Matrix Completion: Theory and Implementations

Liangzu Peng
School of Information Science and Technology
ShanghaiTech University
penglz@shanghaitech.edu.cn

Abstract

This report presents efforts given by Liangzu Peng during the course project Convex Optimization, where some theory and algorithmic implementations for matrix completion are considered. Specifically, we survey the theory that there is no spurious local minimum for positive semidefinite (PSD) matrix completion given by Ge et al. [4]. To further appreciate this theory, we implement a gradient descent algorithm for PSD matrix completion (PSD-GD) as a sanity check, the experiments corresponding to the theory as observed. Moreover, some more advanced algorithms, i.e., direct nuclear norm minimization [3] (NucNorm) and singular value threshold algorithms [2] (SVT) are implemented. From the comparative study between NucNorm and SVT we see the tradeoff between speed and accuracy for optimization algorithms. finally, we conduct an experiment which shows the behavior of SVT implementation corresponds to the theory in [2].

1 Introduction

Matrix Completion is the problem of revealing the matrix from partially observed entries, motivated by many real world applications. For example, One may want to find answers in a partially filled out survey, or, as a more famous example, one would like to infer users' preference for unrated items in recommender systems (The Netflix Problem). Formally speaking, PSD Matrix Completion problem can be formulated as following problem:

$$\underset{X}{\text{minimize}}\, ||P_\Omega(XX^\top - M)||_F^2, \tag{1}$$

where $M = ZZ^\top \in \mathbb{R}^{d\times d}$ is a positive semidefinite matrix, $\Omega \subset [d] \times [d]$ is the set of observed entries and P is the projection operator that maps a matrix A to $P_\Omega(A)$, where $P_\Omega(A)$ has the same values as A on Ω, and 0 outside of Ω. While Problem 1 is in general ill-posed, Candes and Recht [3] proves that Problem 1 can be exactly solved via convex optimization with some probability for low rank matrix with incoherence assumption. Ongie et al. [6] generalizes low-rank matrix completion to a much wider class of data models by modeling data as an algebraic variety, in which way the matrix to complete is probably high rank.

In this report we will study some highlights in Ge et al. (Sec. 2) at first, then show the experimental results (Sec. 3), and finally we conclude this project (Sec. 4)

2 Theory

The main goal or Ge et al. is to prove that PSD matrix completion, i.e., the following function

$$f(X) = \frac{1}{2}||P_\Omega(XX^\top - M)||_F^2 + \lambda R(X) \tag{2}$$

with $ZZ^\top \in \mathbb{R}^{d\times d}$ positive semidefinite, where $R(X) : \mathbb{R}^{d\times d} \to \mathbb{R}$ is a regularization term that will be introduced later, has no spurious local minimum, i.e., all local minima of Problem 2 are global minimum. To prove this, it is enough to show, by Lemma 3.1 in Sun and Luo [7] that the solution matrix $XX^\top$ is close enough to the target matrix $M = ZZ^\top$. Ge et al. [4] start the proof from rank-1 case with solution matrix $xx^\top$ constrained in the incoherence ball

$$\mathcal{B} = \{x \in \mathbb{R}^d : ||x||_\infty < \frac{2\mu}{\sqrt{d}}, ||x|| \leq 1\} \tag{3}$$

for simplicity, while the proof for rank-k case is merely a natural extension of rank-1 case by replacing vector operation with matrix operation. Of all the proofs, the trickiest and the most interesting part is how to extend the proof in the incoherence ball to the entire space. Ge et al. [4] introduce a regularization term to refine the geometry of the objective function $f(X)$, thus making every stationary point incoherent, e.g., in rank-1 case, we have the following results by Lemma 4.7 in the paper [4] (informal). Given a regularizer

$$R(X) = \Sigma_{i=1}^d h(x_i),$$

where $h(t) = (t-\alpha)^4 \mathbb{I}_{t \geq \alpha}$ for some α that will be determined in the proof later (not here), we have

$$||x||_\infty \leq c, \forall x \in \{x \in \mathbb{R}^d : \nabla f(x) + \nabla R(X) = 0\}$$

for some $c \in \mathbb{R}$. While this proof strategy seems to be applicable to other problems, the design of a suitable regularizer remains elusive and unmotivated, and leads to some degree of complexities.

