Open Problem: Average-Case Hardness of Hypergraphic Planted Clique Detection

Yuetian Luo[†]   YLUO86@WISC.EDU

Anru R. Zhang[†*]   ANRUZHANG@STAT.WISC.EDU

[†]*Department of Statistics, University of Wisconsin-Madison*
[*]*Department of Biostatistics and Bioinformatics, Duke University*

Editors: Jacob Abernethy and Shivani Agarwal

Abstract

We note the significance of hypergraphic planted clique (HPC) detection in the investigation of computational hardness for a range of tensor problems. We ask if more evidence for the computational hardness of HPC detection can be developed. In particular, we conjecture if it is possible to establish the equivalence of the computational hardness between HPC and PC detection.

1. Introduction

The analysis of tensors has emerged as an active topic in recent years. Numerous datasets in the form of tensors emerge in various applications, such as collaborative filtering, neuroimaging analysis, hyperspectral imaging. In addition, tensor methods have been applied to many machine learning problems where the observations are not necessarily tensors, such as additive index models, high-order interaction pursuit, topic and latent variable models (Anandkumar et al., 2014). Compared with the inference for vectors or matrices, the tensor problems often possess distinct characteristics that bring significant computational challenges. As pointed out by the seminal work of Hillar and Lim (2013), extensions of many matrix concepts such as operator norm, singular values and eigenvalues are possible but computationally NP-hard. There often exist intrinsic gaps between the computational barriers and statistical limits in many tensor problems, as observed in tensor completion (Barak and Moitra, 2016), tensor PCA (Richard and Montanari, 2014; Hopkins et al., 2015; Wein et al., 2019; Zhang and Xia, 2018; Lesieur et al., 2017; Brennan and Bresler, 2020), high-order clustering (Luo and Zhang, 2020), etc. The analysis of computational barriers has attracted enormous attention because of its crucial role in the understanding of the computational feasibility of a wide range of tensor problems.

In the past decades, several lines of works have been devoted to the analysis of computational complexity in high-dimensional matrix problems. Such works include the average-case reduction (Berthet and Rigollet, 2013), analysis of restricted model of computation (e.g., statistical query Kearns (1998)), analysis of algorithms such as sum-of-squares, belief propagation, approximate message passing, and analysis of optimization landscape. In particular, the average-case reduction requires randomized polynomial-time reduction from the conjectured hard problem in distribution to the target problem up to the conjectured com-

putational barrier (Brennan et al., 2018). Once the average-case reduction is established, all hardness results of the conjectured hard problem can be automatically inherited to the target problem. This provides a one-shot solution to the hardness of the target problem.

It would be ideal to use the same logic and establish the computational barriers for various tensor problems rigorously by average-case reduction from commonly raised conjectures, such as the planted clique (PC) detection or Boolean satisfiability problem (SAT), to the target tensor problem so that all of the hardness results of these well-studied conjectures could be naturally inherited. However, this route is often complicated by the multiway structure of the tensor. To overcome this, an alternative way of average-case reduction from hypergraphic planted clique (HPC) has been proposed in Zhang and Xia (2018); Luo and Zhang (2020). HPC has a more natural tensor structure that enables a more straightforward average-case reduction. A better understanding of the computational hardness relationship between PC detection and HPC detection can make a significant contribution to our knowledge of the average-case complexity in various tensor problems.

2. Planted Clique Detection and Hypergraphic Planted Clique Detection

A d-hypergraph can be seen as an order-d extension of regular graph. In a d-hypergraph $G = (V(G), E(G))$, each hyperedge $e \in E$ includes an unordered group of d different vertices in V. Define $\mathcal{G}_d(N, 1/2)$ as the Erdős-Rényi d-hypergraph with N vertices, where each hyperedge $(i_1, \ldots, i_d)$ is independently included in E with probability $1/2$. Given a d-hypergraph G, define its adjacency tensor $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}}(G) \in (\{0,1\}^N)^{\otimes d}$ as

$$\boldsymbol{\mathcal{A}}_{[i_1,\ldots,i_d]} = \begin{cases} 1, & \text{if } (i_1, \ldots, i_d) \in E; \\ 0, & \text{otherwise.} \end{cases}$$

We define $\mathcal{G}_d(N, 1/2, \kappa)$ as the hypergraphic planted clique (HPC) model with the clique size κ. To generate $G \sim \mathcal{G}_d(N, 1/2, \kappa)$, we sample a random hypergraph from $\mathcal{G}_d(N, 1/2)$, pick κ vertices uniformly at random from $[N]$, denote them as K, and connect all hyperedges e if all vertices of e are in K. The hypergraphic planted clique detection (HPC) can be formulated as the following hypothesis testing problem:

$$H_0 : G \sim \mathcal{G}_d(N, 1/2) \quad \text{v.s.} \quad H_1 : G \sim \mathcal{G}_d(N, 1/2, \kappa). \tag{1}$$

We say a test ϕ solves the HPC detection if $\limsup_{N \to \infty} \mathbb{P}_{H_0}(\phi(\boldsymbol{\mathcal{A}}) = 1) + \mathbb{P}_{H_1}(\phi(\boldsymbol{\mathcal{A}}) = 0) = 0$, i.e., the sum of type-I, II errors converges to zero.

