Convex and Nonconvex Optimization for Low-Rank Matrix Completion: A Review

Zijian Zhao Yichen Ji

March 5, 2024

1 Introduction

Low-rank matrix completion is a pivotal challenge aimed at reconstructing a low-rank matrix from its partially observed and noisy entries. This problem is crucial across numerous fields where high-dimensional data is encapsulated within low-dimensional geometric structures, albeit under limited observations. The quest for such low-rank matrices necessitates algorithms that balance statistical and computational efficiency. Within this realm, convex relaxation and nonconvex optimization emerge as two primary approaches to attain the aforementioned trade-off.

Nonconvex methods, particularly those adopting the Burer-Monteiro factorization, often lead to more computationally efficient solutions. In contrast, convex relaxation, despite its practical efficacy, historically lagged in theoretical justification, especially for large-scale or high-dimensional challenges, rendering it less attractive for such contexts. The turning point came with recent advancements, notably by [5], who bridged the theoretical gap by demonstrating the close alignment between solutions derived from convex and nonconvex optimization, thus offering a theoretical justification to the empirical reliability observed in convex relaxation. This literature review will go through the progress of convex and nonconvex approaches that deal with low-rank matrix completion and aims to unravel the inherent synergies between these two approaches, thereby illuminating their complementary strengths in tackling this complex problem [5].

2 Nonconvex Optimization for Low-Rank Matrix Completion

2.1 Burer-Monteiro Factorization

Nonconvex optimization for low-rank matrix completion aims to solve the following optimization problem

$$\begin{aligned} \min_{Z\in\mathbb{R}^{n_1\times n_2}} \quad & F(Z) \\ \text{s.t.} \quad & \operatorname{rank}(X)\le r \end{aligned} \tag{1}$$

where $F:\mathbb{R}^{n_1\times n_2}\to\mathbb{R}$ is a given convex loss function. Let $Z=LR^T$, where $L\in\mathbb{R}^{n_1\times r}$ and $R\in\mathbb{R}^{n_2\times r}$, we have

$$\min_{L\in\mathbb{R}^{n_1\times r},R\in\mathbb{R}^{n_2\times r}} f(L,R):=F(LR^T)+\operatorname{reg}(L,R) \tag{2}$$

where we added constraint of low rank directly to the loss function and we added a regularization term. This formulation is the Burer-Monteiro factorization [1]. The key change from the original formulation is for low-rank case ($r \ll \min\{n_1, n_2\}$), the size of all the variables is approximately linear (L, R) in $n_1 + n_2$, while originally the variables (X) are quadratic. Now it is possible to design an algorithm of linear time.

2.2 Projected Gradient Descent is Fantastic

To solve this problem, [3] lists three major classes of iterative schemes to find the global optimum: (Projected) gradient descent [2, 11, 12, 4, 16, 15, 17, 13], Alternating minimization [10, 7], and Singular value projection (SVP) [8, 14, 9]. (Projected) gradient descent would be used to a loss function $f(L, R)$ with respect to the factor variables (L, R):

$$L^{t+1} = \mathcal{P}_{\mathcal{L}} \left[L^t - \eta^t \nabla_L f(L^t, R^t) \right] \tag{3}$$

$$R^{t+1} = \mathcal{P}_{\mathcal{R}} \left[R^t - \eta^t \nabla_R f(L^t, R^t) \right] \tag{4}$$

where η^t is the step size and $\mathcal{P}_{\mathcal{L}}, \mathcal{P}_{\mathcal{R}}$ denote the Euclidean projection onto the sets $\mathcal{L}$ and $\mathcal{R}$. Alternating minimization would hold other factors constant and optimize one of the factors alternatively by a convex problem. That is:

$$L^{t+1} = \arg \min_{L \in \mathbb{R}^{n_1 \times r}} f(L, R^t) \tag{5}$$

$$R^{t+1} = \arg \min_{R \in \mathbb{R}^{n_2 \times r}} f(L^{t+1}, R) \tag{6}$$

SVP would do gradient descent of $F(LR^T)$ on the $n_1 \times n_2$ matrix space and then use SVD to project back to the factor space:

$$(L^{t+1}, R^{t+1}) = \mathrm{SVD}_r \left[L^t R^{t^T} - \eta^t \nabla F(L^t R^{t^T}) \right] \tag{7}$$

where $\mathrm{SVD}_r(Z)$ returns the top rank-r factors of Z.

