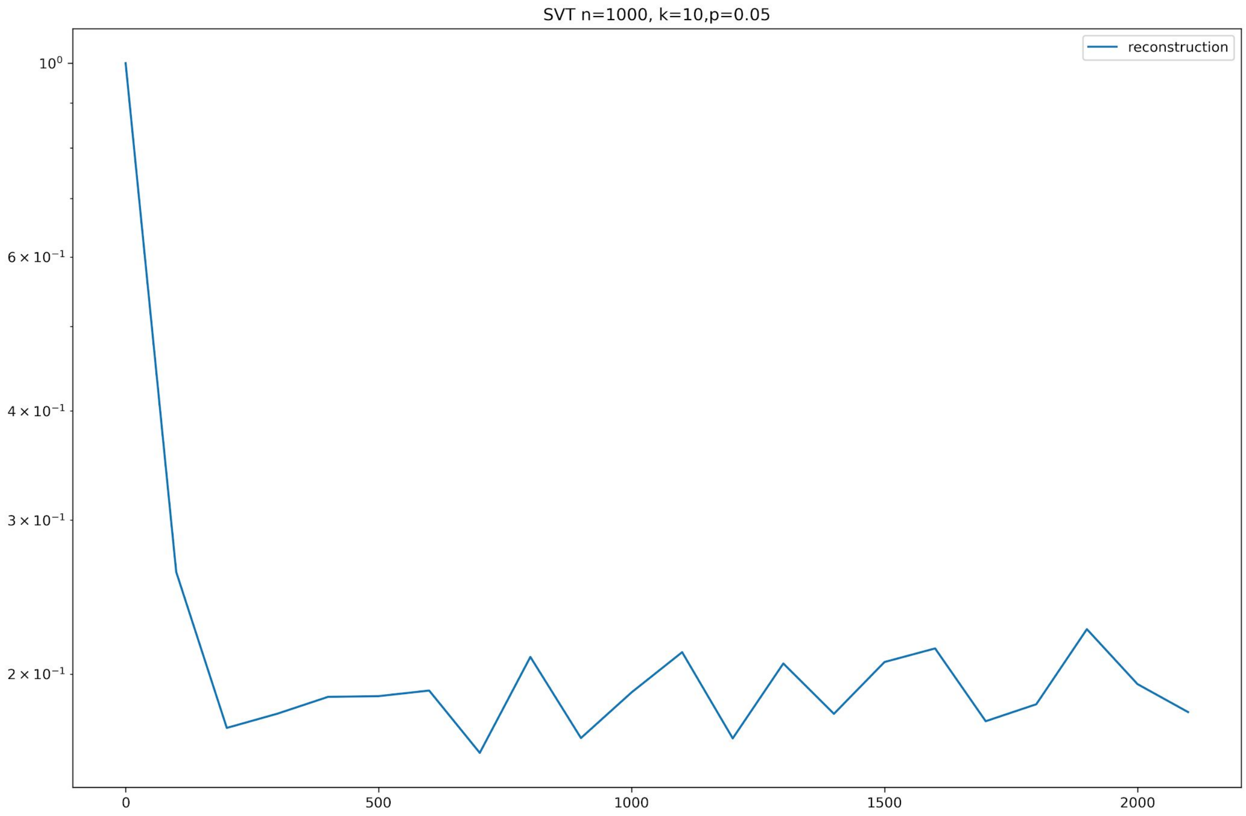
Reconstruction Error vs dim(X)



