Tensor SVD: Statistical and Computational Limits

Abrar Zahin
ASU ID: 1217426678
azahin@asu.edu

Anirudh Rayas
ASU ID: 1218397102
ahrayas@asu.edu

Abstract

In this review we explore the statistical and computational limits for the Tesor SVD problem. We begin with the following paper (1) which outlines a general framework for tensor singular value decomposition. The Tensor SVD problem is completely characterized by this quantity called the signal to noise ratio (SNR) and exhibits three phases, with strong SNR, they show that the classical higher-order orthogonal iteration (HOOI) achieves the minimax optimal rate of convergence in estimation, with weak SNR, the information-theoretical lower bound implies that it is impossible to have consistent estimation in general, however with moderate SNR, they show that the non-convex maximum likelihood estimation (MLE) although having NP-hard computational cost, it provides an optimal solution and also under the hardness hypothesis of hypergraphic planted clique detection, there are no polynomial time algorithms performing consistently in general.

1 Introduction

Singular value decomposition and principal component analysis has been an important tool in multivariate and high dimensional data analysis and have been thoroughly studied in the case of matrices, however they only capture first order interactions and ignore higher order ones, therefore towards this end we intend to study SVD for tensors which are higher order analogues of matrices and we wish to explore the statistical and computational limits for tensor SVD. Tensors have been actively studied in machine learning, electrical engineering and statistics. Some applications involving tensor data includes recommender systems, neuroimaging analysis, computer vision, topic modeling and community detection to name a few. A common objective here is to dig out the underlying high-order low-rank structure, such as the singular subspaces and the whole low-rank tensors, buried in the noisy observations. In this regard, we motivate that tensor SVD is important and therefore it is natural to study it's statistical properties and to complete the picture we wish to study the computational efficiency of the algorithms.

Specifically, we are interested in a low rank tensor $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ which are observed with entrywise corruptions as follows:

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z}$$

where $\mathbf{Z}$ is a third order tensor with iid Gaussian entries, $\mathbf{X}$ is a fixed tensor whose rows and columns lie in a low dimensional subspace, say U_1, U_2, U_3 respectively. The goal of tensor

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

SVD is to estimate U_1, U_2, U_3 and $\mathbf{X}$ from noisy observation $\mathbf{Y}$.

Tensor SVD is interesting than it's two dimensional matrix counterpart for a couple of reasons. Firstly, tensors have a more involved structure, in that they incorporate dependencies in three or more directions while the matrices can incorporate only two. Secondly, many operations for matrices, such as operator norm, singular value decomposition, are either not well defined or computational NP-hard for higher order tensors. Third, high-order tensors often bring about high dimensionality and impose computational challenges since they have significantly more entries.

2 Tensor SVD: Methodology

In this section, we will note some basic notation, preliminaries , and important tensor algebra concepts which will be used in this report. For a more detailed tutorial of tensor algebra, readers are also referred to (2) . For $a, b \in \mathbb{R}$, $a \wedge b = \min\{a, b\}$, $a \wedge b = \max\{a, b\}$. For two sequences $\{a_i\}$, $\{b_i\}$, if there are two constants $C, c > 0$ such that $ca_i \leq b_i \leq Ca_i$ for all $i \geq 1$, we denote $a \asymp b$. C, c, C_0, c_0 are used to denote varying generic constants. For any matrix $A \in \mathbb{R}^{p_1 \times p_2}$ and $\sigma_1(A) \geq \ldots, \geq \sigma_{p_1 \wedge p_2}(A)$ denotes the singular value in non-increasing order. In this report we are interested in the smallest singular value of A: $\sigma_{\min}(A) = \sigma_{p_1 \wedge p_2}(A)$. In addition, the class of matrix Schatten q-norms will be used: $\|A\|_q = \left(\sum_{j=1}^{p_1 \wedge p_2} \sigma_j^q(A)\right)^{1/q}$. $\mathrm{SVD}_r(A)$ denote the leading r singular vectors of A, such that $\mathrm{SVD}_r(A) \in \mathbb{O}_{p_1,r}$. Furthermore, projection operator is defined as $P_A = A\left(A^T A\right)^{\dagger} A^T$, where $(.)^{\dagger}$ represents pseudo-inverse. $\sin\Theta$ distances are used to measure the difference between singular subspaces. Specifically, for any two $p \times r$ matrices with orthonormal columns U and $\hat{U}$, $\Theta\left(U, \hat{U}\right) \triangleq \mathrm{diag}\left(\arccos\left(\sigma_1\right), \ldots, \arccos\left(\sigma_r\right)\right) \in \mathbb{R}^{r \times r}$, where $\sigma_1 \geq \ldots, \geq \sigma_r \geq 0$ are the singular subspaces of $U^T \hat{U}$. The schatten $q-\sin\Theta$-norm is then defined as follows,

