

# Matrix Completion has No Spurious Local Minimum

Meraj Hashemizadeh  
March 2020

# What is matrix completion?

$$M \in \mathbb{R}^{d \times d}$$

Partially observed subset  $\Omega = \{(i, j) \mid M_{ij} \text{ is observed}\}$

Is it possible from the available entries to guess the many entries that are missing?

In general it is an impossible task because the unknown entries could be anything. However, if one knows that the matrix is **low rank** and makes a few reasonable assumptions, then the matrix can be reconstructed.

# Recommender systems

*Item j*

				4		
	3					2
4	1			4		
			2			1
4				4		
			3			
	4				3	

*User i*

We want to recover the missing elements in the user ratings

# What kind of matrices can be recovered?

This question seems difficult to answer

$$\begin{bmatrix} - & 2 & 3 & 4 \\ - & 4 & 6 & 8 \\ - & 6 & 9 & 12 \end{bmatrix}$$

- ❖ The observed entries of the matrix should be selected uniformly at random.

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

No matter how you sample this matrix you need every entry to recover it.

- ❖ There seems a be relation between the sparsity of the matrix and the number of required observed samples.

# Coherence

- (Coherence parameter, [Recht 11]) Let  $U$  be a subspace of  $\mathbb{R}^d$  of dimension  $r$ , and  $P_U$  be the orthogonal projection onto  $U$ . The coherence of  $U$  (with respect to the standard basis,  $\{\hat{e}_1, \dots, \hat{e}_d\}$ ), is defined to be:

$$\mu(U) = \frac{d}{r} \max_{1 \leq i \leq d} \|P_U \hat{e}_i\|^2$$

- ( $\mu_0$ -Incoherence) Given a matrix  $M$  of rank  $r$ , we say that  $M$  is  $\mu_0$ -incoherent if:

$$\max(\mu(U), \mu(V)) \leq \mu_0$$

# Previous Work

Convex relaxation (nuclear norm minimization)

[Recht,Fazel&Parrilo 10][Candes&Recht 08][Candes&Tao 09]

[Recht 11]

$$\min_X ||X||_*$$

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

Optimal sample complexity  $O(dr \log d)$

Main drawback speed

# Previous Work

Low-rank factorization [Burer&Monteiro 03]

$$d \begin{bmatrix} M \end{bmatrix} = \begin{bmatrix} r \\ Z \end{bmatrix} \begin{bmatrix} Z^\top \end{bmatrix}$$

$$\min_{X \in \mathbb{R}^{d \times r}} f(X) := \frac{1}{2} \sum_{(i,j) \in \Omega} \left( \left( \frac{M_{ij}}{Z_{ij}} \right) \left( \frac{X_{ij}}{Z_{ij}} \right) \right)^2$$

# Putting it all together

$$f(X) = \frac{1}{2} \left\| \sum_{(i,j) \in \Omega} (M_{ij} - X_{ij}) \right\|_F^2$$

Partially observed  $\Omega = \{(i, j) : M_{ij} \text{ is observed}\}$

Each entry is observed with probability  $p$

$M$  is  $\mu$ -incoherent,  $\mu = \sqrt{d} \max_i \|Z_i\|_2 / \|Z\|_F$



# This paper's contribution

$$f(X) := \frac{1}{2} \sum_{(i,j) \in \Omega} \left( M_{ij} - (XX^\top)_{ij} \right)^2$$

Main Theorem [Informal]: Under usual assumptions, with high probability over the choice of  $\Omega$ ,

