IEEE International Conf. on Acoustics, Speech and Signal Processing

California Institute of Technology, (Caltech), California, USA

ICASSP'18 - Calgary, Alberta, Canada, 15-20 Apr. 2018

An Improved Initialization for Low-Rank Matrix Completion Based on Rank-1 Updates

Ahmed Douik and Babak Hassibi
{ahmed.douik,hassibi}@caltech.edu

Presented by Christos Thrampoulidis, Massachusetts Institute of Technology.

April 19, 2018

Why matrix completion? Global positioning

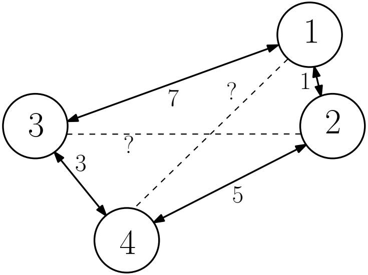

Figure: Graph with partially observable distance.

Why matrix completion? Global positioning

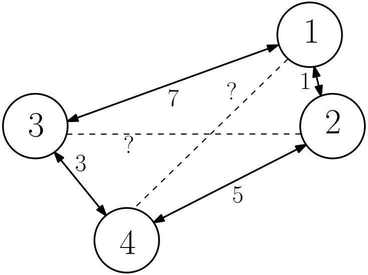

Figure: Graph with partially observable distance.

- Find X such that

$$\begin{bmatrix} 0 & 1 & 7 & X \\ 1 & 0 & X & 5 \\ 7 & X & 0 & 3 \\ X & 5 & 3 & 0 \end{bmatrix}$$

has low-rank.

Why matrix completion? Netflix problem

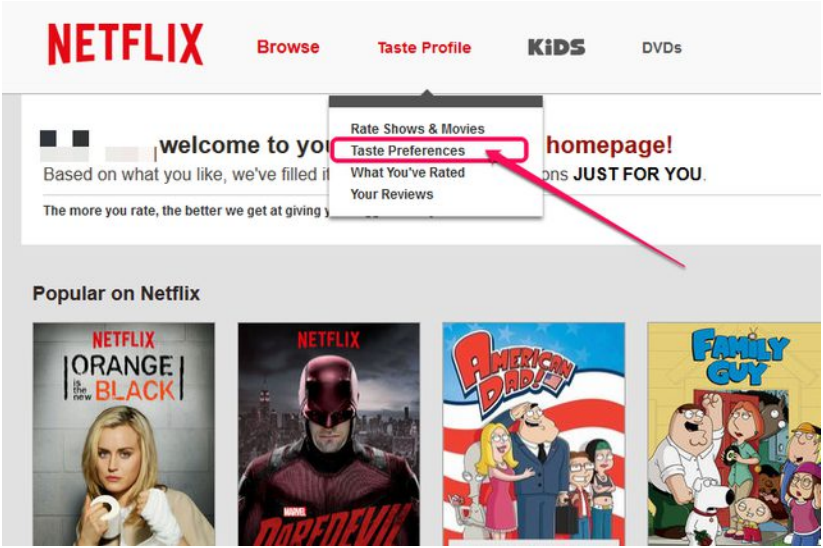

Figure: Netflix recommendation system.

Mathematical formulation

- Let $\mathbf{A}$ be the partially observable matrix and Ω be the set of observable indices.
- The matrix completion problem can be written as:

$$\begin{aligned} \min_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} \quad & \text{Rank}(\mathbf{X}) \\ \text{s.t.} \quad & \mathbf{X}_{ij} = \mathbf{A}_{ij}, \ \forall \ (i,j) \in \Omega. \end{aligned}$$

[1] Rong Ge, Jason D. Lee, and Tengyu Ma “Matrix Completion has No Spurious Local Minimum" in Neural Information Processing Systems (NIPS), 2016.

