Information-Theoretic Guarantees for Recovering Low-Rank Tensors from Symmetric Rank-One Measurements

Eren C. Kızıldağ KIZILDAG@ILLINOIS.EDU
Department of Statistics, University of Illinois at Urbana-Champaign [*]

Editors: Gautam Kamath and Po-Ling Loh

Abstract

We investigate the sample complexity of recovering tensors with low symmetric rank from symmetric rank-one measurements, a setting particularly motivated by the study of higher-order interactions in statistics and the analysis of two-layer polynomial neural networks. Using a covering number argument, we analyze the performance of the symmetric rank minimization program and establish near-optimal sample complexity bounds when the underlying distribution is log-concave. Our measurement model involves random symmetric rank-one tensors, leading to involved probability calculations. To address these challenges, we employ the Carbery-Wright inequality, a powerful tool for studying anti-concentration properties of random polynomials, and leverage orthogonal polynomial expansions. Additionally, we provide a sample complexity lower bound via Fano's inequality, and discuss broader implications of our results for two-layer polynomial networks.

Keywords: Symmetric tensors, tensor recovery, rank minimization, covering numbers, low-rank, log-concave distributions.

1. Extended Abstract

We study the problem of recovering an unknown, order-ℓ tensor $\boldsymbol{\mathcal{T}}^* \in \mathbb{R}^{d\times\cdots\times d}$ from random measurements of the form

$$Y_i = \langle \boldsymbol{\mathcal{T}}^*, \boldsymbol{X}_i^{\otimes \ell}\rangle, \quad i = 1, \dots, N, \tag{1}$$

where N is the sample size and $\boldsymbol{X}_i \in \mathbb{R}^d$ are i.i.d. random vectors with i.i.d. entries sampled from a log-concave distribution on $\mathbb{R}$. We assume $\boldsymbol{\mathcal{T}}^*$ has low symmetric rank, i.e.,

$$\operatorname{rank}_S(\boldsymbol{\mathcal{T}}^*) := \min\{t \ge 1 : \boldsymbol{\mathcal{T}}^* = \textstyle\sum_{i\le t} \lambda_i \boldsymbol{v}_i^{\otimes \ell}, \lambda_1, \dots, \lambda_r \in \mathbb{R}, \boldsymbol{v}_1, \dots, \boldsymbol{v}_r \in \mathbb{R}^d\} \le r$$

for some r. This setting arises naturally in the study of higher-order interactions in statistics (Bien et al., 2013; Basu et al., 2018; Hao et al., 2020), where the unknown tensor often exhibits low-rank structure (Sidiropoulos and Kyrillidis, 2012; Hung et al., 2016; Hao et al., 2020). Moreover, our setting is also closely related to the problem of learning two-layer polynomial neural networks, see, e.g., Soltanolkotabi et al. (2018); Du and Lee (2018); Emschwiller et al. (2020); Sarao Mannelli et al. (2020); Kızıldağ (2022); Martin et al. (2024); Gamarnik et al. (2024).

Our main contributions are summarized as follows:

- *Strong Recovery*: We establish that for $N = \Omega(dr)$, the symmetric rank minimization program, $\min_{\boldsymbol{\mathcal{T}}} \operatorname{rank}_S(\boldsymbol{\mathcal{T}})$ subject to $\langle \boldsymbol{\mathcal{T}}, \boldsymbol{X}_i^{\otimes \ell}\rangle = Y_i, \forall i = 1, \dots, N$, recovers *all* $\boldsymbol{\mathcal{T}}^*$ with probability one. Our proof leverages multiple techniques, including Carbery-Wright inequality for the anti-concentration of random polynomials (Carbery and Wright, 2001), orthogonal polynomial expansions (Lalley; Szegö, 1939), covering number estimates for low-rank tensors (Zhang and Kileel, 2023), and monotonicity of covering numbers (Vershynin, 2018).

[*] Extended abstract. Full version appears as (Kızıldağ, 2025).

© 2025 E.C. Kızıldağ.

