Proceedings of Machine Learning Research vol 196:1–1, 2024 37th Annual Conference on Learning Theory

Fundamental Limits of Non-Linear Low-Rank Matrix Estimation

Pierre Mergny PIERRE.MERGNY@EPFL.CH
IdePHICS laboratory, École Fédérale Polytechnique de Lausanne

Justin Ko JUSTIN.KO@UWATERLOO.CA
Department of Statistics and Actuarial Science, University of Waterloo and UMPA, ENS Lyon and CNRS

Florent Krzakala FLORENT.KRZAKALA@EPFL.CH
IdePHICS laboratory, École Fédérale Polytechnique de Lausanne

Lenka Zdeborová LENKA.ZDEBOROVA@EPFL.CH
SPOC laboratory, École Fédérale Polytechnique de Lausanne

Editors: Shipra Agrawal and Aaron Roth

Abstract

We investigate the problem of estimating a low-rank matrix from non-linear and possibly noisy observations, in the high-dimensional regime $N \gg 1$. Specifically, we examine the scenario where the probabilistic channel may exhibit zero Fisher information, a setting where reconstruction of the signal is impossible for any fixed value of the signal-to-noise ratio λ. In such cases, we show that one can manage to accurately reconstruct the target by appropriately scaling the signal with the dimension.

We prove a strong universality result showing that Bayes-optimal performances are characterized by an equivalent Gaussian model with an effective prior, whose parameters are entirely determined by an expansion of the non-linear function. In particular, we show that to reconstruct the signal accurately, one requires a signal-to-noise ratio growing as $N^{\frac{1}{2}(1-1/k_F)}$, where k_F is the first non-zero Fisher information coefficient of the function. We further show that there exists an explicit entrywise transformation of the observed data, which is completely determined by the structure of the non-linear channel, into a Fisher matrix, which effectively reduces the non-linear spiked matrices, to a *classical* spiked Wigner matrix model.

On the information-theoretic side, we prove a universality result that shows the mutual information of non-linear low-rank matrix estimation corresponds to the one of a Gaussian spiked model with a noise level given by the kth order Fisher information coefficient. The mutual information in the high dimensional limit is computed and expressed as a variational problem. Furthermore, we use the associated spiked matrix characterization, to prove a fixed point characterization for the optimal overlaps between the effective signal x^{k_F} and a sample from the posterior. This gives a fixed point condition of the limiting overlaps, MMSE, of the weak recovery threshold and of the conditions for its existence.

On the algorithmic side, we establish that the standard approximate message-passing algorithm (AMP) for Wigner spiked models applied to the Fisher matrix achieves the MMSE under conditions on the prior distribution of the target vector known for the linear version. We also study the performance of a simpler, off-the-shelve method and, in particular, spectral algorithm. In particular, we show the Fisher matrix is an optimal method for PCA, and displays a transition for the MSE which we characterize. We further demonstrated that optimal denoising of the top eigenvector yields near AMP performances.

In summary, this work provides a unified approach to study non-linear spiked matrices. These results allow us to unify and generalize several recent works on matrix factorization, random matrix theory, PCA, and Bayes optimal estimation based on information exponent.[1]

1. Extended abstract. Full version appears as [arXiv:2403.04234 , v2]