$$\|\sin\Theta\left(U, \hat{U}\right)\|_q = \left(\sum_{i=1}^{r} \sin^q\left(\arccos\left(\sigma_i\right)\right)\right)^{1/q} = \left(\sum_{i=1}^{r}\left(1 - \sigma^2\right)^{q/2}\right)^{1/q}, \quad 1 \leq q \leq +\infty.$$

For any tensor $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$, define it's mode-1 matricization as a $p_1 \times (p_2 p_3)$ matrix $\mathcal{M}_1(\mathbf{X})$ such that

$$[\mathcal{M}_1(\mathbf{X})]_{i,(j-1)p_3+k} = \mathbf{X}_{ijk}.$$

The mode-2 and mode-3 matricizations $\mathcal{M}_2(\mathbf{X}) \in \mathbb{R}^{p_2 \times (p_3 p_1)}$ and $\mathcal{M}_3(\mathbf{X}) \in \mathbb{R}^{p_3 \times (p_1 p_2)}$ are defined similarly Tucker ranks of $\mathbf{X} = (r_1, r_2, r_3)$ where $r_i = \mathrm{rank}(\mathcal{M}_i(\mathbf{X}))$. We define the marginal multiplication as $\times_1 : \mathbb{R}^{p_1 \times p_2 \times p_3} \times \mathbb{R}^{r_1 \times p_1} \to \mathbb{R}^{r_1 \times p_2 \times p_3}$ as

$$\mathbf{X} \times_1 Y = \sum_{i'=1}^{p_1} X_{i'jk} Y_{ii'}.$$

Similarly $\times_2$ and $\times_3$ can be defined. Tucker rank is associated with the following decomposition. Let $U_1 \in \mathbb{O}_{p_1,r_1}$, $U_2 \in \mathbb{O}_{p_2,r_2}$, $U_3 \in \mathbb{O}_{p_3,r_3}$ be the left singular vectors of $\mathcal{M}_1(\mathbf{X})$, $\mathcal{M}_2(\mathbf{X})$, $\mathcal{M}_3(\mathbf{X})$ respectively, then there exists a core tensor $\mathbf{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ such that

$$\mathbf{X} = \mathbf{S} \times_1 U_1 \times_2 U_2 \times_3 U_3$$

Finally, to measure the tensor estimation error, tensor Frobenius norm is defined as follows.

$$\|\mathbf{X}\|_F = \left[\sum_{i,j,k=1}^{p_1,p_2,p_3} \mathbf{X}_{ijk}^2\right]^{1/2}.$$

3 Statistical Limits: Minimax Upper and Lower Bounds

3.1 Tensor SVD: Methodology

Using tucker rank, the original tensor SVD model in Eq. (1) can be formulated as follows,

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z} = \mathbf{S} \times_1 U_1 \times_2 U_2 \times_3 U_3 + \mathbf{Z}, \tag{1}$$

where $\{\mathbf{Z}_{ijk}\}_{i,j,k=1}^{p_1,p_2,p_3} \stackrel{iid}{\sim} \mathcal{N}(0,\sigma^2)$ is a noisy tensor, $U_1 \in \mathbb{O}_{p_1,r_1}$, $U_2 \in \mathbb{O}_{p_2,r_2}$, $U_3 \in \mathbb{O}_{p_3,r_3}$, and $\mathbf{S} \in \mathbb{R}^{r_1\times r_2\times r_3}$. Our goal is to estimate U_1, U_2, U_3 from $\mathbf{X}$ and $\mathbf{Y}$. One approach for achieving this goal is maximum likelihood estimation (MLE) for estimating $\hat{\mathbf{X}}_{\text{mle}}$ and $\hat{U}^k_{\text{mle}}$ for $k = 1, 2, 3$. MLE seeks for best (r_1, r_2, r_3) approximation for $\mathbf{Y}$ in Frobenius norm. Though MLE achieves optimal rate of convergence, MLE for tensor SVD problem is computationally NP-hard and non-convex (3). Thus, MLE may not be applicable in practice.