**All local minima of  $f(X)$  are global minima and recover  $M$ .**

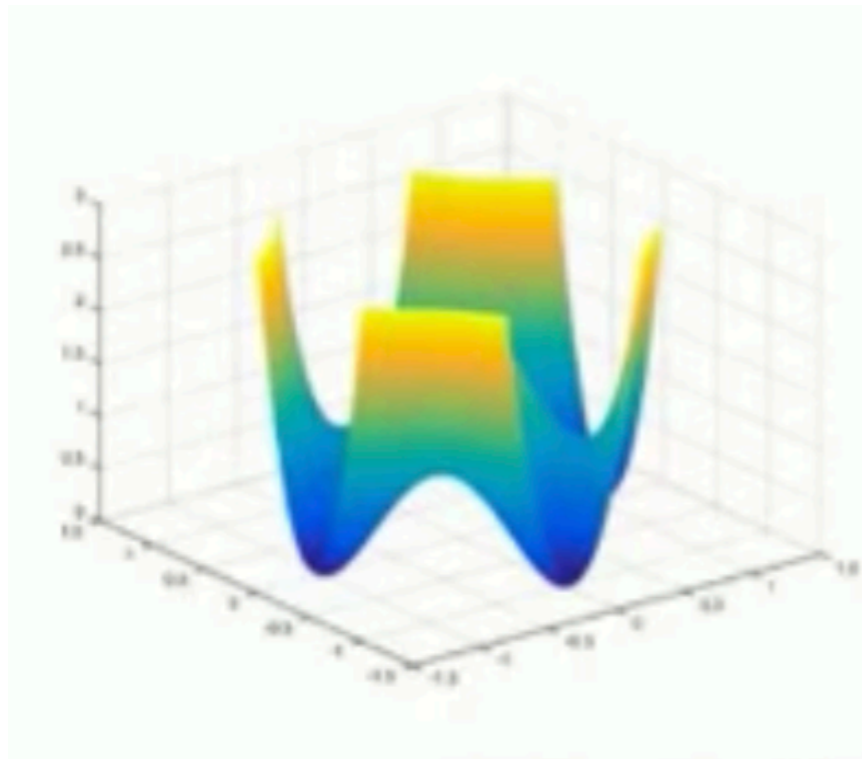
Assumptions:

1.  $M$  is  $\mu$ -incoherent
2.  $M$  is well-conditioned, i.e. has small condition number
3.  $\mathbb{E}\{ |\Omega| \} = pd^2 \gtrsim O(dr^6)$  (suboptimal)

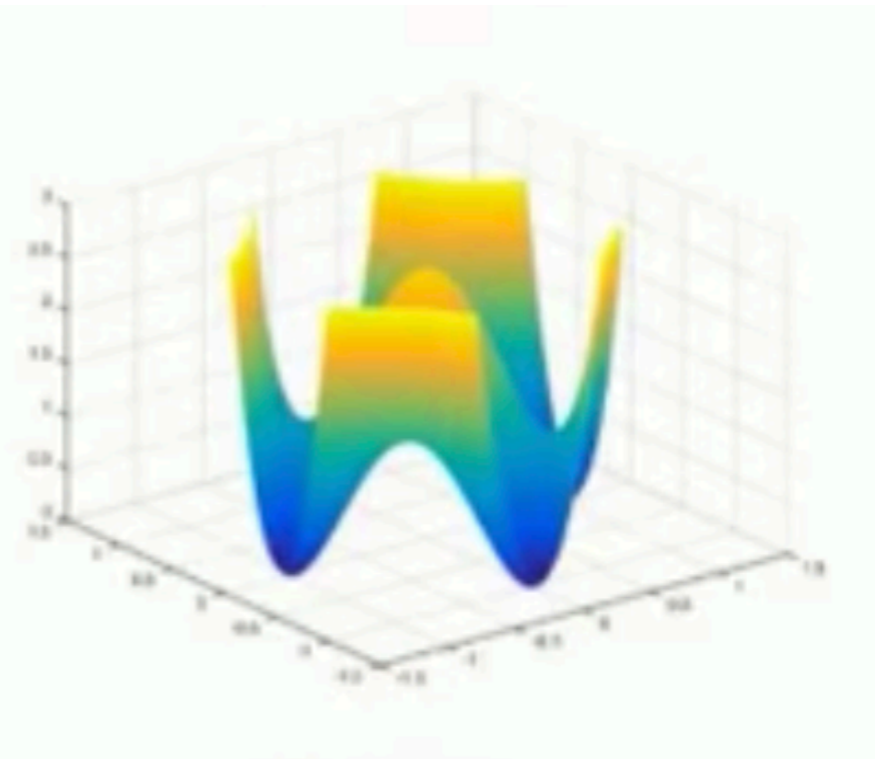
# Proof technique

Observation: geometric properties are preserved after random sampling

Full observation



Partial observation



rank-1, restricted to random 2-dimensional subspace

# Proof technique

Linear properties are preserved by random sampling.

$$\sum_{(i,j) \in \Omega} W_{ij} = \sum_{(i,j) \in [d] \times [d]} \mathbf{1}_{(i,j) \in \Omega} W_{ij} \approx p \cdot \sum_{(i,j) \in [d] \times [d]} W_{ij}$$