When $d = 2$, HPC detection reduces to the well-regard planted clique (PC) detection problem. The PC detection is statistically solvable if $\kappa \geqslant (2 + \epsilon) \log_2 N$ for $\epsilon > 0$ by searching over all subgraphs of size $(2 + \epsilon) \log_2 N$ (Bollobás and Erdös, 1976). However, the best known polynomial-time algorithm requires $\kappa = \Omega(\sqrt{N})$ (Dekel et al., 2014; Deshpande and Montanari, 2015a). It has been widely conjectured that no polynomial-time algorithm can solve the PC detection when $\kappa = o(\sqrt{N})$. This conjecture has been strengthened in various types of computational models (Jerrum, 1992; Hopkins, 2018; Feldman et al., 2017; Deshpande and Montanari, 2015b; Meka et al., 2015; Barak et al., 2019; Gamarnik and Zadik, 2019). As a result, the hardness conjecture of PC detection has become a key assumption in obtaining computational lower bounds for many problems, such as sparse PCA (Berthet and Rigollet, 2013; Wang et al., 2016), submatrix detection (Ma and Wu, 2015), community detection (Hajek et al., 2015), etc.

Compared to PC detection, HPC detection ($d \geqslant 3$) has received far less attention in literature. Bollobás and Erdös (1976) proved that $K_N^d/(d!\log_2(N))^{1/(d-1)} \overset{a.s.}{\to} 1$ if K_N^d is the size of the largest fully connected subgraph in $G \sim \mathcal{G}_d(N, 1/2)$, which implies the HPC detection problem can be solved by exhaustive search in super-polynomial time when $\kappa \geqslant ((d! + \epsilon)\log_2(N))^{1/(d-1)}$ for any $\epsilon > 0$. Zhang and Xia (2018) observed that applying SVD on the matricization/unfolding of $\boldsymbol{\mathcal{A}}$ solves HPC detection efficiently if $\kappa = \Omega(\sqrt{N})$ but fails when $\kappa = N^{1/2-\epsilon}$ for any $\epsilon > 0$. Then they made a version of the following conjecture indicating that $\kappa = \Omega(\sqrt{N})$ is essential for solving HPC detection in polynomial time.

Conjecture 1 (HPC detection) *Suppose $d \geqslant 2$ is a fixed integer. If*
$$\limsup_{N \to \infty} \log \kappa / \log \sqrt{N} \leqslant 1 - \tau \quad \text{for any } \tau > 0,$$
for any sequence of polynomial-time tests $\{\phi\}_N : G \to \{0, 1\}$, $\liminf_{N\to\infty} \mathbb{P}_{H_0}(\phi(G) = 1) + \mathbb{P}_{H_1}(\phi(G) = 0) > 1/2$.

Recently, Luo and Zhang (2020) provides evidence for HPC detection conjecture by showing that *Metropolis algorithms* (Jerrum, 1992) and *low-degree polynomial algorithms* (Hopkins, 2018) can not solve the problem in polynomial-time when $\kappa = O(N^{\frac{1}{2}-\tau})$ for any $\tau > 0$.

3. How hard is HPC detection?

We raise the following challenges on the average-case computational hardness of HPC detection.

Challenge 1 *Can we establish the sharp limit of κ to guarantee that HPC can be solved by SVD on the matricization/unfolding of $\boldsymbol{\mathcal{A}}$? Is it possible to develop another polynomial-time algorithm for HPC that works below this limit?*

Such results for PC detection can be implied by Knowles and Yin (2013) and Deshpande and Montanari (2015a).

Challenge 2 *Can we provide more evidence for HPC detection conjecture under more classes of algorithms, such as statistical query (SQ), sum-of-squares (SOS), belief propagation (BP), approximate message passing (AMP), etc, in addition to the evidence in Luo and Zhang (2020)?*

Challenge 3 *At the threshold level of $\kappa = O(N^{1/2-\tau})$ for $\tau > 0$, what is relationship between HPC detection and PC detection in their computational hardness?*

First, HPC detection is no harder than PC detection. Given the adjacency tensor $\boldsymbol{\mathcal{A}}$ of hypergraph G, by fixing $(d-2)$ indices and varying the remaining two, we get an adjacency matrix $\mathbf{A}$ from $\boldsymbol{\mathcal{A}}$ which is a PC detection problem. So we can do a randomized polynomial-time reduction from HPC detection to PC detection without changing the planted clique size κ.