3 Implementations

3.1 PSD-GD

While the theory in Ge et al. [4] remains technically difficult, we argue that we can believe the theory even if the theoretical details are elusive. The approach is to verify it empirically. This motivates the following algorithm and experiments. We implement gradient descent method for PSD matrix completion, and run it 500 times with random initialization. A crucial observation is that the algorithm converges to the same point in 500 times of running, we can thus to some extend believe that there is no spurious local minimum. The details are given below.

Let $f(X) = \frac{1}{2}||P_\Omega(XX^\top - M)||_F^2$ as in Problem 1 (no regularization for simplicity). Notice that in the presentation, we claim that computing the gradient of$f(X)$ is difficult and as a workaround we compute the gradient of $g(X) = \frac{1}{2}||XX^\top - P_\Omega(M)||_F^2$ by observing that $g(X)$ is an upper bound of $f(X)$. Professor Shi requires us to prove $g(X) \geq f(X)$ in the report, which is almost immediate since

$$f(X) = \frac{1}{2}||P_\Omega(XX^\top - M)||_F^2 = \frac{1}{2}||P_\Omega(XX^\top) - P_\Omega(M)||_F^2 \overset{(*)}{\leq} \frac{1}{2}||XX^\top - P_\Omega(M)||_F^2 = g(X),$$

where $(*)$ holds since the terms outside of Ω in Frobenius norm is nonnegative.

However, it turns out that the gradient of $f(X)$ is given by

$$\nabla f(X) = 2P_\Omega(XX^\top)X - 2P_\Omega(M)X.$$

Based on this formula, we refine the gradient descent algorithm to minimize $f(X)$, obtaining similar results as minimizing $g(X)$ shown in the presentation. We can see from Figure 1 that, in 500 times of running, the algorithm converges to the same function value. It remains to verify whether they are the same minimizer of $f(X)$. Indeed, as shown in Figure 2, up to some numerical errors (± 0.0001), they have the same norm, implying that they are in fact the same minimizer. In these two experiments we empirically verified as a sanity check that there is no spurious local minimum for the function $f(X)$.

3.2 NucNorm and SVT

The experiments in this subsection will conduct a comparative study of NucNorm and SVT algorithm, and gives an empirical verification of the theorems in SVT paper in Cai et al. [2].

Low rank matrix completion problem can be reformulated as the following one as a convex relaxation of rank minimization problem:

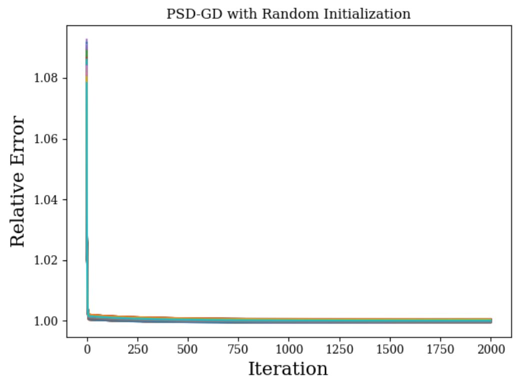

Figure 1: **PSD-GD** runs 500 times for random initialization (note that there are 500 curves in the plot).