Law of large numbers

$$\langle P_{\Omega}(A), P_{\Omega}(B) \rangle = \langle A, P_{\Omega}(B) \rangle \approx p \langle A, B \rangle$$

$$\|P_{\Omega}(A)\|_F^2 \approx p \cdot \|A\|_F^2$$

$\Rightarrow$  Objective  $f$ , its gradient and hessian are linear in  $\mathbf{1}_{\Omega}$

# Proof idea

## rank-1 case

Full observation

$$g(x) = \frac{1}{2} ||zz^\top - xx^\top||_F^2$$

$$\nabla g(x) = 0 \Rightarrow \langle x, z \rangle^2 = ||x||^4$$

$$\nabla g(x) = (zz^\top - xx^\top)x = 0$$

$$\Rightarrow \langle x, \nabla g(x) \rangle = 0$$

$$\Rightarrow \langle x, (zz^\top - xx^\top)x \rangle = 0$$

$$\Rightarrow \langle x, z \rangle^2 = ||x||^4$$

Partial observation

$$f(x) = \frac{1}{2} ||P_\Omega(zz^\top) - P_\Omega(xx^\top)||_F^2$$

$$\nabla f(x) = 0 \Rightarrow \langle x, z \rangle^2 \approx ||x||^4$$

$$\nabla f(x) = (P_\Omega(zz^\top) - P_\Omega(xx^\top))x = 0$$

$$\Rightarrow \langle x, \nabla f(x) \rangle = 0$$

$$\Rightarrow \langle x, (P_\Omega(zz^\top) - P_\Omega(xx^\top))x \rangle = 0$$

$$\Rightarrow \langle x, z \rangle^2 \approx ||x||^4$$

# Proof idea

## rank-1 case

Full observation

$$g(x) = \frac{1}{2} ||zz^\top - xx^\top||_F^2$$

$$\nabla^2 g(x) \succeq 0 \Rightarrow ||x||^2 \geq 1/3$$

Proof:

$$\nabla^2 g(x) \succeq 0$$

$$\Rightarrow \langle z, \nabla^2 g(x)z \rangle \geq 0$$

$\Rightarrow \dots$  (elementary calculation)

$$\Rightarrow ||x||^2 \geq 1/3 ||z||^2 = 1/3$$

Partial observation

$$f(x) = \frac{1}{2} ||P_\Omega(zz^\top) - P_\Omega(xx^\top)||_F^2$$

$$\nabla^2 f(x) \succeq 0 \Rightarrow ||x||^2 \geq 1/4$$

Proof:

$$\nabla^2 f(x) \succeq 0$$

$$\Rightarrow \langle z, \nabla^2 f(x)z \rangle \geq 0$$

$\Rightarrow \dots$  (elementary calculation)

$$\Rightarrow ||x||^2 \gtrsim 1/3 ||z||^2 = 1/4$$

# Proof idea

## rank-1 case

Full observation

$$g(x) = \frac{1}{2} ||zz^\top - xx^\top||_F^2$$

$$\nabla g(x) = 0 \Rightarrow \langle x, z \rangle^2 = ||x||^4$$

$$\nabla^2 g(x) \succeq 0 \Rightarrow ||x||^2 \geq 1/3$$

$$\langle x, z \rangle^2 \geq 1/9$$

$$\langle z, \nabla g(x) \rangle = \langle z, (zz^\top - xx^\top)x \rangle = 0$$

$$\Rightarrow \langle x, z \rangle(1 - ||x||^2) = 0$$

$$\Rightarrow ||x||^2 = 1 \Rightarrow x = \pm z$$

Partial observation

$$f(x) = \frac{1}{2} ||P_\Omega(zz^\top) - P_\Omega(xx^\top)||_F^2$$

$$\nabla f(x) = 0 \Rightarrow \langle x, z \rangle^2 \approx ||x||^4$$

$$\nabla^2 f(x) \succeq 0 \Rightarrow ||x||^2 \geq 1/4$$

$$\langle x, z \rangle^2 \geq 1/16$$

$$\langle z, \nabla f(x) \rangle = \langle z, P_\Omega(zz^\top - xx^\top)x \rangle = 0$$

$$\Rightarrow \dots$$

$$\Rightarrow ||x||^2 \approx 1 \Rightarrow x \approx \pm z$$

# Proof idea

## rank-1 case

Concentration inequalities ( $\approx$ ) only hold for incoherent  $x$ .