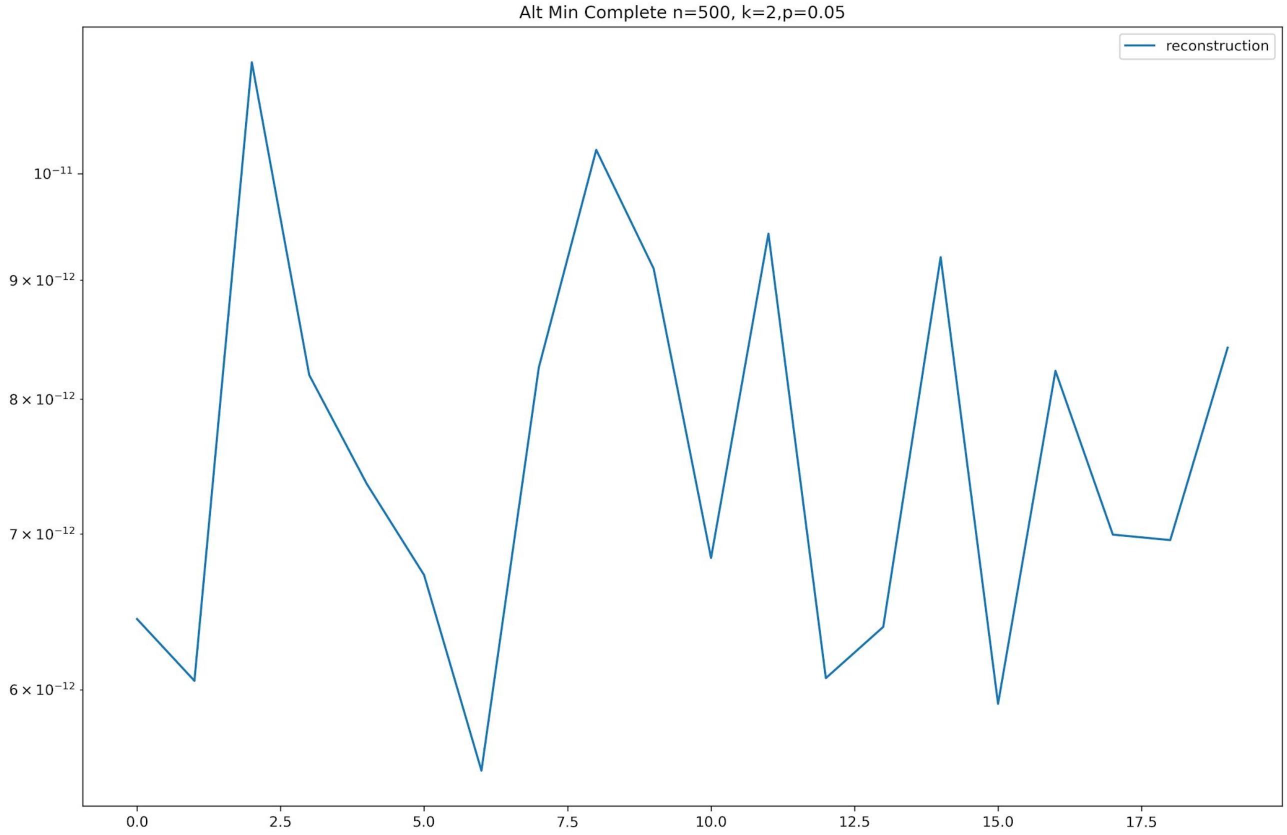
Reconstruction error vs sparsity



SVT Convergence



Alt Min Complete

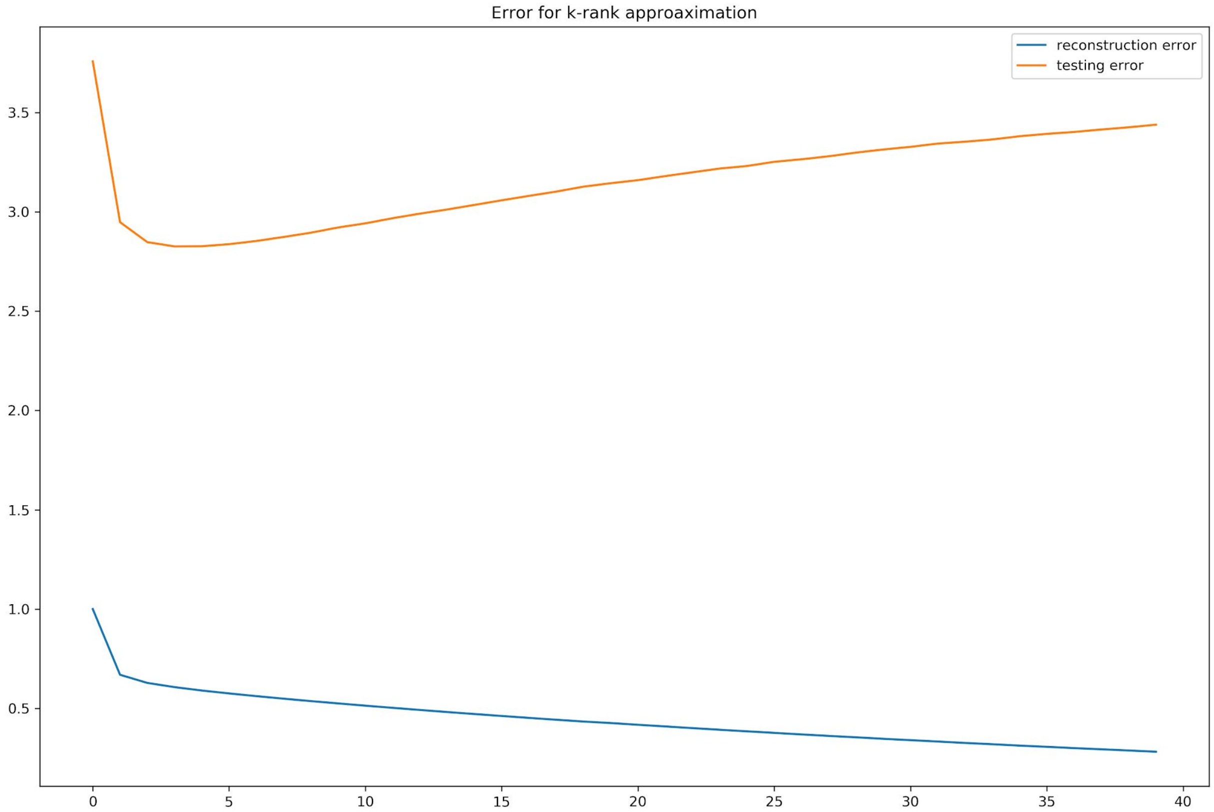


Real World Application

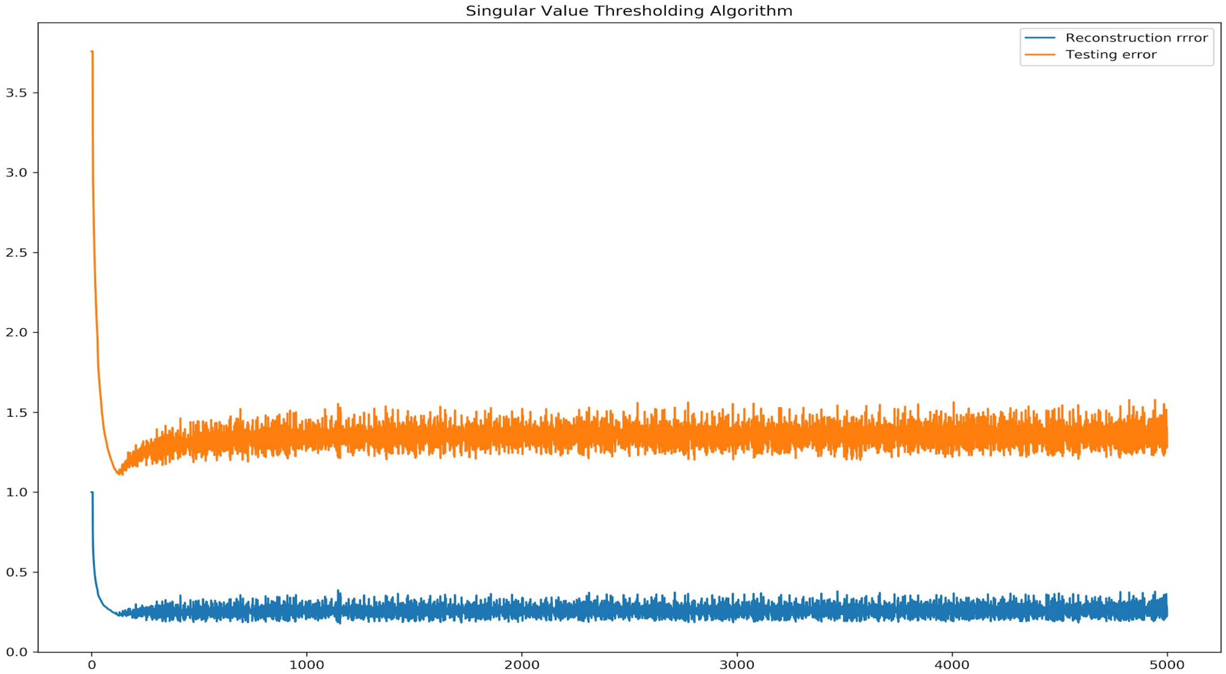
Dataset

- MovieLens (100k)
  - 943 Users
  - 1682 Movies
  - 100000 total ratings
- Test dataset generated by removing ratings of 10 users from original dataset
- Train Error = reconstruction error of train matrix
- Test Error = MSE of predicted ratings

K-rank approximation



SVT



Singular Value Thresholding Algorithm
Reconstruction rrror
Testing error
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
1000
2000
3000
4000
5000

Thanks