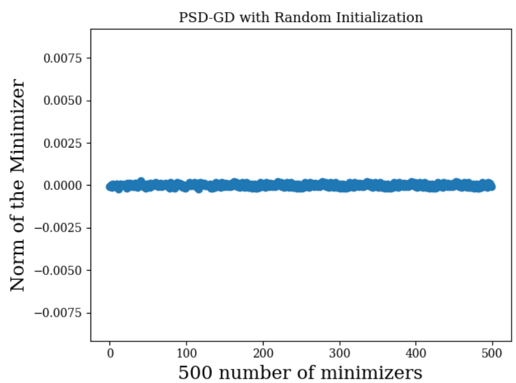

Figure 2: The norm of 500 minimizers after subtracting the average.

$$\begin{array}{ll} \text{minimize} & ||X||_* \\ \text{subject to} & P_\Omega(X) = P_\Omega(M), \end{array}$$

where $||X||_*$ is the sum of singular value of the matrix X. This problem can be solved directly via semidefinite programming and we solved it via SCS solver [5] in CVXPY [1], although SCS solver is generally slow for SDP programs. Alternatively, it can also be solved via SVT algorithm [2]:

$$\begin{cases} X^k = \mathcal{D}_\tau(Y^{k-1}) \\ Y^k = Y^{k-1} + \lambda_k P_\Omega(M - X^k), \end{cases}$$

where $\mathcal{D}_\tau(Y) = \text{prox}_{\tau||\cdot||_*}(Y)$ (as shown in homework 4) and $\lambda_k \in (0, 2)$ is the stepsize. As shown in Table 1, SVT algorithm trades accuracy for speed compared to NucNorm algorithm.

Table 1: Experiments comparing NucNorm and SVT algorithms run on MovieLens 100K ($n \times m = 943 \times 1682, |\Omega| = 10^5$).

Method	NucNorm	SVT
Time	1437 s	242 s
Relative Error	0.0001	0.026

Another experiment we conduct is the correspondence between the behavior of SVT algorithm and Theorem 4.2 in the paper [2]. The result is presented in Figure 3, from which we can see that SVT algorithm converges to a unique solution when the stepsize $\lambda_k \in (0, 2)$. That is exactly Theorem 4.2 tries to announce.

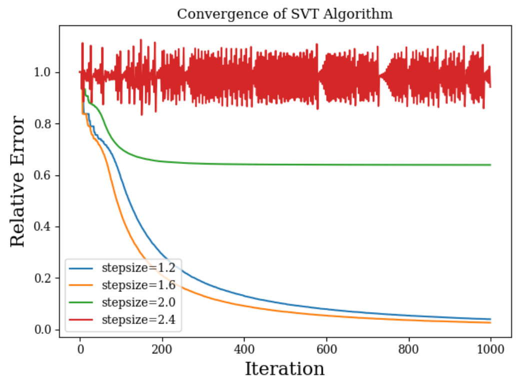

Figure 3: Convergence behavior of SVT algorithm for different stepsizes $(1.2, 1.6, 2.0, 2.4)$.

4 Conclusion and Future Work

In this course project we try to understand the theory behind positive semidefinite matrix completion problem and code some algorithmic implementations for low rank matrix completion. Emphases are put heavily on the interaction and correspondence between the theory and implementations, and on the comparison of different implementations for the same problem. An immediate problem of interest is that why the function $g(X) = \frac{1}{2}||XX^\top - P_\Omega(M)||_F^2$ (Sec. 3.1) has no spurious local minimum as we empirically show in the presentation. Building theoretical justification for this problem may be our future work. Another issue of interest that we want to investigate further is, in the absence of direct and elegant proof, how to design a proper regularizer for any given non-convex problem.

References

[1] Steven Diamond Akshay Agrawal, Robin Verschueren and Stephen Boyd. A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60, 2018.

[2] Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[3] Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717, Apr 2009.

[4] Rong Ge, Jason D. Lee, and Tengyu Ma. Matrix completion has no spurious local minimum. *CoRR*, abs/1605.07272, 2016.

[5] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd. Conic Optimization via Operator Splitting and Homogeneous Self-Dual Embedding. *ArXiv e-prints*, December 2013.

[6] G. Ongie, R. Willett, R. D. Nowak, and L. Balzano. Algebraic Variety Models for High-Rank Matrix Completion. *ArXiv e-prints*, March 2017.

[7] Ruoyu Sun and Zhi-Quan Luo. Guaranteed matrix completion via non-convex factorization. *IEEE Trans. Information Theory*, 62(11):6535–6579, 2016.