In order to overcome the computational difficulties of MLE, a version of higher order orthogonal iteration (HOOI) is considered which includes three main steps: spectral initialization, power iteration, and tensor projection. The first two steps produce optimal estimations of $U1, U2, U3$, and the final step outputs an optimal estimator of the underlying low-rank tensor $\mathbf{X}$.

Spectral initialization: Since U_1, U_2, and U_3 represents the singular subspaces of $\mathcal{M}_1(\mathbf{X})$, $\mathcal{M}_2(\mathbf{X})$, and $\mathcal{M}_3(\mathbf{X})$, respectively. HOOI algorithm initialize $\hat{U}_1^{(0)}, \hat{U}_1^{(0)}, \hat{U}_2^{(0)}$ by performing SVD on three *matricized* version of observed tensor $\mathbf{Y}$.

$$\hat{U}_k^{(0)} = \text{SVD}_{r_k}(\mathcal{M}_k(\mathbf{Y})) = \text{the first } r_k \text{ left singular vectors of } \mathcal{M}_k(Y) \tag{2}$$

Power Iteration: $\hat{U}_1^{(t)}, \hat{U}_2^{(t)}$ and $\hat{U}_3^{(t)}$ are updated by denoising $\mathbf{Y}$ with $\hat{U}_1^{(t-1)}, \hat{U}_2^{(t-1)}$ and $\hat{U}_3^{(t-1)}$ via a projection operation. Thus, for $t = 1, 2, \ldots$ following is computed

$$\hat{U}_1^{(t)} = \text{first } r_1 \text{ left singular vectors of } \mathcal{M}_1\left(\mathbf{Y} \times_2 \left(\hat{U}_2^{(t-1)}\right)^T \times_3 \left(\hat{U}_3^{(t-1)}\right)^T\right),$$
$$\hat{U}_2^{(t)} = \text{first } r_2 \text{ left singular vectors of } \mathcal{M}_2\left(\mathbf{Y} \times_1 \left(\hat{U}_1^{(t)}\right)^T \times_3 \left(\hat{U}_3^{(t-1)}\right)^T\right),$$
$$\hat{U}_3^{(t)} = \text{first } r_3 \text{ left singular vectors of } \mathcal{M}_3\left(\mathbf{Y} \times_1 \left(\hat{U}_1^{(t)}\right)^T \times_3 \left(\hat{U}_3^{(t)}\right)^T\right).$$

The iteration is stopped when no more denoising is possible or a pre-determined maximum number of iterations is reached.

Projection: With the final estimates $\hat{U}_1^{(t)}, \hat{U}_2^{(t)}, \hat{U}_3^{(t)}$, $\mathbf{S}$ and $\mathbf{X}$ are estimated as follows,

$$\hat{\mathbf{S}} = \mathbf{Y} \times_1 \hat{U}_1^T \times_2 \hat{U}_2^T \times_3 \hat{U}_3^T, \quad \hat{\mathbf{X}} = \mathbf{S} \times_1 \hat{U}_1 \times_2 \hat{U}_2 \times_3 \hat{U}_3 = \mathbf{Y} \times_1 P_{\hat{U}_1} \times_2 P_{\hat{U}_2} \times_3 P_{\hat{U}_3}. \tag{3}$$

3.2 Statistical Limits: Minimax Upper and Lower Bounds

In this subsection, we look at the statistical limits for tensor SVD. Specifically, we will note the corresponding upper bounds and lower bounds of HOOI developed by this paper. For any $\mathbf{X} \in \mathbb{R}^{p_1\times p_2\times p_3}$, $\lambda = \min_{k=1,2,3,\sigma_{r_k}}(\mathcal{M}_k(\mathbf{X}))$ as the minimal singular values of each matricization. Suppose the signal-to-noise ratio is $\lambda/\sigma = p^\alpha$, where $p = \min\{p_1, p_2, p_3\}$. Then tensor SVD problem operates in three distinct phases: $\alpha \geq 3/4$ (strong SNR), $\alpha < 1/2$ (weak SNR), and $1/2 \leq \alpha < 3/4$. (moderate SNR).