Mathematical formulation

- Let $\mathbf{A}$ be the partially observable matrix and Ω be the set of observable indices.
- The matrix completion problem can be written as:

$$\begin{aligned} \min_{\mathbf{X}\in\mathbb{R}^{n_1\times n_2}} \quad & \text{Rank}(\mathbf{X}) \\ \text{s.t.} \quad & \mathbf{X}_{ij} = \mathbf{A}_{ij},\ \forall\ (i,j)\in\Omega. \end{aligned}$$

- Unless under specific scenarios[1], the matrix completion problem is non-convex with the presence of local minima.

[1] Rong Ge, Jason D. Lee, and Tengyu Ma "Matrix Completion has No Spurious Local Minimum" in Neural Information Processing Systems (NIPS), 2016.

Assumptions

In the rest of this talk, we assume the following

- The partially observed matrix $\mathbf{A}$ is generated from the multiplication of two i.i.d. Gaussian matrices with zero mean and unit variance, i.e., $\mathbf{A} = \mathbf{U}\mathbf{V}^T$ with $\mathbf{U} \in \mathbb{R}^{n_1 \times r}$ and $\mathbf{V} \in \mathbb{R}^{n_2 \times r}$ i.i.d. $\mathcal{N}(0,1)$.
- The set Ω is sampled according to a Bernoulli model. In other words, each entry (i,j) with $1 \leq i \leq n_1$ and $1 \leq j \leq n_2$ is included in the set Ω with probability p.
- The completion rank r is known apriori.

Related Work and Results

	Convex Relaxation	Non-convex Approach
Objective	$\min_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} \|\|\mathbf{X}\|\|_*^2$	$\min_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} \|\|\mathbf{A} - \mathbf{X}\|\|_{\mathbf{\Omega}}^2$
Constraint	$\mathbf{X} \odot \mathbf{\Omega} = \mathbf{A} \odot \mathbf{\Omega}$	$\text{Rank}(\mathbf{X}) = r$
Dimension	$n_1 n_2$	$(n_1 + n_2)r$
Algorithm(s)	SDP[2]	Riemannian optimization[3] Alternate projection[4]
Guarantees		

[2] Candès, E. J. and Terence, T. “The Power of Convex Relaxation: Near-optimal Matrix Completion" in IEEE Transactions on Information Theory, 2010.

[3] Bart, V. “Low-rank matrix completion by Riemannian optimization" in SIAM Journal on Optimization, 2013.

[4] Prateek, J. and Praneeth, N. and Sujay, S. “Low-rank matrix completion using alternating minimization" in proc. of ACM symposium on Theory of computing, 2013

Is the convex approximation good?

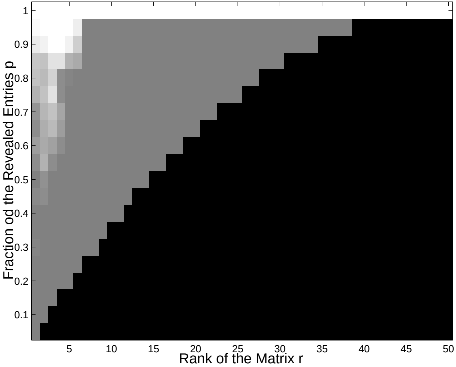

Figure: Performance of the nuclear norm relaxation.

Is the convex approximation good?

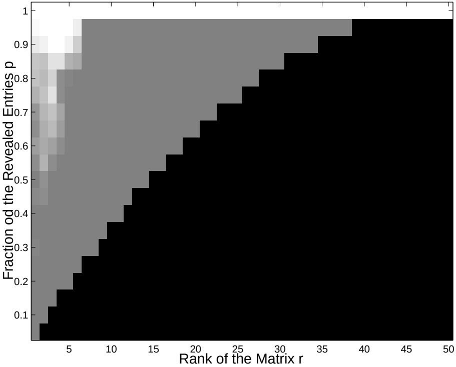

Figure: Performance of the nuclear norm relaxation.

Drawbacks:

- Only works for low-rank,
- Very slow because large SDP.