- *Sample Complexity Lower Bound*: In a different statistical setting where $\boldsymbol{\mathcal{T}}^*$ is drawn from a discrete space, we establish a sample complexity lower bound: any estimator $\widehat{\boldsymbol{\mathcal{T}}}$ for $\boldsymbol{\mathcal{T}}^*$, whether deterministic or randomized, incurs an estimation error of at least some $\delta > 0$, unless $N = \widetilde{\Omega}(dr^{1-\gamma})$, where $\gamma > 0$ is arbitrary. To prove this result, we establish a packing number bound for symmetric tensors with low symmetric rank, potentially of independent interest, using a variant of the Gilbert-Varshamov lemma from coding theory (Gilbert, 1952; Varshamov, 1957) derived via the probabilistic method (Alon and Spencer, 2016).

- *Implications for Neural Networks*: Consider a two-layer neural network of width r, computing $\sum_{1\le j\le r} a_j^*\sigma(\langle \boldsymbol{W}_j^*, \boldsymbol{X}\rangle)$ on input $\boldsymbol{X} \in \mathbb{R}^d$, where $\boldsymbol{W}_j^* \in \mathbb{R}^d$ and $a_j^* \in \mathbb{R}$ are ground-truth weights, and $\sigma(x) = x^\ell$ is the activation function. Our results provide improved sample complexity bounds in the *underparameterized* regime, $r = O(d^{\ell-1})$; they remain competitive in the *overparameterized* regime, $r = \Omega(d^{\ell-1})$, particularly when $a_i^* = \Theta(1)$ or when the spectral norm of $\boldsymbol{W} \in \mathbb{R}^{r\times d}$ with rows $\boldsymbol{W}_j^*$ grows polynomially with $\max\{r, d\}$.

Our work aligns with a broad literature on low-rank matrix and tensor recovery (Candes and Tao, 2005; Candès et al., 2006; Cai et al., 2010; Eldar et al., 2012; Mu et al., 2014; Rauhut et al., 2017; Cai et al., 2020; Ahmed et al., 2020; Grotheer et al., 2022; Luo and Zhang, 2023). Much of this literature adopts measurement models of form $Y_i = \langle \boldsymbol{\mathcal{T}}^*, \mathcal{X}_i\rangle$ for tensors $\mathcal{X}_i$ consisting of i.i.d. sub-Gaussian entries or satisfying the tensor restricted isometry property, simplifying probability estimates. Our approach extends beyond these settings and captures polynomial networks.

Acknowledgments

I would like to thank anonymous ALT 2025 referees and the area chair for their valuable feedback, which helped improve the presentation and strengthen the results.

References

Talal Ahmed, Haroon Raja, and Waheed U Bajwa. Tensor regression using low-rank and sparse tucker decompositions. *SIAM Journal on Mathematics of Data Science*, 2(4):944–966, 2020.

Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2016.

Sumanta Basu, Karl Kumbier, James B Brown, and Bin Yu. Iterative random forests to discover predictive and stable high-order interactions. *Proceedings of the National Academy of Sciences*, 115(8):1943–1948, 2018.

Jacob Bien, Jonathan Taylor, and Robert Tibshirani. A lasso for hierarchical interactions. *Annals of statistics*, 41(3):1111, 2013.

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.

Jian-Feng Cai, Lizhang Miao, Yang Wang, and Yin Xian. Provable near-optimal low-multilinear-rank tensor recovery. *arXiv preprint arXiv:2007.08904*, 2020.

Emmanuel J Candes and Terence Tao. Decoding by linear programming. *IEEE transactions on information theory*, 51(12):4203–4215, 2005.

Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

Anthony Carbery and James Wright. Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. *Mathematical research letters*, 8(3):233–248, 2001.

Simon Du and Jason Lee. On the power of over-parametrization in neural networks with quadratic activation. In *International conference on machine learning*, pages 1329–1338. PMLR, 2018.