2.3 Numerical Proximity as Motivation

Within the realm of methods for low-rank matrix completion, (projected) gradient descent stands out for its ability to often yield globally optimal solutions. This method is favored not only for its straightforwardness and low iteration cost but also for the comprehensive framework it provides for both noiseless and noisy scenarios [3]. Nonconvex optimization, in particular, has demonstrated its efficacy through both practical efficiency and robust theoretical statistical guarantees. Attributes such as local convergence, implicit regularization, and the assurance of global convergence—especially via saddle-escaping algorithms that exploit strict saddle properties—contribute to its high estimation accuracy.

Furthermore, empirical studies reveal a noteworthy proximity between the solutions obtained via convex proximal gradient methods and those derived from nonconvex gradient descent with spectral initialization [6]. This observation suggests that the theoretical foundation established for nonconvex approaches can be instrumental in refining the statistical error bounds associated with convex relaxation methods. We aim to delve deeper into this synergy in the subsequent section, exploring how the insights from nonconvex optimization can enhance our understanding and application of convex relaxation techniques.

3 Convex Relaxation for Low-Rank Matrix Completion

3.1 Bridging convex relaxation with nonconvex Burer-Monteiro

We change the original problem (1) to a penalized version

$$\min_{Z\in\mathbb{R}^{n\times n}} g(Z) := F(Z) + \lambda \text{ rank}(Z) \tag{8}$$

where F is a convex function and $\lambda > 0$ is a regularization parameter. Convex relaxation aims to remove the nonconvex rank function by convex terms. For example:

$$\min_{Z\in\mathbb{R}^{n\times n}} g(Z) := F(Z) + \lambda\|Z\|_* \tag{9}$$

where $\|\cdot\|_*$ is the nuclear norm, the sum of singular values of X as a convex surrogate for the rank function. More specifically, [5] considers the following regularized least-squares convex program:

$$\min_{Z\in\mathbb{R}^{n\times n}} g(Z) := \frac{1}{2}\sum_{(i,j)\in\Omega}(Z_{ij} - M_{ij})^2 + \lambda\|Z\|_* \tag{10}$$

where $\Omega \subseteq \{1,\cdots,n\}\times\{1,\cdots,n\}$ denotes a set of indices, M is the observed noisy matrix, assuming an additive form of noise corruption E over the ground truth M^*, $M = M^\star + E$.

A primary challenge in employing convex relaxation methods for noisy matrix completion is the absence of closed-form solutions that provide reliable guarantees. To establish that convex solutions offer sound theoretical assurances, one must identify a specific nonconvex algorithm capable of yielding solutions with desirable properties. Subsequently, if it can be demonstrated that these nonconvex solutions closely approximate their convex counterparts—achieving a tight approximation—we can then extend statistical guarantees to the convex solutions as well [5].

3.2 Statistical Control

For convenience and brevity of presentation, here we only present the results when the true matrix has bounded rank and condition number, that is, $r, \kappa = O(1)$, delivering near-optimal statistical guarantees. Please refer to [5] Theorem 2 for a more general setting when the rank and the condition number are both allowed to grow with the dimension n.

