Oregon State University

Exact Linear Convergence Rate Analysis for Low-Rank Symmetric Matrix Completion via Gradient Descent

Trung Vu and Raviv Raich
School of EECS, Oregon State University, Corvallis, OR 97331-5501, USA
{vutru,raich}@eecs.oregonstate.edu

2021 TORONTO Canada June 6-11, 2021 Metro Toronto Convention Centre

1. Symmetric Matrix Completion Problem

1	?	?	1
?	?	6	?
?	6	?	2
1	?	2	1

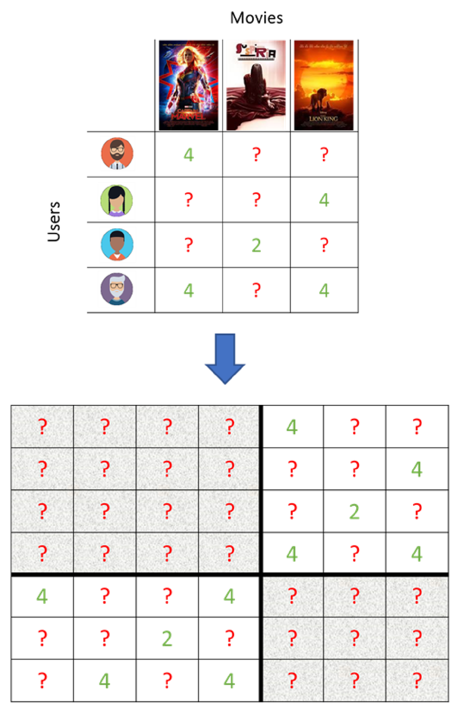
$n = 4, r = 1, s = 8$

- Positive semidefinite (PSD) rank-r matrix $M \in \mathbb{R}^{n\times n}$
  - $M = XX^T$, where $X \in \mathbb{R}^{n\times r}$
- Sampling set Ω with cardinality s
  - Ω is symmetric!

find	M_{ij} for $(i,j) \in \Omega^c$
given	$rank(M) \leq r$
and	M_{ij} for $(i,j) \in \Omega$

2. Motivation

- ❑ Maximum likelihood estimation of covariance matrices in Gaussian graphical models
- ❑ Density matrix completion in quantum state tomography
- ❑ Low-rank approximation of correlation matrices in finance and risk management
- ❑ Solving non-symmetric case using symmetric case via *semidefinite lifting*



$n_1 = 4, n_2 = 3$
$r = 1, \tilde{s} = 5$

$n = 7$
$r = 1, s = 10$

3. Existing Approaches

Approach	Problem formulation	Existing algorithms
Linearly-constrained nuclear norm minimization	$\min_{Z\in\mathbb{R}^{n\times n}} \|Z\|_* \text{ s.t. } \sum_{(i,j)\in\Omega} \|Z_{ij} - M_{ij}\| = 0$	Iterative soft thresholding, i.e., SVT [4], APG [5], CGD [6]
Rank-constrained least squares	$\min_{Z\in\mathbb{R}^{n\times n}} \sum_{(i,j)\in\Omega} (Z_{ij} - M_{ij})^2 \text{ s.t. } rank(Z) \leq r$	Iterative hard thresholding, i.e., SVP [7], NIHT [8], Accelerated IHT [9,10]
Low-rank factorization	$\min_{X\in\mathbb{R}^{n\times r}} \sum_{(i,j)\in\Omega} \left((XX^T)_{ij} - M_{ij}\right)^2$	Gradient descent [11,12], projected gradient descent [13]

4. Gradient Descent (GD)

- SMCP as *unconstrained* *non-convex* optimization

$$\underbrace{\min_{X\in\mathbb{R}^{n\times r}} \frac{1}{4}\|P_\Omega(XX^T - M)\|_F^2}_{f(X)} \qquad [\mathcal{P}_\Omega(Z)]_{ij} = \begin{cases} Z_{ij} & \text{if } (i,j)\in\Omega, \\ 0 & \text{otherwise.}\end{cases}$$

Algorithm 1 (Non-convex) Gradient Descent
Input: $X^0, \mathcal{P}_\Omega(M), \eta$
Output: $\{X^k\}$
1: **for** $k = 0, 1, 2, \ldots$ **do**
2: $\quad X^{k+1} = X^k - \eta \underbrace{\mathcal{P}_\Omega(X^k X^{k^T} - M)X^k}_{\nabla f(X^k)}$

Loose *global* convergence analysis!

5. Local Convergence Analysis - Preliminaries

- Rank-r eigendecomposition of M
  $M = U\Sigma U^T$ → • $U \in \mathbb{R}^{n\times r}$ is a semi-orthogonal matrix • $\Sigma \in \mathbb{R}^{r\times r}$ is a diagonal matrix
- Projection onto the null space of M
  $P_{U_\perp} = I_n - UU^T$
  $$P_r(M+E) - M = \underbrace{E - P_{U_\perp} E P_{U_\perp}}_{\nabla P_r(M)\cdot E} + O(\|E\|_F^2)$$
- Vectorization of the projection onto the tangent plane of the set of rank-r matrices at M
  $P_1 = I_{n^2} - P_{U_\perp} \otimes P_{U_\perp}$ → $\text{vec}(E - P_{U_\perp} E P_{U_\perp}) = P_1 \text{vec}(E)$
- Vectorization of the projection onto the set of symmetric matrices
  $P_2 = \frac{I_{n^2} + T_{n^2}}{2}$ → $\text{vec}\left(\frac{E + E^T}{2}\right) = P_2 \text{vec}(E)$