Performance of the Riemannian approach

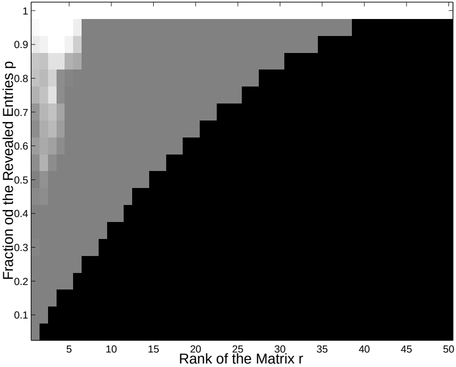

Figure: Performance of the nuclear norm relaxation.

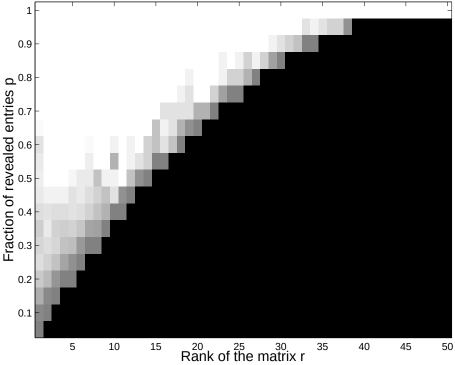

Figure: Riemannian approach with arbitrary initialization.

Pros and cons of the Riemannian approach

Advantages:

- Large convergence region,
- Very fast as compared to solving SDPs.

Drawbacks:

- No convergence theoretical guarantees,
- Performance sensitive to initialization.

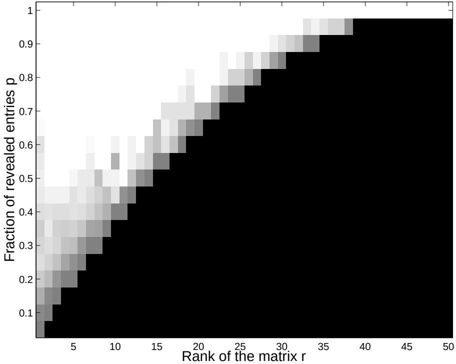

Figure: Riemannian approach with arbitrary initialization.

Matrix completion as norm minimization

- The matrix completion problem can be reformulated as:

$$(\mathbf{P}_r) \min_{\mathbf{X}\in\mathbb{R}^{n_1\times n_2}} ||\mathbf{A}-\mathbf{X}||^2_\Omega = \sum_{(i,j)\in\Omega} (\mathbf{A}_{ij}-\mathbf{X}_{ij})^2$$

$$\text{s.t.} \quad \text{Rank}(\mathbf{X}) = r.$$

Matrix completion as norm minimization

- The matrix completion problem can be reformulated as:

$$(\mathbf{P}_r) \min_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} ||\mathbf{A} - \mathbf{X}||^2_{\Omega} = \sum_{(i,j) \in \Omega} (\mathbf{A}_{ij} - \mathbf{X}_{ij})^2$$

$$\text{s.t.} \quad \text{Rank}(\mathbf{X}) = r.$$

- However, as expected the problem still hard and requires a good initialization.

Successive rank one update

- Starting with $\mathbf{X}_0^* = \mathbf{0}$, we propose a successive rank one update initialization as follows

$$\begin{aligned} \mathbf{X}_{n+1}^* = \text{arg} \min_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} & ||\mathbf{A} - \mathbf{X}||_\Omega^2 \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{X}_n + \mathbf{x}\mathbf{y}^T \\ & \mathbf{x} \in \mathbb{R}^{n_1}, \ \mathbf{y} \in \mathbb{R}^{n_2} \end{aligned}$$

- The above problem is solved using the Riemannian[5] method on the low-rank manifold.
- The algorithm is executed r times to produce a rank r initialization.

[5] Bart, V. "Low-rank matrix completion by Riemannian optimization" in SIAM Journal on Optimization, 2013.

Are we close to the optimal solution?

- The performance of the initialization is difficult to characterize.
- However, a sufficient condition is given extending the SVD.
- The extended SVD (E-SVD), like the SVD, can be computed efficiently (n^4 operations as compared to n^3 for the SVD).