Theorem 1 (Upper bound for HOOI)**.** *Suppose there exist constants $C_0, c_0 > 0$ such that $p_k \leq c_0 p$, $\|\mathbf{X}\|_F \leq C_0\sigma\exp(c_0 p)$, $r_k \leq C_0 p^{1/2}$ for $p = \min\{p_1, p_2, p_3\}$, and $k = 1, 2, 3$. Then there exist absolute constants $C_{gap}, C > 0$, which do not depend on $p - k, r_k, \lambda, \sigma, q$, such that whenever $\lambda/\sigma \geq C_{gap}p^{3/4}$, after at most $t_{\max} = C\left(\log\left(\frac{p}{\lambda}\right) \vee 1\right)$ iterations in HOOI, the following upper bounds hold,*

$$\mathbb{E}_{r_k}^{-1/q}\|sin\Theta\left(\hat{U}_k, U_k\right)\|_q \leq C\frac{\sqrt{p_k}}{\lambda/\sigma}, k = 1, 2, 3, \quad 1 \leq q \leq \infty \tag{4}$$

$$\mathbb{E}\|\hat{\mathbf{X}} - \mathbf{X}\|_F \leq C\sigma^2(p_1r_1 + p_2r_2 + p_3r_3), \ \mathbb{E}\frac{\|\hat{\mathbf{X}} - \mathbf{X}\|_F^2}{\|\mathbf{X}\|_F^2} \leq \left(\frac{(p_1 + p_2 + p_3)}{\frac{\lambda^2}{\sigma^2}} \wedge 1\right). \tag{5}$$

In other words, Theorem 1 confirms that this tensor unfolding mehcanism to apply HOOI algorithm guarantees an upper bound on the error in estimation of low dimensional subspaces and original tensor. Also this paper notes that their result in Theorem 1 outperforms the ones by sum-of-squares based convex relaxations based scheme (4), where an additional logarithmic factor on the assumption of λ is required. One important point to note in Theorem 1 is that the strong SNR assumption is crucial to guratee the performance of HOOI, because λ should be at least of order $p^{3/4}$ to provide a meaningful initializations.

Moreover, the achieved upper bound by the estimators obtained by MLE under the assumption that $\lambda\sigma \geq Cp^{1/2}$ are noted in (1). Furthermore, they established the lower bound for tensor SVD considering the following class of general low-rank tensors,

$$\mathcal{F}_{\mathbf{p},\mathbf{r}}(\lambda) = \{\mathbf{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3} : \mathrm{rank}_k(\mathbf{X}) \leq r_k, \sigma_{r_k}(\mathcal{M}_k(\mathbf{X})) \geq \lambda, k = 1,2,3\} \tag{6}$$

Here $\mathbf{p} = (p_1, p_2 p_3)$, $\mathbf{r} = (r_1, r_2, r_3)$ represent the dimension and rank triplets, λ is the smallest non-zero singular value for each matricization of $\mathbf{X}$, which essentially measures the signal strength of the problem. The following lower bound holds over $\mathcal{F}_{\mathbf{p},\mathbf{r}}(\lambda)$.

Theorem 2 (Lower Bound)**.** *Suppose $p = \min\{p_1, p_2, p_3\}, \max\{p_1, p_2, p_3\} \leq C_0 p, \max\{_1, r_2, r_3\} \leq C_0 \min(r_1, r_2, r_3), 4r_1 \leq r_2 r_3, 4r_2 \leq r_3 r_1, 4r_3 \leq r_1 r_2, 1 \leq r_k \leq p_k/3$ and $\lambda > 0$, then there exists a universal constant $c > 0$ such that for $1 \leq q \leq \infty$,*

$$\inf_{\tilde{U}_k} \sup_{\mathbf{X} \in \mathcal{F}_{\mathbf{p},\mathbf{r}}(\lambda)} \mathbb{E}_{r_k}^{-1/q} \| sin\Theta\left(\tilde{U}_k, U_k\right) \|_q \geq c\left(\frac{\sqrt{p_k}}{\lambda/\sigma} \wedge 1\right), \quad k = 1,2,3, \tag{7}$$

$$\inf_{\hat{\mathbf{X}}} \sup_{\mathbf{X} \in \mathcal{F}_{\mathbf{p},\mathbf{r}}(\lambda)} \mathbb{E}\|\hat{\mathbf{X}} - \mathbf{X}\|_F^2 \geq c\sigma^2 (p_1 r_1 + p_2 r_2 + p_3 r_3), \tag{8}$$

$$\inf_{\hat{\mathbf{X}}} \sup_{\mathbf{X} \in \mathcal{F}_{\mathbf{p},\mathbf{r}}(\lambda)} \mathbb{E}\frac{\|\hat{\mathbf{X}} - \mathbf{X}\|_F^2}{\|\mathbf{X}\|_F^2} \leq c\left(\frac{p_1 + p_2 + p_3}{\lambda^2/\sigma^2} \wedge 1\right). \tag{9}$$