6. A Recursion on the Error

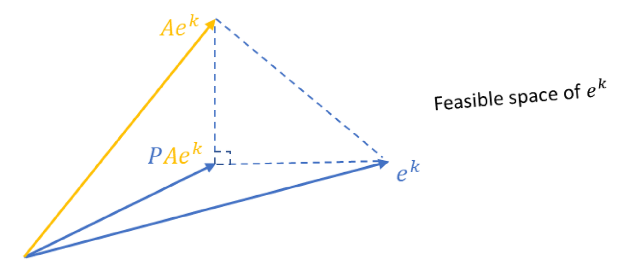
- Recall the GD update
  $$X^{k+1} = X^k - \eta P_\Omega\left(X^k X^{k^T} - M\right) X^k$$
  $M\oplus M = M\otimes I_n + I_n \otimes M$
  $S \in \mathbb{R}^{s\times n^2}: \begin{cases} SS^T = I_s \\ vec(P_\Omega(E)) = S^T S vec(E)\end{cases}$
- Let $E^k = X^k X^{k^T} - M$ be the error matrix
  $$\Rightarrow E^{k+1} = E^k - \eta\left(P_\Omega(E^k)M + MP_\Omega(E^k)\right) + O\left(\|E^k\|_F^2\right)$$
- Let $e^k = \text{vec}(E^k)$ be the error vector
  $$\Rightarrow e^{k+1} = \underbrace{(I_{n^2} - \eta(M\oplus M)S^T S)}_{A} e^k + O\left(\|e^k\|_2^2\right)$$
  $Av = v$ for all v s.t. $Sv = 0$
  $\Rightarrow \rho(A) \geq 1$!

7. Structural Constraints on the Error

Structural properties of the error matrix

$E = XX^T - M$

1. $E = E^T$
2. $P_r(M + E) = M + E$,
   - where P_r is the projection onto the set of rank-r matrices
3. $M + E$ is PSD



Structural properties of the error vector

$e = \text{vec}(E) = \text{vec}(XX^T - M)$

1. $e = P_2 e$
2. $e = P_1 e + O(\|e\|_2^2)$
3. Negligible effect

$$\Rightarrow e = \underbrace{P_1 P_2}_{P} e + O\left(\|e\|_2^2\right)$$

$P_1 = I_{n^2} - P_{U_\perp} \otimes P_{U_\perp}$
$P_2 = \frac{I_{n^2} + T_{n^2}}{2}$

8. Linear Rate of Convergence

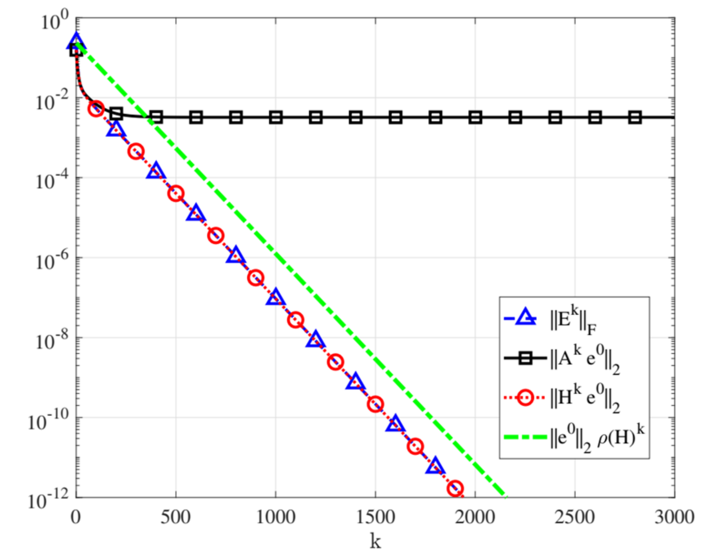
- Integrating structural constraints

$$e^{k+1} = \underbrace{PAP}_{H} e^k + O\left(\|e^k\|_2^2\right)$$

$P = P_1 P_2 = (I_{n^2} - P_{U_\perp} \otimes P_{U_\perp}) \frac{I_{n^2} + T_{n^2}}{2}$
$H = P(I_{n^2} - \eta(M\oplus M)S^T S)P$

Theorem

If $\rho(H) < 1$, then there exists a neighborhood $\mathcal{N}(M)$ such that for any $X^0 X^{0^T} \in \mathcal{N}(M)$:
$$\|e^k\| \leq C\|e^0\|\rho(H)^k$$
for some numerical constant $C > 0$.



9. Conclusions

- ❑ **Local** convergence analysis recovers the **exact** rate of linear convergence of GD for matrix completion.
- ❑ Integrating **structural constraints** is the key to obtain the convergence rate for constrained nonlinear difference equations.
- ❑ It is interesting to extend the analysis to the **non-symmetric** case and make connection to existing works on the **global** convergence.

Check our full paper at https://arxiv.org/abs/2102.02396
Further results on local convergence analysis at https://trungvietvu.github.io/