Theorem 1. Let $M^\star$ be rank-r and μ-incoherent with a condition number κ, where the rank and the condition number satisfy $r, \kappa = O(1)$. Suppose that Assumption 1 holds and take $\lambda = C_\lambda\sigma\sqrt{np}$ in (10) for some large enough constant $C_\lambda > 0$. Assume the sample size obeys $n^2p \geq C\mu^2 n\log^3 n$ for some sufficiently large constant $C > 0$, and the noise satisfies $\sigma \lesssim \sqrt{\frac{np}{\mu^3\log n}}\|M^\star\|_\infty$ for some sufficiently small constant $c > 0$. Then with probability exceeding $1 - O\left(n^{-3}\right)$:

1. Any minimizer Z_{cvx} of (10) obeys

$$\begin{aligned}
&\|Z_{\text{cvx}} - M^\star\|_{\text{F}} \lesssim \frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\|M^\star\|_{\text{F}}\,; \quad \|Z_{\text{cvx}} - M^\star\| \lesssim \frac{\sigma}{\sigma_{\min}}\sqrt{\frac{n}{p}}\|M^\star\|\,; \\
&\|Z_{\text{cux}} - M^\star\|_\infty \lesssim \frac{\sigma}{\sigma_{\min}}\sqrt{\frac{\mu n\log n}{p}}\|M^\star\|_\infty\,.
\end{aligned} \tag{11}$$

2. Letting $Z_{\text{cvx},r} \triangleq \arg\min_{Z:\text{rank}(\boldsymbol{Z})\leq r} \|\boldsymbol{Z} - Z_{\text{cvx}}\|_{\text{F}}$ be the best rank-r approximation of Z_{cux}, we have

$$\|Z_{\text{cvx},r} - Z_{\text{cvx}}\|_{\text{F}} \leq \frac{1}{n^3} \cdot \frac{\sigma}{\sigma_{\min}} \sqrt{\frac{n}{p}} \|M^\star\|,$$

and the error bounds in (11) continue to hold if Z_{cvx} is replaced by $Z_{\text{cvx},r}$.

3.3 Approximate Nonconvex Optimizer

In [5], an iterative nonconvex algorithm is proposed to approximate the primal solution to the convex program (10) that lacks of closed-form primal solution and is challenging to construct a dual certificate and identify a primal-dual pair satisfying the first-order optimality condition. Under some conditions for the regularization parameter, a critical point of the nonconvex regularized least-squares problem (2) can be translated to the unique minimizer of the convex program (10). However, since the convergence behavior of the iterative algorithm e.g. gradient descent to the global optimizer is not clear in the presence of noise, [5] proposes an approximate nonconvex optimizer as a reasonably tight proxy of the convex solution. When initiated from the exact solution (ground truth, denoted as $X^\star$ and $y^\star$) and executed over a sufficient number of iterations, the gradient descent trajectory generated by Algorithm 1 is guaranteed to intersect with at least one approximate stationary point of (2):

Algorithm 1 Construction of an approximate primal solution

1: Initialization: $\boldsymbol{X}^0 = \boldsymbol{X}^\star; \boldsymbol{Y}^0 = \boldsymbol{Y}^\star$.
2: **for** $t = 0, 1, \ldots, t_0 - 1$ **do**
3: $\quad \boldsymbol{X}^{t+1} = \boldsymbol{X}^t - \eta \nabla_{\boldsymbol{X}} f\left(\boldsymbol{X}^t, \boldsymbol{Y}^t\right) = \boldsymbol{X}^t - \frac{\eta}{p}\left(\mathcal{P}_\Omega\left(\boldsymbol{X}^t \boldsymbol{Y}^{t\top} - \boldsymbol{M}\right)\boldsymbol{Y}^t + \lambda \boldsymbol{X}^t\right)$
4: $\quad \boldsymbol{Y}^{t+1} = \boldsymbol{Y}^t - \eta \nabla_{\boldsymbol{Y}} f\left(\boldsymbol{X}^t, \boldsymbol{Y}^t\right) = \boldsymbol{Y}^t - \frac{\eta}{p}\left(\left[\mathcal{P}_\Omega\left(\boldsymbol{X}^t \boldsymbol{Y}^{t\top} - \boldsymbol{M}\right)\right]^\top \boldsymbol{X}^t + \lambda \boldsymbol{Y}^t\right)$
5: **end for**

It is critical to acknowledge, however, that Algorithm 1, in its current form, lacks practicality due to its initiation requirement at the ground truth—a condition primarily adopted to streamline the theoretical exposition. To enhance its practical utility, one could consider integrating spectral initialization, a modification poised to surmount the initial condition barrier, thereby bridging the gap between theoretical elegance and practical applicability.