Theorem 2 implies that under the weak SNR setting the constant term dominates in Eq. (7) and there are no consistent estimators for U_1, U_2, U_3. On the other hand, when moderate and strong SNR are considered, $\frac{\sqrt{p_k}}{\lambda\sigma}$ domiates in Eq. (7) and also provides minimax lower bounds for the estimation error.

4 Computational Limits in Moderate SNR Case

Here we look at the computational limits for Tensor SVD under the moderate SNR regime. More specifically the paper considers the case where $\lambda/\sigma = p^{\alpha}$ for $1/2 \leq \alpha < 3/4$. It has been shown in the paper that every polynomial time algorithm is statistically inconsistent in estimating the singular subspaces and the core tensor based on the computational hardness assumption. The computational lower bounds are established based on the hardness hypothesis of hypergraphic planted clique detection.

4.1 Hypergraphic planted clique detection (HPC)

Let $G = (V, E)$ be a graph where V and $E = \{(i,j) : i, j \in V\}$ are the vertex and edge sets respectively. A 3-hypergraph is a generalization of a graph where each edge can join atmost three vertices. Given a 3-hypergraph $G = (V, E)$ with $V = |N|$, then it's adjacency tensor $A \in \{0,1\}^{N \times N \times N}$ is defined as

$$A_{ijk} = \begin{cases} 1 & \text{if } (i,j,k) \in E \\ 0 & \text{else} \end{cases} \tag{10}$$

The paper assumes that the random hypergraph follows the Erdös Rényi model where each hyperedge is created with probability 1/2, we denote this model as $\mathcal{G}_3(N, 1/2)$. Let $V_1 \subset V$ and $\kappa_N \leq |V_1|$ then we denote by $\mathcal{G}_3(N, \kappa_N, 1/2)$ as the hypergraph where a clique of size κ_N is planted in V_1. The planted clique detection problem is basically given a random hypergraph, is it possible to detect whether a clique is planted or not in the Erdös Rényi model. More formally the HPC problem can be cast as the following detection problem.

Definition 1. *Let G be drawn from either $\mathcal{G}_3(N, 1/2, \kappa_N, V_1)$ or $\mathcal{G}_3(N, 1/2, \kappa_N, V_2)$ where $V_1 = \{1, 2, \dots, N/2\}$ and $V_2 = \{N/2+1, N/2+2, \dots, N\}$. The hypergraphic planted clique detection denoted as $PC_3(N, \kappa_N)$ refers to the following hypothesis testing problem*

$$H_0 : G \sim \mathcal{G}_3(N, 1/2, \kappa_N, V_1) \quad v/s \quad H_1 : G \sim \mathcal{G}_3(N, 1/2, \kappa_N, V_2) \tag{11}$$

Given a hypergraph sampled from either H_0 or H_1 with an adjacency tensor $\mathbf{A} \in \{0,1\}^{N\times N\times N}$. Let $\psi(\cdot) : \{0,1\}^{N\times N\times N} \to \{0,1\}$ be a binary valued function on $\mathbf{A}$ such that $\psi(A) = 1$ indicates rejection of H_0. Then the risk of test ψ is defined as

$$\mathcal{R}_{N,\kappa_N}(\psi) = \mathbb{P}_{H_0}\{\psi(A) = 1\} + \mathbb{P}_{H_1}\{\psi(A) = 0\} \tag{12}$$

Simply put, given a random hypergraph $G \sim H_0$ or H_1 our goal is to identify whether the clique is planted in the first or second half of vertices.