Extended Singular Value Decomposition

- Recall that the SVD of $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ such that $\Sigma =$diag$(\sigma_1 \geq \cdots \geq \sigma_n \geq 0)$, $\mathbf{U} = [\mathbf{u}_1, \ \cdots \ , \ \mathbf{u}_n] \in \mathbb{R}^{n_1 \times n}$ and $\mathbf{V} = [\mathbf{v}_1, \ \cdots \ , \ \mathbf{v}_n] \in \mathbb{R}^{n_2 \times n}$, $n = \min(n_1, n_2)$ satisfying
  1. $\mathbf{U}^T\mathbf{U} = \mathbf{I}_n$.
  2. $\mathbf{V}^T\mathbf{V} = \mathbf{I}_n$.

Extended Singular Value Decomposition

- Recall that the SVD of $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ such that $\Sigma =$diag$(\sigma_1 \geq \cdots \geq \sigma_n \geq 0)$, $\mathbf{U} = [\mathbf{u}_1, \ \cdots \ , \ \mathbf{u}_n] \in \mathbb{R}^{n_1 \times n}$ and $\mathbf{V} = [\mathbf{v}_1, \ \cdots \ , \ \mathbf{v}_n] \in \mathbb{R}^{n_2 \times n}$, $n =$ min(n_1, n_2) satisfying
  1. $\mathbf{U}^T\mathbf{U} = \mathbf{I}_n$.
  2. $\mathbf{V}^T\mathbf{V} = \mathbf{I}_n$.
  3. $\langle \mathbf{u}_i\mathbf{v}_i^T | \mathbf{u}_j\mathbf{v}_j^T \rangle = \delta_{ij}, \ 1 \leq i,j \leq n$

Extended Singular Value Decomposition

- Recall that the SVD of $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ such that $\Sigma =$diag$(\sigma_1 \geq \cdots \geq \sigma_n \geq 0)$, $\mathbf{U} = [\mathbf{u}_1, \ \cdots \ , \ \mathbf{u}_n] \in \mathbb{R}^{n_1 \times n}$ and $\mathbf{V} = [\mathbf{v}_1, \ \cdots \ , \ \mathbf{v}_n] \in \mathbb{R}^{n_2 \times n}$, $n =$ min(n_1, n_2) satisfying
  1. $\mathbf{U}^T\mathbf{U} = \mathbf{I}_n$.
  2. $\mathbf{V}^T\mathbf{V} = \mathbf{I}_n$.
  3. $\langle \mathbf{u}_i\mathbf{v}_i^T | \mathbf{u}_j\mathbf{v}_j^T \rangle = \delta_{ij}, \ 1 \leq i,j \leq n$
- The Extended-SVD of $\mathbf{A}$ given the set of revealed entries Ω is given by $\mathbf{A}\odot\Omega = (\mathbf{U}\Sigma\mathbf{V}^T)\odot\Omega$ with $\Sigma =$diag$(\sigma_1 \geq \cdots \geq \sigma_n \geq 0)$, $\mathbf{U} = [\mathbf{u}_1, \ \cdots \ , \ \mathbf{u}_n] \in \mathbb{R}^{n_1 \times n}$ and $\mathbf{V} = [\mathbf{v}_1, \ \cdots \ , \ \mathbf{v}_n] \in \mathbb{R}^{n_2 \times n}$, $n =$ min(n_1, n_2) satisfying