$$f(X) := \frac{1}{2} \sum_{(i,j) \in \Omega} \left( M_{ij} - (XX^\top)_{ij} \right)^2 + R(X)$$

$$R(X) = \sum_{i=1}^d \text{relu}^4(\|X_i\| - \alpha)$$

$$R(Z) = 0$$

# Proof sketch (general case)

$$\nabla f(X) = 0 \Rightarrow \max_i \|X_i\| \leq 4\alpha$$

$$\nabla^2 f(X) \succeq 0 \Rightarrow \sigma_{\min}(X) \geq \frac{1}{4} \sigma_{\min}(Z)$$

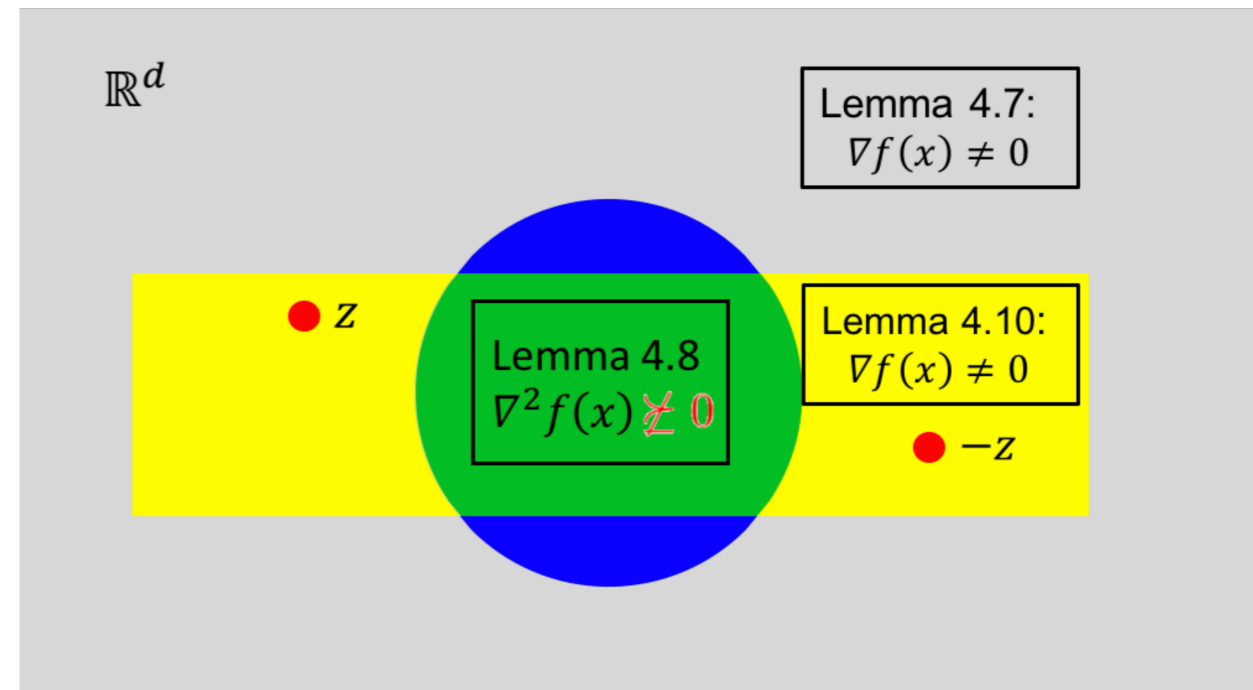
$$\|XX^\top - ZZ^\top\|_F^2 \leq O(\epsilon)$$

Lemma [Sun&Luo16]:

There exists  $U$  such that

$$UU^\top = ZZ^\top, \text{ and}$$

$$\langle -\nabla f(X), U - X \rangle \geq \frac{p}{4} \|M - XX^\top\|_F^2$$





# Results

SGD finds global minimum from arbitrary initialization

Can be extended to asymmetric and noisy cases

The simple proof technique is applicable in other problems with partial observations

# References

- [1] Ge, R., Lee, J.D. and Ma, T., 2016. Matrix completion has no spurious local minimum. In *Advances in Neural Information Processing Systems* (pp. 2973-2981).
- [2] Candès, E.J. and Recht, B., 2009. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6), p.717.
- [3] Recht, B., 2011. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12(Dec), pp.3413-3430.
- [4] Recht, B., Fazel, M. and Parrilo, P.A., 2010. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3), pp.471-501.
- [5] Burer, S. and Monteiro, R.D., 2003. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2), pp.329-357.
- [6] Sun, R. and Luo, Z.Q., 2016. Guaranteed matrix completion via non-convex factorization. *IEEE Transactions on Information Theory*, 62(11), pp.6535-6579.