Understanding Non-convex Optimization for Matrix Completion

Ruoyu Sun
Facebook AI Research

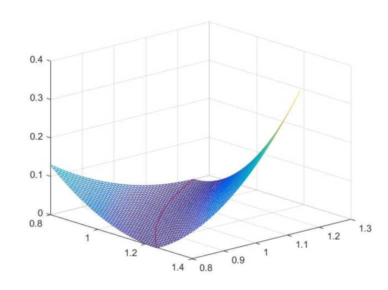
Joint work with Zhi-Quan Luo (U of Minnesota and CUHK-SZ)

Sep 23, 2016.

What I Will Talk About

We will prove: a non-convex factorization formulation can be solved to global optima.

Matrix factorization has a certain geometrical property:



Introduction

Motivation: Recommendation System

- Recommendation systems (personalized)

Richard	5	1	4
Mary	?	2	5
Steve	4	3	2

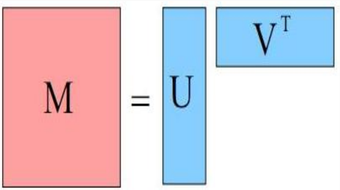
- Question: predict missing ratings?
- Assumption: low-rank.
- **Matrix Completion** (MC): recover a low-rank matrix M from partial data [Candes,Recht-09].

MC and big data

- Low-rank models are ubiquitous
  - “small” information from “big” data
- **Idea of MC**: To extract “small” information, a few samples enough
- Other applications:
  - Structure-from-motion in computer vision [Tomasi,Kanade-92],[Chen,Suter-04]
  - System identification in control [Liu,Vandenberghe-08]
  - Collaborative ranking (search, advertisements, marketing, etc.) [Yi et al.-11],[Lu,Neghaban-14]
  - Genomic data integration [Cai-Cai-Zhang-14]

Fill in the gap?

- Matrix factorization (non-convex): $M = UV^T$.



- Interestingly, always global-min for MC, if M low-rank.
- Long been used in recommendation systems [Funk'06], [Koren'09]
- **Question**: any theory?

Summary of Results

Theorem 1 (local geometry) For a properly regularized non-convex formulation, there is no bad stationary point in a certain neighborhood of the global optima.

Theorem 2 (global convergence) Starting from certain initial point, most standard algorithms converge to global optima.

Background and Formulation

Brief History of Matrix Completion

Convex (08-10): Theory for convex formulation [Candes-Tao'08] [Candes-Recht'09], [Gross'09], [Recht'09], [Chen'14].

Non-convex Grassman manifold: [Keshavan-Montanari-Oh'09]

Non-convex Resampling AltMin (12-14): [Jain et al.'12], [Keshavan'12], [Hardt'13], [Hardt-Wooters'14].

- Subtle issue: Theory does not match algorithm.

Our work: Geometry of Matrix Factorization [Sun-Luo'14].

Factorization is Fundamental

Previous understanding: extension of CS/LASSO

Current: "factorization" is crucial.

Recent "trend": Matrix factorization.

- **Industry**: embedding at, say, facebook
- **Theory**: PCA [Musco,Musco-15] , Laplacian linear system [Kyng,Sacheva'16]
- **Non-convex optimization**: tensor decomposition, phase retrieval, dictionary learning, etc.

Factorization for Matrix Completion

- Notations:
  - $M \in \mathbb{R}^{n\times n}$, rank $r \ll n$ [1]
  - Ω: set of sampled positions.
- Matrix factorization formulation (non-convex):

$$\mathrm{P}_0: \min_{X\in\mathbb{R}^{n\times r}, Y\in\mathbb{R}^{n\times r}} \frac{1}{2}\|\mathcal{P}_\Omega(M - XY^T)\|_F^2. \tag{1}$$

where $\mathcal{P}_\Omega$ is a sampling operator

$$\mathcal{P}_\Omega(Z)_{ij} = \begin{cases} Z_{ij} & (i,j)\in\Omega, \\ 0 & (i,j)\notin\Omega. \end{cases}$$

[1] M can be rectangular; assume square just for simplicity.

Algorithms to solve MF

- Recall: $\min_{X,Y} \|\mathcal{P}_\Omega(M - XY^T)\|_F^2$.
- Algorithmic idea: solve smaller sub-problems
  - Coordinate Descent for MC [Koren-09],[Wen-Yin-Zhang-12],[Yu et al. 12][Hastie et al. 14]
  - SGD for MC [Koren-09], [Funk-06], [Gemulla et al. 11], [Recht-Re-13], [Zhuang et al. 13] .
- SGD solves Netflix problem (500k $\times$ 20k) in $\leq$ 3 mins! [Recht-Re-13]

Local Geometry

Full-Observation Case: Low-rank Approximation

- **Matrix factorization** formulation (non-convex):

$$\min_{X,Y \in \mathbb{R}^{n\times r}} \quad \frac{1}{2}\|M - XY^T\|_F^2. \tag{2}$$

- **Claim**: Optimal solution = best rank-r approximation.

- Is this problem hard? Why?