Extended Singular Value Decomposition

- Recall that the SVD of $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ such that $\Sigma =$diag$(\sigma_1 \geq \cdots \geq \sigma_n \geq 0)$, $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_n] \in \mathbb{R}^{n_1 \times n}$ and $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_n] \in \mathbb{R}^{n_2 \times n}$, $n = \min(n_1, n_2)$ satisfying
  1. $\mathbf{U}^T\mathbf{U} = \mathbf{I}_n$.
  2. $\mathbf{V}^T\mathbf{V} = \mathbf{I}_n$.
  3. $\langle \mathbf{u}_i\mathbf{v}_i^T | \mathbf{u}_j\mathbf{v}_j^T \rangle = \delta_{ij},\ 1 \leq i,j \leq n$
- The Extended-SVD of $\mathbf{A}$ given the set of revealed entries Ω is given by $\mathbf{A}\odot\Omega = (\mathbf{U}\Sigma\mathbf{V}^T)\odot\Omega$ with $\Sigma =$diag$(\sigma_1 \geq \cdots \geq \sigma_n \geq 0)$, $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_n] \in \mathbb{R}^{n_1 \times n}$ and $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_n] \in \mathbb{R}^{n_2 \times n}$, $n = \min(n_1, n_2)$ satisfying
  1. $\mathbf{U}^T\mathbf{U} =$ diag$(\gamma_1, \cdots, \gamma_n)$.

Extended Singular Value Decomposition

- Recall that the SVD of $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ such that $\Sigma =$diag$(\sigma_1 \geq \cdots \geq \sigma_n \geq 0)$, $\mathbf{U} = [\mathbf{u}_1, \ \cdots \ , \ \mathbf{u}_n] \in \mathbb{R}^{n_1 \times n}$ and $\mathbf{V} = [\mathbf{v}_1, \ \cdots \ , \ \mathbf{v}_n] \in \mathbb{R}^{n_2 \times n}$, $n =$ min(n_1, n_2) satisfying
  1. $\mathbf{U}^T\mathbf{U} = \mathbf{I}_n$.
  2. $\mathbf{V}^T\mathbf{V} = \mathbf{I}_n$.
  3. $\langle \mathbf{u}_i\mathbf{v}_i^T | \mathbf{u}_j\mathbf{v}_j^T \rangle = \delta_{ij}, \ 1 \leq i,j \leq n$
- The Extended-SVD of $\mathbf{A}$ given the set of revealed entries Ω is given by $\mathbf{A}\odot\Omega = (\mathbf{U}\Sigma\mathbf{V}^T)\odot\Omega$ with $\Sigma =$diag$(\sigma_1 \geq \cdots \geq \sigma_n \geq 0)$, $\mathbf{U} = [\mathbf{u}_1, \ \cdots \ , \ \mathbf{u}_n] \in \mathbb{R}^{n_1 \times n}$ and $\mathbf{V} = [\mathbf{v}_1, \ \cdots \ , \ \mathbf{v}_n] \in \mathbb{R}^{n_2 \times n}$, $n =$ min(n_1, n_2) satisfying
  1. $\mathbf{U}^T\mathbf{U} =$ diag$(\gamma_1, \ \cdots \ , \ \gamma_n)$.
  2. $\mathbf{V}^T\mathbf{V} = \mathbf{I}_n$.

Extended Singular Value Decomposition

- Recall that the SVD of $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ such that $\Sigma =$diag$(\sigma_1 \geq \cdots \geq \sigma_n \geq 0)$, $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_n] \in \mathbb{R}^{n_1 \times n}$ and $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_n] \in \mathbb{R}^{n_2 \times n}$, $n = \min(n_1, n_2)$ satisfying
  1. $\mathbf{U}^T\mathbf{U} = \mathbf{I}_n$.
  2. $\mathbf{V}^T\mathbf{V} = \mathbf{I}_n$.
  3. $\langle \mathbf{u}_i\mathbf{v}_i^T | \mathbf{u}_j\mathbf{v}_j^T \rangle = \delta_{ij}, \ 1 \leq i,j \leq n$
- The Extended-SVD of $\mathbf{A}$ given the set of revealed entries Ω is given by $\mathbf{A} \odot \Omega = (\mathbf{U}\Sigma\mathbf{V}^T) \odot \Omega$ with $\Sigma =$diag$(\sigma_1 \geq \cdots \geq \sigma_n \geq 0)$, $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_n] \in \mathbb{R}^{n_1 \times n}$ and $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_n] \in \mathbb{R}^{n_2 \times n}$, $n = \min(n_1, n_2)$ satisfying
  1. $\mathbf{U}^T\mathbf{U} =$ diag$(\gamma_1, \cdots, \gamma_n)$.
  2. $\mathbf{V}^T\mathbf{V} = \mathbf{I}_n$.
  3. $\langle \mathbf{u}_i\mathbf{v}_i^T | \mathbf{u}_j\mathbf{v}_j^T \rangle_{\mathbf{\Omega}} = \delta_{ij}, \ 1 \leq i,j \leq n$.