4 Conclusion and Future Work

[5] contributes a refined statistical analysis for the established convex program (10), circumventing the previously requisite spikiness constraint. The theoretical exploration conducted within this work exposes a fascinating synergy between convex relaxation and nonconvex optimization. This discovery is posited to have significant implications, extending its relevance to a multitude of challenges beyond the problem of matrix completion, with extensions including robust PCA, blind deconvolution, structured matrix completion, etc.

References

[1] Samuel Burer and Renato D.C. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, February 2003. 2.1

[2] Samuel Burer and Renato D.C. Monteiro. Local Minima and Convergence in Low-Rank Semidefinite Programming. *Mathematical Programming*, 103(3):427–444, July 2005. 2.2

[3] Yudong Chen and Yuejie Chi. Harnessing Structures in Big Data via Guaranteed Low-Rank Matrix Estimation, May 2018. 2.2, 2.3

[4] Yudong Chen and Martin J. Wainwright. Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees, September 2015. 2.2

[5] Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, and Yuling Yan. Noisy Matrix Completion: Understanding Statistical Guarantees for Convex Relaxation via Nonconvex Optimization, October 2019. 1, 3.1, 3.1, 3.2, 3.3, 4

[6] Yuejie Chi, Yue M Lu, and Yuxin Chen. Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Transactions on Signal Processing*, 67(20):5239–5269, 2019. 2.3

[7] Moritz Hardt. Understanding Alternating Minimization for Matrix Completion. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 651–660, October 2014. 2.2

[8] Prateek Jain, Raghu Meka, and Inderjit Dhillon. Guaranteed Rank Minimization via Singular Value Projection. In *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010. 2.2

[9] Prateek Jain and Praneeth Netrapalli. Fast Exact Matrix Completion with Finite Samples. In *Proceedings of The 28th Conference on Learning Theory*, pages 1007–1034. PMLR, June 2015. 2.2

[10] Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '13, pages 665–674, New York, NY, USA, June 2013. Association for Computing Machinery. 2.2

[11] Raghunandan Keshavan, Andrea Montanari, and Sewoong Oh. Matrix Completion from Noisy Entries. In *Advances in Neural Information Processing Systems*, volume 22. Curran Associates, Inc., 2009. 2.2

[12] Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix Completion From a Few Entries. *IEEE Transactions on Information Theory*, 56(6):2980–2998, June 2010. 2.2

[13] Cong Ma, Kaizheng Wang, Yuejie Chi, and Yuxin Chen. Implicit Regularization in Nonconvex Statistical Estimation: Gradient Descent Converges Linearly for Phase Retrieval and Matrix Completion. In *Proceedings of the 35th International Conference on Machine Learning*, pages 3345–3354. PMLR, July 2018. 2.2

[14] Praneeth Netrapalli, Niranjan U N, Sujay Sanghavi, Animashree Anandkumar, and Prateek Jain. Non-convex Robust PCA. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. 2.2

[15] Ruoyu Sun and Zhi-Quan Luo. Guaranteed Matrix Completion via Non-Convex Factorization. *IEEE Transactions on Information Theory*, 62(11):6535–6579, November 2016. 2.2

[16] Ke Wei, Jian-Feng Cai, Tony F. Chan, and Shingyu Leung. Guarantees of Riemannian Optimization for Low Rank Matrix Recovery. *SIAM Journal on Matrix Analysis and Applications*, 37(3):1198–1222, January 2016. 2.2

[17] Qinqing Zheng and John Lafferty. Convergence Analysis for Rectangular Matrix Completion Using Burer-Monteiro Factorization and Gradient Descent, November 2016. 2.2