Non-convex matrix sensing: Breaking the quadratic rank barrier in the sample complexity

Dominik Stöger DOMINIK.STOEGER@KU.DE
Mathematical Institute for Machine Learning and Data Science (MIDS)
KU Eichstätt-Ingolstadt

Yizhe Zhu YIZHEZHU@USC.EDU
Department of Mathematics, University of Southern California

Editors: Nika Haghtalab and Ankur Moitra

Abstract

For the problem of reconstructing a low-rank matrix from a few linear measurements, two classes of algorithms have been widely studied in the literature: convex approaches based on nuclear norm minimization, and non-convex approaches that use factorized gradient descent. Under certain statistical model assumptions, it is known that nuclear norm minimization recovers the ground truth as soon as the number of samples scales linearly with the number of degrees of freedom of the ground-truth. In contrast, while non-convex approaches are computationally less expensive, existing recovery guarantees assume that the number of samples scales at least quadratically with the rank r of the ground-truth matrix.

In this paper, we close this gap by showing that the non-convex approaches can be as efficient as nuclear norm minimization in terms of sample complexity. Namely, we consider the problem of reconstructing a positive semidefinite matrix $\mathbf{X}_\star \in \mathbb{R}^{d \times d}$ from a few Gaussian measurements of the form

$$\mathbf{y}_i = \frac{1}{\sqrt{m}} \text{trace}\left(\mathbf{A}_i \mathbf{X}_\star\right) \qquad \text{for } i = 1, 2, \ldots, m.$$

We show that factorized gradient descent with spectral initialization converges to the ground truth with a linear rate as soon as the number of samples scales with $\Omega(rd\kappa^2)$, where d is the dimension, and κ is the condition number of the ground truth matrix. This improves the previous rank-dependence in the sample complexity of non-convex matrix factorization from quadratic to linear. Our proof relies on a probabilistic decoupling argument, where we show that the gradient descent iterates are only weakly dependent on the individual entries of the measurement matrices. We expect that our proof technique will be of independent interest to other non-convex problems.[1]

Keywords: non-convex optimization, matrix sensing, sample complexity, virtual sequences

Acknowledgments

D.S. is grateful to Mahdi Soltanolkotabi for fruitful discussions, in particular regarding Theorem 10, and to Felix Krahmer for helpful comments. Y.Z. was partially supported by NSF-Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning and an AMS-Simons Travel Grant.

1. Extended abstract. Full version appears as [arxiv:2408.13276v3].

References

Ji Chen, Dekai Liu, and Xiaodong Li. Nonconvex rectangular matrix completion via gradient descent without ℓ_2, ∞ regularization. *IEEE Trans. Inf. Theory*, 66(9):5806–5841, 2020. ISSN 0018-9448. doi: 10.1109/TIT.2020.2992234.

Yuejie Chi, Yue M. Lu, and Yuxin Chen. Nonconvex optimization meets low-rank matrix factorization: an overview. *IEEE Trans. Signal Process.*, 67(20):5239–5269, 2019. ISSN 1053-587X. doi: 10.1109/TSP.2019.2937282.

Rong Ge, Jason D Lee, and Tengyu Ma. Matrix completion has no spurious local minimum. *Advances in Neural Information Processing Systems*, 29, 2016.

Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from a few entries. *IEEE Trans. Inf. Theory*, 56(6):2980–2998, 2010. ISSN 0018-9448. doi: 10.1109/TIT.2010.2046205.

Yuanzhi Li, Tengyu Ma, and Hongyang Zhang. Algorithmic regularization in over-parameterized matrix sensing and neural networks with quadratic activations. In *Conference On Learning Theory*, pages 2–47. PMLR, 2018.

Cong Ma, Kaizheng Wang, Yuejie Chi, and Yuxin Chen. Implicit regularization in nonconvex statistical estimation: gradient descent converges linearly for phase retrieval, matrix completion, and blind deconvolution. *Found. Comput. Math.*, 20(3):451–632, 2020. ISSN 1615-3375. doi: 10.1007/s10208-019-09429-9.

Mahdi Soltanolkotabi, Dominik Stöger, and Changzhi Xie. Implicit balancing and regularization: Generalization and convergence guarantees for overparameterized asymmetric matrix sensing. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 5140–5142. PMLR, 2023.

Dominik Stöger and Mahdi Soltanolkotabi. Small random initialization is akin to spectral learning: Optimization and generalization guarantees for overparameterized low-rank matrix reconstruction. *Advances in Neural Information Processing Systems*, 34:23831–23843, 2021.

Ruoyu Sun and Zhi-Quan Luo. Guaranteed matrix completion via non-convex factorization. *IEEE Trans. Inf. Theory*, 62(11):6535–6579, 2016. ISSN 0018-9448. doi: 10.1109/TIT.2016.2598574.

Stephen Tu, Ross Boczar, Max Simchowitz, Mahdi Soltanolkotabi, and Ben Recht. Low-rank solutions of linear matrix equations via procrustes flow. In *International Conference on Machine Learning*, pages 964–973. PMLR, 2016.

Xingyu Xu, Yandi Shen, Yuejie Chi, and Cong Ma. The power of preconditioning in overparameterized low-rank matrix sensing. In *International Conference on Machine Learning*, pages 38611–38654. PMLR, 2023.

Qinqing Zheng and John Lafferty. Convergence analysis for rectangular matrix completion using burer-monteiro factorization and gradient descent. *arXiv preprint arXiv:1605.07051*, 2016.