Performance of the proposed initialization

Recall the matrix completion problem

$$(\mathbf{P}_r) \min_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} ||\mathbf{A} - \mathbf{X}||_{\Omega}^2$$
$$\text{s.t.} \quad \text{Rank}(\mathbf{X}) = r.$$

Performance of the proposed initialization

Recall the matrix completion problem

$$(\mathbf{P}_r) \min_{\mathbf{X}\in\mathbb{R}^{n_1\times n_2}} ||\mathbf{A}-\mathbf{X}||_\Omega^2 \quad \text{s.t.} \quad \text{Rank}(\mathbf{X}) = r.$$

Theorem

A sufficient condition for the output $\mathbf{X}_r$ after r iterations to serve as a good initialization to $(\mathbf{P}_r)$, in the sense that it is closer in the Frobenius norm to the optimal solution than the all zeros matrix, is:

$$(1-\alpha)||\mathbf{A}-\mathbf{U}_r\Sigma_r\mathbf{V}_r^T||_{\mathbf{\Omega}} \leq \sqrt{1+\alpha^2}||\mathbf{A}||_{\mathbf{\Omega}}.$$

where $\mathbf{U}_r\Sigma_r\mathbf{V}_r^T$ is the truncated E-SVD of $\mathbf{A}$, and $\alpha = \sqrt{\frac{1-p}{p}}$, with p being the probability that an entry is revealed.

Performance of the proposed initialization

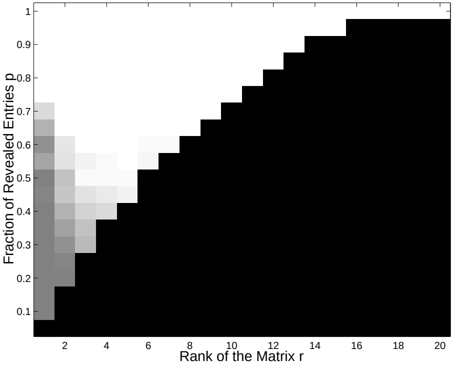

Figure: Region in which the sufficient condition of Theorem 1 is satisfied for a 20×20 matrix.

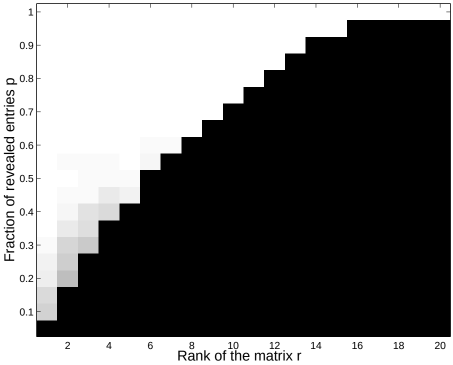

Figure: Region in which the initialization is close to the optimal solution for a 20×20 matrix.

How to deal with local minima?

- Being close to the optimal is good but not sufficient.
- Presence of local minima in the search space.
- Use multiple norms, denoted by Ψ, derived from the Ω-norm to randomize the location of local minima while preserving the position of the global one.

How to deal with local minima?

- Being close to the optimal is good but not sufficient.
- Presence of local minima in the search space.
- Use multiple norms, denoted by Ψ, derived from the Ω-norm to randomize the location of local minima while preserving the position of the global one.