When the hyperedges are replaced by edges the HPC problem reduces to the well studied planted clique detection problem. The difficulty of the planted clique detection depends on the size of the clique. The statistical limit for the clique size was shown to be $\kappa_N = o(\log N)$ in (5), that is it is impossible to determine whether a planted clique exists or not because the $\mathcal{G}_2(N, 1/2)$ random graph model contains a clique of size $2\log N$ with high probability. The computational limit is as follows, if $\kappa_N \geq C\sqrt{N}$ then a planted clique can be located by a polynomial time algorithm (6). However when $\log N < \kappa_N < \sqrt{N}$ then it is widely conjectured that no polynomial time exists (7). It has also been argued in (8) that the HPC problem is atleast as difficult as the planted clique detection problem. Towards this the paper presents the following computational hardness assumption on hypergraphic planted clique detection.

Hypothesis H(τ): For any sequence $\{\kappa_N\}$ such that $\limsup_{N\to\infty} \frac{\log \kappa_N}{\log \sqrt{N}} \leq 1 - \tau$ and any sequence of polynomial time tests $\{\psi_N\}$

$$\liminf_{N\to\infty} \mathcal{R}_{N,\kappa_N}(\psi_N) \geq \frac{1}{2} \tag{13}$$

4.2 Computational lower bounds of Tensor SVD

To establish computational lower bounds the paper considers an average case reduction. Average case reduction is a one shot solution once established all hardness results of the conjectured hard problem can be inherited to the target problem. It is ideal to do average case reduction from commonly raised conjectures like planted clique since these problems have been widely studied and conjectured that no polynomial algorithm exists under settings. In this case the paper considers an average case reduction to hypergraphic planted clique detection since this has a natural tensor structure. The following theorem establishes the computational lower bound.

Theorem 3. *Suppose the hypergraphic planted clique assumption $H(\tau)$ holds for some $\tau \in (0,1)$. Then there exists absolute constants $c_0, c_1 > 0$ such that if $\frac{\lambda}{\sigma} \leq c_0 \left[\frac{p^{\frac{3}{4}(1-\tau)}}{\sqrt{\log 3p}}\right]$ then for any integers $r_1, r_2, r_3 \geq 1$ and any polynomial time estimator $\widehat{U}_k^{(p)}, \widehat{\boldsymbol{X}}^{(p)}$, the following inequalities hold*

$$\liminf_{p\to\infty} \sup_{X\in\mathcal{F}_{p,r}(\lambda)} \mathbb{E}\left[\|\sin\Theta(\widehat{U}_k^{(p)}, U_k)\|^2\right] \geq c_1, \quad k = 1,2,3 \tag{14}$$

$$\liminf_{p\to\infty} \sup_{X\in\mathcal{F}_{p,r}(\lambda)} \frac{\|\widehat{\boldsymbol{X}}^{(p)} - \boldsymbol{X}\|_F^2}{\|\boldsymbol{X}\|_F^2} \geq c_1 \tag{15}$$

In other words suppose the hypergraphic planted clique hypothesis holds for some $\tau \in (0,1)$, then there exists constants such that if $\frac{\lambda}{\sigma} \lesssim p^{\frac{3}{4}(1-\tau)}$, then any polynomial time algorithm incurs an error which is uniformly bounded by the constants

References

[1] A. Zhang and D. Xia, "Tensor svd: Statistical and computational limits," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 7311–7338, 2018.

[2] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[3] C. J. Hillar and L.-H. Lim, "Most tensor problems are np-hard," *Journal of the ACM (JACM)*, vol. 60, no. 6, pp. 1–39, 2013.

[4] S. B. Hopkins, J. Shi, and D. Steurer, "Tensor principal component analysis via sum-of-square proofs," in *Conference on Learning Theory.* PMLR, 2015, pp. 956–1006.

[5] B. Bollobás and P. Erdös, "Cliques in random graphs," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 80, no. 3. Cambridge University Press, 1976, pp. 419–427.

[6] N. Alon, M. Krivelevich, and B. Sudakov, "Finding a large hidden clique in a random graph," *Random Structures & Algorithms*, vol. 13, no. 3-4, pp. 457–466, 1998.

[7] V. Feldman, E. Grigorescu, L. Reyzin, S. S. Vempala, and Y. Xiao, "Statistical algorithms and a lower bound for detecting planted cliques," *Journal of the ACM (JACM)*, vol. 64, no. 2, pp. 1–37, 2017.

[8] Y. Luo and A. R. Zhang, "Open problem: Average-case hardness of hypergraphic planted clique detection," in *Conference on Learning Theory.* PMLR, 2020, pp. 3852–3856.