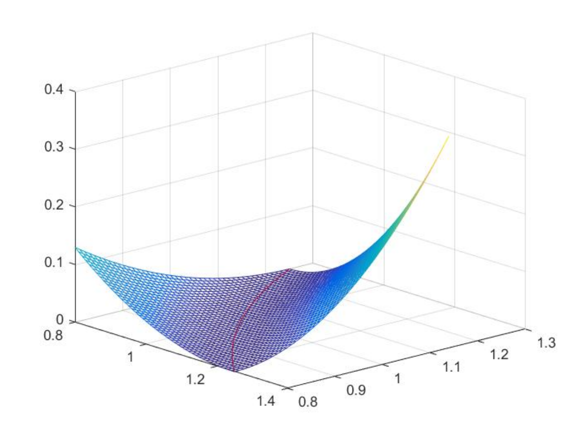
- AltMin equivalent to power method [Szlam,Tulloch,Tygert'16]
  What about other algorithms?

Simplest Case

Consider $n = r = 1$, i.e.

$$\min_{x,y} (xy - 1)^2.$$

Nice geometry: locally no other critical points.

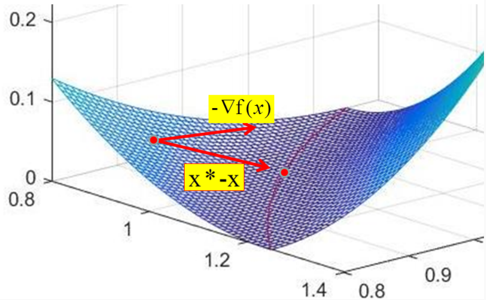


Step 1: Local Geometry of $\|M - XY^T\|_F^2$

Lemma[2]: In a neighborhood of M, $\forall x = (X, Y)$, $\exists$ optimal x^* s.t.

$$\langle \nabla f(x), x - x^* \rangle \geq c\|x - x^*\|^2.$$

Interpretation: local direction aligns with global direction $x^* - x$.



Corollary: Locally no bad critical point!

[2] Related to Polyak inequality, Lojaswicz inequality, error bound, etc.

Step 2: Geometry of $\|\mathcal{P}_\Omega(M - XY^T)\|_F^2$

Question: why $f(x, y) = (xy - 1)^2$ has nice geometry?

- "**Answer**": strong convexity of $(z - 1)^2$ preserved after $z \to xy$

[3] Techniques: new perturbation analysis + probability tools

Step 2: Geometry of $\|\mathcal{P}_\Omega(M - XY^T)\|_F^2$

Question: why $f(x, y) = (xy - 1)^2$ has nice geometry?

- "**Answer**": strong convexity of $(z - 1)^2$ preserved after $z \to xy$

Challenge: $\|\mathcal{P}_\Omega(M - Z)\|_F^2$ not strongly convex!

Lemma 2[S-L'14] (RIP): In a local "incoherent" region,[3]

$$\|\mathcal{P}_\Omega(M - Z)\|_F^2 \geq c\|M - Z\|^2.$$

"Incoherent' means $Z = XY^T$ where X, Y have bounded row norms.

[3] Techniques: new perturbation analysis + probability tools

Step 3: Geometry with Constraints/Regularizers

Need to add regularizer G_1 to force incoherence.

- Issue: ∇G_1 not aligned with global direction $x^* - x$.

Need to add one more regularizer G_2 to correct G_1 [4].

Theorem [S-Luo]: Any stationary point of $F + G_1 + G_2$ in a local incoherent region is a global-min, i.e., the original matrix, under standard assumptions on M, Ω.

[4] Technique: Perturbation analysis for "pre-conditioning", rather involved.

Formulation with Regularizers

- Our formulation

$$\min F(X, Y) + G_1(X, Y) + G_2(X, Y),$$

G_1, G_2 guarantee (X, Y) close to

$$(X, Y) \in K_1 = \{\|X^{(i)}\| \le \beta_1, \quad \|Y^{(i)}\| \le \beta_1, \ \forall\ i.\}$$
$$(X, Y) \in K_2 = \{\|X\|_F \le \beta_T, \quad \|Y\|_F \le \beta_T\},$$

- K_1: incoherence (row-norm)
- K_2: boundedness

Simulation: push # of samples to fundamental limit $\approx 2nr$.

Summary

- **Goal**: Understand why MF works
- **Result**: guarantee for non-convex matrix completion
- **Fundamental question**: Why low-rank approximation is "easy"? Go beyond power method!
- **Take-away**: MF has nice geometry
  - with missing entries, regularization needed

References

- [Sun-Luo-14] Ruoyu Sun and Zhi-Quan Luo, "**Guaranteed Matrix Completion via Non-convex Factorization**,"
  –2015 INFORMS Optimization Society Student Paper Competition, Honorable Mention.
  –*FOCS* (Foudation of Computer Science) 2015.
  –To appear in *IEEE Transaction on Information Theory*,2016.

- [Sun-15] Ruoyu Sun, "**Matrix Completion via Non-convex Factorization: Algorithms and Theory**," Ph.D. Dissertation, University of Minnesota, 2015.

- Remark: [Sun-Luo-14] focuses on theory;
  My thesis [Sun-15] presents new numerical findings supporting theory.

Thank You!