In other words, for positive and random Ψ_{ij}'s, we solve the problem

$$(\mathbf{P}_r) \min_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} ||\mathbf{A} - \mathbf{X}||_{\Psi}^2 = \sum_{(i,j) \in \Omega} \Psi_{ij} (\mathbf{A}_{ij} - \mathbf{X}_{ij})^2$$
$$\text{s.t.} \quad \text{Rank}(\mathbf{X}) = r.$$

Landscape change with Ψ

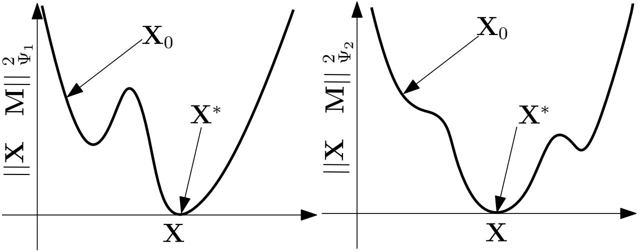

Figure: Effect of the random norm Ψ.

Performance of the proposed method

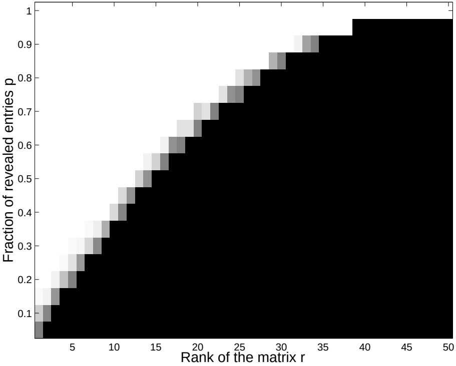

Figure: Riemannian approach with improved initialization.

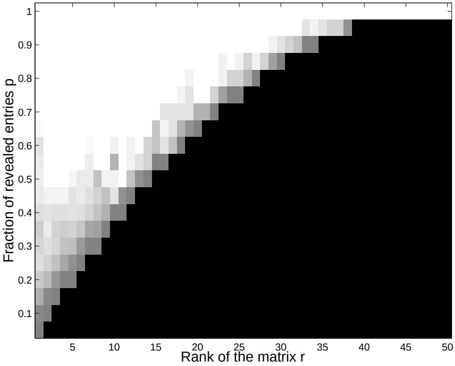

Figure: Riemannian approach with arbitrary initialization.

Much faster convergence

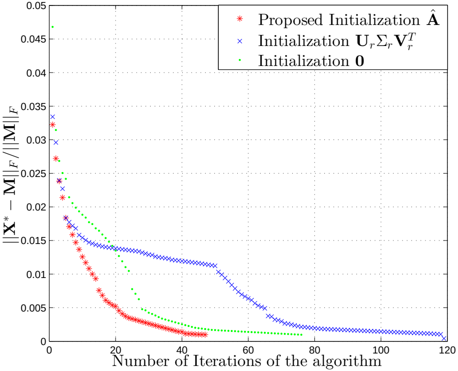

Figure: Speed of convergence for $n = 50$ and $r = 29$.

- Impressive convergence region.
- Very close to the information theoretical bound.
- Highly efficient use of the computation resources.

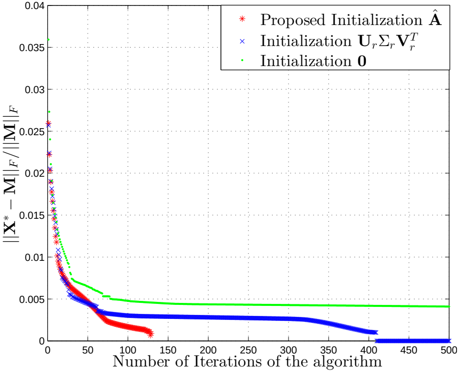

Figure: Speed of convergence for $n = 50$ and $r = 33$.

- Impressive convergence region.
- Very close to the information theoretical bound.
- Highly efficient use of the computation resources.
- Solve previously impossible to solve configuration.

Conclusion

- This work propose an efficient method, with theoretical guarantees, to find an improved initialization to the matrix completion problem.
- To mitigate the effect of the local minima, a new class of norms is introduced to random the location of local minima.
- Simulation results shows a two-fold improvement:
  1. Larger convergence region.
  2. Better convergence speed.
- Extension of the work to an online setting is of high interest to industry.

THANK YOU

For more questions, please email

ahmed.douik@caltech.edu