Yonina C Eldar, Deanna Needell, and Yaniv Plan. Uniqueness conditions for low-rank matrix recovery. *Applied and Computational Harmonic Analysis*, 33(2):309–314, 2012.

Matt Emschwiller, David Gamarnik, Eren C Kızıldağ, and Ilias Zadik. Neural networks and polynomial regression. demystifying the overparametrization phenomena. *arXiv preprint arXiv:2003.10523*, 2020.

David Gamarnik, Eren C Kızıldağ, and Ilias Zadik. Stationary points of a shallow neural network with quadratic activations and the global optimality of the gradient descent algorithm. *Mathematics of Operations Research*, 2024.

Edgar N Gilbert. A comparison of signalling alphabets. *The Bell system technical journal*, 31(3): 504–522, 1952.

Rachel Grotheer, Shuang Li, Anna Ma, Deanna Needell, and Jing Qin. Iterative hard thresholding for low cp-rank tensor models. *Linear and Multilinear Algebra*, 70(22):7452–7468, 2022.

Botao Hao, Anru R Zhang, and Guang Cheng. Sparse and low-rank tensor estimation via cubic sketchings. In *International conference on artificial intelligence and statistics*, pages 1319–1330. PMLR, 2020.

Hung Hung, Yu-Ting Lin, Penweng Chen, Chen-Chien Wang, Su-Yun Huang, and Jung-Ying Tzeng. Detection of gene–gene interactions using multistage sparse and low-rank regression. *Biometrics*, 72(1):85–94, 2016.

Eren C Kızıldağ. *Algorithms and Algorithmic Barriers in High-Dimensional Statistics and Random Combinatorial Structures*. PhD thesis, Massachusetts Institute of Technology, 2022.

Eren C. Kızıldağ. Information-theoretic guarantees for recovering low-rank tensors from symmetric rank-one measurements. *arXiv:2502.05134*, 2025.

Steven Lalley. Orthogonal Polynomials. https://galton.uchicago.edu/~lalley/Courses/386/OrthogonalPolynomials.pdf.

Yuetian Luo and Anru R Zhang. Low-rank tensor estimation via riemannian gauss-newton: Statistical optimality and second-order convergence. *The Journal of Machine Learning Research*, 24 (1):18274–18321, 2023.

Simon Martin, Francis Bach, and Giulio Biroli. On the impact of overparameterization on the training of a shallow neural network in high dimensions. In *International Conference on Artificial Intelligence and Statistics*, pages 3655–3663. PMLR, 2024.

Cun Mu, Bo Huang, John Wright, and Donald Goldfarb. Square deal: Lower bounds and improved relaxations for tensor recovery. In *International conference on machine learning*, pages 73–81. PMLR, 2014.

Holger Rauhut, Reinhold Schneider, and Željka Stojanac. Low rank tensor recovery via iterative hard thresholding. *Linear Algebra and its Applications*, 523:220–262, 2017.

Stefano Sarao Mannelli, Eric Vanden-Eijnden, and Lenka Zdeborová. Optimization and generalization of shallow neural networks with quadratic activation functions. *Advances in Neural Information Processing Systems*, 33:13445–13455, 2020.

Nicholas D Sidiropoulos and Anastasios Kyrillidis. Multi-way compressed sensing for sparse low-rank tensors. *IEEE Signal Processing Letters*, 19(11):757–760, 2012.

Mahdi Soltanolkotabi, Adel Javanmard, and Jason D Lee. Theoretical insights into the optimization landscape of over-parameterized shallow neural networks. *IEEE Transactions on Information Theory*, 65(2):742–769, 2018.

Gabor Szegö. *Orthogonal polynomials*, volume 23. American Mathematical Soc., 1939.

Rom Rubenovich Varshamov. Estimate of the number of signals in error correcting codes. *Docklady Akad. Nauk, SSSR*, 117:739–741, 1957.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Yifan Zhang and Joe Kileel. Covering number of real algebraic varieties and beyond: Improved bounds and applications. *arXiv e-prints*, pages arXiv–2311, 2023.