Now Challenge 3 becomes, how easier is HPC detection compared to PC detection? A one-shot approach to tackle this challenge is to do randomized polynomial-time reduction from PC to HPC. But this seems to be a very difficult task, upon which we have made several

failed attempts. If reduction from PC to HPC could be proved, we would be able to arrive at the conclusion that PC and HPC detection are equivalently hard and all hardness results established in PC detection can be applied to HPC. The resulting connection between PC and HPC would enable solutions to a huge class of tensor computational hardness problems through existing matrix arguments. Such a connection can also provide solid evidence for the computational hardness assumptions of generalized planted clique problems proposed in (Brennan and Bresler, 2020), based on which we can establish various computational lower bounds more easily.

References

Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *The Journal of Machine Learning Research*, 15(1):2773–2832, 2014.

Boaz Barak and Ankur Moitra. Noisy tensor completion via the sum-of-squares hierarchy. In *Conference on Learning Theory*, pages 417–445, 2016.

Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019.

Quentin Berthet and Philippe Rigollet. Complexity theoretic lower bounds for sparse principal component detection. In *Conference on Learning Theory*, pages 1046–1066, 2013.

Béla Bollobás and Paul Erdös. Cliques in random graphs. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 80, pages 419–427. Cambridge University Press, 1976.

Matthew Brennan and Guy Bresler. Reducibility and statistical-computational gaps from secret leakage. *arXiv preprint arXiv:2005.08099*, 2020.

Matthew Brennan, Guy Bresler, and Wasim Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference On Learning Theory*, pages 48–166, 2018.

Yael Dekel, Ori Gurel-Gurevich, and Yuval Peres. Finding hidden cliques in linear time with high probability. *Combinatorics, Probability and Computing*, 23(1):29–49, 2014.

Yash Deshpande and Andrea Montanari. Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Foundations of Computational Mathematics*, 15(4):1069–1128, 2015a.

Yash Deshpande and Andrea Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In *Conference on Learning Theory*, pages 523–562, 2015b.

Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *Journal of the ACM (JACM)*, 64(2):8, 2017.

David Gamarnik and Ilias Zadik. The landscape of the planted clique problem: Dense subgraphs and the overlap gap property. *arXiv preprint arXiv:1904.07174*, 2019.

Bruce Hajek, Yihong Wu, and Jiaming Xu. Computational lower bounds for community detection on random graphs. In *Conference on Learning Theory*, pages 899–928, 2015.

Christopher J Hillar and Lek-Heng Lim. Most tensor problems are np-hard. *Journal of the ACM (JACM)*, 60(6):45, 2013.

Samuel Hopkins. Statistical inference and the sum of squares method. *Cornell thesis*, 2018.

Samuel B Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-square proofs. In *Conference on Learning Theory*, pages 956–1006, 2015.

Mark Jerrum. Large cliques elude the metropolis process. *Random Structures & Algorithms*, 3(4):347–359, 1992.

Michael Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the ACM (JACM)*, 45(6):983–1006, 1998.

Antti Knowles and Jun Yin. The isotropic semicircle law and deformation of wigner matrices. *Communications on Pure and Applied Mathematics*, 66(11):1663–1749, 2013.

Thibault Lesieur, Léo Miolane, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Statistical and computational phase transitions in spiked tensor estimation. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 511–515. IEEE, 2017.

Yuetian Luo and Anru R Zhang. Tensor clustering with planted structures: Statistical optimality and computational limits. *arXiv preprint arXiv:2005.10743*, 2020.

Zongming Ma and Yihong Wu. Computational barriers in minimax submatrix detection. *The Annals of Statistics*, 43(3):1089–1116, 2015.

Raghu Meka, Aaron Potechin, and Avi Wigderson. Sum-of-squares lower bounds for planted clique. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 87–96. ACM, 2015.

Emile Richard and Andrea Montanari. A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems*, pages 2897–2905, 2014.

Tengyao Wang, Quentin Berthet, and Richard J Samworth. Statistical and computational trade-offs in estimation of sparse principal components. *The Annals of Statistics*, 44(5): 1896–1930, 2016.

Alexander S Wein, Ahmed El Alaoui, and Cristopher Moore. The kikuchi hierarchy and tensor pca. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1446–1468. IEEE, 2019.

Anru Zhang and Dong Xia. Tensor SVD: Statistical and computational limits. *IEEE Transactions on Information Theory*, 64(11):7311–7338, 2018.